Electronic Journal of Statistics
Vol. 13 (2019) 4035–4078
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1620

Maximum smoothed likelihood component density estimation in mixture models with known mixing proportions

Tao Yu*

Department of Statistics and Applied Probability
National University of Singapore, 117546, Singapore
e-mail: stayt@nus.edu.sg

Pengfei Li†

Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
e-mail: pengfei.li@uwaterloo.ca

and

Jing Qin

National Institute of Allergy and Infectious Diseases
National Institutes of Health, Maryland 20892, U.S.A.
e-mail: jingqin@niaid.nih.gov

Abstract: Mixture models appear in many research areas. In genetic and epidemiology applications, sometimes the mixture proportions may vary but are known. For such data, the existing methods for the underlying component density estimation may produce undesirable results: negative values in the density estimates. In this paper, we propose a maximum smoothed likelihood method to estimate these component density functions. The proposed estimates maximize a smoothed log likelihood function which can inherit all the important properties of probability density functions. A majorization-minimization algorithm is suggested to compute the proposed estimates numerically. We show that, starting from any initial value, the algorithm converges. Furthermore, we establish the asymptotic convergence rate of the L_1 errors of our proposed estimators. Our method provides a general framework for dealing with many similar mixture model problems. An adaptive procedure is suggested for choosing the bandwidths in our estimation procedure. Simulation studies show that the proposed method is very promising and can be much more efficient than the existing method in terms of the L_1 errors. A malaria data application shows the advantages of our method over others.

MSC 2010 subject classifications: Primary 62G07; Secondary 62G20.

*Supported in part by Singapore Ministry of Education Academic Research Fund Tier 1.
†Supported in part by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-06592).

Keywords and phrases: EM-like algorithm, empirical process, M-estimators, majorization-minimization algorithm, mixture data, smoothed likelihood function.

Received September 2018.

Contents

1. Introduction

In this paper, we study data sets with the following mixture structure. Let $\{X_i, \boldsymbol{\alpha}_i\}$, $i = 1, \ldots, n$, be independent and identically distributed (i.i.d.) copies of $\{X, \boldsymbol{\alpha}\}$. For every $i = 1, \ldots, n$, X_i comes from one of the M subpopulations with probability density functions (pdfs) $f_1(x), \ldots, f_M(x)$. Denote by $\alpha_{i,j}$ the probability that X_i is from the jth subpopulation and let $\boldsymbol{\alpha}_i = (\alpha_{i,1}, \ldots, \alpha_{i,M})^\tau$. Clearly $\alpha_{i,j} \geq 0$ and $\sum_{j=1}^M \alpha_{i,j} = 1$. The pdf of X_i conditioning on $\boldsymbol{\alpha}_i$ is given by

$$X_i | \boldsymbol{\alpha}_i \sim \sum_{j=1}^M \alpha_{i,j} f_j(x). \tag{1.1}$$

Practically, $\boldsymbol{\alpha}_i$ is known, observable, or can be reliably estimated from other sources. That is, conditioning on $\boldsymbol{\alpha}_i$, X_i follows a mixture model with known mixing proportions. To make this model identifiable, we need some condition on $\boldsymbol{\alpha}$. The details of such a condition is given in Condition 4 of Section 4; a stronger but more intuitive condition is that $\boldsymbol{\alpha}$ is a continuous random vector, or it is a discrete random vector with at least M supports. Our main interest in this paper is to estimate $f_1(x), \dots, f_M(x)$ nonparametrically.

Data with the mixture structure in (1.1) have been frequently identified in the literature and in practice. Acar and Sun (2013) provided one example of such data. In genetic association studies of single nucleotide polymorphisms (SNPs), the corresponding genotypes of the SNPs are usually not deterministic; in the resulting data, they are typically delivered as genotype probabilities from various genotype calling or imputation algorithms (see for example Li et al., 2009 and Carvalho et al., 2010). Ma and Wang (2012) summarized two types of genetic epidemiology studies in which such mixture data are collected. These studies are kin-cohort studies (Wang et al., 2008) and quantitative trait locus studies (Lander and Botstein, 1989; Wu et al., 2007); see also Wang et al. (2012) and the references therein. Section 7 also gives an example of such data in the context of malaria.

