Title

Flexible estimation of a semiparametric two-component mixture model with one parametric component

Journal

Electronic Journal of Statistics, 9(1)

ISSN

1935-7524

Authors

Ma, Y

Yao, W

Publication Date

2015

DOI

10.1214/15-EJS1008

Peer reviewed

Electronic Journal of Statistics
Vol. 9 (2015) 444–474
ISSN: 1935-7524
DOI: 10.1214/15-EJS1008

Flexible estimation of a semiparametric two-component mixture model with one parametric component

Yanyuan Ma

Department of Statistics
University of South Carolina
Columbia, SC 29208
e-mail: yanyuan.ma@stat.sc.edu

and

Weixin Yao

Department of Statistics
University of California, Riverside
Riverside, CA 92521
e-mail: weixin.yao@ucr.edu

Abstract: We study a two-component semiparametric mixture model where one component distribution belongs to a parametric class, while the other is symmetric but otherwise arbitrary. This semiparametric model has wide applications in many areas such as large-scale simultaneous testing/multiple testing, sequential clustering, and robust modeling. We develop a class of estimators that are surprisingly simple and are unique in terms of their construction. A unique feature of these methods is that they do not rely on the estimation of the nonparametric component of the model. Instead, the methods only require a working model of the unspecified distribution, which may or may not reflect the true distribution. In addition, we establish connections between the existing estimator and the new methods and further derive a semiparametric efficient estimator. We compare our estimators with the existing method and investigate the advantages and cost of the relatively simple estimation procedure.

Keywords and phrases: Efficiency, large-scale simultaneous testing, mixture models, multiple testing, robust statistics, semiparametric estimator.

Received August 2014.

1. Introduction

Mixture models have a long history. The most classical mixture model is a parametric model of the form

$$g(x) = \sum_{j=1}^{J} p_j f_j(x; \boldsymbol{\alpha}_j),$$

where p_j's are positive values satisfying $\sum_{j=1}^{J} p_j = 1$ and each f_j is a probability density function (pdf) known up to the parameters $\boldsymbol{\alpha}_j$ for $j = 1, \ldots, J$.

The interest is in estimating p_j as well as $\boldsymbol{\alpha}_j$, for $j = 1, \ldots, J$. Because it is a fully parametric model, as long as the problem is identifiable, the familiar maximum likelihood estimator will provide the best estimation for the parameters. Please see for example, Lindsay (1995), Böhning (1999), McLachlan and Peel (2000), and Frühwirth-Schnatter (2006) for more detailed illustration of parametric mixture models.

At least two forms of extensions from the original parametric mixture model have attracted considerable interest recently. In the first form, the mixture components f_j's are assumed to be identical, symmetric but otherwise unspecified. Hence this is a nonparametric extension of the original mixture model. In this extension, the parameter $\boldsymbol{\alpha}_j$ is the center of the jth component, written as μ_j, and the model has the form

$$g(x) = \sum_{j=1}^{J} p_j f(x - \mu_j). \tag{1.1}$$

Identifiability and estimation for p_j, μ_j, $j = 1, \ldots, J$ are studied extensively for the extended nonparametric mixture model, see for example, Bordes et al. (2006a) and Hunter et al. (2007). Hall and Zhou (2003) and Hall et al. (2005) considered a multivariate version of (1.1) for which the component distributions have independent nonparametric components.

The second extension allows one of the f_j's to be symmetric with unknown center and otherwise unspecified, while keeping the other components parametric. Hence this is a semiparametric extension of the original mixture model. Because the combination of all the parametric components can itself be viewed as a parametric model, this extension has the general form

$$g(x; \boldsymbol{\beta}, \eta) = (1 - p) f(x; \boldsymbol{\alpha}) + p\eta(x - \mu), \tag{1.2}$$

where $0 < p < 1$. Here, $\boldsymbol{\alpha}$ is an unknown parameter of the parametric component f, while we use η to denote the symmetric nonparametric component. We collect all the parameters of interest in the d dimensional parameter vector $\boldsymbol{\beta} \equiv (\mu, p, \boldsymbol{\alpha}^{\mathrm{T}})^{\mathrm{T}}$. In this article, we restrict our interest in the semiparametric mixture model (1.2).

Model (1.2) arises naturally in large-scale simultaneous testing and multiple hypothesis testing problems. For example, in detecting differentially expressed genes under two or more conditions in microarray data, one naturally encounters (1.2). Specifically, suppose we construct a test statistic for each gene. Assume that under the null hypothesis, the test statistic has a distribution $f(x; \boldsymbol{\alpha})$ that is either completely known or known up to a parameter $\boldsymbol{\alpha}$. Then the test statistics collected from all genes automatically have a mixture distribution of $f(x; \boldsymbol{\alpha})$, representing the null distribution, and $\eta(x - \mu)$, representing the unknown alternative distribution. This directly leads to (1.2). Indeed, there has been much effort in utilizing the parametric mixture model, i.e., further assuming a parametric form for $\eta(x - \mu)$, in the multiple hypothesis testing problems emerged from bioinformatics. Please see, for example, Allison et al. (2002), Pounds and Morris

(2003), Efron (2004), Genovese and Wasserman (2004), Langaas et al. (2005), and McLachlan et al. (2006). Model (1.2) is also used in robust statistics (Hampel et al., 1986; Huber and Ronchetti, 2009) to describe data that are contaminated, where the uncontaminated part of the data is described by $f(x;\boldsymbol{\alpha})$, and the contamination is captured with a minimal restriction via $\eta(x-\mu)$. Compared to the contaminated normal mixture model used in classical robust statistics literature, where η is typically assumed to have a normal distribution with a large variance, the model (1.2) is certainly much more flexible. In fact, the more relaxed form of the semiparametric mixture distribution in (1.2) allows us to check any parametric assumption, such as the assumption of normality, about $\eta(\cdot)$.

Model (1.2) is also used in some other areas in practice. For example, Song et al. (2010) applied it in sequential clustering. Bordes et al. (2013) investigated a regression setting of (1.2). Please see, for example, Bordes et al. (2006b) and Song et al. (2010) for more applications of (1.2).

When f is completely known, hence $\boldsymbol{\alpha}$ does not appear, Bordes et al. (2006b) proved the identifiability of model (1.2) under certain conditions and proposed a symmetrization based distance estimator for μ and p. Bordes and Vandekerkhove (2010) further established that the estimator is root-n consistent and asymptotically normal. Hohmann and Holzmann (2013) applied a similar symmetrization based distance estimator to model (1.2) when a location shift parameter $\boldsymbol{\alpha}$ appears in the model. Xiang, Yao, and Wu (2014) proposed estimating the model (1.2) based on minimum profile Hellinger distance. When η is not assumed to be symmetric and μ does not appear, Song et al. (2010) proposed a kernel type EM algorithm where η is estimated with nonparametric kernel density estimation, and Ma et al. (2011) proposed nonparametric maximum likelihood estimators with discretized non-parametric component.

The extended model given in (1.2), which allows both $\boldsymbol{\alpha}$ and μ to be unknown, is the focus of our study. In Section 2, we give some conditions for model (1.2) to be identifiable. Given that model (1.2) is identifiable, it is still not obvious how to estimate the parameter of interest $\boldsymbol{\beta}$. To this end, we develop a class of estimators that are surprisingly simple and are unique in terms of their construction. These estimators are easy to construct because they completely bypass the estimation of the nonparametric density function η or its corresponding cumulative distribution function (cdf). Thus, the nonparametric nature of (1.2) is in a way eluded operationally. Instead of estimating η or its cdf, the class simply adopts a working model in its place. Regardless of whether the working model is correct or not correct, or regardless whether it approximates the true pdf reasonably well or badly, the resulting estimator is always consistent. The effect of the working model is mainly reflected in the variability of the parameter estimation, and we generally favor a working model that yields smaller estimation variability. In Section 5, we further develop an estimator that involves faithfully estimating η to achieve the minimum estimation variance among all possible consistent estimators.

The construction of the estimator class heavily relies on the symmetry of η. In fact, exploiting the symmetric nature of η, we are able to write a much wider class of estimators. However, the class of estimators we recommend has the

additional advantage of being more explicit and tractable. It also has a nice connection to the most efficient estimator for $\boldsymbol{\beta}$ in the sense of semiparametric efficiency given by Bickel et al. (1993). We further show that the symmetrization based distance estimator proposed by Bordes et al. (2006b) also belongs to this class. Thus, as an alternative to the derivation in Bordes and Vandekerkhove (2010), one can also obtain the asymptotic properties of the estimator based on the general results established here in Section 3.

The rest of the article is organized as the following. In Section 2, we discuss the identifiability of (1.2). In Section 3, we investigate the symmetry of η to reveal a general approach to construct estimators. We then propose a family of explicit estimators and study their properties. We investigate the link between our estimators and the existing methods in Section 4 and explain how to achieve the optimal efficiency in Section 5. Finite sample performance of the estimators are illustrated in Section 6. We conclude the article with some discussion in Section 7 and collect all the technical details in an Appendix.

2. Identifiability

Without any constraints on η, model (1.2) is generally not identifiable. This is easily seen since we can exclude $(1-p)f(x;\boldsymbol{\alpha})$ from $g(x)$ for different $p, \boldsymbol{\alpha}$ values and view the remaining component as $p\eta(x-\mu)$. Even under the symmetry constraints of η, the identifiability of model (1.2) is still not guaranteed automatically, as we illustrate in the following two examples.

Example 1. If $\mathbf{f}(x;\boldsymbol{\alpha})$ is symmetric about zero and $\eta(x) = f(x;\boldsymbol{\alpha})$, then

$$(1-p)f(x;\boldsymbol{\alpha}) + pf(x-\mu;\boldsymbol{\alpha}) = (1-2p)f(x;\boldsymbol{\alpha}) + (2p)\widetilde{\eta}(x-\mu/2),$$

where $0 < p < 1/2, \widetilde{\eta}(x) = \{f(x+\mu/2;\boldsymbol{\alpha}) + f(x-\mu/2;\boldsymbol{\alpha})\}/2$, hence the problem is not identifiable.

In fact, when we fix $\mu = 2$, Example 1 reduces to the second non-identifiable example given in Bordes et al. (2006b).

Example 2. If $\eta(x)$ and $f(x;\boldsymbol{\alpha})$ are symmetric about zero, then

$$(1-p)f(x;\boldsymbol{\alpha}) + p\eta(x) = (1-\widetilde{p})f(x;\boldsymbol{\alpha}) + \widetilde{p}\widetilde{\eta}(x),$$

where $\widetilde{\eta}(x) = \{(\widetilde{p}-p)f(x;\boldsymbol{\alpha}) + p\eta(x)\}/\widetilde{p}$ and $p < \widetilde{p} < 1$, hence the problem is not identifiable.

We next give some simple sufficient conditions for model (1.2) to be identifiable when $\eta(x)$ is symmetric about 0 and $\boldsymbol{\alpha}$ does not appear.

Proposition 1. *Suppose $\eta(x)$ is symmetric about 0 and $\boldsymbol{\alpha}$ does not appear. Model (1.2) is identifiable if either of the following two conditions is satisfied*

1. *$f(x) > 0$. In addition, $\lim_{x\to\infty} \frac{\eta(x-\delta)}{f(x)} = 0$ or $\lim_{x\to-\infty} \frac{\eta(x-\delta)}{f(x)} = 0$ for any δ.*
2. *$\eta(x) > 0$. In addition, $\lim_{x\to\infty} \frac{f(x)}{\eta(x-\delta)} = 0$ or $\lim_{x\to-\infty} \frac{f(x)}{\eta(x-\delta)} = 0$ for any δ.*

Based on Proposition 1, we can see that model (1.2) is identifiable if $\eta(x)$ and $f(x)$ have different tail properties. In addition, note that the identifiable results in Proposition 3 of Bordes et al. (2006b) are special cases of Proposition 1. Compared to Bordes et al. (2006b), Proposition 1 does not require moment assumptions of $f(\cdot)$ and $\eta(\cdot)$; in addition, Proposition 1 does not require $f(x)$ to have bounded support when $\eta(x)$ has heavier tails than $f(x)$. The proof of Proposition 1 is in Appendix A.1.

