



Semiparametric modelling of two-component mixtures with stochastic dominance

Jingjing Wu¹ · Tasnima Abedin² · Qiang Zhao³

Received: 11 October 2021 / Revised: 21 March 2022 / Accepted: 13 April 2022 /

Published online: 24 May 2022

© The Institute of Statistical Mathematics, Tokyo 2022

Abstract

In this work, we studied a two-component mixture model with stochastic dominance constraint, a model arising naturally from many genetic studies. To model the stochastic dominance, we proposed a semiparametric modelling of the log of density ratio. More specifically, when the log of the ratio of two component densities is in a linear regression form, the stochastic dominance is immediately satisfied. For the resulting semiparametric mixture model, we proposed two estimators, maximum empirical likelihood estimator (MELE) and minimum Hellinger distance estimator (MHDE), and investigated their asymptotic properties such as consistency and normality. In addition, to test the validity of the proposed semiparametric model, we developed Kolmogorov–Smirnov type tests based on the two estimators. The finite-sample performance, in terms of both efficiency and robustness, of the two estimators and the tests were examined and compared via both thorough Monte Carlo simulation studies and real data analysis.

Keywords Two-component semiparametric mixture model · Stochastic dominance · Maximum empirical likelihood estimator · Minimum Hellinger distance estimator · Asymptotic normality and robustness

✉ Jingjing Wu
jinwu@ucalgary.ca

¹ Department of Mathematics and Statistics, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

² Clinical Research Unit and Translational Laboratories, Alberta Health Services, 1331 29 Street NW, Calgary, AB T2N 4N2, Canada

³ School of Mathematics and Statistics, Shandong Normal University, No.1 University Road, Science Park, Changqing District, Jinan 250358, Shandong, China

1 Introduction

We consider the two-sample model introduced by Abedin (2018). Specifically, suppose there is a sample from a two-component mixture population $H = (1 - \lambda)F + \lambda G$, where the mixing proportion λ is strictly between 0 and 1, and F , G and H are cumulative distribution functions (c.d.f.) such that $F \neq G$ and F and G satisfy the stochastic dominance constraint $F \geq G$. In addition, we have a separate sample identified as from the first component F . As a result, the data and model are formularized as

$$\begin{aligned} X_1, \dots, X_m &\stackrel{\text{i.i.d.}}{\sim} f(x), \\ Y_1, \dots, Y_n &\stackrel{\text{i.i.d.}}{\sim} h(x) = (1 - \lambda)f(x) + \lambda g(x), \quad x \in \mathbb{R}, \end{aligned} \quad (1)$$

where f , g and h are the corresponding probability density functions (p.d.f.) of F , G and H , respectively. Here f , g and λ are all unknown. The problem of our interest is to make inferences for λ , treating f and g as nuisance parameters, and to estimate the likelihood of a new observation coming from a particular component.

A motivating example that made model (1) arise is the problem to identify differentially expressed genes in case-control studies (e.g., contract vs. not contract COVID-19) of genetic data, such as in Chen and Wu (2013), Kharchenko et al. (2014), Lu et al. (2015) and Ficklin et al. (2017), among many others. In order to solve the problem, frequently a proposed test is applied repeatedly to each single gene. The distribution of those thousands of test statistic values is thus a two-component mixture $(1 - \lambda)f + \lambda g$, where f is the p.d.f. of the statistic under the null hypothesis that a particular gene is not differentially expressed while g is that under the alternative hypothesis that a particular gene is differentially expressed, and λ is the proportion of differentially expressed genes. Usually f is much easier to derive theoretically than g , otherwise if f is unknown in practice, in many studies pathologists or experts can confidently identify some genes that are not differentially expressed so that we have a sample (test statistic values of those genes) from f . The latter case is exactly model (1). The dominance constraint $F \geq G$ is very intuitive in many cases where the statistic values of marker genes are likely larger (or smaller) than those of non-marker genes (e.g., Student's t , ANOVA test statistic). In this motivating example one may argue that the genes, and thus the values of a particular statistic, are not i.i.d. as assumed in model (1). Nevertheless, the distribution of the statistic over all genes is assumed nonparametrically unknown which will provide enough flexibility to weaken, if not remove, the effect of dependence. By Bayes' rule the probability of a gene with test statistic value y being a marker gene is given by

$$p(y) := P(\text{marker gene} | y) = \frac{\lambda g(y)}{(1 - \lambda)f(y) + \lambda g(y)}. \quad (2)$$

Once λ , f and g are estimated, one can estimate according to (2) the probability that a particular gene is differentially expressed.

Besides the motivating example, model (1) could also be used to model many other real data structure. For more examples readers are referred to Abedin (2018)

and Wu and Abedin (2021) which, to our best knowledge, are the only works in literature on model (1). In these two works, the authors proposed and studied two estimators. The first one is based on c.d.f. estimations with use of the dominance inequality, while the second is a maximum likelihood estimator (MLE) based on multinomial approximation. Another work on a model closely related to model (1) is given in Smith and Vounatsou (1997). However their model doesn't assume the stochastic dominance constraint but instead makes a generally stronger assumptions on the p.d.f.s f and g .

This paper is organized in the following way. In Sect. 2, we introduce a semiparametric mixture model in order to accommodate the stochastic dominance constraint. A maximum empirical likelihood estimator (MELE) and a minimum Hellinger distance estimator (MHDE) of the unknown parameters are proposed in Sects. 3 and 4, respectively. Their asymptotic properties such as consistency and asymptotic normality are also presented in these two sections. Section 5 is devoted to testing the validity of the proposed semiparametric model with use of the MELE and MHDE. The finite-sample performance of the two proposed estimators and the goodness-of-fit tests are assessed in Sect. 6 via thorough Monte Carlo simulation studies, while the implementation of the proposed methods are demonstrated in Sect. 7 through two real data examples. Finally the concluding remarks are presented in Sect. 8. Some conditions for the theoretical results are deferred to Appendix, while the derivations and proofs of the theorems and lemmas are given in a separate supplementary document to save space.

2 A semiparametric modelling

Abedin (2018) and Wu and Abedin (2021) proposed and studied two consistent estimators for model (1). However, due to the non-identifiability of the model in general, it is difficult to obtain an estimator with good asymptotic properties such as asymptotic normality. Therefore, we introduce in what follows a semiparametric model which will be proved identifiable and will ensure the stochastic dominance.

Let Z denote a binary response variable and Y the associated covariate. Then the logistic regression model is given by

$$P(Z = 1|Y = y) = \frac{\exp [\alpha^* + \beta^\top r(y)]}{1 + \exp [\alpha^* + \beta^\top r(y)]},$$

where $r(y) = (r_1(y), \dots, r_p(y))^\top$ is a given $p \times 1$ vector of functions of y , α^* is the intercept parameter and $\beta = (\beta_1, \dots, \beta_p)^\top$ is the $p \times 1$ coefficient parameter vector. In case-control studies data are collected retrospectively. For example, a random sample of subjects with disease $Z = 1$ ('case') and a separate random sample of subjects without disease $Z = 0$ ('control') are selected with Y observed for each subject. Let $\pi = P(Z = 1) = 1 - P(Z = 0)$. Let $f(y)$ and $g(y)$ denote the conditional p.d.f.s of Y given $Z = 0$ and $Z = 1$ respectively, then it follows from the Bayes' rule that

$$g(y) = \exp[\alpha + \beta^\top r(y)]f(y), \quad (3)$$

where $\alpha = \alpha^* + \log[(1 - \pi)/\pi]$. Now model (1) is reduced to the semiparametric mixture model

$$\begin{aligned} X_1, \dots, X_m &\stackrel{\text{i.i.d.}}{\sim} f(x), \\ Y_1, \dots, Y_n &\stackrel{\text{i.i.d.}}{\sim} h_\theta(x) := h(x) = \{(1 - \lambda) + \lambda \exp[\alpha + \beta^\top r(x)]\}f(x), \end{aligned} \quad (4)$$

with $f \in \mathcal{F}$ and the parameter vector of interest $\theta = (\lambda, \alpha, \beta^\top)^\top \in \Theta \subseteq \mathbb{R}^{p+2}$ such that $\int e^{\alpha + \beta^\top r(x)} f(x) dx = 1$, where

$$\begin{aligned} \Theta &= \{(\lambda, \alpha, \beta^\top)^\top : \lambda \in (0, 1), \beta \neq 0\}, \\ \mathcal{F} &:= \left\{ f : f \geq 0, \int f(x) dx = 1, \int \exp[\beta^\top r(x)] f(x) dx < \infty, f \text{ is continuous} \right\}. \end{aligned} \quad (5)$$

The relationship (3) between two p.d.f.s was first proposed by Anderson (1972). It essentially assumes that the log-likelihood ratio of the two p.d.f.s is linear in the observations. With $r(x) = x$ or $r(x) = (x, x^2)^\top$, it has wide applications in logistic discriminant analysis (Anderson, 1972, 1979) and case-control studies (Prentice and Pyke, 1979; Breslow and Day, 1980). For $r(x) = x$, (3) encompasses many common distributions, including two exponential distributions with different means and two normal distributions with common variance but different means. Model (3) with $r(x) = (x, x^2)^\top$ also coincides with the exponential family of densities considered in Efron and Tibshirani (1996). Moreover, model (3) can be viewed as a biased sampling model with the ‘tilt’ weight function $\exp[\alpha + \beta^\top r(x)]$ depending on the unknown parameters α and β . Note that the test of equality of f and g can be regarded as a special case of model (4) with $\beta = 0$.

Qin and Zhang (1997) discussed a goodness-of-fit test for logistic regression based on case-control data where the first sample comes from the control group f and independently the second sample comes from the case group g . They proposed a Kolmogorov–Smirnov type statistic to test the validity of (3) with $r(y) = y$. When data from both the mixture and the two individual components satisfying (3) are available, Qin (1999) developed an empirical likelihood ratio-based statistic for constructing confidence intervals of the mixing proportion. For the same model and data structure, Zhang (2002) proposed an EM algorithm to calculate the MELE while (Zhang, 2006) proposed a score statistic to test the mixing proportion. Chen and Wu (2013) employed (3) to model differentially expressed genes of acute lymphoblastic leukemia patients and acute myeloid leukemia patients.

For model (4) with $r(y) = y$, if $\beta > 0$ then we can easily check that $p(y)$ in (2), the probability of y being from g , is a monotonic increasing function. Further we can prove in the next theorem that if $\beta > 0$, then the stochastic dominance constraint $F \geq G$ is implied by (3). We call $(1, r(y))^\top$ linearly independent on the support, say Ω , of f if $(1, r(y))^\top$, as a vector of functions of y , is linearly independent over Ω .

Theorem 1 *If $(1, r(y))^T$ is linearly independent on the support of f , then model (4) with parameter space (5) is identifiable. Specially, (4) is identifiable when $r(y) = y$; if further $\beta > 0$, then $F \geq G$.*

Even though Theorem 1 tells us that the condition (3) is stronger than the original stochastic dominance constraint, the resulting semiparametric mixture model (4) is identifiable and has better interpretation than the nonparametric mixture model (1). In addition, the estimation of (4) may possess better asymptotic properties than those of (1). One may also consider higher order polynomial for $r(y)$ and find sufficient conditions for $F \geq G$. For model simplicity, we focus on model (4) with $r(x) = x$ and $\beta > 0$ (to ensure the dominance $F \geq G$) throughout this paper. We first propose two estimators for this model.

3 MELE of the parameters

Let $(T_1, \dots, T_{m+n})^T = (X_1, \dots, X_m, Y_1, \dots, Y_n)^T$ be the pooled data. In MELE, F is assumed a step function with $p_i = dF(T_i)$ (jump at T_i) such that $\sum_{i=1}^{m+n} p_i = 1$. As a result, $dH(Y_j) = p_{m+j}[(1 - \lambda) + \lambda e^{\alpha + \beta Y_j}]$. Consequently, the empirical likelihood function of (4) with $r(y) = y$ and $\beta > 0$ is

$$L(\lambda, \alpha, \beta) = \prod_{i=1}^m dF(X_i) \prod_{j=1}^n dH(Y_j) = \prod_{i=1}^{m+n} p_i \prod_{j=1}^n [(1 - \lambda) + \lambda e^{\alpha + \beta Y_j}],$$

subject to constraints $\beta > 0$, $0 < \lambda < 1$, $p_i \geq 0$, $\sum_{i=1}^{m+n} p_i = 1$, and $\sum_{i=1}^{m+n} p_i e^{\alpha + \beta T_i} = 1$. The constraint $\beta > 0$ won't limit the use of model (4) and proposed estimation methods, as we can always switch F with G if $\beta < 0$ in model (4). Though in this work we assume both F and G are two continuous populations, we can see that the above likelihood is also valid for discrete populations. To find the MELE, we use the Lagrange multipliers and maximize

$$\sum_{i=1}^{m+n} \log p_i + \sum_{j=1}^n \log [(1 - \lambda) + \lambda e^{\alpha + \beta Y_j}] - t_1 \left[\sum_{i=1}^{m+n} p_i - 1 \right] - t_2 \left[\sum_{i=1}^{m+n} p_i e^{\alpha + \beta T_i} - 1 \right].$$

By taking partial derivatives we obtain (details in the supplementary document)

$$p_i = \frac{1}{(m+n)[1 + \rho_N \lambda (e^{\alpha + \beta T_i} - 1)]}, \quad (6)$$

where $\rho_N = n/(m+n)$ with $N = m+n$. Therefore ignoring a constant, the empirical log-likelihood function is

$$l(\lambda, \alpha, \beta) \propto \sum_{j=1}^n \log [(1 - \lambda) + \lambda e^{\alpha + \beta Y_j}] - \sum_{i=1}^{m+n} \log [1 + \rho_N \lambda (e^{\alpha + \beta T_i} - 1)]. \quad (7)$$