Under the mixture model (1.1), statistical methods for estimating the component cumulative distribution functions (cdfs) have been investigated; see Ma and Wang (2012) and the references therein. Ma and Wang observed that the classical maximum empirical likelihood estimators of these component cdfs are either highly inefficient or inconsistent, and they proposed a class of weighted least square estimators. The estimation of the pdfs has received less attention. Ma et al. (2011) proposed a family of kernel-based weighted least squares estimators for the component pdfs under the assumption that $\boldsymbol{\alpha}_i$ is continuous. However, their approach has two limitations: (1) the estimates do not inherit the nonnegativity property of a regular density function; as is well known, this property is often important in downstream density-based studies. (2) Their method depends on an initial guess for the underlying densities; as a result, their estimators are locally efficient only when this initial guess is correct.

The maximum smoothed likelihood was introduced by Eggermont and LaRiccia (1995a). It is analogous to the maximum parametric likelihood, but in the nonparametric setup for density estimation. It inherits many of the good properties of the parametric likelihood estimation. This method has been used successfully to solve many difficult inverse convolution problems; see for example Eggermont and LaRiccia (1995b). It has also been widely applied for the estimation of the density and distribution functions in various statistical problems. For example, it has been applied to estimate smooth monotone and unimodal densities (Eggermont and LaRiccia, 2000), the density and hazard rate of the event time distribution (Groeneboom et al., 2010), the component densities in multivariate mixture model (Levine et al., 2011), the cumulative distribution function for the interval censoring model (Groeneboom, 2014), the densities in two-sample problem with likelihood ratio ordering (Yu et al., 2017). We incorporate this idea in our method. Our proposed estimators for $f_1, \dots, f_M$, namely

$\widehat{f}_1, \ldots, \widehat{f}_M$, inherit all the important properties of pdfs and can handle data with continuous or discrete $\boldsymbol{\alpha}_i$'s. We also propose a majorization-minimization algorithm that computes these density estimates numerically. We show that for finite samples, starting from any initial value, this algorithm not only increases the smoothed likelihood function but also leads to estimates that maximize the smoothed likelihood function.

Another contribution of this paper is to establish the L_1 asymptotic consistency and the corresponding convergence rate for our density estimates. Because of the properties (see Section 4) of the nonlinear operator "$\mathcal{N}_h$" defined in Section 2 and the complicated form of the smoothed log-likelihood function, the development of asymptotic theory for nonparametric density estimates in the framework of mixture models is technically challenging and still lacking in the literature. We solve this problem by employing advanced theory from empirical processes (see van der Vaart and Wellner 1996, Kosorok 2008, and the references therein). We expect that the technical tools established in this paper will benefit the future study of asymptotic theory for nonparametric density estimates in other mixture models; see for example Levine et al. (2011).

The rest of the paper is organized as follows. Section 2 presents our proposed density estimates based on the smoothed likelihood principle. Section 3 suggests a majorization-minimization algorithm to numerically compute these density estimates and establishes the finite-sample convergence properties of this algorithm. Section 4 studies the asymptotic behaviour of our density estimators. Section 5 proposes a bandwidth selection procedure that is easily imbedded into the majorization-minimization algorithm. Section 6 presents simulation studies, which show that the proposed method is more efficient than existing methods in terms of the integrated square error. Section 7 applies our method to a real-data example, and Section 8 ends the paper with some discussion. The technical details are relegated to the Appendix.

2. Maximum smoothed likelihood estimation

With the observed data $\{X_i, \boldsymbol{\alpha}_i\}_{i=1}^n$ from Model (1.1), we propose a maximum smoothed likelihood method for estimating $f_1, \ldots, f_M$. We consider the set of functions

$$\mathcal{C} = \{(f_1, \ldots, f_M) : f_j \text{ is a pdf, } j = 1, \ldots, M\}\,.$$

Furthermore, we assume that the f_j's have the common support S_x.