3. Semiparametric estimators and properties

3.1. General estimators

When constructing the simplest generalized moment type of estimators, the usual practice is to solve an estimating equation of the form $\sum_{i=1}^{n} \mathbf{a}(x_i;\boldsymbol{\beta}) = \mathbf{0}$. As long as $E\{\mathbf{a}(X;\boldsymbol{\beta})\} = \mathbf{0}$ and the estimating equation does not degenerate, the solution $\widehat{\boldsymbol{\beta}}$ will then be a root-n consistent estimator of $\boldsymbol{\beta}$. Thus, finding the mean zero estimating function is critical for constructing generalized moment estimators. If we start with an arbitrary function $\mathbf{a}(x;\boldsymbol{\beta})$ that does not necessarily have mean zero, an obvious way to correct it is through subtracting its mean. This requires to calculate $E\{\mathbf{a}(X;\boldsymbol{\beta})\}$. Even though $\mathbf{a}(x;\boldsymbol{\beta})$ is explicitly chosen, its expectation involves $g(x;\boldsymbol{\beta},\eta)$, and since η is not known, $E\{\mathbf{a}(X;\boldsymbol{\beta})\}$ is generally uncomputable without an approximated η function. However, the fact that η is a symmetric function in (1.2) leads to some unexpected construction as we now explain.

The explicit calculation of $E\{\mathbf{a}(X;\boldsymbol{\beta})\}$ yields

$$\begin{aligned}
E\{\mathbf{a}(X;\boldsymbol{\beta})\} &= \int \mathbf{a}(x;\boldsymbol{\beta})\{(1-p)f(x;\boldsymbol{\alpha}) + p\eta(x-\mu)\}dx \\
&= \int \mathbf{a}(t+\mu;\boldsymbol{\beta})\{(1-p)f(t+\mu;\boldsymbol{\alpha}) + p\eta(t)\}dt \\
&= \int_0^\infty \mathbf{a}(t+\mu;\boldsymbol{\beta})\{(1-p)f(t+\mu;\boldsymbol{\alpha}) + p\eta(t)\}dt \\
&\quad + \int_0^\infty \mathbf{a}(\mu-t;\boldsymbol{\beta})\{(1-p)f(\mu-t;\boldsymbol{\alpha}) + p\eta(-t)\}dt \\
&= (1-p)\int_0^\infty \{\mathbf{a}(\mu+t;\boldsymbol{\beta})f(t+\mu;\boldsymbol{\alpha}) + \mathbf{a}(\mu-t;\boldsymbol{\beta})f(\mu-t;\boldsymbol{\alpha})\}dt \\
&\quad + p\int_0^\infty \{\mathbf{a}(\mu+t;\boldsymbol{\beta}) + \mathbf{a}(\mu-t;\boldsymbol{\beta})\}\eta(t)dt.
\end{aligned}$$

The first component of the above expression is a computable quantity at any parameter value of $\boldsymbol{\beta}$, while the only difficulty due to the presence of η is reflected in the second component. Thus, critically, if we choose the function $\mathbf{a}(x;\boldsymbol{\beta})$ so that $\mathbf{a}(x;\boldsymbol{\beta}) + \mathbf{a}(2\mu - x;\boldsymbol{\beta}) = \mathbf{0}$ for all x and $\boldsymbol{\beta}$, then the effect of η vanishes and an estimating function can be readily obtained. Under such choice of $\mathbf{a}(x;\boldsymbol{\beta})$,

we have

$$\begin{aligned} & \int_0^\infty \{\mathbf{a}(\mu+t;\boldsymbol{\beta})f(t+\mu;\boldsymbol{\alpha}) + \mathbf{a}(\mu-t;\boldsymbol{\beta})f(\mu-t;\boldsymbol{\alpha})\}dt \\ = & \int_0^\infty \mathbf{a}(\mu+t;\boldsymbol{\beta})\{f(t+\mu;\boldsymbol{\alpha}) - f(\mu-t;\boldsymbol{\alpha})\}dt \\ = & \frac{1}{2}\int \mathbf{a}(\mu+t;\boldsymbol{\beta})\{f(t+\mu;\boldsymbol{\alpha}) - f(\mu-t;\boldsymbol{\alpha})\}dt. \end{aligned}$$

This simple yet crucial observation leads to a very general class of estimators obtained by solving

$$n^{-1}\sum_{i=1}^n \mathbf{a}(x_i;\boldsymbol{\beta}) - \frac{1-p}{2}\int \mathbf{a}(\mu+t;\boldsymbol{\beta})\{f(t+\mu;\boldsymbol{\alpha}) - f(\mu-t;\boldsymbol{\alpha})\}dt = \mathbf{0}, \quad (3.1)$$

where $\mathbf{a}(x;\boldsymbol{\beta}) \in \mathbb{R}^d$ satisfies $\mathbf{a}(x;\boldsymbol{\beta}) + \mathbf{a}(2\mu - x;\boldsymbol{\beta}) = \mathbf{0}$. This implies that any d-dimensional odd function reparameterized to be a function of $\mu + x$ is a qualified choice for $\mathbf{a}$.

3.2. A specific family of estimators

The consideration above provides many choices of $\mathbf{a}(x;\boldsymbol{\beta})$, hence many potential estimators of $\boldsymbol{\beta}$. In this article, we restrict our attention to a specific family of $\mathbf{a}(x;\boldsymbol{\beta})$ functions that is arguably the most interesting. Motivated by the optimal efficiency of the maximum likelihood estimator in the parametric model framework, we consider the score function $\mathbf{S}_{\boldsymbol{\beta}} = (S_\mu, S_p, \mathbf{S}_{\boldsymbol{\alpha}}^{\mathrm{T}})^{\mathrm{T}}$, where

$$S_\mu = \frac{-\eta'(x-\mu)}{g(x;\boldsymbol{\beta},\eta)}, \quad S_p = \frac{\eta(x-\mu) - f(x;\boldsymbol{\alpha})}{g(x;\boldsymbol{\beta},\eta)}, \quad \mathbf{S}_{\boldsymbol{\alpha}} = \frac{f'_{\boldsymbol{\alpha}}(x;\boldsymbol{\alpha})}{g(x;\boldsymbol{\beta},\eta)}.$$

While the score function in its original form does not qualify as one choice of the $\mathbf{a}(x;\boldsymbol{\beta})$ function, we modify them by making the denominator symmetric around μ and the numerator anti-symmetric around μ. This yields $\mathbf{a} = (a_\mu, a_p, \mathbf{a}_{\boldsymbol{\alpha}}^{\mathrm{T}})^{\mathrm{T}}$, where

$$\begin{aligned} a_\mu &= \frac{\eta'(x-\mu)}{g(x;\boldsymbol{\beta},\eta) + g(2\mu-x;\boldsymbol{\beta},\eta)}, \\ a_p &= \frac{f(x;\boldsymbol{\alpha}) - f(2\mu-x;\boldsymbol{\alpha})}{g(x;\boldsymbol{\beta},\eta) + g(2\mu-x;\boldsymbol{\beta},\eta)}, \\ \mathbf{a}_{\boldsymbol{\alpha}} &= \frac{f'_{\boldsymbol{\alpha}}(x;\boldsymbol{\alpha}) - f'_{\boldsymbol{\alpha}}(2\mu-x;\boldsymbol{\alpha})}{g(x;\boldsymbol{\beta},\eta) + g(2\mu-x;\boldsymbol{\beta},\eta)}. \end{aligned}$$

While closely related to the score functions, $\mathbf{a}$ cannot be directly used because it relies on the unknown pdf η. A simple alternative is to replace η with an arbitrary working version of η, such as normal density. Denoting the working function η^*, then we obtain a practically implementable function $\mathbf{a}^*$, where $\mathbf{a}^*$ is defined analogously as $\mathbf{a}$ except that η is replaced by η^*, a working symmetric pdf. Specifically, the estimation procedure is the following.

Step 1: Select a working model for η. Denote the working model η^*. Write $g^*(x;\boldsymbol{\beta}) = (1-p)f(x;\boldsymbol{\alpha}) + p\eta^*(x-\mu)$.

Step 2: Using numerical integration to approximate the functions of $\boldsymbol{\beta}$:

$$r_\mu^*(\boldsymbol{\beta}) = \frac{1-p}{2}\int \frac{\eta^{*'}(t)\{f(\mu+t;\boldsymbol{\alpha}) - f(\mu-t;\boldsymbol{\alpha})\}}{g^*(\mu-t;\boldsymbol{\beta}) + g^*(\mu+t;\boldsymbol{\beta})}dt,$$

$$r_p^*(\boldsymbol{\beta}) = \frac{1-p}{2}\int \frac{\{f(\mu+t;\boldsymbol{\alpha}) - f(\mu-t;\boldsymbol{\alpha})\}^2}{g^*(\mu-t;\boldsymbol{\beta}) + g^*(\mu+t;\boldsymbol{\beta})}dt,$$

$$\mathbf{r}_\alpha^*(\boldsymbol{\beta}) = \frac{1-p}{2} \times \int \frac{\{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha}) - \mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})\}\{f(\mu+t;\boldsymbol{\alpha}) - f(\mu-t;\boldsymbol{\alpha})\}}{g^*(\mu-t;\boldsymbol{\beta}) + g^*(\mu+t;\boldsymbol{\beta})}dt.$$

Write $\mathbf{r}^*(\boldsymbol{\beta}) = \{r_\mu^*(\boldsymbol{\beta}), r_p^*(\boldsymbol{\beta}), \mathbf{r}_{\boldsymbol{\alpha}}^*(\boldsymbol{\beta})^{\mathrm{T}}\}^{\mathrm{T}}$.

Step 3: Define

$$\begin{aligned} a_\mu^*(x_i;\boldsymbol{\beta}) &= \frac{\eta^{*'}(x_i-\mu)}{g^*(x_i) + g^*(2\mu - x_i)}, \\ a_p^*(x_i;\boldsymbol{\beta}) &= \frac{f(x_i) - f(2\mu - x_i)}{g^*(x_i) + g^*(2\mu - x_i)}, \\ \mathbf{a}_{\boldsymbol{\alpha}}^*(x_i;\boldsymbol{\beta}) &= \frac{\mathbf{f}'_\alpha(x_i;\boldsymbol{\alpha}) - \mathbf{f}'_\alpha(2\mu - x_i;\boldsymbol{\alpha})}{g^*(x_i) + g^*(2\mu - x_i)}. \end{aligned}$$

Let $\mathbf{a}^*(x_i;\boldsymbol{\beta}) = \{a_\mu^*(x_i;\boldsymbol{\beta}), a_p^*(x_i;\boldsymbol{\beta}), \mathbf{a}_{\boldsymbol{\alpha}}^*(x_i;\boldsymbol{\beta})^{\mathrm{T}}\}^{\mathrm{T}}$ and $\mathbf{h}_{\boldsymbol{\beta}}^*(x_i;\boldsymbol{\beta}) = \mathbf{a}^*(x_i;\boldsymbol{\beta}) - \mathbf{r}^*(\boldsymbol{\beta})$.

Step 4: Solve the estimating equation $\sum_{i=1}^n \mathbf{h}_{\boldsymbol{\beta}}^*(x_i;\boldsymbol{\beta}) = \mathbf{0}$ to obtain $\widehat{\boldsymbol{\beta}}$.

The estimating equation in Step 4 can be solved via the standard Newton-Raphson procedure. One nice feature of the above class of estimator is its simplicity. We do not need to estimate the unspecified component η at all. Instead, only a working model is adopted in its place. Thus, nonparametric estimation is completely bypassed. Another advantage of the family is in its richness. Different choice of the working model η^* will yield different estimators for $\boldsymbol{\beta}$. Thus, the family includes many different consistent estimators. Finally, the family is robust, in the sense that regardless of the choice of η^*, the resulting estimator of $\boldsymbol{\beta}$ is always guaranteed to be root-n consistent and has the usual asymptotic normal distribution. We summarize the asymptotic properties of the estimator in Theorem 1 and write the proof in Appendix A.3.

Theorem 1. *Whether or not $\eta^*(t) = \eta(t)$, under the regularity conditions listed in Appendix A.2, the estimator obtained through the above procedure satisfies $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1^{\mathrm{T}}})$ in distribution when $n \to \infty$, where*

$$\mathbf{A} = E(\partial \mathbf{h}_{\boldsymbol{\beta}}^*/\partial \boldsymbol{\beta}^{\mathrm{T}}), \quad \mathbf{B} = E(\mathbf{h}_{\boldsymbol{\beta}}^* \mathbf{h}_{\boldsymbol{\beta}}^{*^{\mathrm{T}}}).$$

Here, all the convergence and expectations are under the measure defined by the true distribution.

Remark 1. In practice, there are various choices of working models for η^*. To determine which working model is favorable, a simple practice is to utilize the result of Theorem 1 to assess the estimation variability. One can then choose the working model that yields the smallest estimation variability, for example, smallest trace of $\text{var}(\widehat{\boldsymbol{\beta}})$, as the favorable working model and proceed with the analysis.

3.3. Extension

In practice, when we select a model η^*, especially when we are selecting a model with the hope that it captures the true pdf η sufficiently well, we may want to leave some aspects of the model unspecified. For example, when we use a normal working model for η^*, we may want to leave the variance of the model unspecified, although the mean is predetermined to be zero due to the symmetry requirement of the working model. This implies that instead of a fully specified function $\eta^*(t)$, we use a model $\eta^*(t;\boldsymbol{\gamma})$. Generally speaking, as long as we can estimate $\boldsymbol{\gamma}$ at a root-n rate, then the resulting $\widehat{\boldsymbol{\beta}}$ is still root-n consistent and asymptotically normal. Of course, the variability might be affected when different $\boldsymbol{\gamma}$ values are used.