Let $\hat{\theta}_{\text{MELE}} = (\hat{\lambda}_{\text{MELE}}, \hat{\alpha}_{\text{MELE}}, \hat{\beta}_{\text{MELE}})^{\top}$ denote the MELE of θ , i.e., the maximizer of the log-likelihood function l in (7). In our numerical studies, we used `Optim` in the R-package to find $\hat{\theta}_{\text{MELE}}$. Then the MELE of $p(y)$ in (2) is

$$\hat{p}_{\text{MELE}}(y) = \frac{\hat{\lambda}_{\text{MELE}} \exp[\hat{\alpha}_{\text{MELE}} + \hat{\beta}_{\text{MELE}}y]}{(1 - \hat{\lambda}_{\text{MELE}}) + \hat{\lambda}_{\text{MELE}} \exp[\hat{\alpha}_{\text{MELE}} + \hat{\beta}_{\text{MELE}}y]}. \quad (8)$$

We now examine the asymptotic properties of the MELE $\hat{\theta}_{\text{MELE}}$. Define

$$S_N = -\frac{1}{N} \begin{pmatrix} \frac{\partial^2 l}{\partial \lambda^2} & \frac{\partial^2 l}{\partial \lambda \partial \alpha} & \frac{\partial^2 l}{\partial \lambda \partial \beta} \\ \frac{\partial^2 l}{\partial \alpha \partial \lambda} & \frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \beta} \\ \frac{\partial^2 l}{\partial \beta \partial \lambda} & \frac{\partial^2 l}{\partial \alpha \partial \beta} & \frac{\partial^2 l}{\partial \beta^2} \end{pmatrix} \quad (9)$$

with l given in (7), and let

$$S = \int \left(\frac{\partial w_1(y)}{\partial \theta} \right) \left(\frac{\partial w_1(y)}{\partial \theta} \right)^{\top} \frac{f}{w_1 w_2}(y) dy = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{12} & S_{22} & S_{23} \\ S_{13} & S_{23} & S_{33} \end{pmatrix}, \quad (10)$$

where

$$\begin{aligned} S_{11} &= \int (e^{\alpha+\beta y} - 1)^2 \cdot \frac{f}{w_1 w_2}(y) dy, \\ S_{22} &= \lambda^2 \int e^{2\alpha+2\beta y} \cdot \frac{f}{w_1 w_2}(y) dy, \\ S_{33} &= \lambda^2 \int y^2 e^{2\alpha+2\beta y} \cdot \frac{f}{w_1 w_2}(y) dy, \\ S_{12} &= \lambda \int e^{\alpha+\beta y} (e^{\alpha+\beta y} - 1) \cdot \frac{f}{w_1 w_2}(y) dy, \\ S_{13} &= \lambda \int y e^{\alpha+\beta y} (e^{\alpha+\beta y} - 1) \cdot \frac{f}{w_1 w_2}(y) dy, \\ S_{23} &= \lambda^2 \int y e^{2\alpha+2\beta y} \cdot \frac{f}{w_1 w_2}(y) dy \end{aligned}$$

with

$$w_1(y) = 1 - \lambda + \lambda e^{\alpha+\beta y} \quad \text{and} \quad w_2(y) = 1 - \rho\lambda + \rho\lambda e^{\alpha+\beta y}. \quad (11)$$

Since $f \in \mathcal{F}$ and $r(x) = x$, implying $\int e^{\beta y} f(y) dy < \infty$, it is easy to show that all the S_{ij} 's defined above are finite.

Lemma 1 Assume $\rho_N \rightarrow \rho$ as $N \rightarrow \infty$, with $0 < \rho < 1$. Then $S_N \xrightarrow{P} \rho(1 - \rho)S$ as $N \rightarrow \infty$, where S_N and S are defined in (9) and (10) respectively and \xrightarrow{P} denotes convergence in probability.

The condition $\rho_N \rightarrow \rho$ with $0 < \rho < 1$, as $N \rightarrow \infty$, in Lemma 1 requires that the sample sizes m and n converge to infinity at the same order. With the results in Lemma 1, the following theorem gives the asymptotic normality of the MELE $\hat{\theta}_{\text{MELE}}$. Let

$$\begin{aligned} V &= \int \left(\frac{\partial w_1(y)}{\partial \theta} \right) \left(\frac{\partial w_1(y)}{\partial \theta} \right)^T \frac{f}{w_1 w_2}(y) dy - \int \frac{\partial w_1(y)}{\partial \theta} \frac{f}{w_2}(y) dy \int \left(\frac{\partial w_1(y)}{\partial \theta} \right)^T \frac{f}{w_2}(y) dy \\ &= S - \int \frac{\partial w_1(y)}{\partial \theta} \frac{f}{w_2}(y) dy \int \left(\frac{\partial w_1(y)}{\partial \theta} \right)^T \frac{f}{w_2}(y) dy \\ &= \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{12} & V_{22} & V_{23} \\ V_{13} & V_{23} & V_{33} \end{pmatrix}, \end{aligned} \quad (12)$$

where

$$\begin{aligned} V_{11} &= \int (e^{\alpha+\beta y} - 1)^2 \frac{f}{w_1 w_2}(y) dy - \left[\int (e^{\alpha+\beta y} - 1) \frac{f}{w_2}(y) dy \right]^2, \\ V_{22} &= \lambda^2 \int e^{2\alpha+2\beta y} \frac{f}{w_1 w_2}(y) dy - \lambda^2 \left[\int e^{\alpha+\beta y} \frac{f}{w_2}(y) dy \right]^2, \\ V_{33} &= \lambda^2 \int y^2 e^{2\alpha+2\beta y} \frac{f}{w_1 w_2}(y) dy - \lambda^2 \left[\int y e^{\alpha+\beta y} \frac{f}{w_2}(y) dy \right]^2, \\ V_{12} &= \lambda \int e^{\alpha+\beta y} (e^{\alpha+\beta y} - 1) \frac{f}{w_1 w_2}(y) dy - \lambda \int (e^{\alpha+\beta y} - 1) \frac{f}{w_2}(y) dy \int e^{\alpha+\beta y} \frac{f}{w_2}(y) dy, \\ V_{13} &= \lambda \int y e^{\alpha+\beta y} (e^{\alpha+\beta y} - 1) \frac{f}{w_1 w_2}(y) dy - \lambda \int (e^{\alpha+\beta y} - 1) \frac{f}{w_2}(y) dy \int y e^{\alpha+\beta y} \frac{f}{w_2}(y) dy, \\ V_{23} &= \lambda^2 \int y e^{2\alpha+2\beta y} \frac{f}{w_1 w_2}(y) dy - \lambda^2 \int e^{\alpha+\beta y} \frac{f}{w_2}(y) dy \int y e^{\alpha+\beta y} \frac{f}{w_2}(y) dy, \end{aligned}$$

with w_1 and w_2 defined in (11).

Theorem 2 Assume $\rho_N \rightarrow \rho$ as $N \rightarrow \infty$ with $0 < \rho < 1$ and S defined in (10) is invertible. Then under model (4) with parameter space (5) and some regularity conditions (for MELE in general; see Qin and Lawless, 1994),

$$\sqrt{N} \begin{pmatrix} \hat{\lambda}_{\text{MELE}} - \lambda \\ \hat{\alpha}_{\text{MELE}} - \alpha \\ \hat{\beta}_{\text{MELE}} - \beta \end{pmatrix} \xrightarrow{D} N(0, \Sigma),$$

where $\Sigma = \frac{1}{\rho(1-\rho)} S^{-1} V S^{-1}$ with S and V defined in (10) and (12) respectively, and \xrightarrow{D} denotes convergence in distribution. In addition, V and further Σ are positive definite.

In Theorem 2, the assumption of S being invertible is generally true except for some special cases such as $\lambda = 0$, or $\alpha = \beta = 0$, or f is a singleton. When $0 < \lambda < 1$, $\beta \neq 0$ and f is continuous (or an even weaker condition that the support of f contains at least

two points), we would expect S being invertible. With the results in Theorem 2, one can easily make MELE-based inferences, such as Wald test, about the parameter when the unknown θ in Σ is replaced with some consistent estimators such as the MELE.

4 MHDE of the parameters

Though MELE is efficient, it is generally non-robust against outliers and model misspecifications. As a robust alternative, we propose in this section an MHDE for the unknown parameters in model (4).

The Hellinger distance between two functions f_1 and f_2 is defined as $\|f_1^{1/2} - f_2^{1/2}\|$, the L^2 -norm of root functions. For a fully parametric model $\{h_\theta : \theta \in \Theta\}$ with Θ the parameter space, the MHDE of θ is defined as

$$\hat{\theta}_{\text{MHDE}} = \arg \min_{t \in \Theta} \|h_t^{1/2} - \hat{h}^{1/2}\|, \quad (13)$$

where \hat{h} is an appropriate nonparametric estimator of h_θ . MHDE was first introduced by Beran (1977) for fully parametric models. Beran (1977) showed that the MHDE for parametric model has both full efficiency and good robustness properties. Lindsay (1994) outlined the comparison between MHDE and MLE in terms of robustness and efficiency and showed that MHDE and MLE are members of a larger class of efficient estimators with various second-order efficiency properties. However, the literature on MHDE for mixture models is not redundant. Lu et al. (2003) considered the MHDE for mixture of Poisson regression models. MHDE of mixture complexity for finite mixture models was investigated by Woo and Sriram (2006, 2007). Recently, MHDE has been extended from parametric models to semiparametric models. Wu et al. (2010) proposed an MHDE for two-sample case-control data under model (3) and investigated the asymptotic properties and robustness of the proposed estimator. Xiang et al. (2014) proposed a minimum profile Hellinger distance estimator (MPHDE) for the two-component semiparametric mixture model studied by Bordes et al. (2006) where one component is known and the other is an unknown symmetric function with unknown location parameter. Wu et al. (2017) and Wu and Zhou (2018) proposed algorithms for calculating the MPHDE and the MHDE, respectively, for the two-component semiparametric location-shifted mixture model. Inspired by these works, we propose to use the MHDE to estimate the parameters in (4).

In model (4), even though α and β can possibly take any value on the real line, we can essentially use intervals that are large enough to cover their true values. So for practical purpose, we can assume that $\theta \in \Theta$ with Θ a compact subset of \mathbb{R}^3 . To give the MHDE for model (4), note that the MHDE defined in (13) is not available in practice since the f in $h_t(x) = (1 - t_1 + t_1 e^{t_2 + t_3 x})f(x)$, with $t = (t_1, t_2, t_3)^\top$, is unknown. Intuitively, we can use an appropriate nonparametric density estimator, say f_m , based on X_1, \dots, X_m to replace f and apply the plug-in rule to estimate h_t by

$$\hat{h}_t(x) = (1 - t_1 + t_1 e^{t_2 + t_3 x}) f_m(x). \quad (14)$$

With another appropriate nonparametric density estimator, say h_n , of h_θ based on Y_1, \dots, Y_n , we define the MHDE of $\theta = (\lambda, \alpha, \beta)^\top$ as

$$\hat{\theta}_{\text{MHDE}} = T(f_m, h_n) = \arg \min_{t \in \Theta} \left\| \hat{h}_t^{1/2} - h_n^{1/2} \right\|. \quad (15)$$

That is, $\hat{\theta}_{\text{MHDE}}$ is the minimizer t of the Hellinger distance between the estimated parametric model \hat{h}_t and the nonparametric density estimator h_n . In this work, we use kernel density estimators

$$\text{and } f_m(x) = \frac{1}{mb_m} \sum_{i=1}^m K_0 \left(\frac{x - X_i}{b_m} \right), \quad (16)$$

$$h_n(x) = \frac{1}{nb_n} \sum_{j=1}^n K_1 \left(\frac{x - Y_j}{b_n} \right), \quad (17)$$

where K_0 and K_1 are kernel p.d.f.s and bandwidths b_n and b_m are positive sequences such that $b_m \rightarrow 0$ as $m \rightarrow \infty$ and $b_n \rightarrow 0$ as $n \rightarrow \infty$. The MHDE of $p(y)$ in (2) is given by

$$\hat{p}_{\text{MHDE}}(y) = \frac{\hat{\lambda}_{\text{MHDE}} \exp[\hat{\alpha}_{\text{MHDE}} + \hat{\beta}_{\text{MHDE}} y]}{(1 - \hat{\lambda}_{\text{MHDE}}) + \hat{\lambda}_{\text{MHDE}} \exp[\hat{\alpha}_{\text{MHDE}} + \hat{\beta}_{\text{MHDE}} y]}. \quad (18)$$

Note that in (15) we do not impose any restriction on \hat{h}_t to make it a density function. The reason behind this is that a $t \in \Theta$ that makes \hat{h}_t not a density can still make h_t a density. The true parameter value θ may not make \hat{h}_θ a density, but it is not reasonable to exclude θ as the estimate $\hat{\theta}_{\text{MHDE}}$ of itself. As the explicit expression of $\hat{\theta}_{\text{MHDE}}$ does not exist, one needs to use iterative methods such as Newton-Raphson to numerically calculate it. Karunamuni and Wu (2011) has shown for parametric models that with an appropriate initial value, even one-step iteration works well and gives a quite accurate approximation of MHDE.

We now examine the asymptotic properties of the MHDE $\hat{\theta}_{\text{MHDE}}$ given in (15). Let \mathcal{H} be the set of all p.d.f.s with respect to Lebesgue measure on the real line. The conditions (D1)–(D3) and (C0)–(C11) are listed in Appendix. The following theorem gives the existence, Hellinger-metric consistency, and uniqueness of the proposed MHDE.