Given Model (1.1) and the observations $\{X_i, \boldsymbol{\alpha}_i\}_{i=1}^n$, the conditional log-likelihood is given by

$$\widetilde{l}_n(f_1, \ldots, f_M) = \sum_{i=1}^n \log\left\{\sum_{j=1}^M \alpha_{i,j} f_j(X_i)\right\}.$$

However, as is well known, this log-likelihood function is unbounded in $\mathcal{C}$; see p. 25 in Silverman (1986) and p. 111 in Eggermont and LaRiccia (2001). Therefore, the corresponding maximum likelihood estimates do not exist. This unboundedness problem can be solved by incorporating the smoothed likelihood approach (Eggermont and LaRiccia, 1995a). Specifically, we define the smoothed log-likelihood of $f_1(x), \dots, f_M(x)$ to be

$$l_n(f_1, \dots, f_M) = \sum_{i=1}^{n} \log \left\{ \sum_{j=1}^{M} \alpha_{i,j} \mathcal{N}_{h_j} f_j(X_i) \right\}, \tag{2.1}$$

where $\mathcal{N}_h f(x)$ is the nonlinear smoothing operator for a density function f, represented by

$$\mathcal{N}_h f(x) \quad = \quad \exp \left\{ \int_{\mathbb{R}} K_h(u - x) \log f(u) du \right\}. \tag{2.2}$$

Here $K_h(x) = \frac{1}{h} K(x/h)$, $K(\cdot)$ is a kernel function supported on $[-L, L]$, and h is the bandwidth for the nonlinear smoothing operator. By convention, we define $0 \log(0) = 0$, $\log(0) = -\infty$, and $\exp(-\infty) = 0$.

Our proposed maximum smoothed likelihood estimators for $f_1, \dots, f_M$ are given by

$$(\widehat{f}_1, \dots, \widehat{f}_M) = \text{argmax}_{(f_1, \dots, f_M) \in \mathcal{C}} l_n(f_1, \dots, f_M). \tag{2.3}$$

We observe that the smoothed likelihood function defined in (2.1) has the following properties. First, based on Lemma 3.1(iii) of Eggermont (1999), $l_n(\cdot)$ is concave in $\mathcal{C}$, and $\mathcal{C}$ is a convex set of functions. Second, if the kernel function $K(t)$ is bounded and $h_j > 0$, $j = 1, \dots, M$ are fixed, then $l_n(\cdot)$ is also bounded in $\mathcal{C}$, since for every x and $(f_1, \dots, f_M) \in \mathcal{C}$,

$$\mathcal{N}_{h_j} f_j(x) \quad \leq \quad \exp \left[\log \left\{ \int_{\mathbb{R}} K_{h_j}(u - x) f(u) du \right\} \right] \leq \sup_t K(t)/h_j.$$

Therefore, the maximizer of $l_n(\cdot)$ exists, i.e. the optimization problem (2.3) is well defined. Furthermore, if we assume that for every $j = 1, \dots, M$, the X_i's corresponding to $\alpha_{i,j} > 0$ are dense in S_x, then $l_n(\cdot)$ is strictly concave in $\mathcal{C}$ and thus the solution to the optimization problem (2.3) is unique. Here, "dense" means that for every $j = 1, \dots, M$, and $x \in S_x$, the interval $[x - Lh_j, x + Lh_j]$ contains at least one observation X_i such that the corresponding $\alpha_{i,j} > 0$.

3. The majorization-minimization algorithm

In this section, we propose an algorithm that numerically calculates $\widehat{f}_1, \dots, \widehat{f}_M$ with given bandwidths $h_1, \dots, h_M$ and we study the finite-sample convergence property of this algorithm. The proposed algorithm, called the majorization-minimization algorithm, is in spirit similar to the majorization-minimization algorithm in Levine et al. (2011) and the EM-like algorithm in Hall et al. (2005).