As an example, consider the case when γ is the standard deviation of η and is undetermined. The relation

$$E(X^2) = (1-p)\int x^2 f(x;\boldsymbol{\alpha})dx + p(\mu^2+\gamma^2)$$

leads to the estimating equation $\sum_{i=1}^n h_\gamma(x;\boldsymbol{\beta},\gamma) = 0$, where

$$h_\gamma(x;\boldsymbol{\beta},\gamma) = x_i^2 - (1-p)\int x^2 f(x;\boldsymbol{\alpha})dx - p(\mu^2+\gamma^2).$$

This yields an explicit solution

$$\widehat{\gamma}^2(\boldsymbol{\beta}) = \frac{1}{p}\left\{n^{-1}\sum_{i=1}^n x_i^2 - (1-p)\int x^2 f(x;\boldsymbol{\alpha})dx - p\mu^2\right\}.$$

In practice, we can either profile out γ using the above relation or solve for $\boldsymbol{\beta},\gamma$ jointly from $\sum_{i=1}^n \mathbf{h}^*(x;\boldsymbol{\beta},\gamma) = \mathbf{0}$, where $\mathbf{h}^*(x;\boldsymbol{\beta},\gamma) = \{\mathbf{h}^*_{\boldsymbol{\beta}}(x;\boldsymbol{\beta},\gamma)^{\mathrm{T}}, h_\gamma(x;\boldsymbol{\beta},\gamma)\}^{\mathrm{T}}$. Regardless of the computational procedure of obtaining the estimator $\widehat{\boldsymbol{\beta}}$, we have the asymptotic properties of $\widehat{\boldsymbol{\beta}}$ stated in Theorem 2, with its proof given in Appendix A.4.

Theorem 2. *If the estimator $\widehat{\boldsymbol{\beta}}$ and $\widehat{\gamma}$ are obtained by solving $\sum_{i=1}^n \mathbf{h}(x_i;\boldsymbol{\beta},\gamma) = \mathbf{0}$, then under the regularity conditions listed in Appendix A.2, when $n\to\infty$, $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \to N(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1\mathrm{T}})$ in distribution, where $\mathbf{A}$, $\mathbf{B}$ are defined in Theorem 1.*

Remark 2. The result in Theorem 2 implies an unusual property of our estimator. Because the estimation variance of $\widehat{\boldsymbol{\beta}}$ does not change from using the

true parameter value γ to using the parameter value estimated together with $\boldsymbol{\beta}$, it indicates that the estimation of the additional parameter γ does not cost anything. In other words, if we had known the true variance γ^2 of η, and used it in place of $\eta^*(t)$ in forming $\mathbf{h}^*_{\boldsymbol{\beta}}(x;\boldsymbol{\beta},\gamma)$, we will not further improve the estimation efficiency of $\widehat{\boldsymbol{\beta}}$. From the proof of Theorem 2 in the Appendix, we can see that this is a direct result of the robustness property of $\mathbf{h}^*_{\boldsymbol{\beta}}$, in that it has mean zero at the true $\boldsymbol{\beta}$ value regardless which η^* function is used or which scalar parameter of η^* is used.

Remark 3. When the additional parameter $\boldsymbol{\gamma}$ includes parameters other than the standard deviation, for example, when $\boldsymbol{\gamma}$ contains the degrees of freedom in a student t distribution (Liu and Rubin, 1995), a moment type of estimator may not always apply. A simple approach to estimate the general $\boldsymbol{\gamma}$ is through maximum likelihood estimation. Specifically, we can treat $g^*(\boldsymbol{\beta},\boldsymbol{\gamma}) \equiv (1-p)f(x;\boldsymbol{\alpha}) + p\eta^*(x-\mu;\boldsymbol{\gamma})$ as a parametric model, and construct estimator for $\boldsymbol{\gamma}^*$ through maximizing the loglikelihood, in combination with solving $\sum_{i=1}^n \mathbf{h}^*_{\boldsymbol{\beta}}(x_i;\boldsymbol{\beta}) = \mathbf{0}$. Computationally, we can either perform a profile likelihood type of procedure or solve joint estimating equations. Specifically, in the profiling approach, we obtain $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta})$ from maximizing $g^*(\boldsymbol{\beta},\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$, and insert the relation to $\sum_{i=1}^n \mathbf{h}^*_{\boldsymbol{\beta}}(x_i;\boldsymbol{\beta}) = \mathbf{0}$ to obtain $\widehat{\boldsymbol{\beta}}$. In the estimating equation approach, we take derivative of the loglikelihood of the model $g^*(\boldsymbol{\beta},\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$ to obtain a set of estimating equations, and join with $\sum_{i=1}^n \mathbf{h}^*_{\boldsymbol{\beta}}(x_i;\boldsymbol{\beta}) = \mathbf{0}$ to form the complete set of estimating equations. We then solve the complete set of estimating equations to obtain $(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\gamma}})$. Because $g^*(\boldsymbol{\beta},\boldsymbol{\gamma})$ may not be a correct model, the corresponding $\widehat{\boldsymbol{\gamma}}$ converges to a value $\boldsymbol{\gamma}$ that minimizes the Kullback-Leibler distance between $g^*(\boldsymbol{\beta},\boldsymbol{\gamma})$ and $g(\boldsymbol{\beta},\eta)$ (White, 2002). Replacing all the γ instances in the proof of Theorem 2 in Appendix A.4 with $\boldsymbol{\gamma}$, we can obtain the same result as in Theorem 2. Specifically, using $\widehat{\boldsymbol{\gamma}}$, the resulting asymptotic variability of $\widehat{\boldsymbol{\beta}}$ is the same as using $\boldsymbol{\gamma}$.

4. Relation to existing estimator

When f is fully specified and centered, Bordes et al. (2006b) proposed an estimator that minimizes a distance norm between two functions. While they used a general L_q distance in their description, they analysed the L_2 norm distance in their implementation. Thus, we focus on the L_2 norm here. Write the corresponding cdf of f and g as F and G respectively, and use $\widehat{G}(x) = n^{-1}\sum_{i=1}^n I(x_i \le x)$ to indicate the empirical version of $G(x)$. They propose to estimate μ through minimizing the squared distance $\int\{H_1(x;\mu) - H_2(x;\mu)\}^2 dx$ with respect to μ, where

$$
\begin{aligned}
H_1(x;\mu) &= n^{-1}\sum_{i=1}^n \{\mu I(x_i \le \mu + x) + (x_i - \mu)F(\mu + x)\},\\
H_2(x;\mu) &= n^{-1}\sum_{i=1}^n \{x_i - \mu I(x_i \le \mu - x) + (\mu - x_i)F(\mu - x)\},
\end{aligned}
$$

and then obtain $\widehat{p} = n^{-1}\sum_{i=1}^{n} x_i/\widehat{\mu}$. This procedure is equivalent to solving

$$\begin{aligned}\int \{H_1(x;\mu) - H_2(x;\mu)\}\frac{\partial\{H_1(x;\mu) - H_2(x;\mu)\}}{\partial\mu}dx &= 0\\ n^{-1}\sum_{i=1}^{n}(x_i-\mu) - (p-1)\mu &= 0 \qquad (4.1)\end{aligned}$$

jointly. The function $x-\mu$ is certainly a qualified choice of $\mathbf{a}$ in (3.1). In addition, $E\{X-\mu-(p-1)\mu\} = 0$. Hence the second equation in (4.1) belongs to the general family described in Section 3.1. Let $b(x;\mu) = E[\partial\{H_1(x;\mu)-H_2(x;\mu)\}/\partial\mu]$. We write the first equation in (4.1) equivalently as $n^{-1}\sum_{i=1}^{n} a(x_i;\mu) + o_p(n^{-1/2}) = 0$, where

$$\begin{aligned}a(x_i;\mu) &= \int \mu\{I(x_i \le \mu + x) + I(x_i \le \mu - x) - 1\}\, b(x;\mu)dx\\ &\quad + \int (x_i-\mu)\{F(\mu+x) + F(\mu-x) - 1\}\, b(x;\mu)dx.\end{aligned}$$

It is easy to check that $a(x;\mu) + a(2\mu - x;\mu) = 0$. In addition,

$$\begin{aligned}E\{a(X;\mu)\} &= \mu\int\{G(\mu+x) + G(\mu-x) - 1\}\, b(x;\mu)dx\\ &\quad - \mu\int(1-p)\{F(\mu+x) + F(\mu-x) - 1\}\, b(x;\mu)dx\\ &= p\mu\int\left\{\int_{-\infty}^{x}\eta(t)dt + \int_{-\infty}^{-x}\eta(t)dt - 1\right\} b(x;\mu)dx\\ &= 0,\end{aligned}$$

where the first equality used the fact that f is centered, and the last equality used the symmetry of η. This shows that the first estimating equation of (4.1) also belongs to the general family described in Section 3.1. Thus, the estimator by Bordes et al. (2006b) is indeed a special member of our estimator family described in (3.1). As a byproduct, this analysis shows that we can use the result in Theorem 1 to obtain the asymptotic result of the estimator of Bordes et al. (2006b), which is an alternative approach to Bordes and Vandekerkhove (2010).

5. Semiparametric efficient estimator

Having derived a general class of estimators, it is natural to further ask whether this class contains the efficient estimator and how to achieve the optimal efficiency. We thus need to obtain the efficient score functions. To derive the efficient score, we calculate the tangent space with respect to the nuisance function η, denoted Λ and calculate the residual of the projection of the score function $\mathbf{S}_{\boldsymbol{\beta}}$ on Λ. In Appendix A.5 and A.6, we derive the form of Λ and its orthogonal complement $\Lambda^{\perp}$, and further derive the efficient score.

Theorem 3. *Assuming the regularity conditions listed in the Appendix hold. The efficient score is* $\mathbf{S}_{\text{eff}}(x-\mu) = \{S_{\text{eff},\mu}(x-\mu), S_{\text{eff},p}(x-\mu), \mathbf{S}^{\mathrm{T}}_{\text{eff},\boldsymbol{\alpha}}(x-\mu)\}^{\mathrm{T}}$,

where

$$\begin{aligned}
S_{\text{eff}\,\mu}(x-\mu) &= -2p\frac{\eta'(x-\mu)-g(2\mu-x)c_\mu}{g(2\mu-x)+g(x)}\\
S_{\text{eff}\,p}(x-\mu) &= -\frac{f(x)-f(2\mu-x)-g(2\mu-x)c_p}{g(x)+g(2\mu-x)}\\
\mathbf{S}_{\text{eff}\,\boldsymbol{\alpha}}(x-\mu) &= (1-p)\frac{\mathbf{f}'_\alpha(x;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(2\mu-x;\boldsymbol{\alpha})-g(2\mu-x)\mathbf{c}_\alpha}{g(2\mu-x)+g(x)}.
\end{aligned}$$

Here,

$$\begin{aligned}
c_\mu &= E\left\{\frac{\eta'(x-\mu)}{g(x)+g(2\mu-x)}\right\}\left[E\left\{\frac{g(2\mu-x)}{g(x)+g(2\mu-x)}\right\}\right]^{-1},\\
c_p &= E\left\{\frac{f(x)-f(2\mu-x)}{g(x)+g(2\mu-x)}\right\}\left[E\left\{\frac{g(2\mu-x)}{g(x)+g(2\mu-x)}\right\}\right]^{-1},\\
\mathbf{c}_\alpha &= E\left\{\frac{\mathbf{f}'_\alpha(x;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(2\mu-x;\boldsymbol{\alpha})}{g(x)+g(2\mu-x)}\right\}\left[E\left\{\frac{g(2\mu-x)}{g(x)+g(2\mu-x)}\right\}\right]^{-1}.
\end{aligned}$$