Theorem 3 Suppose Θ is a compact parameter space and (D1) holds for all $t \in \Theta$. Then

- (i) For every $\varphi \in \mathcal{H}$ and f_m defined in (16) with K_0 compactly supported, there exist $T(f, \varphi)$ and $T(f_m, \varphi)$ in Θ satisfying (15).

- (ii) Suppose that $m, n \rightarrow \infty$ as $N \rightarrow \infty$ and $\theta = T(f, \varphi)$ is unique for $\varphi \in \mathcal{H}$. Then $\theta_N = T(f_m, \varphi_n) \rightarrow \theta$ as $N \rightarrow \infty$ for any density sequences f_m and φ_n such that $\|\varphi_n^{1/2} - \varphi^{1/2}\| \rightarrow 0$ and $\sup_{t \in \Theta} \|\hat{h}_t^{1/2} - h_t^{1/2}\| \rightarrow 0$ as $N \rightarrow \infty$ with \hat{h}_t given in (14).
- (iii) $T(f, h_\theta) = \theta$ uniquely for any $\theta \in \Theta$.

With the results given in Theorem 3, the consistency of our proposed MHDE can be proved and is presented in the following theorem.

Theorem 4 Let $m, n \rightarrow \infty$ as $N \rightarrow \infty$. Suppose that (D1) holds for any $\theta \in \Theta$ and the bandwidths b_m and b_n in (16) and (17) respectively satisfy $b_m, b_n \rightarrow 0$ and $mb_m, nb_n \rightarrow \infty$ as $N \rightarrow \infty$. Further suppose that either (D2) or (D3) holds. Then $\|f_m^{1/2} - f^{1/2}\| \rightarrow 0$, $\|h_n^{1/2} - h_\theta^{1/2}\| \rightarrow 0$ and $\sup_{t \in \Theta} \|\hat{h}_t^{1/2} - h_t^{1/2}\| \rightarrow 0$ as $N \rightarrow \infty$. Furthermore, $\hat{\theta}_{\text{MHDE}} \rightarrow \theta$ as $N \rightarrow \infty$, where $\hat{\theta}_{\text{MHDE}}$ is defined in (15) with f_m, h_n and \hat{h}_t given in (16), (17) and (14), respectively.

We derive in the next theorem an expression of the bias term $\hat{\theta}_{\text{MHDE}} - \theta$. Define symmetric matrix

$$\Delta(\theta) = \int \left(\frac{\partial w_1(x)}{\partial \theta} \right) \left(\frac{\partial w_1(x)}{\partial \theta} \right)^\top \frac{f}{w_1}(x) dx = \begin{pmatrix} \Delta_{11}(\theta) & \Delta_{12}(\theta) & \Delta_{13}(\theta) \\ \Delta_{12}(\theta) & \Delta_{22}(\theta) & \Delta_{23}(\theta) \\ \Delta_{13}(\theta) & \Delta_{23}(\theta) & \Delta_{33}(\theta) \end{pmatrix}, \quad (19)$$

where

$$\begin{aligned} \Delta_{11}(\theta) &= \int \left(\frac{\partial w_1(x)}{\partial \lambda} \right)^2 \frac{f}{w_1}(x) dx = \int (e^{\alpha+\beta x} - 1)^2 \frac{f}{w_1}(x) dx, \\ \Delta_{22}(\theta) &= \int \left(\frac{\partial w_1(x)}{\partial \alpha} \right)^2 \frac{f}{w_1}(x) dx = \lambda^2 \int e^{2\alpha+2\beta x} \frac{f}{w_1}(x) dx, \\ \Delta_{33}(\theta) &= \int \left(\frac{\partial w_1(x)}{\partial \beta} \right)^2 \frac{f}{w_1}(x) dx = \lambda^2 \int x^2 e^{2\alpha+2\beta x} \frac{f}{w_1}(x) dx, \\ \Delta_{12}(\theta) &= \int \frac{\partial w_1(x)}{\partial \lambda} \frac{\partial w_1(x)}{\partial \alpha} \frac{f}{w_1}(x) dx = \lambda \int e^{\alpha+\beta x} (e^{\alpha+\beta x} - 1) \frac{f}{w_1}(x) dx, \\ \Delta_{13}(\theta) &= \int \frac{\partial w_1(x)}{\partial \lambda} \frac{\partial w_1(x)}{\partial \beta} \frac{f}{w_1}(x) dx = \lambda \int x e^{\alpha+\beta x} (e^{\alpha+\beta x} - 1) \frac{f}{w_1}(x) dx, \\ \Delta_{23}(\theta) &= \int \frac{\partial w_1(x)}{\partial \alpha} \frac{\partial w_1(x)}{\partial \beta} \frac{f}{w_1}(x) dx = \lambda^2 \int x e^{2\alpha+2\beta x} \frac{f}{w_1}(x) dx. \end{aligned}$$

Let

$$A_N(\theta) = \int \frac{\partial w_1}{\partial \theta}(x) \left[\frac{f_m^{1/2} h_n^{1/2}}{w_1^{1/2}}(x) - f_m(x) \right] dx = \begin{pmatrix} A_{N1}(\theta) \\ A_{N2}(\theta) \\ A_{N3}(\theta) \end{pmatrix}, \quad (20)$$

where

$$\begin{aligned}
A_{N1}(\theta) &= \int (e^{\alpha+\beta x} - 1) \left[\frac{f_m^{1/2} h_n^{1/2}}{w_1^{1/2}}(x) - f_m(x) \right] dx, \\
A_{N2}(\theta) &= \int \lambda e^{\alpha+\beta x} \left[\frac{f_m^{1/2} h_n^{1/2}}{w_1^{1/2}}(x) - f_m(x) \right] dx, \\
A_{N3}(\theta) &= \int \lambda x e^{\alpha+\beta x} \left[\frac{f_m^{1/2} h_n^{1/2}}{w_1^{1/2}}(x) - f_m(x) \right] dx.
\end{aligned}$$

Theorem 5 Suppose $\theta \in \text{int}(\Theta)$, $\Delta(\theta)$ defined in (19) is invertible, (C2) and assumptions in Theorem 4 hold. Then it follows that

$$\hat{\theta}_{\text{MHDE}} - \theta = 2[\Delta^{-1}(\theta) + R_N]A_N(\theta), \quad (21)$$

where $\hat{\theta}_{\text{MHDE}}$ is defined by (15) and R_N is a 3×3 matrix with elements tending to zero in probability as $N \rightarrow \infty$.

From this result, we obtain immediately the asymptotic distribution of the MHDE.

Theorem 6 Suppose that $\hat{\theta}_{\text{MHDE}}$ defined in (15) satisfies (21). Further suppose that conditions (C0) – (C9) hold. Then the asymptotic distribution of $\sqrt{N}(\hat{\theta}_{\text{MHDE}} - \theta)$ is $N(0, \Sigma)$, where Σ is defined by

$$\begin{aligned}
\Sigma &= \Delta^{-1}(\theta) \left[\frac{1}{1-\rho} \bar{\Delta}(\theta) + \frac{1}{\rho} \Delta(\theta) \right] \Delta^{-1}(\theta) \\
&= \frac{1}{\rho(1-\rho)} \Delta^{-1}(\theta) [\Delta(\theta) + \rho(\bar{\Delta}(\theta) - \Delta(\theta))] \Delta^{-1}(\theta)
\end{aligned}$$

with

$$\begin{aligned}
\bar{\Delta}(\theta) &= \int \frac{\partial w_1}{\partial \theta}(x) \left[\frac{\partial w_1}{\partial \theta}(x) \right]^T f(x) dx, \\
\Delta(\theta) &= \int \frac{\partial w_1}{\partial \theta}(x) \left[\frac{\partial w_1}{\partial \theta}(x) \right]^T \frac{f}{w_1}(x) dx.
\end{aligned}$$

In Theorems 5 and 6, the assumption of $\Delta(\theta)$ being invertible is generally true except for some special cases such as $\lambda = 0$, or $\alpha = \beta = 0$, or f is a singleton.

Since MELE is a likelihood-based method, we expect MELE asymptotically more efficient than MHDE. Looking at the asymptotic covariance matrices given in Theorems 2 and 6 for MELE and MHDE respectively, we note that roughly speaking $|S_{ij}|$ is smaller than $|\Delta_{ij}|$ while $|\bar{\Delta}_{ij}|$ is much bigger than both, due to the fact that the denominator w_1 or w_2 can control the exponential terms in the integrands. As a result, the asymptotic variances of the MELEs of λ , α and β are smaller than those of the corresponding MHDEs.

To see the asymptotic relative efficiency of the MHDE with respect to the MELE, we take the normal mixture $(1-\lambda)N(0, 1) + \lambda N(\mu, 1)$, a model examined also in our later simulation studies, as an example and Table 1 below presents the

Table 1 Asymptotic covariance matrices of the MELE and MHDE under normal mixture $(1 - \lambda)N(0, 1) + \lambda N(\mu, 1)$ with $\rho = 0.5$

μ	λ	MELE	MHDE
$\mu = .5$ $\left(\begin{array}{l} \alpha = -.125 \\ \beta = .5 \end{array} \right)$.05	$\begin{bmatrix} 153 & 631 & -1258 \\ 631 & 2852 & -5705 \\ -1258 & -5705 & 11,416 \end{bmatrix}$	$\begin{bmatrix} 157 & 556 & -1260 \\ 556 & 4456 & -5714 \\ -1260 & -5714 & 11,434 \end{bmatrix}$
	.20	$\begin{bmatrix} 164 & 179 & -357 \\ 179 & 213 & -428 \\ -357 & -428 & 861 \end{bmatrix}$	$\begin{bmatrix} 169 & 167 & -363 \\ 166 & 317 & -435 \\ -363 & -435 & 876 \end{bmatrix}$
	.50	$\begin{bmatrix} 163 & 78 & -155 \\ 78 & 40 & -80 \\ -155 & -80 & 166 \end{bmatrix}$	$\begin{bmatrix} 174 & 79 & -167 \\ 79 & 58 & -86 \\ -167 & -86 & 178 \end{bmatrix}$
	.80	$\begin{bmatrix} 140 & 44 & -91 \\ 44 & 15 & -31 \\ -91 & -31 & 67 \end{bmatrix}$	$\begin{bmatrix} 159 & 50 & -105 \\ 50 & 24 & -36 \\ -105 & -36 & 77 \end{bmatrix}$
	.95	$\begin{bmatrix} 119 & 33 & -69 \\ 33 & 10 & -20 \\ -69 & -20 & 45 \end{bmatrix}$	$\begin{bmatrix} 142 & 39 & -83 \\ 39 & 16 & -25 \\ -83 & -25 & 54 \end{bmatrix}$
$\mu = 1$ $\left(\begin{array}{l} \alpha = -.5 \\ \beta = 1 \end{array} \right)$.05	$\begin{bmatrix} 14 & 171 & -169 \\ 171 & 2548 & -2559 \\ -169 & -2559 & 2573 \end{bmatrix}$	$\begin{bmatrix} 18 & 104 & -179 \\ 104 & 4298 & -2711 \\ -179 & -2711 & 2728 \end{bmatrix}$
	.20	$\begin{bmatrix} 17 & 63 & -63 \\ 63 & 285 & -292 \\ -63 & -292 & 302 \end{bmatrix}$	$\begin{bmatrix} 23 & 63 & -79 \\ 63 & 459 & -369 \\ -79 & -369 & 383 \end{bmatrix}$
	.50	$\begin{bmatrix} 16 & 29 & -30 \\ 29 & 65 & -69 \\ -30 & -69 & 77 \end{bmatrix}$	$\begin{bmatrix} 26 & 44 & -50 \\ 44 & 123 & -115 \\ -50 & -115 & 126 \end{bmatrix}$
	.80	$\begin{bmatrix} 11 & 15 & -16 \\ 15 & 24 & -27 \\ -16 & -27 & 33 \end{bmatrix}$	$\begin{bmatrix} 21 & 30 & -34 \\ 30 & 57 & -57 \\ -34 & -57 & 67 \end{bmatrix}$
	.95	$\begin{bmatrix} 7 & 8 & -10 \\ 8 & 14 & -16 \\ -10 & -16 & 21 \end{bmatrix}$	$\begin{bmatrix} 14 & 20 & -23 \\ 20 & 37 & -38 \\ -23 & -38 & 46 \end{bmatrix}$
$\mu = 2$ $\left(\begin{array}{l} \alpha = -2 \\ \beta = 2 \end{array} \right)$.05	$\begin{bmatrix} 2 & 78 & -40 \\ 78 & 4566 & -2422 \\ -40 & -2422 & 1296 \end{bmatrix}$	$\begin{bmatrix} 14 & 441 & -268 \\ 441 & 29,913 & -14,795 \\ -268 & -14,795 & 7744 \end{bmatrix}$
	.20	$\begin{bmatrix} 3 & 37 & -20 \\ 37 & 830 & -474 \\ -20 & -474 & 277 \end{bmatrix}$	$\begin{bmatrix} 34 & 561 & -309 \\ 561 & 11,232 & -5998 \\ -309 & -5998 & 3242 \end{bmatrix}$
	.50	$\begin{bmatrix} 3 & 18 & -10 \\ 18 & 235 & -146 \\ -10 & -146 & 96 \end{bmatrix}$	$\begin{bmatrix} 51 & 537 & -298 \\ 537 & 5969 & -3304 \\ -298 & -3304 & 1842 \end{bmatrix}$

Table 1 (continued)

μ	λ	MELE	MHDE
	.80	$\begin{bmatrix} 1 & 8 & -5 \\ 8 & 93 & -63 \\ -5 & -63 & 47 \end{bmatrix}$	$\begin{bmatrix} 37 & 367 & -209 \\ 367 & 3759 & -2140 \\ -209 & -2140 & 1229 \end{bmatrix}$
	.95	$\begin{bmatrix} 1 & 3 & -2 \\ 3 & 49 & -35 \\ -2 & -35 & 29 \end{bmatrix}$	$\begin{bmatrix} 14 & 191 & -112 \\ 191 & 2639 & -1540 \\ -112 & -1540 & 909 \end{bmatrix}$

asymptotic covariance matrices of the MELE and MHDE for $\rho = 0.5$ and various parameter settings $\mu = .5, 1, 2$ (i.e., $(\alpha, \beta) = (-.125, .5), (-.5, 1), (-2, 2)$ respectively) and $\lambda = .05, .20, .50, .80, .95$. From Table 1 we see that, as expected, the elements in the covariance matrices of the MELE are almost always smaller than those of the MHDE, in absolute values. The difference in the asymptotic variances of the MELE and MHDE of λ or β is very small for smaller μ ($= .5$) and smaller λ values. The relative efficiency of the MHDE with respect to the MELE deteriorates when λ increases or, especially when μ increases. We also observe from Table 1 that the asymptotic variances of both the MELE and MHDE of α and β decrease when λ increases, which is expected by the fact that the information of α and β is contained in the second component only and the amount of such information increases when the mixing proportion λ of the second component increases. Comparatively, when λ increases, the asymptotic variances of both the MELE and MHDE of λ increase first, reach their maxima around $\lambda = 0.5$ and then decrease. This phenomenon can be explained by the fact that the estimation of λ is associated with a Bernoulli random variable with parameter λ and the variance of this random variable, $\lambda(1 - \lambda)$, as a function of λ has the same increasing-decreasing trend.