To facilitate our theoretical development, we define the majorization-minimization updating operator $\mathcal{G}$ on $\mathcal{C}$ as follows. For any $(f_1,\ldots,f_M)\in\mathcal{C}$, let

$$\mathcal{G}(f_1,\ldots,f_M)=(f_1^{\mathcal{G}},\ldots,f_M^{\mathcal{G}}), \tag{3.1}$$

where

$$\begin{aligned} f_j^{\mathcal{G}}(x) &= \frac{\sum_{i=1}^n w_{i,j}K_{h_j}(x-X_i)}{\sum_{i=1}^n w_{i,j}}, \\ \text{with}\quad w_{i,j} &= \frac{\alpha_{i,j}\mathcal{N}_{h_j}f_j(X_i)}{\sum_{k=1}^M \alpha_{i,k}\mathcal{N}_{h_k}f_k(X_i)}. \end{aligned} \tag{3.2}$$

Note that in every updating step, the updated $f_j^{\mathcal{G}}(\cdot)$ is essentially obtained by maximizing the minorant

$$l_n(f_1,\ldots,f_M)+\sum_{i=1}^n\sum_{j=1}^M w_{i,j}\left\{\log\mathcal{N}_{h_j}g_j(X_i)-\log\mathcal{N}_{h_j}f_j(X_i)\right\}$$

with respect to $g_j(\cdot)$.

We first show that $\mathcal{G}$ is capable of increasing the smoothed log-likelihood function l_n at every updating step.

Theorem 1. *For every $(f_1,\ldots,f_M)\in\mathcal{C}$, we have*

$$l_n\left(\mathcal{G}(f_1,\ldots,f_M)\right)\geq l_n\left(f_1,\ldots,f_M\right).$$

Theorem 1 immediately leads to our proposed majorization-minimization algorithm as follows. Given initial values $(f_1^0,\ldots,f_M^0)\in\mathcal{C}$, for $s=0,1,2,\cdots$, we iteratively update from $(f_1^s,\ldots,f_M^s)$ to $(f_1^{s+1},\ldots,f_M^{s+1})$ via

$$(f_1^{s+1},\ldots,f_M^{s+1})=\mathcal{G}(f_1^s,\ldots,f_M^s).$$

Clearly, Theorem 1 above ensures that for every $s=0,1,\ldots$, we have

$$l_n\left(f_1^{s+1},\ldots,f_M^{s+1}\right)\geq l_n\left(f_1^s,\ldots,f_M^s\right).$$

Furthermore, since for any $(f_1,\ldots,f_M)\in\mathcal{C}$, $\mathcal{G}(f_1,\ldots,f_M)$ belongs to the class of functions

$$\mathcal{F}_n=\left\{(f_1,\ldots,f_M):f_j(x)=\frac{\sum_{i=1}^n w_{i,j}K_{h_j}(x-X_i)}{\sum_{i=1}^n w_{i,j}};0\leq w_{i,j}\leq 1\right\}, \tag{3.3}$$

we have $(f_1^s,\ldots,f_M^s)\in\mathcal{F}_n$ for $s\geq 1$. Next, we study the finite-sample convergence property of this majorization-minimization algorithm; we observe that the technical development of this property is nontrivial. We first present a necessary and sufficient condition under which $(\widehat{f}_1,\ldots,\widehat{f}_M)\in\mathcal{C}$ is a solution of the optimization problem (2.3).

Theorem 2. *Assume that* $\sum_{i=1}^{n} \alpha_{i,j} > 0$ *for every* j*. For* $(\widehat{f}_1, \dots, \widehat{f}_M) \in \mathcal{C}$*, we have*

$$l_n(\widehat{f}_1, \dots, \widehat{f}_M) = \sup_{(f_1,\dots,f_M)\in\mathcal{C}} l_n(f_1, \dots, f_M)$$

if and only if $(\widehat{f}_1, \dots, \widehat{f}_M) = \mathcal{G}(\widehat{f}_1, \dots, \widehat{f}_M)$ *almost surely under the Lebesgue measure.*

The following corollary results from an immediate application of Theorem 2; the straightforward proof is omitted.