Theorem 3 reveals that to achieve optimal efficiency, one is obliged to have the luck of using the true model η as the working model or to estimate the nuisance function η and its derivative η'. In addition, it is crucial to realize that the expectation in the denominators of $c_\mu, c_p, \mathbf{c}_{\boldsymbol{\alpha}}$ need to be computed using numerical approximation, instead of using sample average as an approximation to the expectation. In fact, if we use the sample average, the estimating equation will become a member in the family described in Section 3.2, corresponding to the working model $\eta^* = \eta$ and $g^* = g$, and the efficiency will not be achieved in theory. This is different from many semiparametric models, where replacing an expectation with its estimate via sample average is almost a standard practice. We can estimate $g(x)$ and its first derivative nonparametrically to obtain $\widetilde{g}(x)$ and $\widetilde{g}'(x)$. Taking into account that $\eta'(x)$ is an odd function. we can then obtain $\widehat{\eta}'(x) = [\widetilde{g}'(\mu+x) - \widetilde{g}'(\mu-x) - (1-p)\{f'(\mu+x;\boldsymbol{\alpha}) - f'(\mu-x;\boldsymbol{\alpha})\}]/(2p)$, and further refine the nonparametric estimator of $g(x)$ as $\widehat{g}(x) = \{\widetilde{g}(x) + \widetilde{g}(2\mu - x)\}/2 + (1-p)\{f(x;\boldsymbol{\alpha}) - f(2\mu-x;\boldsymbol{\alpha})\}/2$. We can then form

$$\begin{aligned}
\widehat{c}_\mu &= \kappa^{-1}\int\frac{\widehat{\eta}'(t)\widehat{g}(\mu+t)}{\widehat{g}(\mu-t)+\widehat{g}(\mu+t)}dt = \frac{1-p}{2\kappa}\int\frac{\widehat{\eta}'(t)\{f(\mu+t)-f(\mu-t)\}}{\widehat{g}(\mu-t)+\widehat{g}(\mu+t)}dt,\\
\widehat{c}_p &= \kappa^{-1}\int\frac{\{f(\mu+t)-f(\mu-t)\}\widehat{g}(\mu+t)}{\widehat{g}(\mu-t)+\widehat{g}(\mu+t)}dt\\
&= \frac{1-p}{2\kappa}\int\frac{\{f(\mu+t)-f(\mu-t)\}^2}{\widehat{g}(\mu-t)+\widehat{g}(\mu+t)}dt,\\
\widehat{\mathbf{c}}_\alpha &= \kappa^{-1}\int\frac{\{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})\}\widehat{g}(\mu+t)}{\widehat{g}(\mu-t)+\widehat{g}(\mu+t)}dt\\
&= \frac{1-p}{2\kappa}\int\frac{\{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})\}\{f(\mu+t)-f(\mu-t)\}}{\widehat{g}(\mu-t)+\widehat{g}(\mu+t)}dt,
\end{aligned}$$

where

$$\kappa = \int \frac{\widehat{g}(\mu - t)\widehat{g}(\mu + t)}{\widehat{g}(\mu - t) + \widehat{g}(\mu + t)} dt.$$

Finally, we replace $\eta', g, c_\mu, c_p, \mathbf{c}_{\boldsymbol{\alpha}}$ in $\mathbf{S}_{\text{eff}}$ with $\widehat{\eta}', \widehat{g}, \widehat{c}_\mu, \widehat{c}_p, \widehat{\mathbf{c}}_{\boldsymbol{\alpha}}$ to obtain the efficient score that can be implemented in practice.

A natural choice of the starting value of the efficient estimator is the result from a local estimator. The need to estimate the nuisance function η and its first derivative η' makes the efficient estimator more computationally intensive in comparison with the estimators proposed in Section 3.2 and 3.3. In practice, we have seen the computation time of the efficient estimator about four to five times that of the other estimators. Our experience is that in finite samples, the optimal efficiency often does not exhibit a clear gain. Because of these considerations, unless sample size is very large and one is willing to perform nonparametric estimation, we would recommend carrying out the estimation under a reasonable working model η^*.

6. Numerical analysis

6.1. Simulations

To investigate the finite sample performance of the estimator family, we conducted two simulation studies. In the first simulation, f is completely known and centered, and is set to be the standard normal pdf, while the true η is a student t-distribution with four degrees-of-freedom recentered to $\mu \neq 0$. Since both $f(x)$ and $\eta(x)$ are symmetric about 0, Proposition 1 ensures that model (1.2) is identifiable.

We generated 1000 data sets each with sample size 1000, and experimented with three classes of different estimators. In the first class, we used the normal distribution as a working model. In the second class, we used the student t-distribution with the wrong degrees-of-freedom three as a working model. Finally, in the last class, we used the student t-distribution with the correct degrees of freedom as a working model. In each class, we experimented with a wrong variance parameter in η^*, which is set to be half of the true variance, as well as setting the variance parameter as unknown and estimating it together with the parameters of interest. For comparison, we also implemented the oracle estimator where the true η function is used as a working model, as well as the efficient estimator, where nonparametric estimation is carried out. Since when f is completely known and centered, the symmetrization method (Symm) proposed in Bordes et al. (2006b) can also be used, hence we also included it for comparison.

The results are summarized in Table 1. In all our estimators, the estimated standard deviations are computed from the asymptotic results, while we replace the expectation using the sample average and the true parameter values using their estimates. The confidence interval coverages are based on the asymptotic normality results, where the confidence intervals are constructed using the estimated standard deviations as well. Similarly, we report the estimated standard deviations and confidence intervals for Symm using the asymptotic results of

TABLE 1

Simulation 1 results. The true value (μ, p), the average estimates ($\widehat{\mu}$, $\widehat{p}$), the sample standard errors ("sd"), the mean of the estimated standard errors ($\widehat{\text{sd}}$) and the 95% confidence interval of 9 different estimators are reported. "Normal", "t_3", "t_4" are the different working models η^*. Either a fixed wrong variance (fix var) or an estimated variance (est var) is used in the working model η^*

	Normal		t_3		t_4		oracle	eff	Symm
	fix var	est var	fix var	est var	fix var	est var			
μ	3	3	3	3	3	3	3	3	3
$\widehat{\mu}$	2.9955	2.9943	2.9936	2.9976	2.9944	2.9976	2.9962	2.9979	2.9897
sd	0.0850	0.1041	0.1048	0.0839	0.0947	0.0833	0.0832	0.0885	0.0966
$\widehat{\text{sd}}$	0.0842	0.1539	0.1047	0.0846	0.0943	0.0832	0.0823	0.0916	0.0826
CI	94.2%	94.8%	92.1%	94.1%	92.7%	94.4%	94.4%	92.6%	94.2%
p	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
$\widehat{p}$	0.4017	0.4021	0.4021	0.4012	0.4019	0.4012	0.4015	0.4026	0.4020
sd	0.0216	0.0232	0.0259	0.0214	0.0237	0.0213	0.0212	0.0211	0.0242
$\widehat{\text{sd}}$	0.0218	0.0309	0.0260	0.0216	0.0237	0.0213	0.0212	0.0221	0.0217
CI	94.7%	93.9%	94.8%	94.5%	95.0%%	94.9%	94.7%	95.0%	95.1%

Bordes and Vandekerkhove (2010). From the table, we can see that our proposed method gives reasonable estimates regardless of whether the working model is correct or wrong. When the working model is close to the true model and the estimated variance is used, the proposed method yields better performance than the symmetrization method and provides similar results to the oracle one and the proposed efficient estimator. In addition, the estimated standard errors and the sample standard errors are reasonably close in all situations. Furthermore, the coverage percentages of the estimated confidence intervals based on the proposed methods are very close to the nominal level, which demonstrates the effectiveness of the inference tools of the proposed methods.

Our second simulation extends the first one by allowing the scale parameter of f to be unknown and treating its logarithm as the α parameter. Based on (A.1), we can see that model (1.2) is still identifiable in this case. We implemented three classes of estimators, as well as the oracle estimator and the efficient estimator as in simulation one, and reported the results in Table 2. Because Bordes et al. (2006b) assumes that the first component density f is completely known, the symmetrization method is not applicable in this case. From the table, we can see that our proposed method provides reasonable estimates for all working models. In addition, the estimated standard errors and the average coverage percentages of the constructed confidence interval are all close to the their nominal values.

A by-produce of the efficient estimator is the density estimation of both the nuisance function $\eta(x)$ and the mixture density $g(x)$. We provide the median estimated curves as well as the 95% confidence bands from both simulations in Figure 1.

It is often observed in semiparametric models that although an efficient estimator is asymptotically optimal and should perform as well as the oracle estimator, its practical gain over some other sub-optimal estimators can sometimes be substantial, sometimes not, and its performance may or may not be sufficiently close to that of the oracle estimator. This is mainly due to the difference between

TABLE 2

Simulation 2 results. The true value (μ, p, α), the average estimates ($\widehat{\mu}$, $\widehat{p}$, $\widehat{\alpha}$), the sample standard errors ("sd"), the mean of the estimated standard errors ($\widehat{\text{sd}}$) and the 95% confidence interval of 8 different estimators are reported. "Normal", "t_3", "t_4" are the different working models η^*. Either a fixed wrong variance (fix var) or an estimated variance (est var) is used in the working model η^*

	Normal		t_3		t_4		oracle	eff
	fix var	est var	fix var	est var	fix var	est var		
μ	3	3	3	3	3	3	3	3
$\widehat{\mu}$	2.9854	2.9911	2.9933	2.9966	2.9930	2.9960	2.9939	2.9960
sd	0.1112	0.1146	0.1153	0.0875	0.1024	0.0861	0.0858	0.0925
$\widehat{\text{sd}}$	0.1083	0.1810	0.1153	0.0888	0.1021	0.0873	0.0863	0.0993
CI	93.9%	95.5%	91.9%	93.3%	92.4%	94.3%	94.2%	93.0%
p	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
$\widehat{p}$	0.4039	0.4032	0.4023	0.4015	0.4024	0.4017	0.4022	0.4032
sd	0.0346	0.0259	0.0297	0.0230	0.0266	0.0227	0.0227	0.0231
$\widehat{\text{sd}}$	0.0290	0.0422	0.0296	0.0231	0.0264	0.0228	0.0227	0.0243
CI	94.8%	96.0%	93.6%	94.4%	94.6%	95.1%	94.7%	94.6%
α	0	0	0	0	0	0	0	0
$\widehat{\alpha}$	-0.0141	-0.0044	-0.0018	-0.0022	-0.0028	-0.0026	-0.0031	-0.0069
sd	0.1081	0.0470	0.0484	0.0401	0.0450	0.0398	0.0399	0.0414
$\widehat{\text{sd}}$	0.0825	0.1075	0.0493	0.0411	0.0454	0.0409	0.0409	0.0422
CI	95.8%	96.2%	94.8%	95.5%	94.1%	95.5%	95.7%	95.1%

finite sample performance and asymptotic properties. In practice, the bias and variance in estimating the nuisance parameters, such as η here, may have an effect in the second or higher order terms, and in finite samples, the second or higher order effect can sometimes overwhelm the first order property. The finite sample performance also depends on the problem setting. The same estimator may perform well in one problem while not as well in another problem, although theoretically, the property of the estimator is the same in both models.

6.2. Real data analysis

We further apply the new estimation procedure to three data examples to demonstrate the applications of the semiparametric mixture model (1.2) in sequential clustering problems (Song and Nicolae, 2009) and in multiple testing problems. Because of the superiority of the efficient estimator demonstrated in the previous sections, for simplicity of the presentation, we will only report the results from the efficient estimator.

To illustrate the application of the proposed estimation procedure to sequential clustering algorithm, we apply the proposed method to the well known Iris flower data. The Iris flower data has been analyzed by Fisher and many other researchers and is a popular benchmark data for clustering and classification applications. The data contains four attributes: sepal length (in cm), sepal width (in cm), petal length (in cm), and petal width (in cm). There are 3 classes of 50 instances each, where each class refers to a type of Iris plant.