5 Test of the semiparametric model

In this section we discuss the validity of the semiparametric mixture model (4) or equivalently the relationship (3) between f and g , with $r(x) = x$. Several goodness-of-fit test statistics for testing the relationship (3) in case-control studies are available in literature; see, for example, Qin and Zhang (1997), Zhang (1999, 2001, 2006), and Deng et al. (2009). For our semiparametric mixture model (4), we will construct Kolmogorov–Smirnov (K–S) type statistics based on the proposed MELE and MHDE.

The idea of K–S test statistic is to use the discrepancy between two c.d.f. estimates, one with the model assumption and the other without, to assess the validity of a model. For model (4) with $r(x) = x$, we can use the empirical c.d.f. based on the first sample X_i 's as the first estimate and the MELE or MHDE based on both samples X_i 's and Y_i 's exploiting (4) as the second. We first look at the special case of testing $\beta = 0$ in model (4). When $\beta = 0$, we also have $\alpha = 0$, which implies the equality of the two components F and G and further the equality of F and H . The two-sample K–S statistic for testing the equality of F and H is

$$\sup_i |\hat{F}(t) - \hat{H}(t)| = \frac{N}{n} \sup_i |\hat{F}(t) - \tilde{F}_0(t)|,$$

where $N = m + n$, \hat{F} , \hat{H} and \tilde{F}_0 denote the empirical c.d.f.s based on X_i 's, Y_i 's and T_i 's respectively, and $(T_1, \dots, T_N)^\top = (X_1, \dots, X_m, Y_1, \dots, Y_n)^\top$ is the pooled sample. Note that \hat{F} and \hat{H} are the nonparametric MLEs of F and H respectively without the assumption of $F = H$, whereas \tilde{F}_0 is the nonparametric MLE of F with the assumption of $F = H$. Now consider the general case of testing (4) with any β . Motivated by the construction of the K–S statistic for testing $\beta = 0$, to test the validity of model (4) with $r(x) = x$, we propose to use the test statistic

$$KS = N^{1/2} \sup_t |\hat{F}(t) - \tilde{F}(t)|, \quad (22)$$

where \tilde{F} is an estimator of F based on either the MELE or the MHDE of $\theta = (\lambda, \alpha, \beta)^\top$ with model assumption (4). Recall in Sect. 3, with $p_i = dF(T_i)$ and $\rho_N = n/N$, the MELE \hat{p}_i of p_i is given by (6). Then an estimator \tilde{F} of F under model (4) is given by

$$\tilde{F}(t) = \sum_{i=1}^N \hat{p}_i I(T_i \leq t) = \frac{1}{N} \sum_{i=1}^N \frac{I(T_i \leq t)}{1 - \rho_N \hat{\lambda} + \rho_N \hat{\lambda} e^{\hat{\alpha} + \hat{\beta} T_i}}. \quad (23)$$

If the $\hat{\theta} = (\hat{\lambda}, \hat{\alpha}, \hat{\beta})^\top$ in (6) and (23) is the MELE $\hat{\theta}_{MELE}$ that we constructed in Sect. 3, then the resulting \tilde{F}_{MELE} is the actual MELE of F under (4) and we denote the corresponding test statistic in (22) as KS_{MELE} . This test statistic is similar to that in Qin and Zhang (1997), but they used it for case-control data instead of our more complicated mixture model (4). Intuitively, we can also use the MHDE $\hat{\theta}_{MHDE}$ given in (15) for $\hat{\theta}$, and then we denote the resulting \tilde{F} in (23) and KS in (22) as \tilde{F}_{MHDE} and KS_{MHDE} , respectively.

In our later numerical studies, we use bootstrap procedure to find the approximated distributions and critical values for KS_{MELE} and KS_{MHDE} . To generate bootstrapping data, we randomly select independent samples X_i^{**} 's from $d\tilde{F}(x)$ and Y_i^{**} 's from $(1 - \hat{\lambda} + \hat{\lambda} e^{\hat{\alpha} + \hat{\beta} x})d\tilde{F}(x)$, where $\hat{\theta}$ and \tilde{F} are either the MELEs $\hat{\theta}_{MELE}$ and \tilde{F}_{MELE} or the MHDEs $\hat{\theta}_{MHDE}$ and \tilde{F}_{MHDE} , respectively. Note that both X_i^{**} 's and Y_i^{**} 's are selected from the pooled data (T_1, \dots, T_N) but with different probability distribution

function. This means that some of the selected X_i^{**} 's could be values in the original second sample Y_i 's and some of the selected Y_i^{**} 's could be values in the original first sample X_i 's. Let (T_1^*, \dots, T_N^*) denote the combined bootstrapping sample and $\hat{\theta}^* = (\hat{\lambda}^*, \hat{\alpha}^*, \hat{\beta}^*)^\top$ be either the MELE or the MHDE based on the bootstrapping samples X_i^{**} 's and Y_i^{**} 's. Then we can calculate the empirical function \hat{F} based on X_i^{**} 's, the quantities in (6) and the function in (23) based on T_i^{**} 's and $\hat{\theta}^*$, with results denoted by \hat{F}^* , \hat{p}_i^* and \tilde{F}^* , respectively. Then finally the bootstrapping KS test statistic is

$$KS^* = N^{1/2} \sup_t \left| \hat{F}^*(t) - \tilde{F}^*(t) \right|.$$

We can generate 1000 bootstrapping samples to produce 1000 bootstrapping KS statistic values for both KS_{MELE} and KS_{MHDE} at the same time. Then the distributions, and further the critical values, of KS_{MELE} and KS_{MHDE} can be estimated by these respective 1000 statistic values.

6 Simulation studies

In this section we examine through Monte Carlo simulation studies the finite-sample performance of the proposed MELE and MHDE of the parameters in model (4) and the proposed K–S tests of this model. In our simulation study, we consider mixture of normals, Poissons or uniforms given in Table 2. We can easily check that all the five models satisfy the stochastic dominance condition, however only models M1–M4 satisfy relationship (4). Model M5 does not satisfy (4) since the two components have different support. Even though the focus of this paper is on continuous mixture models, we also want to check the performance of the proposed methods for discrete mixture models such as M3 and M4. For each model, the true values of α and β are calculated and listed in Table 2.

6.1 Efficiency of the estimators

For each of the mixture models in Table 2, we consider different λ values varying from .05 to .95. Under each model, we take $N = 1000$ repeated random samplings

Table 2 Mixture models considered in simulation study

Model	Form	α	β	Note
M1	$(1 - \lambda)N(0, 1) + \lambda N(1, 1)$	-.5	1	Mixture of normals that are close
M2	$(1 - \lambda)N(0, 1) + \lambda N(5, 1)$	-12.5	5	Mixture of normals that are apart
M3	$(1 - \lambda)Po(2) + \lambda Po(4)$	-2	.693	Mixture of Poissons that are close
M4	$(1 - \lambda)Po(2) + \lambda Po(6)$	-4	1.099	Mixture of Poissons that are apart
M5	$(1 - \lambda)U(0, 4) + \lambda U(2, 6)$	NA	NA	Mixture of uniforms

and for each sampling we use two different sample sizes $(m, n) = (30, 30)$ and $(100, 100)$. In the kernel density estimators f_m and h_n given in (16) and (17) respectively, standard normal p.d.f. is used as both kernels K_0 and K_1 , and the bandwidths b_m and b_n are chosen the same as in Silverman (1986) and Wu and Abedin (2021). The effect of different choice of kernel on kernel estimator is trivial, while the bandwidth has more influence on the finite-sample performance and the convergence rate of the nonparametric kernel estimator. Nevertheless, Theorems 4–6 indicate that the MHDE has the consistency and asymptotic normality if the bandwidths are of the form $b_m, b_n = O(N^{-r})$ with r from a subinterval of $(0, 1)$. This subinterval may depend on the population of the two components. With the component population unknown, we choose $r = 1/5$ in our simulation as in Silverman (1986) and Wu and Abedin (2021) which achieves the optimal mean square error of the kernel estimator. In addition, the bandwidths b_m and b_n of our choice are adaptive in the sense that they involve a factor of robust scale estimate to incorporate the different variation of f and h .

To compute the estimates numerically, we use λ_+ , an estimator based on odds ratio, given by Smith and Vounatsou (1997) as the initial of λ . Initial values of α and β are calculated by exploiting the relationship (3), or equivalently

$$\log \frac{h(x)/f(x) - (1 - \lambda)}{\lambda} = \alpha + \beta x.$$

Thus for each T_i in the pooled sample, we generate the pair (T_i, R_i) , where $R_i = \log \frac{h_n(T_i)/f_m(T_i) - (1 - \lambda_+)}{\lambda_+}$. Then we use (T_i, R_i) , $i = 1, \dots, N$, to fit a least-squares regression line and the fitted coefficients are used as the initials of α and β . To evaluate the finite-sample performance of an estimator $\hat{\theta}$ of θ , we calculate the bias and mean squared error (MSE) over $N = 1000$ replications. The results are presented in Tables 3, 4 and 5.

Table 3 gives the biases and MSEs of $\hat{\lambda}_{\text{MELE}}$ and $\hat{\lambda}_{\text{MHDE}}$ for all the five models with varying λ values and sample size $m = n = 30$ whereas Table 4 is for $m = n = 100$. For the purpose of comparison, we also report the results of the two estimators $\hat{\lambda}$ and $\hat{\lambda}_L$ proposed in Wu and Abedin (2021) for model (1) directly, without assuming the relationship (3). The $\hat{\lambda}$ is based on cumulative distribution estimation while $\hat{\lambda}_L$ is the MLE of the multinomial approximation of model (1). From Tables 3 and 4 we can see that all the four methods are very competitive when λ is concerned. The nonparametric estimators $\hat{\lambda}$ and $\hat{\lambda}_L$ perform slightly better than the semiparametric estimators $\hat{\lambda}_{\text{MELE}}$ and $\hat{\lambda}_{\text{MHDE}}$ when the two components are close to each other (such as M1 and M3) or the relationship (3) does not holds (M5), while $\hat{\lambda}_{\text{MELE}}$ and $\hat{\lambda}_{\text{MHDE}}$ perform better than $\hat{\lambda}$ and $\hat{\lambda}_L$ when the two components are relatively far away (such as M2). Both the MELE and the MHDE perform surprisingly well for model M5, even though the assumed relationship (3) is not valid for M5.