Principal component analysis shows that the first principal component accounts for 92.46% of the total variability, so it is a good practice to con-

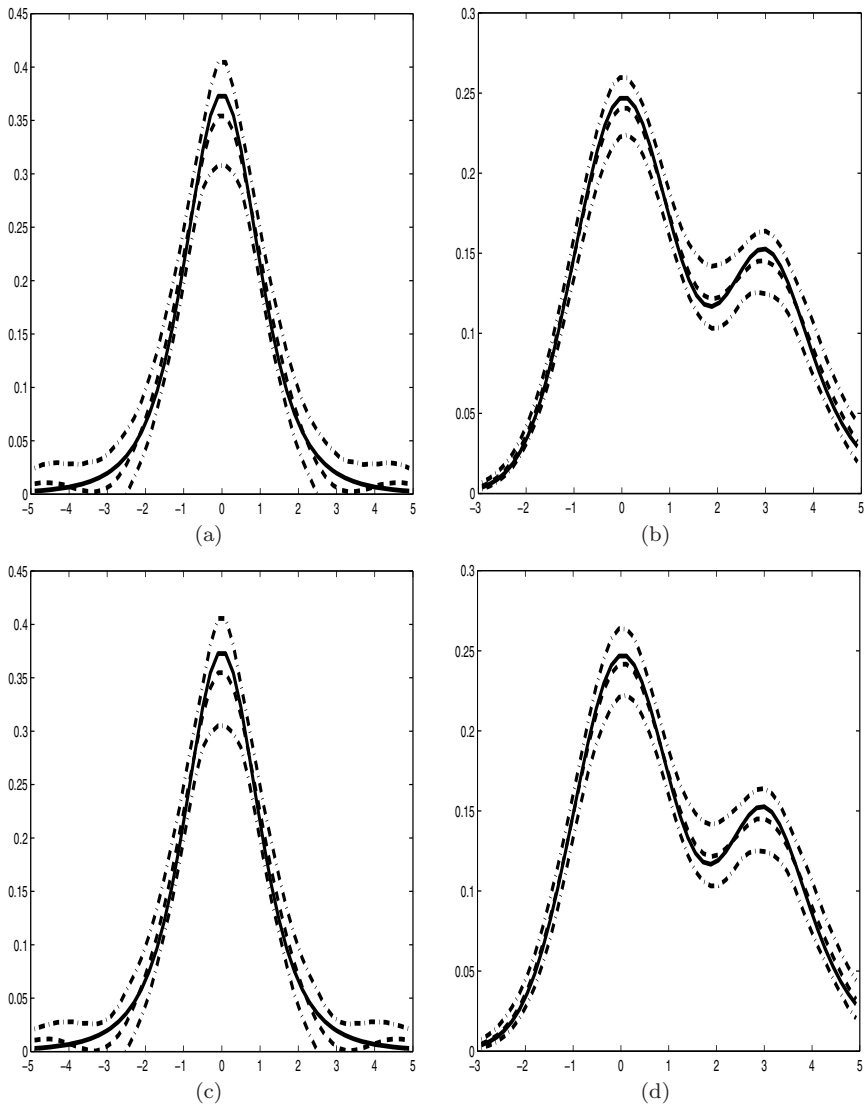

FIG 1. *True function (solid line), median estimation (dashed line) and 95% confidence band (dash-dotted line) of $\eta(x)$ (left) and $g(x)$ (right) in simulation 1 (upper) and simulation 2 (lower).*

sider the one-dimensional subspace of the four-dimensional sample space represented by the first component. The first principal component loading vector is $(0.36, -0.08, 0.86, 0.35)$.

Suppose we want to perform clustering based on the first principal component of the Iris flower data without using the class indictors. Song and Nicolae (2009) proposed a sequential clustering algorithm to perform clustering by find-

TABLE 3
Parameter estimates and their estimated standard errors ("sd") for Iris flower data analysis

	μ	p	α
True	3.9469	0.6667	-1.5105
Eff	3.9602	0.6832	-1.7180
$\widehat{\text{sd}}$	0.2489	0.0457	0.3264

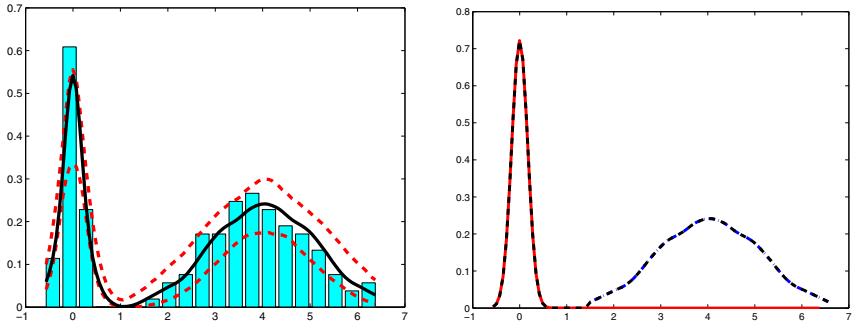

FIG 2. *Left panel: plot of the estimated density function (solid line) with 95% confidence band (dashed line) of $g(x)$ overlay with the histogram of the data. Right panel: plot of null (solid line) and non-null components (dash-dotted line) (weighted respectively by $(1-\hat{p})$ and $\hat{p}$) overlay with $\hat{g}(x)$ (dashed line) in the Iris data analysis.*

ing clusters sequentially. The sequential clustering algorithm does not require specifying the number of clusters and it allows some objects not to be assigned to any clusters. The algorithm starts with finding a local center of a cluster first, and then identifies whether an object belongs to that cluster or not. It iterates the above procedure until no new cluster is found. Based on Song and Nicolae (2009), observation 8 is selected as the center of the first cluster. We adjust all observations by subtracting observation 8 from each observation. The first cluster can then be considered as one component that has normal density with centered mean 0 and unknown variance, and the rest of data can be considered from the other mixture component with unknown density. Therefore, by fitting the semiparametric mixture model (1.2), we can classify whether each observation belongs to the first cluster or not. One might also use a mixture of normals to approximate the nonparametric component $\eta(\cdot)$. However, using the semiparametric mixture model (1.2) avoids the selection of the number of mixture components.

The results of the analysis on the Iris data are summarized in Table 3. The parameter α is the logarithm of the standard deviation of the first component. Noting that the true proportion is 2/3 and the true values for μ and α are calculated using the class indicators. From Table 3, we can see that the proposed estimator works quite well. We also plotted the estimated mixture distribution density, along with the confidence intervals in the left panel of Figure 2. The

TABLE 4
Parameter estimates and their estimated standard errors ("sd") for Breast cancer data analysis

	μ	p
Eff	1.7007	0.3137
$\widehat{\text{sd}}$	0.1084	0.0285

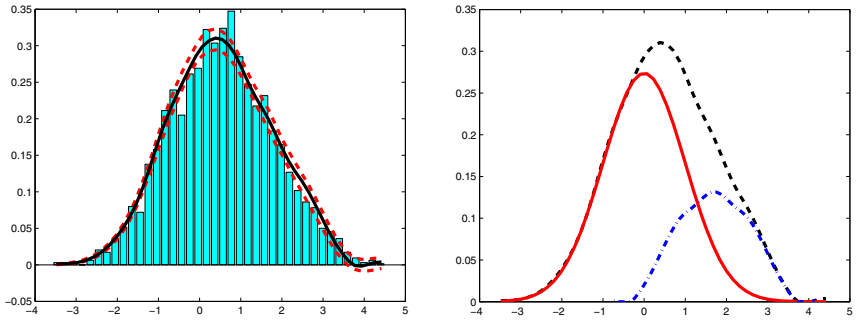

FIG 3. *Left panel: plot of the estimated density function (solid line) with 95% confidence band (dashed line) of $g(x)$ overlay with the histogram of the data. Right panel: plot of null (solid line) and non-null components (dash-dotted line) (weighted respectively by $(1-\hat{p})$ and $\hat{p}$) overlay with $\hat{g}(x)$ (dashed line) in the breast cancer data analysis.*

right panel of Figure 2 provides the corresponding estimated densities $\widehat{g}(x)$, $(1-\widehat{p})f(x)$ and $\widehat{p}\widehat{\eta}(x-\widehat{\mu})$ functions. In this data set, f and $\widehat{\eta}$ are clearly separated and $\widehat{g}$ has two distinctive modes. In particular, $\widehat{g}$ overlays with $\widehat{p}\widehat{\eta}(x-\widehat{\mu})$ and $(1-\widehat{p})f(x,\widehat{\mu})$ completely in two regions, corresponding to the two distinctive mixture components.

To illustrate the application of the new method in multiple hypothesis testing, we consider the detection of differentially expressed genes based on the breast cancer data of Hedenfalk et al. (2001). They examined gene expressions in breast cancer tissues from women who were carriers of the hereditary BRCA1 or BRCA2 gene mutations, predisposing to breast cancer. The data consist of 3,226 genes on 7 BRCA1 arrays and 8 BRCA2 arrays. Based on Storey and Tibshirani (2003), if any gene had one or more measurements exceeding 20, then this gene was eliminated. This leaves us 3,170 genes. The p-values are calculated based on permutation tests (Storey and Tibshirani, 2003). We then obtain the z-scores by the probit transformation of the p-values, given by $z_i = \Phi^{-1}(1-p_i)$ (McLachlan et al., 2006).

The results are reported in Table 4 and Figure 3. Specifically, Figure 3 indicates an adequate description of the data using the semiparametric mixture model (1.2). Based on Table 4, the proportion of genes satisfying the alternative hypothesis is around 31%, which is consistent with the results reported in Langaas et al. (2005). We now further explain how to perform multiple hypothesis

TABLE 5
Estimated FDR for various levels of the threshold c applied to the posterior probability of nondifferentially expression for the breast cancer data

c	N_r	$\widehat{FDR}$
0.05	69	0.0413
0.1	186	0.0607
0.2	342	0.1005
0.3	512	0.1496
0.4	689	0.2015
0.5	854	0.2499

testing using the fitted semiparametric mixture model (1.2). Let

$$\hat{\tau}_i = (1-\hat{p})\hat{f}(z_i)/\{(1-\hat{p})\hat{f}(z_i) + \hat{p}\hat{\eta}(z_i - \hat{\mu})\}$$

be the classification probability that the ith gene is not differentially expressed. The gene-specific posterior probabilities $\hat{\tau}_i$ is also referred to as the local false discovery rate (local FDR) by Efron and Tibshirani (2002) and can be viewed as an empirical Bayes version of the Benjamini and Hochberg (1995) methodology (Efron, 2004). We can select all genes with

$$\hat{\tau}_i \leq c$$

to be differentially expressed. The cut point c can be selected by controlling the false discovery rate (FDR)(Benjamini and Hochberg, 1995). Based on McLachlan et al. (2006), the FDR can be estimated by

$$\widehat{FDR} = \frac{1}{N_r}\sum_i \hat{\tau}_i I_{[0,c]}(\hat{\tau}_i),$$

where $N_r = \sum_i I_{[0,c]}(\hat{\tau}_i)$ is the total number of differentially expressed genes identified, and $I_A(x)$ is an indicator function which is one if $x \in A$ and zero otherwise. The results are reported in Table 5 based on the proposed efficient estimator. For example, if $c = 0.05$, then the estimated FDR is 0.04 and $N_r = 69$ genes would be declared to be differentially expressed. If the threshold value c is increased to 0.1 then $N_r = 186$ genes would be declared to be differentially expressed with the estimated FDR increased to 0.06. Compared to $c = 0.05$, $c = 0.1$ might be a better choice since it can detect almost three times as many differentially expressed genes with only slightly larger estimated FDR. McLachlan and Wockner (2010) applied the normal mixture model on this data by assuming a normal distribution for $\eta(\cdot)$ and obtained similar results.

As a last example, we apply the proposed estimation procedure to a lipid metabolism data of Callow et al. (2000), where the effect of knocking out the gene apolipoprotein AI gene were investigated. Smyth et al. (2005) analyzed this data set based on the theory presented in Smyth (2004). The data consist

TABLE 6
Parameter estimates and their estimated standard errors ("sd") for Lipid metabolism data analysis

	μ	p
Eff	0.9895	0.1275
$\widehat{\text{sd}}$	0.0365	0.0141

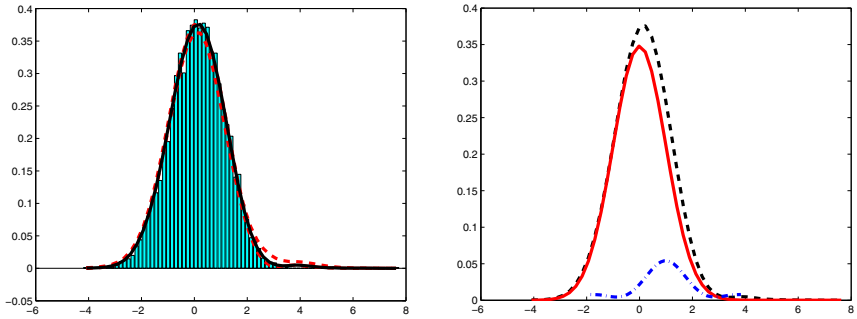

FIG 4. *Left panel: plot of the estimated density function (solid line) with 95% confidence band (dashed line) of $g(x)$ overlay with the histogram of the data. Right panel: plot of null (solid line) and non-null components (dash-dotted line) (weighted respectively by $(1-\hat{p})$ and $\hat{p}$) overlay with $\hat{g}(x)$ (dashed line) in the lipid metabolism data analysis.*

of 6,384 genes and the $p-$values were calculated based on the comparison of knockout mice with normal mice. We then obtain the z-scores given by $z_i = \Phi^{-1}(1-p_i)$ and fit the semiparametric model (1.2) on the z-scores.

The results are reported in Table 6 and Figure 4. Based on Figure 4, we can see that the semiparametric mixture model (1.2) fit the data very well. Compared to the breast cancer data analysis, lipid metabolism data has much smaller proportion (13%) of genes that satisfy the alternative hypothesis. The estimate of p is consistent with the four estimates introduced in Langaas et al. (2005).

In addition, similar to the previous example, we report the multiple hypothesis testing results by controlling the FDR in Table 7. For example, if $c = 0.55$, then $N_r = 11$ genes would be declared to be differentially expressed with the estimated FDR=0.0814; if $c = 0.3$, then $N_r = 10$ genes would be declared to be differentially expressed with a much smaller estimated FDR value (0.0346). In this example, compared to $c = 0.55$, $c = 0.3$ might be a better choice since it only detects one less differentially expressed gene but has much smaller estimated FDR.

7. Discussion

In this article, we proposed a new class of estimators for a two-component semiparametric mixture model where one component distribution belongs to a para-

TABLE 7
Estimated FDR for various levels of the threshold c applied to the posterior probability of nondifferentially expression for the lipid metabolism data

c	N_r	$\widehat{FDR}$
0.1	8	0.0004
0.2	9	0.0157
0.3	10	0.0346
0.55	11	0.0814
0.57	12	0.1217
0.58	13	0.1564

metric class, while the other is symmetric with unknown center but otherwise arbitrary. The semiparametric model can be used for large-scale simultaneous testing/multiple testing, sequential clustering, or robust modeling. The simulation studies and real data applications demonstrate the effectiveness of the proposed methods.

It will be interesting to know whether the proposed estimators can be extended to some other semiparametric mixture models, such as model (1.1). In addition, it will also be interesting to know whether the proposed estimators can be extended to the semiparametric mixture of regression models where one component is known while the other component is unknown (Vandekerkhove, 2012) or all the component error densities are assumed to be unknown (Hunter and Young, 2012).

Our proposed estimators make use of the symmetry nature of η, which is also assumed by Bordes et al. (2006b). However, in some applications, it might be infeasible to impose the symmetry assumption on η. Of course, identifiability of model (1.2) will then require other structures to be imposed on η. Estimation procedures will inevitably rely on the specific structure, and will need to be studied in each case.

Acknowledgment

The authors thank the editor, the associate editor, and reviewers for their constructive comments that have led to a dramatic improvement of the earlier version of this article. The research is supported by NSF grants DMS-1206693, 1461677 and NIH grant R01-NS073671. The authors also thank Dr. Vandekerkhove and Dr. Bordes for generously providing us the R codes they used in their paper Bordes and Vandekerkhove (2010).

Appendix

A.1. Proof of Proposition 1

We first give two Lemmas which will be used later.

Lemma 1. *If $\eta(x)$ is symmetric about 0, then model (1.2) is non-identifiable if and only if there exist vectors $(\widetilde{p}, \widetilde{\mu}, \widetilde{\boldsymbol{\alpha}}) \neq (p, \mu, \boldsymbol{\alpha})$ such that*

$$\begin{aligned} &\eta(x-\widetilde{\mu}+\mu)-\eta(x+\widetilde{\mu}-\mu) \qquad \text{(A.1)}\\ &=\frac{1-p}{p}\{f(x+\widetilde{\mu};\boldsymbol{\alpha})-f(-x+\widetilde{\mu};\boldsymbol{\alpha})\}-\frac{1-\widetilde{p}}{p}\{f(x+\widetilde{\mu};\widetilde{\boldsymbol{\alpha}})-f(-x+\widetilde{\mu};\widetilde{\boldsymbol{\alpha}})\} \end{aligned}$$

for all $x \in \mathbb{R}$.

Proof. Suppose the model (1.2) is non-identifiable and there is another set of parameter values $\widetilde{p}, \widetilde{\mu}, \widetilde{\boldsymbol{\alpha}}, \widetilde{\eta}$ such that it yields the same pdf $g(x)$, i.e.

$$(1-\widetilde{p})f(x;\widetilde{\boldsymbol{\alpha}})+\widetilde{p}\widetilde{\eta}(x-\widetilde{\mu})=(1-p)f(x;\boldsymbol{\alpha})+p\eta(x-\mu)$$

for all x. This implies

$$\widetilde{p}\widetilde{\eta}(x)=(1-p)f(x+\widetilde{\mu};\boldsymbol{\alpha})-(1-\widetilde{p})f(x+\widetilde{\mu};\widetilde{\boldsymbol{\alpha}})+p\eta(x+\widetilde{\mu}-\mu).$$

If $(\widetilde{p}, \widetilde{\mu}, \widetilde{\boldsymbol{\alpha}}) = (p, \mu, \boldsymbol{\alpha})$, then $\eta(x) = \widetilde{\eta}(x)$ and the model (1.2) is still identifiable. Therefore, $(\widetilde{p}, \widetilde{\mu}, \widetilde{\boldsymbol{\alpha}}) \neq (p, \mu, \boldsymbol{\alpha})$. Note that $\widetilde{\eta}$ is symmetric about 0, hence we have

$$\begin{aligned} &(1-p)f(x+\widetilde{\mu};\boldsymbol{\alpha})-(1-\widetilde{p})f(x+\widetilde{\mu};\widetilde{\boldsymbol{\alpha}})+p\eta(x+\widetilde{\mu}-\mu)\\ &=(1-p)f(-x+\widetilde{\mu};\boldsymbol{\alpha})-(1-\widetilde{p})f(-x+\widetilde{\mu};\widetilde{\boldsymbol{\alpha}})+p\eta(x-\widetilde{\mu}+\mu). \end{aligned}$$

Recall that p and $\tilde{p}$ belong to $]0, 1[$. This leads to

$$\begin{aligned} &\eta(x-\widetilde{\mu}+\mu)-\eta(x+\widetilde{\mu}-\mu)\\ &=(\frac{1}{p}-1)\{f(x+\widetilde{\mu};\boldsymbol{\alpha})-f(-x+\widetilde{\mu};\boldsymbol{\alpha})\}-(\frac{1}{p}-\frac{\widetilde{p}}{p})\{f(x+\widetilde{\mu};\widetilde{\boldsymbol{\alpha}})-f(-x+\widetilde{\mu};\widetilde{\boldsymbol{\alpha}})\}. \end{aligned}$$

All the above statements are reversible. Therefore, the non-identifiable condition is necessary and sufficient condition. □

Lemma 2. *When f is completely known, i.e., $\boldsymbol{\alpha}$ does not appear, (1.2) reduces to the model considered by Bordes et al. (2006b). In this special case, for a symmetric $\eta(x)$, the condition in Lemma (1) can be simplified to*

$$\eta(x-\widetilde{\mu}+\mu)-\eta(x+\widetilde{\mu}-\mu)=r\{f(x+\widetilde{\mu})-f(-x+\widetilde{\mu})\} \qquad \text{(A.2)}$$

for all $x \in \mathbb{R}$, where $\widetilde{\mu} \neq 0$, $r \neq 0$ and $-1 < r < (1-p)/p$.

Proof. If model (1.2) is non-identifiable, based on Lemma (1), we have

$$(\widetilde{p}-p)\{f(x+\widetilde{\mu})-f(-x+\widetilde{\mu})\}=p\{\eta(x-\widetilde{\mu}+\mu)-\eta(x+\widetilde{\mu}-\mu)\}.$$

If $p = \widetilde{p}$, then $\mu = \widetilde{\mu}$ and $\widetilde{\eta}(x) = \eta(x)$ and thus the model is identifiable. Therefore, $p \neq \widetilde{p}$ and

$$\eta(x-\widetilde{\mu}+\mu)-\eta(x+\widetilde{\mu}-\mu)=r\{f(x+\widetilde{\mu})-f(-x+\widetilde{\mu})\},$$

where $-1 < r = (\widetilde{p}-p)/p < (1-p)/p$ and $r \neq 0$. □

Next, we prove Proposition 1.

Based on the proof in Lemma 2, if there is another set of parameter values $\widetilde{p}, \widetilde{\mu}, \widetilde{\eta}$ such that

$$(1-\widetilde{p})f(x) + \widetilde{p}\widetilde{\eta}(x-\widetilde{\mu}) = (1-p)f(x) + p\eta(x-\mu)$$

for all x, then $r = (\widetilde{p}-p)/p \neq 0$ is required for the non-identifiability of model (1.2). This implies $p \neq \widetilde{p}$. Therefore, if p is estimable, the model (1.2) is identifiable.

Note that $g(x) = (1-p)f(x) + p\eta(x-\mu)$. If $f(x) > 0$ and $\lim_{x\to\infty} \frac{\eta(x-\delta)}{f(x)} = 0$ then $p = 1 - \lim_{x\to\infty} g(x)/f(x)$. If $\lim_{x\to-\infty} \frac{\eta(x-\delta)}{f(x)} = 0$, then $p = 1 - \lim_{x\to-\infty} g(x)/f(x)$. Therefore, p is identified and model (1.2) is identifiable.

Similarly, we can prove that model (1.2) is identifiable if $\eta(x) > 0$ and $\lim_{x\to\infty} \frac{f(x)}{\eta(x-\delta)} = 0$ or $\lim_{x\to-\infty} \frac{f(x)}{\eta(x-\delta)} = 0$.

A.2. Regularity conditions

C1. $f(\cdot, \boldsymbol{\alpha})$ is continuous and twice differentiable with respect to $\boldsymbol{\alpha}$.
C2. $\eta(\cdot)$ and $\eta^*(\cdot)$ are continuous and twice differentiable with respect to all the function arguments.
C3. The expectations of $\partial \mathbf{h}^*_{\boldsymbol{\beta}}/\partial \boldsymbol{\beta}^{\mathrm{T}}$ and $\mathbf{h}^*_{\boldsymbol{\beta}} \mathbf{h}^{*\,\mathrm{T}}_{\boldsymbol{\beta}}$ exist, are bounded and nonsingular.

A.3. Proof of Theorem 1

Standard Taylor expansion yields

$$\begin{aligned}
\mathbf{0} &= n^{-1/2} \sum_{i=1}^{n} \mathbf{h}^*_{\boldsymbol{\beta}}(x_i, \widehat{\boldsymbol{\beta}}) \\
&= n^{-1/2} \sum_{i=1}^{n} \mathbf{h}^*_{\boldsymbol{\beta}}(x_i, \boldsymbol{\beta}) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \mathbf{h}^*_{\boldsymbol{\beta}}(x_i, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&= n^{-1/2} \sum_{i=1}^{n} \mathbf{h}^*_{\boldsymbol{\beta}}(x_i, \boldsymbol{\beta}) + \left[E\left\{ \frac{\partial \mathbf{h}^*_{\boldsymbol{\beta}}(X_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \right\} + o_p(1) \right] \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}),
\end{aligned}$$

where $\boldsymbol{\beta}^*$ is a point on the line connecting $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$. Thus, we obtain

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = -\mathbf{A}^{-1} n^{-1/2} \sum_{i=1}^{n} \mathbf{h}^*_{\boldsymbol{\beta}}(x_i, \boldsymbol{\beta}) + o_p(1).$$

Making use of the properties that $\mathbf{a}(x; \boldsymbol{\beta}) + \mathbf{a}(2\mu - x; \boldsymbol{\beta}) = \mathbf{0}$, and η is symmetric, we can easily verify that

$$\begin{aligned}
& E\{\mathbf{h}^*_{\boldsymbol{\beta}}(X; \boldsymbol{\beta})\} \\
= & \int \mathbf{a}^*(x+\mu; \boldsymbol{\beta}) g(x+\mu) dx - \mathbf{r}^*(\boldsymbol{\beta})
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty \mathbf{a}^*(x+\mu;\boldsymbol{\beta})g(x+\mu)dx + \int_0^\infty \mathbf{a}^*(\mu-x;\boldsymbol{\beta})g(\mu-x)dx - \mathbf{r}^*(\boldsymbol{\beta}) \\
&= \int_0^\infty \mathbf{a}^*(x+\mu;\boldsymbol{\beta})\{g(x+\mu)-g(\mu-x)\}dx - \mathbf{r}^*(\boldsymbol{\beta}) \\
&= (1-p)\int_0^\infty \mathbf{a}^*(x+\mu;\boldsymbol{\beta})\{f(x+\mu;\boldsymbol{\alpha})-f(\mu-x;\boldsymbol{\alpha})\}dx - \mathbf{r}^*(\boldsymbol{\beta}) \\
&= \frac{1-p}{2}\int \mathbf{a}^*(x+\mu;\boldsymbol{\beta})\{f(x+\mu;\boldsymbol{\alpha})-f(\mu-x;\boldsymbol{\alpha})\}dx - \mathbf{r}^*(\boldsymbol{\beta}) \\
&= \mathbf{0}
\end{aligned}
$$

due to the definition of $\mathbf{r}^*(\boldsymbol{\beta})$. Here in the second last equality, we use the fact that the integrant is an even function of x. Hence we obtain