The simulation results for estimating β are given in Table 5. Since (3) is not assumed in Wu and Abedin (2021), their methods are for estimating λ only and thus

Table 3 Bias (MSE) of $\hat{\lambda}_{\text{MELE}}$, $\hat{\lambda}_{\text{MHDE}}$, $\hat{\lambda}$ and $\hat{\lambda}_L$ with $m = n = 30$

Model	λ	$\hat{\lambda}_{\text{MELE}}$	$\hat{\lambda}_{\text{MHDE}}$	$\hat{\lambda}$	$\hat{\lambda}_L$
M1 $\alpha = -.5$ $\beta = 1$.05	.139 (.091)	.126 (.066)	.052 (.031)	.020 (.031)
	.20	.093 (.106)	.069 (.041)	.093 (.074)	.032 (.068)
	.50	.035 (.102)	-.052 (.087)	.067 (.078)	-.039 (.078)
	.80	-.024 (.056)	-.059 (.059)	-.009 (.047)	-.144 (.089)
	.95	-.062 (.032)	-.099 (.040)	-.052 (.029)	-.181 (.090)
M2 $\alpha = -12.5$ $\beta = 5$.05	.029 (.005)	.054 (.007)	.053 (.022)	.022 (.031)
	.20	.019 (.007)	.047 (.017)	.095 (.032)	.055 (.071)
	.50	.004 (.014)	.055 (.017)	.082 (.015)	.029 (.066)
	.80	-.008 (.014)	.009 (.009)	.061 (.014)	-.009 (.036)
	.95	.002 (.002)	.004 (.001)	.034 (.002)	-.011 (.009)
M3 $\alpha = -2$ $\beta = .693$.05	.058 (.055)	.142 (.054)	.203 (.089)	.002 (.015)
	.20	-.019 (.089)	.091 (.060)	.133 (.064)	-.017 (.046)
	.50	-.091 (.148)	.026 (.062)	.019 (.039)	-.076 (.010)
	.80	-.109 (.119)	-.059 (.043)	-.096 (.034)	-.128 (.101)
	.95	-.139 (.093)	-.116 (.039)	-.152 (.042)	-.163 (.102)
M4 $\alpha = -4$ $\beta = 1.099$.05	-.003 (.012)	.067 (.016)	.211 (.092)	.003 (.016)
	.20	-.084 (.037)	.027 (.024)	.178 (.071)	-.007 (.047)
	.50	-.092 (.087)	.016 (.021)	.117 (.034)	-.028 (.084)
	.80	-.027 (.051)	-.032 (.014)	.019 (.010)	-.033 (.059)
	.95	-.027 (.021)	-.054 (.009)	-.026 (.004)	-.048 (.037)
M5	.05	.166 (.177)	.237 (.190)	.249 (.136)	.026 (.032)
	.20	-.002 (.114)	.178 (.127)	.151 (.078)	.049 (.069)
	.50	-.121 (.135)	.076 (.063)	.067 (.048)	.025 (.065)
	.80	-.055 (.076)	.046 (.029)	.031 (.022)	-.001 (.029)
	.95	-.022 (.018)	.014 (.005)	.009 (.006)	-.012 (.010)

in Table 5 we compare $\hat{\beta}_{\text{MELE}}$ and $\hat{\beta}_{\text{MHDE}}$ only. In addition, we omit the results for α and model M5 as α depends on β through normalization and M5 doesn't satisfy (3). From Table 5 we observe that the MHDE performs better than the MELE in terms of both smaller bias and MSE. However, the MHDE tends to give very large bias and MSE when λ is small, such as M2 with $\lambda = .05$. We also observe from Table 5 that the MSEs of both MHDE and MELE of β decrease as the true λ value increases, which is very consistent with the observations from Table 1 on their asymptotic variances. This phenomenon is justified by the fact that when λ increases the expected number of observations from the second component g gets larger, and as a result the data contains more and more information about β . This is also the reason why the MHDE and the MELE give large biases and inflated MSEs for small λ , especially when sample size is small. For example, for model M2 with $\lambda = 0.05$ and $m = n = 30$ (or 100), we would expect only $n\lambda = 1.5$ (or 5) observations on average

Table 4 Bias (MSE) of $\hat{\lambda}_{\text{MELE}}$, $\hat{\lambda}_{\text{MHDE}}$, $\hat{\lambda}$ and $\hat{\lambda}_L$ with $m = n = 100$

Model	λ	$\hat{\lambda}_{\text{MELE}}$	$\hat{\lambda}_{\text{MHDE}}$	$\hat{\lambda}$	$\hat{\lambda}_L$
M1 $\alpha = -.5$ $\beta = 1$.05	.066 (.036)	.059 (.026)	.049 (.019)	.022 (.009)
	.20	.053 (.059)	.039 (.044)	.085 (.052)	.011 (.016)
	.50	.051 (.062)	.039 (.057)	.067 (.052)	-.048 (.026)
	.80	-.006 (.029)	-.018 (.031)	.008 (.026)	-.121 (.033)
	.95	-.041 (.014)	-.058 (.019)	-.027 (.012)	-.149 (.035)
M2 $\alpha = -12.5$ $\beta = 5$.05	.008 (.001)	.013 (.001)	.048 (.015)	.026 (.009)
	.20	-.003 (.004)	.001 (.006)	.065 (.018)	.033 (.016)
	.50	.001 (.005)	.029 (.004)	.049 (.005)	.017 (.019)
	.80	.002 (.002)	.005 (.002)	.057 (.009)	.000 (.010)
	.95	.001 (.001)	.004 (.001)	.036 (.002)	-.002 (.003)
M3 $\alpha = -2$ $\beta = .693$.05	.019 (.019)	.062 (.018)	.035 (.005)	-.000 (.004)
	.20	-.026 (.045)	.049 (.029)	.034 (.013)	-.022 (.013)
	.50	-.022 (.078)	.029 (.039)	-.015 (.001)	-.073 (.029)
	.80	-.023 (.036)	-.043 (.028)	-.084 (.190)	-.121 (.038)
	.95	-.057 (.018)	-.088 (.024)	-.119 (.023)	-.143 (.034)
M4 $\alpha = -4$ $\beta = 1.099$.05	-.024 (.004)	.003 (.004)	.044 (.006)	.000 (.004)
	.20	-.080 (.025)	-.005 (.012)	.071 (.012)	-.000 (.012)
	.50	-.018 (.025)	.002 (.011)	.048 (.009)	-.009 (.021)
	.80	.004 (.007)	-.028 (.006)	.009 (.006)	-.018 (.013)
	.95	-.007 (.002)	-.037 (.004)	-.006 (.002)	-.023 (.005)
M5	.05	.082 (.099)	.186 (.170)	.018 (.002)	.033 (.009)
	.20	-.093 (.040)	.093 (.056)	.017 (.005)	.035 (.014)
	.50	-.132 (.057)	.037 (.028)	.016 (.008)	.011 (.019)
	.80	-.048 (.032)	.063 (.012)	.012 (.006)	-.001 (.008)
	.95	-.006 (.005)	.033 (.002)	.012 (.002)	-.005 (.002)

from g that contain information about β , which thus produces estimates of β with large bias and MSE.

In order to examine the effect of unbalanced sample sizes on the performance of MELE and MHDE, in Table 6 we present the simulation results for sample sizes $(m, n) = (50, 150)$ and $(150, 50)$. Comparing Table 6 with Table 4, we observe that when λ is concerned, the MELE and MHDE under unbalanced sample sizes $(m, n) = (50, 150)$ and $(150, 50)$ perform slightly worse, though still comparable, than under balanced sample sizes $(m, n) = (100, 100)$, especially in terms of MSE. When β is concerned and Table 6 is compared with Table 5, we observe that the performance of the MELE and MHDE under unbalanced sample sizes is significantly worse than that under balanced sample sizes, especially in terms of MSE. These phenomenon can be possibly explained by the fact that both the asymptotic covariance matrices of MELE and MHDE given in Theorems 2 and 6 respectively have a factor $\frac{1}{\rho(1-\rho)}$ in front. Other places in the covariance matrix expressions also involve ρ , though not as influencing as this factor, which

Table 5 Bias (MSE) of $\hat{\beta}_{\text{MELE}}$ and $\hat{\beta}_{\text{MHDE}}$

Model	λ	$m = n = 30$		$m = n = 100$	
		$\hat{\beta}_{\text{MELE}}$	$\hat{\beta}_{\text{MHDE}}$	$\hat{\beta}_{\text{MELE}}$	$\hat{\beta}_{\text{MHDE}}$
M1 $\alpha = -.5$ $\beta = 1$.05	.006 (.436)	-.117 (.359)	.025 (.427)	-.127 (.349)
	.20	.174 (.469)	.023 (.387)	.174 (.407)	.034 (.342)
	.50	.225 (.376)	.110 (.328)	.094 (.198)	-.007 (.193)
	.80	.200 (.245)	.062 (.223)	.095 (.113)	-.011 (.121)
	.95	.219 (.205)	.059 (.119)	.116 (.082)	-.002 (.092)
M2 $\alpha = -12.5$ $\beta = 5$.05	.255 (1.125)	-1.504 (5.266)	.594 (.654)	-.789 (3.278)
	.20	.581 (.648)	.456 (1.009)	.564 (.560)	.534 (.598)
	.50	.564 (.594)	.632 (.641)	.411 (.441)	.535 (.519)
	.80	.500 (.532)	.552 (.592)	.316 (.386)	.463 (.434)
	.95	.428 (.471)	.468 (.599)	.255 (.331)	.423 (.385)
M3 $\alpha = -2$ $\beta = .693$.05	.063 (.641)	-.299 (.265)	.078 (.536)	-.229 (.219)
	.20	.427 (.734)	-.094 (.168)	.399 (.529)	-.025 (.104)
	.50	.504 (.599)	.044 (.097)	.246 (.296)	.016 (.068)
	.80	.335 (.331)	.072 (.077)	.119 (.091)	.035 (.052)
	.95	.301 (.241)	.085 (.068)	.130 (.055)	.034 (.043)
M4 $\alpha = -4$ $\beta = 1.099$.05	.029 (.705)	-.242 (.610)	.201 (.565)	-.155 (.579)
	.20	.512 (.492)	.059 (.292)	.450 (.389)	.093 (.187)
	.50	.337 (.271)	.098 (.085)	.115 (.097)	.042 (.066)
	.80	.144 (.109)	.090 (.073)	.053 (.039)	.039 (.049)
	.95	.131 (.075)	.089 (.070)	.066 (.033)	.033 (.048)

makes the level of deterioration different for λ and β when $m/(m+n)$ deviates away from 1/2 (i.e., balanced samples).

To examine the performance of \hat{p}_{MELE} and \hat{p}_{MHDE} given in (8) and (18) respectively, we calculate misclassification rates (MR) based on these estimates and a simple classification rule with use of .5 as the hard threshold, i.e., an individual with observation y is classified as from G if $\hat{p}(y) > .5$ and as from F if otherwise, where \hat{p} is either \hat{p}_{MELE} or \hat{p}_{MHDE} . Considering the fact that MR is higher for some models, such as M1 when the two components significantly overlap, than some other models such as M2, we use the optimal misclassification rate (OMR) in Wu and Abedin (2021) as the baseline to compare with. The OMR is the misclassification rate calculated assuming $p(y)$ is completely known (ideal scenario) and the same value .5 is used as the classification threshold. The simulation results are presented in Table 7. For the purpose of comparison, we also report the results of this classification rule but based on the nonparametric estimators $\hat{\lambda}$ and $\hat{\lambda}_L$ in Wu and Abedin (2021).

From Table 7 we observe that $\hat{\lambda}_L$ performs worst, while other three methods are quite competitive. Particularly, $\hat{\lambda}_{\text{MELE}}$ and $\hat{\lambda}_{\text{MHDE}}$ performs much better than $\hat{\lambda}$ in model M2, while $\hat{\lambda}$ has a little better performance in M1, M3 and M5. Even though the assumed relationship (3) is not valid for M5, the classification results based on

Table 6 Bias (MSE) of $\hat{\lambda}_{\text{MLE}}$, $\hat{\beta}_{\text{MLE}}$, $\hat{\lambda}_{\text{MHDE}}$ and $\hat{\beta}_{\text{MHDE}}$ for unbalanced sample sizes

Model	λ	$(m, n) = (50, 150)$				$(m, n) = (150, 50)$			
		$\hat{\lambda}_{\text{MLE}}$	$\hat{\beta}_{\text{MLE}}$	$\hat{\lambda}_{\text{MHDE}}$	$\hat{\beta}_{\text{MHDE}}$	$\hat{\lambda}_{\text{MLE}}$	$\hat{\beta}_{\text{MLE}}$	$\hat{\lambda}_{\text{MHDE}}$	$\hat{\beta}_{\text{MHDE}}$
M1	.05	-.049 (.002)	2.445 (5.981)	.033 (.006)	-.163 (.634)	.154 (.098)	-.440 (.694)	.039 (.008)	-.242 (.553)
	.20	.121 (.098)	.073 (.595)	-.048 (.014)	.229 (.460)	-.098 (.061)	-.676 (.852)	.038 (.015)	.146 (.363)
	.50	.078 (.082)	.211 (.414)	-.076 (.028)	.157 (.200)	.084 (.088)	.169 (.353)	-.070 (.034)	.138 (.145)
	.80	-.016 (.033)	.189 (.235)	-.098 (.026)	.105 (.123)	-.009 (.041)	.154 (.198)	-.104 (.037)	.101 (.076)
	.95	-.050 (.017)	.184 (.163)	-.101 (.022)	.088 (.110)	-.063 (.025)	.167 (.145)	-.122 (.034)	.085 (.063)
M2	.05	.001 (.000)	.773 (.839)	-.013 (.003)	.013 (1.833)	.001 (.001)	.250 (1.189)	-.002 (.003)	-.786 (3.547)
	.20	-.001 (.001)	.803 (.761)	-.063 (.017)	.492 (1.119)	-.009 (.005)	.568 (.514)	-.059 (.020)	.333 (1.269)
	.50	-.007 (.006)	.725 (.690)	-.020 (.025)	.735 (.785)	-.008 (.011)	.489 (.446)	-.055 (.051)	.598 (.926)
	.80	-.007 (.008)	.640 (.600)	.005 (.001)	.727 (.710)	-.003 (.005)	.413 (.412)	-.010 (.019)	.636 (.676)
	.95	-.011 (.012)	.518 (.519)	.003 (.000)	.715 (.690)	-.005 (.006)	.384 (.376)	-.002 (.007)	.164 (.615)
M3	.05	.139 (.148)	.137 (.740)	.093 (.061)	.037 (.558)	.198 (.202)	-.108 (.567)	-.028 (.005)	-.585 (.879)
	.20	.047 (.124)	.481 (.740)	.055 (.099)	.286 (.566)	.092 (.154)	.289 (.579)	.083 (.101)	.171 (.451)
	.50	-.013 (.099)	.310 (.368)	-.005 (.110)	.194 (.378)	.008 (.114)	.216 (.283)	.014 (.094)	.132 (.254)
	.80	-.028 (.041)	.175 (.157)	-.072 (.067)	.098 (.180)	-.028 (.048)	.122 (.109)	-.111 (.094)	.153 (.231)
	.95	-.067 (.022)	.173 (.112)	-.098 (.040)	.063 (.094)	-.063 (.023)	.119 (.070)	-.152 (.091)	.133 (.175)
M4	.05	.064 (.081)	.122 (.658)	-.001 (.007)	.423 (.766)	.117 (.129)	-.127 (.676)	-.020 (.005)	-.585 (.879)
	.20	-.051 (.053)	.410 (.421)	-.048 (.024)	.463 (.624)	-.043 (.058)	.305 (.385)	-.005 (.026)	.226 (.428)
	.50	-.032 (.059)	.154 (.192)	-.050 (.039)	.266 (.351)	-.052 (.078)	.149 (.189)	.010 (.031)	.088 (.189)
	.80	.004 (.017)	.040 (.071)	-.046 (.019)	.169 (.177)	-.001 (.023)	.042 (.056)	-.027 (.019)	.066 (.112)
	.95	-.005 (.004)	.061 (.044)	-.052 (.009)	.148 (.142)	-.007 (.005)	.053 (.033)	-.042 (.007)	.052 (.084)

Table 7 Misclassification rate (MR, in %) of $p(y)$ in (2) with threshold .5 based on $\hat{\theta}_{\text{MELE}}$, $\hat{\theta}_{\text{MHDE}}$, $\hat{\lambda}$ and $\hat{\lambda}_L$

Model	λ	$m = n = 30$				$m = n = 100$				OMR
		$\hat{\theta}_{\text{MELE}}$	$\hat{\theta}_{\text{MHDE}}$	$\hat{\lambda}$	$\hat{\lambda}_L$	$\hat{\theta}_{\text{MELE}}$	$\hat{\theta}_{\text{MHDE}}$	$\hat{\lambda}$	$\hat{\lambda}_L$	
M1 $\alpha = -.5$ $\beta = 1$.05	18.53	14.90	8.03	5.60	11.04	8.81	7.37	5.00	4.99
	.20	31.20	28.13	26.30	22.13	27.04	24.97	26.09	20.16	18.61
	.50	38.04	37.93	36.63	46.77	37.29	37.98	37.05	44.37	30.85
	.80	24.50	24.03	22.83	41.57	22.23	23.04	21.29	27.39	18.61
	.95	10.53	12.40	9.07	25.17	8.14	9.04	6.80	11.73	4.99
M2 $\alpha = -12.5$ $\beta = 5$.05	1.13	2.37	2.20	4.97	.69	1.67	2.28	5.00	.24
	.20	1.24	1.70	3.97	20.57	.90	.97	3.02	19.90	.48
	.50	1.27	1.33	2.73	39.13	.91	.89	2.12	25.61	.62
	.80	.53	1.03	6.93	19.97	.61	.61	6.21	4.56	.48
	.95	.37	.63	7.40	4.53	.29	.29	4.16	2.21	.24
M3 $\alpha = -2$ $\beta = .693$.05	13.63	12.27	28.40	5.40	7.61	6.89	5.04	5.01	4.76
	.20	25.47	25.00	32.27	22.03	21.75	20.73	16.69	20.09	13.90
	.50	36.13	34.50	34.13	43.67	34.89	34.06	15.54	44.03	19.05
	.80	26.97	25.00	24.17	35.53	22.87	23.68	25.20	29.74	13.35
	.95	15.73	14.73	14.40	22.57	9.55	11.27	14.24	14.27	6.06
M4 $\alpha = -4$ $\beta = 1.099$.05	6.70	6.30	26.30	5.43	4.41	4.51	3.03	5.01	3.14
	.20	15.53	12.30	25.27	21.60	11.14	10.87	9.61	20.08	7.03
	.50	19.00	16.97	21.60	37.97	16.49	15.85	15.54	35.49	10.19
	.80	14.23	13.17	13.77	24.20	13.00	12.94	13.34	14.68	7.82
	.95	6.07	7.03	5.73	10.03	5.18	5.96	5.81	5.41	3.21
M5	.05	25.97	25.53	30.83	22.07	15.31	22.18	3.09	5.01	2.50
	.20	23.04	30.87	29.90	20.80	16.59	21.61	10.99	20.13	10.00
	.50	29.83	31.90	30.43	45.10	26.38	28.77	24.24	37.34	25.00
	.80	22.47	18.70	16.20	23.07	20.16	15.77	12.76	13.38	10.00
	.95	8.23	5.13	4.33	5.70	5.92	4.84	3.87	3.62	2.50

MELE and MHDE for M5 are surprisingly reasonable and the MRs don't deviate from OMR too much for larger sample sizes.

6.2 Robustness of the estimators

In this subsection, we examine the robustness properties of the proposed MELE and MHDE. In order to compare with the two nonparametric estimators in Wu and

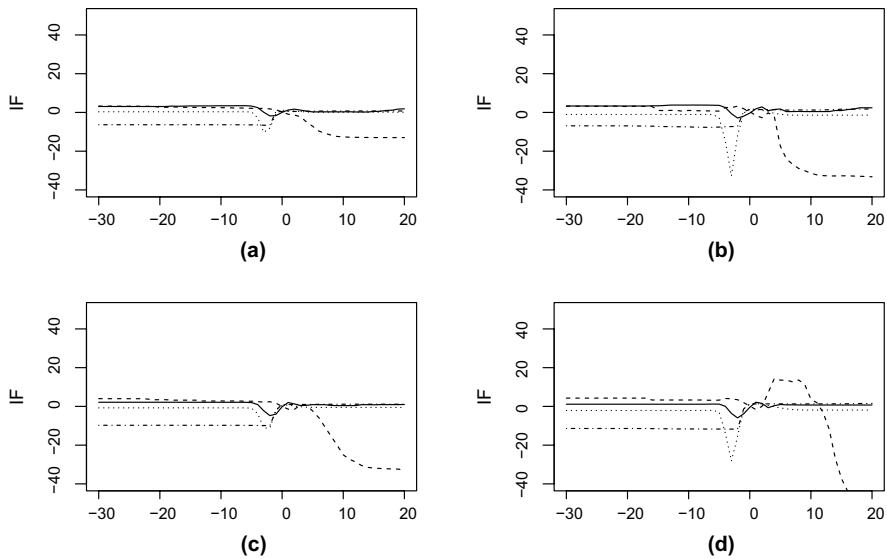


Fig. 1 The α -IFs of $\hat{\lambda}_{\text{MELE}}$ (dashed), $\hat{\lambda}_{\text{MHDE}}$ (solid), $\hat{\lambda}$ (dotted) and $\hat{\lambda}_L$ (dot-dashed) for model M1: **a** $\lambda = .15$ and $m = n = 30$; **b** $\lambda = .15$ and $m = n = 100$; **c** $\lambda = .55$ and $m = n = 30$; **d** $\lambda = .55$ and $m = n = 100$

Abedin (2021), we check the performance of estimators of λ only but not α and β that Wu and Abedin (2021) didn't introduce.

Specifically, we examine the behavior of the four estimators, $\hat{\lambda}_{\text{MELE}}$, $\hat{\lambda}_{\text{MHDE}}$, $\hat{\lambda}$ and $\hat{\lambda}_L$ when data are contaminated by a single outlying observation. Presence of several outliers will be similar and thus omitted here. Note that the outlying observation can be in either the first sample from $f(x)$ or the second sample from the mixture $h(x)$. We report results for the latter case and similar results are observed for the former case. We look at the change in estimate before and after data contamination. For this purpose, the α -influence function (α -IF) given in Beran (1977) is an appropriate measure of the change in estimate. However its application in mixture context is very difficult, as discussed in Karlis and Xekalaki (1998). Therefore, we employ an adaptive version of the α -IF as in Lu et al. (2003) which uses the change in estimate, before and after outlying observations are included, divided by contamination rate. Taking model M1 as an example, after drawing two independent samples with one from $N(0, 1)$ and the other from $(1 - \lambda)N(0, 1) + \lambda N(1, 1)$, we replace the last observation generated from the mixture with a single outlier, an integer from the interval $[-30, 20]$. Then the α -IF is given by

$$IF(x) = \frac{W((X_i)_{i=1}^m, (x, Y_i)_{i=1}^{n-1}) - W((X_i)_{i=1}^m, (Y_i)_{i=1}^n)}{1/N}$$

averaged over 100 replications, where the total sample size $N = 60$ or 200 and W is an estimator of λ . In our study, W is either $\hat{\lambda}_{\text{MELE}}$, $\hat{\lambda}_{\text{MHDE}}$, $\hat{\lambda}$ or $\hat{\lambda}_L$. The results are

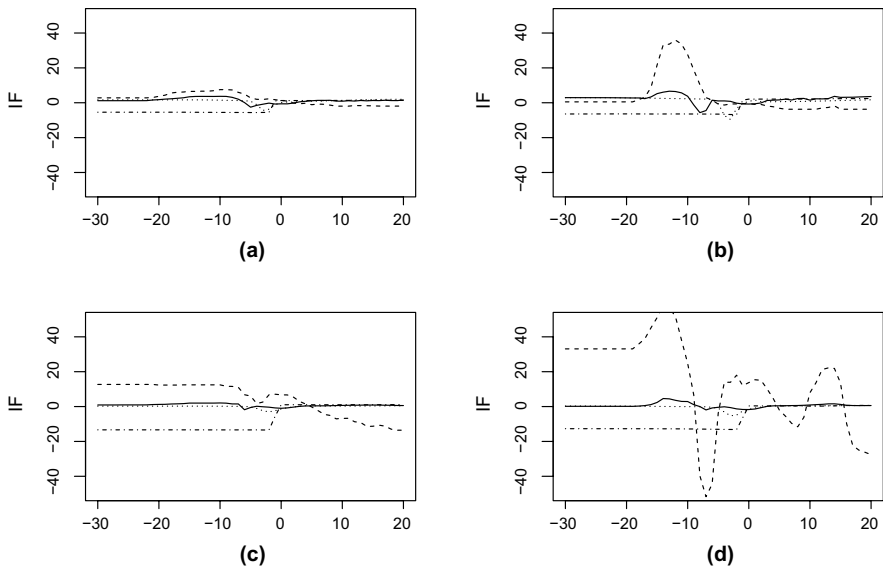


Fig. 2 The α -IFs of $\hat{\lambda}_{\text{MELE}}$ (dashed), $\hat{\lambda}_{\text{MHDE}}$ (solid), $\hat{\lambda}$ (dotted) and $\hat{\lambda}_L$ (dot-dashed) for model M2: **a** $\lambda = .25$ and $m = n = 30$; **b** $\lambda = .25$ and $m = n = 100$; **c** $\lambda = .75$ and $m = n = 30$; **d** $\lambda = .75$ and $m = n = 100$

presented in Figs. 1, 2 and 3. Figure 1 is for M1 with $\lambda = .15$ and $.55$, Figure 2 for M2 with $\lambda = .25$ and $.75$, and Fig. 3 for M3 with $\lambda = .25$ and $.75$. The results for other models and λ values are very similar and thus omitted to save space.

From Figs. 1, 2 and 3 we observe that no matter for which model and what sample size, $\hat{\lambda}_{\text{MHDE}}$ always performs the best and $\hat{\lambda}_{\text{MELE}}$ performs the worst. The α -IF of $\hat{\lambda}_{\text{MELE}}$ is generally unbounded while that of $\hat{\lambda}_{\text{MHDE}}$, $\hat{\lambda}$ and $\hat{\lambda}_L$ seems bounded when the outlying observation increases in both directions for mixture of normals and in the right direction for mixture of Poissons. This indicates that $\hat{\lambda}_{\text{MELE}}$ is generally not resistant to outliers whereas $\hat{\lambda}_{\text{MHDE}}$, $\hat{\lambda}$ and $\hat{\lambda}_L$ are. The bad performance of $\hat{\lambda}_{\text{MELE}}$ is mostly for when the outlying observation is bigger than 10. When the outlying observation is less than 10, the performance of $\hat{\lambda}_{\text{MELE}}$ is generally reasonable and is similar to that of other three estimators. The α -IFs of $\hat{\lambda}_{\text{MHDE}}$, $\hat{\lambda}$ and $\hat{\lambda}_L$ are almost constants for outliers beyond the range $[-10, 5]$ for mixture of normals and $[0, 5]$ for mixture of Poissons, though the constants are of different magnitude, and they fluctuate within the respective ranges. When $\hat{\lambda}_{\text{MHDE}}$, $\hat{\lambda}$ and $\hat{\lambda}_L$

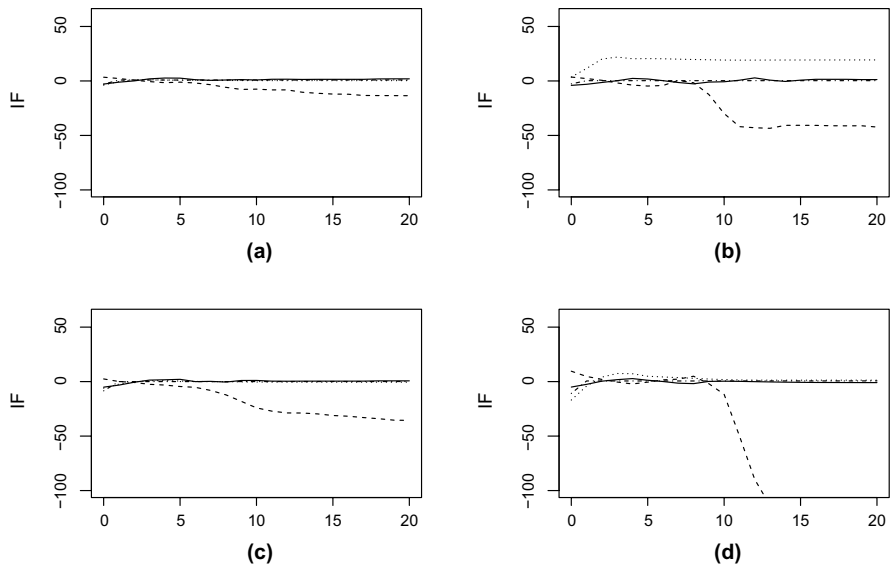


Fig. 3 The α -IFs of $\hat{\lambda}_{\text{MELE}}$ (dashed), $\hat{\lambda}_{\text{MHDE}}$ (solid), $\hat{\lambda}$ (dotted) and $\hat{\lambda}_L$ (dot-dashed) for model M3: **a** $\lambda = .25$ and $m = n = 30$; **b** $\lambda = .25$ and $m = n = 100$; **c** $\lambda = .75$ and $m = n = 30$; **d** $\lambda = .75$ and $m = n = 100$

are compared, $\hat{\lambda}_L$ behaves the worst in terms of having largest α -IF for mixture of normals and $\hat{\lambda}$ behaves the worst for mixture of Poisson distributions.