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \to N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1^{\mathrm{T}}}\right)$$

in distribution when $n \to \infty$. □

A.4. Proof of Theorem 2

Because $E\{\mathbf{h}^*_{\boldsymbol{\beta}}(X;\boldsymbol{\beta})\} = \mathbf{0}$ for η^* function, it is certainly true that $E\{\mathbf{h}^*_{\boldsymbol{\beta}}(X;\boldsymbol{\beta},\boldsymbol{\gamma})\} = \mathbf{0}$. Because $E(h_\gamma(\mathbf{X};\boldsymbol{\beta},\gamma)\} = 0$ as well, so we obtain $E\{\mathbf{h}^*(X;\boldsymbol{\beta},\gamma)\} = \mathbf{0}$. Using the same derivation as in the proof of Theorem 1, we thus obtain $(\boldsymbol{\beta}^{\mathrm{T}},\gamma)^{\mathrm{T}} \to N(\mathbf{0}, \mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1^{\mathrm{T}}})$ in distribution when $n \to \infty$. Here $\mathbf{C} = \{E(\partial\mathbf{h}^*/\partial\boldsymbol{\beta}^{\mathrm{T}}), E(\partial\mathbf{h}^*/\partial\gamma)\}$, $\mathbf{D} = E(\mathbf{h}^*\mathbf{h}^{*^{\mathrm{T}}})$. To extract the variance of $\widehat{\boldsymbol{\beta}}$, we note that because $E\{\mathbf{h}^*_{\boldsymbol{\beta}}(X;\boldsymbol{\beta},\gamma)\} = \mathbf{0}$ for all γ, hence $E\{\partial\mathbf{h}^*_{\boldsymbol{\beta}}(X;\boldsymbol{\beta},\gamma)/\partial\gamma\} = \mathbf{0}$. Thus,

$$
\begin{aligned}
\mathbf{C} &= \begin{bmatrix} E\left\{\partial\mathbf{h}^*_{\boldsymbol{\beta}}(X;\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\boldsymbol{\beta}^{\mathrm{T}}\right\} & \mathbf{0} \\ E\left\{\partial h_\gamma(X;\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\boldsymbol{\beta}^{\mathrm{T}}\right\} & E\left\{\partial h_\gamma(X;\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\gamma\right\} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ E\left\{\partial h_\gamma(X;\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\boldsymbol{\beta}^{\mathrm{T}}\right\} & E\left\{\partial h_\gamma(X;\boldsymbol{\beta},\boldsymbol{\gamma})/\partial\gamma\right\} \end{bmatrix},
\end{aligned}
$$

where $\mathbf{A}$ is defined in Theorem 1. Note that that upper left $d \times d$ block of $\mathbf{D}$ is $\mathbf{B}$ defined in Theorem 1. Thus, simple matrix calculation directly yields that the upper left $d \times d$ block of $\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1^{\mathrm{T}}}$ matrix is $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1^{\mathrm{T}}}$, hence the result is shown. □

A.5. Derivation of Λ and $\Lambda^\perp$

To simplify notation, we omit the parameters here whenever it does not cause confusion. We show that

$$\Lambda = \left\{\frac{\mathbf{a}(x-\mu)}{g(x)} : \mathbf{a}(t) = \mathbf{a}(-t), \int \mathbf{a}(t)dt = \mathbf{0}\right\},$$

$$\Lambda^{\perp} = \left\{ \mathbf{b}(x-\mu) : \mathbf{b}(t) + \mathbf{b}(-t) + \frac{\int_0^\infty \mathbf{b}(t)\{g(\mu+t) - g(\mu-t)\}dt}{\int_0^\infty g(\mu-t)dt} = \mathbf{0} \right\}.$$

Proof. To obtain Λ, we can easily obtain that elements in Λ must be of the form $\mathbf{a}/g(x)$, where $\mathbf{a}$ has to be even and integrates to zero. This implies Λ.

To prove the form of $\Lambda^{\perp}$, use A to denote the set in the right hand side of the above expression for $\Lambda^{\perp}$. First, we show $A \subset \Lambda^{\perp}$. It is easy to check that any element $\mathbf{b}(x-\mu) \in A$ satisfies $\int \mathbf{a}(t)\mathbf{b}^{\mathrm{T}}(t)dt = \mathbf{0}$. Further,

$$\begin{aligned}
& \int \mathbf{b}(x-\mu)g(x)dx \\
= & \int \mathbf{b}(t)g(\mu+t)dt \\
= & \int_0^\infty \mathbf{b}(t)g(\mu+t)dt + \int_0^\infty \mathbf{b}(-t)g(\mu-t)dt \\
= & \int_0^\infty \mathbf{b}(t)g(\mu+t)dt - \int_0^\infty \mathbf{b}(t)g(\mu-t)dt \\
& - \int_0^\infty g(\mu-t)dt \frac{\int_0^\infty \mathbf{b}(t)\{g(\mu+t) - g(\mu-t)\}dt}{\int_0^\infty g(\mu-t)dt} \\
= & \int_0^\infty \mathbf{b}(t)\{g(\mu+t)dt - g(\mu-t)\}dt - \int_0^\infty \mathbf{b}(t)\{g(\mu+t) - g(\mu-t)\}dt \\
= & \mathbf{0}.
\end{aligned}$$

Hence $\mathbf{b}(x-\mu) \in \Lambda^{\perp}$. This shows $A \in \Lambda^{\perp}$. We now show $\Lambda^{\perp} \subset A$. If $\mathbf{b}(x-\mu) \in \Lambda^{\perp}$, then

$$\mathbf{0} = \int \mathbf{a}(x-\mu)\mathbf{b}^{\mathrm{T}}(x-\mu)dx = \int \mathbf{a}(t)\mathbf{b}^{\mathrm{T}}(t)dt$$

for $\mathbf{a} \in \Lambda$. Since $\mathbf{a}$ is even, this implies

$$\mathbf{0} = \int \mathbf{a}(-t)\mathbf{b}^{\mathrm{T}}(-t)dt = \int \mathbf{a}(t)\mathbf{b}^{\mathrm{T}}(-t)dt$$

as well. Summing the above two displays, we have

$$\mathbf{0} = \int \mathbf{a}(t)\{\mathbf{b}^{\mathrm{T}}(t) + \mathbf{b}^{\mathrm{T}}(-t)\}dt.$$

Now let $\mathbf{a}(t) = \mathbf{b}(t) + \mathbf{b}(-t) - \int\{\mathbf{b}(t) + \mathbf{b}(-t)\}dt \in \Lambda$, we have

$$\int \{\mathbf{b}(t) + \mathbf{b}(-t)\}\{\mathbf{b}(t) + \mathbf{b}(-t)\}^{\mathrm{T}}dt = \int \{\mathbf{b}(t) + \mathbf{b}(-t)\}dt \int \{\mathbf{b}(t) + \mathbf{b}(-t)\}^{\mathrm{T}}dt.$$

Hence $\mathbf{b}(t) + \mathbf{b}(-t)$ is a constant, say $-\mathbf{c}$. Combined with $\int \mathbf{b}(t)g(\mu+t)dt = \mathbf{0}$, we obtain

$$\mathbf{0} = \int_0^\infty \mathbf{b}(t)g(\mu+t)dt + \int_0^\infty \mathbf{b}(-t)g(\mu-t)dt$$

$$\begin{aligned} &= \int_0^\infty \mathbf{b}(t)g(\mu+t)dt + \int_0^\infty \{-\mathbf{c}-\mathbf{b}(t)\}g(\mu-t)dt \\ &= \int_0^\infty \mathbf{b}(t)\{g(\mu+t)-g(\mu-t)\}dt - \mathbf{c}\int_0^\infty g(\mu-t)dt. \end{aligned}$$

Thus,

$$\mathbf{c} = \frac{\int_0^\infty \mathbf{b}(t)\{g(\mu+t)-g(\mu-t)\}dt}{\int_0^\infty g(\mu-t)dt}$$

and $\mathbf{b}(t)+\mathbf{b}(-t)+\mathbf{c}=\mathbf{0}$. This shows $\Lambda^\perp \subset A$. □

A.6. Proof of Theorem 3

To prove the result concerning $S_{\text{eff},\mu}$, we first check $S_{\text{eff},\mu}(x-\mu)\in\Lambda^\perp$. To this end,

$$\begin{aligned} S_{\text{eff},\mu}(t) &= -2p\frac{\eta'(t)-g(\mu-t)c_\mu}{g(\mu-t)+g(\mu+t)}, \\ S_{\text{eff},\mu}(-t) &= -2p\frac{-\eta'(t)-g(\mu+t)c_\mu}{g(\mu-t)+g(\mu+t)} \\ S_{\text{eff},\mu}(t)+S_{\text{eff},\mu}(-t) &= 2pc_\mu \end{aligned}$$

and

$$\begin{aligned} &\int_0^\infty S_{\text{eff},\mu}(t)\{g(\mu+t)-g(\mu-t)\}dt \\ =\ & -2p\int_0^\infty \frac{\eta'(t)-g(\mu-t)c_\mu}{g(\mu-t)+g(\mu+t)}\{g(\mu+t)-g(\mu-t)\}dt \\ =\ & -2p\int_0^\infty \frac{\eta'(t)\{g(\mu+t)-g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}dt \\ & + 2pc_\mu\int_0^\infty \frac{g(\mu-t)\{g(\mu+t)-g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}dt \\ =\ & -p\int \frac{\eta'(t)\{g(\mu+t)-g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}dt \\ & + 2pc_\mu\int_0^\infty \frac{g(\mu-t)\{g(\mu+t)-g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}dt \\ =\ & -p\int \frac{\eta'(t)\{g(\mu+t)-g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}dt + 2pc_\mu\int \frac{g(\mu-t)g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt \\ & - 2pc_\mu\int_0^\infty g(\mu-t)dt \\ =\ & -2pc_\mu\int_0^\infty g(\mu-t)dt. \end{aligned}$$

Thus

$$\begin{aligned} & S_{\text{eff},\mu}(t) + S_{\text{eff},\mu}(-t) + \frac{\int_0^\infty S_{\text{eff},\mu}(t)\{g(\mu+t) - g(\mu-t)\}dt}{\int_0^\infty g(\mu-t)dt} \\ = \quad & 2pc_\mu + \frac{-2pc_\mu \int_0^\infty g(\mu-t)dt}{\int_0^\infty g(\mu-t)dt} = 0. \end{aligned}$$

Hence $S_{\text{eff},\mu}(x-\mu) \in \Lambda^\perp$.

Next we prove $S_\mu(x-\mu) - S_{\text{eff},\mu}(x-\mu) \in \Lambda$. We have

$$S_\mu(x-\mu) - S_{\text{eff},\mu}(x-\mu) = \frac{-p\eta'(x-\mu)}{g(x)} + 2p\frac{\eta'(x-\mu) - g(2\mu-x)c_\mu}{g(2\mu-x)+g(x)}.$$

Define $a(x-\mu) = g(x)\{S_\mu(x-\mu) - S_{\text{eff},\mu}(x-\mu)\}$. Then

$$\begin{aligned} a(t) &= -p\eta'(t) + 2p\frac{\eta'(t)g(\mu+t) - g(\mu+t)g(\mu-t)c_\mu}{g(\mu-t)+g(\mu+t)} \\ &= p\frac{\eta'(t)\{g(\mu+t) - g(\mu-t)\} - 2g(\mu+t)g(\mu-t)c_\mu}{g(\mu-t)+g(\mu+t)}. \end{aligned}$$

Obviously, $a(t)$ is an even function. We also have $\int a(t)dt = 0$ due to the definition of c_μ. Thus, we showed $S_\mu(x-\mu) - S_{\text{eff},\mu}(x-\mu) \in \Lambda$.

To prove the result concerning $S_{\text{eff},p}$, we first check $S_{\text{eff},p}(x-\mu) \in \Lambda^\perp$. To this end,

$$\begin{aligned} S_{\text{eff},p}(t) &= -\frac{f(\mu+t) - f(\mu-t) - g(\mu-t)c_p}{g(\mu-t)+g(\mu+t)}, \\ S_{\text{eff},p}(-t) &= -\frac{f(\mu-t) - f(\mu+t) - g(\mu+t)c_p}{g(\mu-t)+g(\mu+t)}, \\ S_{\text{eff},p}(t) + S_{\text{eff},p}(-t) &= c_p, \end{aligned}$$

and

$$\begin{aligned} & \int_0^\infty S_{\text{eff},p}(t)\{g(\mu+t) - g(\mu-t)\}dt \\ = \quad & \int_0^\infty \frac{f(\mu-t) - f(\mu+t) + g(\mu-t)c_p}{g(\mu-t)+g(\mu+t)}\{g(\mu+t) - g(\mu-t)\}dt \\ = \quad & \int_0^\infty \frac{\{f(\mu-t) - f(\mu+t)\}\{g(\mu+t) - g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}dt \\ & + c_p\int_0^\infty \frac{g(\mu-t)\{g(\mu+t) - g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}dt \\ = \quad & \int \frac{\{f(\mu-t) - f(\mu+t)\}g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt \\ & + c_p\int_0^\infty \frac{g(\mu-t)\{g(\mu+t) - g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}dt \end{aligned}$$

$$
\begin{aligned}
&= \int \frac{\{f(\mu-t)-f(\mu+t)\}g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt + c_p\int_0^\infty \frac{2g(\mu-t)g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt \\
&\quad - c_p\int_0^\infty g(\mu-t)dt \\
&= \int \frac{\{f(\mu-t)-f(\mu+t)\}g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt + c_p\int \frac{g(\mu-t)g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt \\
&\quad - c_p\int_0^\infty g(\mu-t)dt \\
&= -c_p\int_0^\infty g(\mu-t)dt.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&S_{\text{eff},p}(t) + S_{\text{eff},p}(-t) + \frac{\int_0^\infty S_{\text{eff},p}(t)\{g(\mu+t)-g(\mu-t)\}dt}{\int_0^\infty g(\mu-t)dt} \\
&= c_p - \frac{c_p\int_0^\infty g(\mu-t)dt}{\int_0^\infty g(\mu-t)dt} = \mathbf{0}.
\end{aligned}
$$

Hence $S_{\text{eff},p}(x-\mu) \in \Lambda^\perp$.