6.3 Test of model

In this part, we examine the performance of the K–S tests we proposed in Sect. 5 for testing the validity of model (3). We consider model (3) with $r(x) = (x, x^2)^\top$ as the collection of all possible models under consideration. Then we test whether the reduced model (3) with $r(x) = x$ is the actual true model or not. For demonstration purpose, we consider mixture of normals $H(x) = (1 - \lambda)F(x) + \lambda G(x)$ with $F \sim N(0, 1)$ and $G \sim N(\mu, \sigma^2)$. Then $f(x)$ and $h(x)$ are related by

$$h_\theta(x) =: h(x) = \left(1 - \lambda + \lambda e^{\alpha + \beta x + \gamma x^2}\right) f(x), \quad (24)$$

where $\alpha = -\frac{1}{2} \left(\log \sigma^2 + \frac{\mu^2}{\sigma^2} \right)$, $\beta = \frac{\mu}{\sigma^2}$ and $\gamma = \frac{1}{2} \left(1 - \frac{1}{\sigma^2} \right)$. Note that (24) is a special case of (3) when $r(x) = (x, x^2)^\top$. If $\sigma = 1$, then $\gamma = 0$ and thus model (3) holds with $r(x) = x$. So testing the validity of model (3) with $r(x) = x$ is equivalent to testing the null hypothesis $H_0 : \gamma = 0$ under model (24). We consider $\gamma = 0, -0.9$ and -1.5 , $\lambda = .35$ and $.65$, and sample sizes $m = n = 30$ and $m = n = 100$. For simplicity, we just fix $\mu = 1$ and as a result $\sigma = 1, .6$ and $.5$ for $\gamma = 0, -0.9$ and -1.5 , respectively. For each λ, γ and sample size considered, we use 500 total number of replications for our calculation. Within each replication, we use totally 1000 bootstrapping

Table 8 Estimated significance level and power of KS_{MELE} and KS_{MHDE}

λ	γ	Sig- nificance level	$m = n = 30$		$m = n = 100$	
			KS_{MELE}	KS_{MHDE}	KS_{MELE}	KS_{MHDE}
0.35	0	.10	.040	.104	.156	.186
		.05	.030	.014	.122	.084
		.01	.002	.000	.002	.002
		.10	.950	.860	.956	.870
	-.9	.05	.904	.802	.910	.710
		.01	.734	.410	.578	.184
		.10	.948	.966	.958	.998
	-1.5	.05	.898	.912	.910	.984
		.01	.716	.580	.536	.846
	0.65	.10	.036	.388	.096	.136
		.05	.030	.170	.122	.056
		.01	.008	.010	.002	.006
0.65	0	.10	.970	.910	.894	.928
		.05	.888	.818	.708	.758
		.01	.464	.282	.158	.120
		.10	.956	.876	.990	.990
	-.9	.05	.858	.762	.908	.944
		.01	.424	.302	.174	.396

samples to estimate the distribution and critical values of the test statistics KS_{MELE} and KS_{MHDE} . We choose different levels of significance $\alpha = .10, .05$ and $.01$. The simulation results are presented in Table 8. Note that $\gamma = 0$ means model (3) with $r(x) = x$ is correct and thus the correspondingly calculated values in Table 8 are the estimated significance levels. When $\gamma \neq 0$, model (3) with $r(x) = x$ is not correct and thus the correspondingly calculated values in Table 8 are the estimated powers at that value of γ .

From Table 8 we can see that the two test statistics KS_{MELE} and KS_{MHDE} are quite competitive in terms of achieved significance level and power. The achieved levels of significance are quite close to the true levels for most of the cases except for the case of KS_{MHDE} with $\lambda = 0.65$ and $m = n = 30$. The power of KS_{MHDE} becomes larger when γ is away from 0 except for the case of $\lambda = 0.65$ and $m = n = 30$. Surprisingly, the power of KS_{MELE} becomes smaller when γ is away from 0 except for the case of $\lambda = 0.65$ and $m = n = 100$. Note that KS_{MELE} and KS_{MHDE} use respectively the MELE and MHDE of β , which have been shown in Sect. 6.1 to have large bias and MSE when sample sizes are small. Consequently, the results for $m = n = 30$ have some abnormalities while the results for $m = n = 100$ are much better and more reliable. As expected, when the significance level α decreases, both the observed significance level and power decrease.

Table 9 Estimation of λ for the grain data

Method	Estimate	95% confidence interval
$\hat{\lambda}_{\text{MELE}}$.75	.61–.88 (.60–.89)
$\hat{\lambda}_{\text{MHDE}}$.76	.64–.92(.65–.88)
$\hat{\lambda}$ (Wu and Abedin, 2021)	.78	.68–.87
$\hat{\lambda}_L$ (Wu and Abedin, 2021)	.79	.58–.88
Poisson mixture (Smith et al., 1986)	.77	.00–.91
Two-by-two table (Smith and Vounatsou, 1997)	.20	.00–1.00
Logistic power (Smith and Vounatsou, 1997)	.61	.58–.64
Monotone logistic (Smith and Vounatsou, 1997)	.74	.61–1.00
Latent class (Smith and Vounatsou, 1997)	.73	.63–.83

For both KS_{MELE} and KS_{MHDE} , the powers are generally high for significance levels $\alpha = 0.10$ and 0.05 .

7 Real data examples

In this section we consider two real-life data examples and demonstrate the applications of our proposed methods.

Example 1: Grain data.

Smith and Vounatsou (1997) analyzed a data to determine the proportion λ of cells in the test sample with size $n = 94$ that were exposed to radioactive materials. The cells in control group with size $m = 94$ were not exposed to radioactivity. The number of grains shown in autoradiograph of cells can measure the amount of radioactive material, however grains can appear in autoradiograph due to radioactive material or due to background fogging. This dataset and more details of the data are available in Smith and Vounatsou (1997), and Wu and Abedin (2021) showed that the dominance constraint $F \geq G$ is valid for this dataset. We apply the two proposed estimators to this data and compare them with the estimators in Wu and Abedin (2021), Smith and Vounatsou (1997) and Smith et al. (1986). The results are given in Table 9, in which bootstrap method with 1000 repetitions are used to calculate the confidence intervals of λ .

From Table 9 we can see that our two proposed methods produce similar point estimates of λ to most other methods available in literature except for the two-by-two table

Table 10 Estimation of λ for the malaria data

Method	Wet Season	Dry Season
$\hat{\lambda}_{\text{MELE}}$.461 (.093)	.330 (.191)
$\hat{\lambda}_L$.435 (.083)	.349 (.102)
Bayesian	.444 (.054)	.305 (.118)

and Logistic power methods. When interval estimation is concerned, the two proposed estimators produce confidence intervals with width similar to those of $\hat{\lambda}$, $\hat{\lambda}_L$ and the latent class method but much smaller than those of Poisson mixture, two-by-two table and monotone logistic methods. In Table 9, the confidence intervals in the parentheses for $\hat{\lambda}_{\text{MELE}}$ and $\hat{\lambda}_{\text{MHDE}}$ are calculated using the asymptotic covariance matrices given in Theorems 2 and 6, respectively. From the results we see that bootstrap approximation is quite accurate. So when making inferences about λ , computationally intensive bootstrap is not necessary for $\hat{\lambda}_{\text{MELE}}$ and $\hat{\lambda}_{\text{MHDE}}$ while it is necessary for all other methods due to the unavailability of results on their asymptotic distribution. This is another great benefit of using $\hat{\lambda}_{\text{MELE}}$ and $\hat{\lambda}_{\text{MHDE}}$ over other methods.

Example 2: Malaria data.

We also examine a clinical malaria dataset first described by Kitua et al. (1996) and further studied by Vounatsou et al. (1998). Parasite densities in children with fever can be modelled as a two-component mixture, where one component represents the parasite densities in children without clinical malaria (F) while the other with clinical malaria (G), and λ is the proportion of children whose fever is attributable to malaria. In this dataset, level of parasitaemia were observed on children between 6 and 9 months in a village in Tanzania for both the wet season ($m = 94$, $n = 251$) and the dry season ($m = 122$, $n = 245$). Wu and Abedin (2021) showed that the dominance constraint $F \geq G$ is valid for both wet and dry seasons.

We apply the proposed $\hat{\lambda}_{\text{MELE}}$ to both seasons and compare the results with $\hat{\lambda}_L$ in Wu and Abedin (2021) and the Bayesian approach in Vounatsou et al. (1998). Note that this is a discretized data (parasitaemia levels were grouped into 10 categories), so kernel smoothing and hence $\hat{\lambda}_{\text{MHDE}}$ and $\hat{\lambda}$ are not applicable for this data. Table 10 presents the estimation results where the numbers in parentheses are estimated standard errors based on 500 bootstrapping samples. From this table we observe that $\hat{\lambda}_{\text{MELE}}$, $\hat{\lambda}_L$ and Bayesian approach produce similar estimates.

8 Concluding remarks

In this work, we proposed a two-component semiparametric mixture model (4) to accommodate the stochastic dominance constraint. We not only proposed and studied two estimators, the MELE and the MHDE, but also proposed K–S statistics to test the validity of the model (4). For both estimators, we proved theoretically their consistency and asymptotic normality. The finite-sample performance, in terms of both efficiency and robustness, of the proposed estimators and tests of the semiparametric model were examined through simulation studies and real data analysis.

The introduction of the two-component semiparametric mixture model (4) has several advantages over the fully nonparametric mixture model (1). First, the semiparametric model automatically solves the unidentifiability issue that the fully nonparametric model generally has. Second, the semiparametric model easily accommodates the stochastic dominance constraint with an equivalent and natural

positivity constraint on the parameter β , while the dominance constraint in the original nonparametric model is imposed on two functions which is much harder to handle and assess directly. Third, the semiparametric model has much better interpretation than the nonparametric model. The β value quantifies the difference between the two components, and the higher the β value the larger the difference. Fourth, for the semiparametric model we proposed two estimators with proved asymptotic normality and derived asymptotic covariance matrices, and these asymptotic results are in lack in literature for the nonparametric model. As a result, we can easily conduct statistical inferences for the mixing proportion λ (and β) based on these derived asymptotics for the semiparametric model, while for the nonparametric model some computationally intensive method, such as bootstrap, is necessary in order to make inferences about λ , due to the unavailability of asymptotic properties of available estimators. Of course, using the semiparametric model instead of the fully nonparametric model will lose some flexibility which we think will not be significant given the fact that many commonly used population families satisfy the exponential tilt relationship (3). This loss is worth, considering the aforementioned advantages of introducing the semiparametric model.

When the two proposed estimators for the semiparametric mixture model (4) are compared, we observe the following preference from our numerical studies. When the estimation of the mixing proportion λ or classification is our interest, MELE and MHDE perform equivalently well, in terms of bias and MSE or misclassification rate, and no one is dominantly better than the other. When the estimation accuracy of β is our focus, MHDE is highly preferred over MELE when λ is moderate or large while MELE is preferred when λ is small. When the presence of outliers is our concern, MHDE is much more robust and thus is highly preferred over MELE.

Since the proposed model and methods were motivated by the example in case-control genetic studies based on test statistic values (one dimension), this paper only focused on univariate case. Nevertheless, we can see that the proposed MELE and MHDE methods can be easily extended to multidimensional case. On the other side, the kernel estimations used in MHDE method may not be able to handle high dimensionality well.

For future work, we may consider to use minimum profile Hellinger distance estimation (MPHDE) for the semiparametric mixture model (4). Wu and Karunamuni (2015) first introduced the profile Hellinger distance particularly for semiparametric models and investigated the MPHDE for semiparametric model of general form. Wu and Karunamuni (2015) proved that the MPHDE is as robust as MHDE and achieves full efficiency at the true model. Xiang et al. (2014) used the MPHDE for a two-component semiparametric mixture model where one component is known up to some unknown parameters while the other component is unspecified. Wu et al. (2017) and Wu and Zhou (2018) applied the MPHDE for semiparametric location-shifted mixture models. Another direction to approach the estimation problem of model (4) could be Bayesian method. Vounatsou et al. (1998) applied Gibbs sampling approach to estimate the unknown mixing proportion in a two-component mixture model with discretized data.

Appendix: Conditions in Theorems 3–6

We decompose the parameter vector θ into two parts $\theta = (\lambda, \theta_r^\top)^\top$, where $\theta_r = (\alpha, \beta)^\top$ represents the regression coefficient parameters in (4). Note that $g_{\theta_r}(x) = e^{\alpha + \beta x} f(x)$ is essentially the g in (3). Parallely for each $t \in \Theta$ we write $t = (t_1, t_r^\top)^\top$ with $t_r = (t_2, t_3)^\top$ and $g_{t_r}(x) = e^{t_2 + t_3 x} f(x)$.

- (D1) There exists an ϵ -neighbourhood $B(\theta_r, \epsilon)$ of θ_r for some $\epsilon \geq 0$ such that $g_{t_r} - g_{\theta_r}$ is bounded by an integrable function for any $t_r \in B(\theta_r, \epsilon)$.
- (D2) f and K_0 in (4) and (16) respectively have compact supports.
- (D3) f in (4) has infinite support, K_0 in (16) is a bounded symmetric density with support $[-a_0, a_0]$ for some $0 < a_0 < \infty$, the second derivative of f exists and there exists a sequence α_m of positive numbers such that as $m \rightarrow \infty$, $\alpha_m \rightarrow \infty$ and

$$\sup_{\theta \in \Theta} \int I_{\{|x| > \alpha_m\}} h_\theta(x) dx \rightarrow 0, \quad (25)$$

$$b_m^2 \sup_{\theta \in \Theta} \int I_{\{|x| > \alpha_m\}} h_\theta(x) \sup_{|t| \leq a_0} \frac{|f^{(2)}(x + tb_m)|}{f(x)} dx \rightarrow 0, \quad (26)$$

$$m^{-1} b_m^{-1} \sup_{\theta \in \Theta} \int I_{\{|x| \leq \alpha_m\}} h_\theta(x) \sup_{|t| \leq a_0} \frac{f(x + tb_m)}{f^2(x)} dx \rightarrow 0, \quad (27)$$

$$b_m^4 \sup_{\theta \in \Theta} \int I_{\{|x| \leq \alpha_m\}} h_\theta(x) \sup_{|t| \leq a_0} \left[\frac{f^{(2)}(x + tb_m)}{f(x)} \right]^2 dx \rightarrow 0, \quad (28)$$

where $f^{(k)}$ denotes that k^{th} derivative of f .

Condition (D3) is satisfied by probability families such as normal and exponential. When f is a normal or an exponential population, (25)–(28) hold with any sequence α_m such that $\alpha_m = o(\log[mb_m])$.

Let $\{\alpha_N\}$ be a sequence of positive numbers such that $\alpha_N \rightarrow \infty$ as $N \rightarrow \infty$.

- (C0) f has infinite support $(-\infty, \infty)$.
- (C1) The second derivative of f exists.
- (C2) $\frac{n}{N} \rightarrow \rho \in (0, 1)$ as $N \rightarrow \infty$.
- (C3) K_0 and K_1 in (16) and (17) respectively are bounded symmetric densities with support $[-a_0, a_0]$ and $[-a_1, a_1]$ respectively, where $0 < a_0, a_1 < \infty$.
- (C4) All the elements in both $\Delta(\theta)$ and $\bar{\Delta}(\theta)$ are finite, where $\Delta(\theta)$ is defined in (19) and $\bar{\Delta}(\theta) = \int \left(\frac{\partial w_1(x)}{\partial \theta} \right) \left(\frac{\partial w_1(x)}{\partial \theta} \right)^\top f(x) dx$.
- (C5) The second derivative of f exists and satisfies for $i = 1, 2, 3$ that as $N \rightarrow \infty$,

$$b_m^2 \int \epsilon_{Ni}^2(x) \frac{f}{w_1}(x) \sup_{|t| \leq a_0} \frac{f^{(2)}(x + tb_m)}{f(x)} dx = O(1),$$

$$\text{where } \epsilon_N(x) = \frac{\partial w_1(x)}{\partial \theta} I_{\{|x| > \alpha_N\}}.$$

(C6) As $N \rightarrow \infty$,

$$\begin{aligned} N \cdot P(|X_1| > \alpha_N - a_0 b_m) &\longrightarrow 0, \\ N \cdot P(|Y_1| > \alpha_N - a_1 b_n) &\longrightarrow 0. \end{aligned}$$

(C7) The second derivatives of f exists and satisfies, as $N \rightarrow \infty$,

$$\begin{aligned} N^{-1/2} b_n^{-1} \int |\delta_N(x)| w_1^{-1}(x) \sup_{|t| \leq a_1} \frac{h_\theta(x + tb_n)}{h_\theta(x)} dx &\longrightarrow 0, \\ N^{1/2} b_n^4 \int |\delta_N(x)| f(x) \sup_{|t| \leq a_1} \left[\frac{h_\theta^{(2)}(x + tb_n)}{h_\theta(x)} \right]^2 dx &\longrightarrow 0, \\ N^{-1/2} b_m^{-1} \int |\delta_N(x)| \sup_{|t| \leq a_0} \frac{f(x + tb_m)}{f(x)} dx &\longrightarrow 0, \\ N^{1/2} b_m^4 \int |\delta_N(x)| f(x) \sup_{|t| \leq a_0} \left[\frac{f^{(2)}(x + tb_m)}{f(x)} \right]^2 dx &\longrightarrow 0, \end{aligned}$$

$$\text{where } \delta_N(x) = \frac{\partial w_1(x)}{\partial \theta} I_{\{|x| \leq \alpha_N\}}.$$

(C8) The second derivatives of f exists and satisfies, as $N \rightarrow \infty$,

$$\begin{aligned} N^{1/2} b_n^2 \int |\delta_N(x)| f(x) \sup_{|t| \leq a_1} \frac{|h_\theta^{(2)}(x + tb_n)|}{h_\theta(x)} dx &\longrightarrow 0, \\ N^{1/2} b_m^2 \int |\delta_N(x)| f(x) \sup_{|t| \leq a_0} \frac{|f^{(2)}(x + tb_m)|}{f(x)} dx &\longrightarrow 0. \end{aligned}$$

(C9) As $N \rightarrow \infty$,

$$\begin{aligned} \sup_{|x| \leq \alpha_N} \sup_{|t| \leq a_1} \frac{h_\theta(x + tb_n)}{h_\theta(x)} &= O(1), \\ \sup_{|x| \leq \alpha_N} \sup_{|t| \leq a_0} \frac{f(x + tb_m)}{f(x)} &= O(1). \end{aligned}$$

(C10) As $N \rightarrow \infty$,

$$\begin{aligned} b_n^2 \int I_{\{|x| \leq \alpha_N\}} h_\theta(x) \sup_{|t| \leq a_1} \left[\frac{\partial^2 \log w_1(y)}{\partial \theta \partial y} \Big|_{y=x+tb_n} \right]^2 dx &\longrightarrow 0, \\ b_m^2 \int I_{\{|x| \leq \alpha_N\}} f(x) \sup_{|t| \leq a_0} \left[\frac{\partial^2 w_1(y)}{\partial \theta \partial y} \Big|_{y=x+tb_m} \right]^2 dx &\longrightarrow 0. \end{aligned}$$

(C11) The second derivative of f exists and satisfies, as $N \rightarrow \infty$,

$$N^{1/2} b_m^2 \int |\epsilon_N(x)| f(x) \sup_{|t| \leq a_0} \frac{|f^{(2)}(x + tb_m)|}{f(x)} dx \longrightarrow 0.$$

Conditions (C4)–(C11) hold for probability families such as normal and exponential. When f is a normal population, (C4)–(C11) hold with $b_m, b_n \asymp N^{-r}$, $1/4 < r < 1/2$, and any sequence α_N such that $\alpha_N = o(N^r \log N)$ and $\log N = o(\alpha_N^2)$. When f is an exponential distribution $f(x) = ce^{-cx}$ with some $c > 0$ and $2\beta - c < 0$,

(C4)–(C11) hold with $b_m, b_n \asymp N^{-r}$, $1/4 < r < 1/2$, and any sequence α_N such that $\alpha_N = o(N^r \log N)$ and $\log N = o(\alpha_N)$.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10463-022-00835-5>.

Acknowledgements The authors thank the Chief Editor, the Associate Editor and the referees for their valuable comments and suggestions that have led to significant improvements in the manuscript. The authors acknowledge with gratitude the support for this research via Discovery Grants project RGPIN-2018-04328 from Natural Sciences and Engineering Research Council (NSERC) of Canada and NSF project ZR2021MA048 of Shandong Province of China.

References

- Abedin, T. (2018). *Inferences for two-component mixture models with stochastic dominance*. Ph.D. thesis, Department of Mathematics and Statistics, University of Calgary.
- Anderson, J. A. (1972). Separate sample logistic discrimination. *Biometrika*, 59, 19–35.
- Anderson, J. A. (1979). Multivariate logistic compounds. *Biometrika*, 66, 17–26.
- Beran, R. (1977). Minimum Hellinger distance estimators for parametric models. *The Annals of Statistics*, 5, 445–463.
- Bordes, L., Delmas, C., Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model where one component is known. *Scandinavian Journal of Statistics*, 33, 733–752.
- Breslow, N. E., Day, N. E. (1980). *Statistical Methods in Cancer Research: the Analysis of Case-Control Studies*. Lyon: IARC Scientific Publications.
- Chen, G., Wu, J. (2013). Molecular classification of acute leukemia. *Some recent advances in mathematics and statistics—Proceedings of statistics 2011 Canada/IMST 2011-FIM XX*, 60–74.
- Deng, X., Wan, S., Zhang, B. (2009). An improved Goodness-of-Fit test for logistic regression models based on case-control data by random partition. *Communications in Statistics: Simulation and Computation*, 38, 233–243.
- Efron, B., Tibshirani, R. (1996). Using specially designed exponential families for density estimation. *The Annals of Statistics*, 24, 2431–2461.
- Ficklin, S. P., Dunwoodie, L. J., Poehlman, W. L., Watson, C., Roche, K. E., Feltus, F. A. (2017). Discovering condition-specific gene co-expression patterns using gaussian mixture models: A cancer case study. *Scientific Reports*, 7, 8617.
- Karlis, D., Xekalaki, E. (1998). Minimum Hellinger distance estimation for Poisson mixtures. *Computational Statistics & Data Analysis*, 29, 81–103.
- Karunamuni, R. J., Wu, J. (2011). One-step minimum Hellinger distance estimation. *Computational Statistics & Data Analysis*, 55, 3148–3164.
- Kharchenko, P. V., Silberstein, L., Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nature Methods*, 11, 740–742.
- Kitua, A. Y., Smith, T., Alonso, P. L., Masanja, H., Urassa, H., Menendez, C., Kimario, J., Tanner, M. (1996). Plasmodium falciparum malaria in the first year of life in an area of intense and perennial transmission. *Tropical Medicine & International Health*, 1, 475–484.
- Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *The Annals of Statistics*, 22, 1081–1114.
- Lu, R., Smith, R. M., Seweryn, M., Wang, D., Hartmann, K., Webb, A., Sadee, W., Rempala, G. A. (2015). Analyzing allele specific RNA expression using mixture models. *BMC Genomics*, 16, 566.
- Lu, Z., Hui, Y. V., Lee, A. H. (2003). Minimum Hellinger distance estimation for finite mixtures of Poisson regression models and its applications. *Biometrics*, 59, 1016–1026.
- Prentice, R. L., Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66, 403–411.

- Qin, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. *The Annals of Statistics*, 27, 1368–1384.
- Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22, 300–325.
- Qin, J., Zhang, B. (1997). A goodness of fit test for logistic regression models based on case-control data. *Biometrika*, 84, 609–618.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- Smith, T., Vounatsou, P. (1997). Logistic regression and latent class models for estimating positives in diagnostic assays with poor resolution. *Communications in Statistics: Theory and Methods*, 26, 1677–1700.
- Smith, T. A., Smith, A. G., Hooper, M. L. (1986). Selection of a mouse embryonal carcinoma clone resistant to the inhibition of metabolic cooperation by retinoic acid. *Experimental Cell Research*, 165, 417–430.
- Vounatsou, P., Smith, T., Smith, A. F. M. (1998). Bayesian analysis of two-component mixture distributions applied to estimating malaria attributable fractions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47, 575–587.
- Woo, M.-J., Sriram, T. N. (2006). Robust estimation of mixture complexity. *Journal of the American Statistical Association*, 101, 1475–1486.
- Woo, M.-J., Sriram, T. N. (2007). Robust estimation of mixture complexity for count data. *Computational Statistics & Data Analysis*, 51, 4379–4392.
- Wu, J., Abedin, T. (2021). A two-component nonparametric mixture model with stochastic dominance. *Journal of the Korean Statistical Society*, 50, 1029–1057.
- Wu, J., Karunamuni, R. J. (2015). Profile Hellinger distance estimation. *Statistics*, 49, 711–740.
- Wu, J., Zhou, X. (2018). Minimum Hellinger distance estimation for a semiparametric location-shifted mixture model. *Journal of Statistical Computation and Simulation*, 88, 2507–2527.
- Wu, J., Karunamuni, R. J., Zhang, B. (2010). Minimum Hellinger distance estimation in a two-sample semiparametric model. *Journal of Multivariate Analysis*, 101, 1102–1122.
- Wu, J., Yao, W., Xiang, S. (2017). Computation of an efficient and robust estimator in a semiparametric mixture model. *Journal of Statistical Computation and Simulation*, 87, 2128–2137.
- Xiang, S., Yao, W., Wu, J. (2014). Minimum profile Hellinger distance estimation for a semiparametric mixture model. *Canadian Journal of Statistics*, 42, 246–267.
- Zhang, B. (1999). A chi-squared goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 86, 531–539.
- Zhang, B. (2001). An information matrix test for logistic regression models based on case-control data. *Biometrika*, 88, 921–932.
- Zhang, B. (2002). An EM algorithm for a semiparametric finite mixture model. *Journal of Statistical Computation and Simulation*, 72, 791–802.
- Zhang, B. (2006). A score test under a semiparametric finite mixture model. *Journal of Statistical Computation and Simulation*, 76, 691–703.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.