Next we prove $S_p(x-\mu) - S_{\text{eff},p}(x-\mu) \in \Lambda$. We have

$$
S_p(x-\mu) - S_{\text{eff},p}(x-\mu) = \frac{\eta(x-\mu)-f(x)}{g(x)} - \frac{f(2\mu-x)-f(x)+g(2\mu-x)c_p}{g(x)+g(2\mu-x)}.
$$

Define $a(x-\mu) = g(x)\{S_p(x-\mu) - S_{\text{eff},p}(x-\mu)\}$. Then

$$
\begin{aligned}
a(t) &= \eta(t) - f(\mu+t) - \frac{\{f(\mu-t)-f(\mu+t)+g(\mu-t)c_p\}g(\mu+t)}{g(\mu-t)+g(\mu+t)} \\
&= \eta(t) - \frac{\{f(\mu+t)g(\mu-t)+f(\mu-t)g(\mu+t)\}+g(\mu+t)g(\mu-t)c_p}{g(\mu-t)+g(\mu+t)}.
\end{aligned}
$$

Obviously, $a(t)$ is an even function. We also have

$$
\begin{aligned}
\int a(t)dt &= 1 - \int \frac{f(\mu+t)g(\mu-t)+f(\mu-t)g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt \\
&\quad - c_p\int \frac{g(\mu+t)g(\mu-t)}{g(\mu-t)+g(\mu+t)}dt \\
&= \int \frac{\{f(\mu+t)-f(\mu-t)\}g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt - c_p\int \frac{g(\mu+t)g(\mu-t)}{g(\mu-t)+g(\mu+t)}dt \\
&= 0
\end{aligned}
$$

due to the definition of c_p. Thus, we showed $S_p(x-\mu) - S_{\text{eff},p}(x-\mu) \in \Lambda$.

Finally, to prove the result regarding $\mathbf{S}_{\text{eff},\boldsymbol{\alpha}}$, we first check $\mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(x-\mu) \in \Lambda^\perp$. To this end,

$$
\mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(t) = (1-p)\frac{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha}) - \mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha}) - g(\mu-t)\mathbf{c}_\alpha}{g(\mu-t)+g(\mu+t)},
$$

$$\begin{aligned}\mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(-t) &= (1-p)\frac{\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})-g(\mu+t)\mathbf{c}_\alpha}{g(\mu-t)+g(\mu+t)},\\ \mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(t)+\mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(-t) &= -(1-p)\mathbf{c}_\alpha,\end{aligned}$$

and

$$\begin{aligned}&\int_0^\infty \mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(t)\{g(\mu+t)-g(\mu-t)\}dt\\ &=(1-p)\int_0^\infty \frac{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})-g(\mu-t)\mathbf{c}_\alpha}{g(\mu-t)+g(\mu+t)}\{g(\mu+t)-g(\mu-t)\}dt\\ &=(1-p)\int_0^\infty \frac{\{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})\}\{g(\mu+t)-g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}\\ &\quad-(1-p)\mathbf{c}_\alpha\int_0^\infty \frac{g(\mu-t)\{g(\mu+t)-g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}dt\\ &=(1-p)\int \frac{\{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})\}g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt\\ &\quad-(1-p)\mathbf{c}_\alpha\int_0^\infty \frac{g(\mu-t)\{g(\mu+t)-g(\mu-t)\}}{g(\mu-t)+g(\mu+t)}dt\\ &=(1-p)\int \frac{\{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})\}g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt\\ &\quad-(1-p)\mathbf{c}_\alpha\int_0^\infty \frac{2g(\mu-t)g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt+(1-p)\mathbf{c}_\alpha\int_0^\infty g(\mu-t)dt\\ &=(1-p)\mathbf{c}_\alpha\int_0^\infty g(\mu-t)dt.\end{aligned}$$

Thus

$$\begin{aligned}&\mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(t)+S_{\text{eff}\boldsymbol{\alpha}}(-t)+\frac{\int_0^\infty \mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(t)\{g(\mu+t)-g(\mu-t)\}dt}{\int_0^\infty g(\mu-t)dt}\\ =\;& -(1-p)\mathbf{c}_\alpha+\frac{(1-p)\mathbf{c}_\alpha\int_0^\infty g(\mu-t)dt}{\int_0^\infty g(\mu-t)dt}=\mathbf{0}.\end{aligned}$$

Hence $\mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(x-\mu)\in\Lambda^\perp$.

Next we prove $\mathbf{S}_{\boldsymbol{\alpha}}(x-\mu)-\mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(x-\mu)\in\Lambda$. We have

$$\begin{aligned}&\mathbf{S}_{\boldsymbol{\alpha}}(x-\mu)-\mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(x-\mu)\\ =\;&(1-p)\left\{\frac{\mathbf{f}'_\alpha(x;\boldsymbol{\alpha})}{g(x)}-\frac{\mathbf{f}'_\alpha(x;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(2\mu-x;\boldsymbol{\alpha})-g(2\mu-x)\mathbf{c}_\alpha}{g(2\mu-x)+g(x)}\right\}.\end{aligned}$$

Define $\mathbf{a}(x-\mu)=g(x)\{\mathbf{S}_{\boldsymbol{\alpha}}(x-\mu)-\mathbf{S}_{\text{eff},\boldsymbol{\alpha}}(x-\mu)\}$. Then

$$\mathbf{a}(t)=(1-p)\frac{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})g(\mu-t)+\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})g(\mu+t)+g(\mu+t)g(\mu-t)\mathbf{c}_\alpha}{g(\mu-t)+g(\mu+t)}.$$

Obviously, $a(t)$ is an even function. We also have

$$\begin{aligned}
&\int \mathbf{a}(t)dt \\
&= (1-p)\int \frac{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})g(\mu-t)+\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})g(\mu+t)+g(\mu+t)g(\mu-t)\mathbf{c}_\alpha}{g(\mu-t)+g(\mu+t)}dt \\
&= (1-p)\int \frac{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})g(\mu-t)+\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt \\
&\quad + (1-p)\int \frac{\{\mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})-\mathbf{f}'_\alpha(\mu-t;\boldsymbol{\alpha})\}g(\mu+t)}{g(\mu-t)+g(\mu+t)}dt \\
&= (1-p)\int \mathbf{f}'_\alpha(\mu+t;\boldsymbol{\alpha})dt \\
&= \mathbf{0},
\end{aligned}$$

due to the definition of $\mathbf{c}_\alpha$ and $\mathbf{f}$ is a valid pdf. Thus, we showed $S_{\boldsymbol{\alpha}}(x-\mu) - S_{\text{eff}\,\boldsymbol{\alpha}}(x-\mu) \in \Lambda$. □

References

Allison, D. B., Gadbury, G. L., Heo, M., Fernandez, J. R., Lee, C. K., Prolla, T. A. and Weindruck, R. (2002). A mixture model approach for the analysis of microarray gene expression data. *Computational Statistics and Data Analysis*, **39**, 1–20.

Benjamini, Y. and Hochberg, Y. (1995). Controllling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of Royal Statistical Society*, **B57**, 289–300. MR1325392

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: The Johns Hopkins University Press. MR1245941

Böhning, D. (1999). *Computer-Assisted Analysis of Mixtures and Applications*. Boca Raton, FL: Chapman and Hall/CRC. MR1684363

Bordes, L., Mottelet S. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model. *Annals of Statistics*, **34**, 1204–1232. MR2278356

Bordes, L., Delmas, C. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model where one component is known. *Scandinavian Journal of Statistics*, **33**, 733–752. MR2300913

Bordes, L., Kojadinovic, I. and Vandekerkhove, P. (2013). Semiparametric estimation of a mixture of two linear regressions where one component is known. *Electronic Journal of Statistics*, **7**, 2603–2644. MR3121625

Bordes, L. and Vandekerkhove., P. (2010). Semiparametric two-component mixture model when a component is known: An asymptotically normal estimator. *Mathematical Methods of Statistics*, **19**, 22–41. MR2682853

Callow, M. J., Dudoit, S., Gong, E. L., Speed, T. P. and Rubin, E. M. (2000). Microarray expression profiling identifies genes with altered expression in hdl-deficient mice. *Genome Research*, **10**, 2022–2029.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association*, **99**, 96–104. MR2054289

Efron, B. and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genetic Epidemiology*, **23**, 70–86.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer. MR2265601

Genovese, C. R. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Annals of Statistics*, **32**, 1035–1061. MR2065197

Hall, P. and Zhou, X. H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Annals of Statistics*, **31**, 201–224. MR1962504

Hall, P., Neeman, A., Pakyari, R. and Elmore, R. (2005). Nonparametric inference in multivariate mixtures. *Biometrika*, **92**, 667–678. MR2202653

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley. MR0829458

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O. P., Wilfond, B., Borg, A. and Trent, J. (2001). Gene-expression profiles in hereditary breast cancer. *New England Journal of Medicine*, **344**, 539–548.

Hohmann, D. and Holzmann, H. (2013). Semiparametric location mixtures with distinct components. *Statistics*, **47**, 348–362. MR3043704

Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics*. New York: Wiley. MR2488795

Hunter, D. R., Wang, S. and Hettmanserger, T. P. (2007). Inference for mixtures of symmetric distributions. *Annals of Statistics*, **35**, 224–251. MR2332275

Hunter, D. R. and Young, D. S. (2012). Semiparametric mixtures of regressions. *Journal of Nonparametric Statistics*, **24**, 19–38. MR2885823

Langaas, M., Lindqvist, B. H. and Ferkingstad, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *Journal of Royal Statistical Society*, **B67**, 555–572. MR2168204

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry, and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics v 5, Hayward, CA: Institute of Mathematical Statistics.

Liu, C. and Rubin, D. B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME, *Statisica Sinica*, **5**, 19–39. MR1329287

Ma, J., Gudlaugsdottir, S. and Wood, G. (2011). Generalized EM estimation for semiparametric mixture distributions with discretized non-parametric component. *Statistics and Computing*, **21**, 601–612. MR2826695

McLachlan, G. J., Bean, R. W. and Ben-Tovim Jones, L. (2006). A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrys. *Bioinformatics*, **22**, 1608–1615.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley. MR1789474

McLachlan, G. J. and Wockner, L. (2010). Use of mixture models in multiple hypothesis testing with applications in Bioinformatics. In Hermann Locarek-Junge and Claus Weihs (Ed.), *Classification as a Tool for Research: Proceedings of the 11th IFCS Biennial Conference and 33rd Annual Conference of the Gesellschaft für Klassifikation* (pp. 177–184). Heidelberg, Germany: Springer-Verlag. MR2722135

Pounds, S. and Morris, S. W. (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics*, **19**, 1236–1242.

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, **3**(1), article 3. MR2101454

Smyth, G. K., Thorne, N. and Wettenhall, J. (2005). *LIMMA: Linear Models for Microarray Data, User's Guide*. Melbourne: Walter and Eliza Hall Institute of Medical Research.

Song, J. and Nicolae, D. (2009). A sequential clustering algorithm with applications to gene expression data. *Journal of the Korean Statistical Society*, **38**, 175–184. MR2649690

Song, S., Nicolae, D. L. and Song, J. (2010). Estimating the mixing proportion in a semiparametric mixture model. *Computational Statistics & Data Analysis*, **54**, 2276–2283. MR2720488

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of National Academy of Sciences of USA*, **100**, 3889–3894. MR1994856

Vandekerkhove, P. (2012). Estimation of a semiparametric mixture of regressions model. *Journal of Nonparametric Statistics*, **25**, 181–208. MR3039977

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25. MR0640163

Xiang, S., Yao, W. and Wu, J. (2014). Minimum profile Hellinger distance estimation for a semiparametric mixture model. *The Canadian Journal of Statistics*, **42**, 246–267. MR3208338