Corollary 1. *Assume that* $\sum_{i=1}^{n} \alpha_{i,j} > 0$ *for every* j*. Let* $(\widehat{f}_1, \dots, \widehat{f}_M)$ *be a solution of the optimization problem (2.3); then* $(\widehat{f}_1, \dots, \widehat{f}_M) \in \mathcal{F}_n$ *almost surely under the Lebesgue measure.*

Corollary 1 is useful for our technical development of asymptotic theory for $\widehat{f}_1, \dots, \widehat{f}_M$ in Section 4. It indicates that the solution of (2.3) is equivalent to the solution of

$$(\widehat{f}_1, \dots, \widehat{f}_M) = \mathrm{argmax}_{(f_1,\dots,f_M)\in\mathcal{F}_n} l_n(f_1, \dots, f_M), \tag{3.4}$$

provided the stated condition $\sum_{i=1}^{n} \alpha_{i,j} > 0$ for every j is satisfied. This condition is quite reasonable since if $\sum_{i=1}^{n} \alpha_{i,j} = 0$ for some j then the jth subpopulation does not appear in the data, and we can delete the corresponding $f_j(x)$ from the mixture model (1.1). Therefore, developing asymptotic theory for $\widehat{f}_1, \dots, \widehat{f}_M$ from (2.3) is equivalent to developing it from (3.4).

Using Theorem 2, we show that the updating sequence $l_n(f_1^s, \dots, f_M^s)$ converges to its global maximum, which implies the convergence of the proposed majorization-minimization algorithm.

Theorem 3. *Assume that* $\sup_t K(t) < \infty$*. Then we have*

$$\lim_{s\to\infty} l_n(f_1^s, \dots, f_M^s) = l_n(\widehat{f}_1, \dots, \widehat{f}_M),$$

where $(\widehat{f}_1, \dots, \widehat{f}_M) \in \mathcal{F}_n$ *is a solution of the optimization problem (2.3).*

Using Theorem 3, if we do not impose further conditions on the data, $l_n(\cdot)$ is not necessarily strictly concave. Therefore, we can show only that the updating sequence $l_n(f_1^s, \dots, f_M^s)$ converges to the maximum of $l_n(f_1, \dots, f_M)$. Note that this does not guarantee the convergence of $(f_1^s, \dots, f_M^s)$ to $(\widehat{f}_1, \dots, \widehat{f}_M)$, i.e. the maximizer of $l_n(f_1, \dots, f_M)$, because this maximizer may not be uniquely defined. Instead, referring to the proof of this theorem, we have shown that there exists at least a subsequence of $(f_1^s, \dots, f_M^s)$ converging to a maximizer of $l_n(f_1, \dots, f_M)$. Furthermore, if we impose a technical condition to ensure that $l_n(\cdot)$ is strictly concave, then $(\widehat{f}_1, \dots, \widehat{f}_M)$ is uniquely defined by (2.3). We can immediately show $\lim_{s\to\infty}(f_1^s, \dots, f_M^s) = (\widehat{f}_1, \dots, \widehat{f}_M)$ for every $x \in S_x$. We refer to the discussion at the end of Section 2 for a sufficient condition ensuring that l_n is strictly concave in $\mathcal{C}$.

We end this section with the following remark about the proposed majorization-minimization algorithm above.

Remark 1. *Ma et al. (2011) discussed an EM-like algorithm in their discussion section to obtain nonnegative component density estimates. In particular, they suggested defining*

$$w_{i,j} = \frac{\alpha_{i,j} f_j(X_i)}{\sum_{k=1}^{M} \alpha_{i,k} f_k(X_i)},$$

and using a similar way as (3.1) to update the resultant density estimates in their paper. Yet, the corresponding theoretical properties as well as the numerical performance of these estimates are left unknown. As commented by Levine et al. (2011), algorithms of this kind do not minimize/maximize any particular objective function; this may impose difficulty in the subsequent technical development. We refer to Levine et al. (2011) for more discussion of such a method.

4. Asymptotic properties for $(\widehat{f}_1, \ldots, \widehat{f}_M)$

In this section, we investigate the asymptotic behaviour of $(\widehat{f}_1, \ldots, \widehat{f}_M)$ given in (2.3). First, we consider the consistency of $\widehat{p}(x, \boldsymbol{\alpha}) = \sum_{j=1}^{M} \alpha_j \mathcal{N}_{h_j} \widehat{f}_j(x)$ under the Hellinger distance, where the Hellinger distance between nonnegative functions $m_1(x, \boldsymbol{\alpha})$ and $m_2(x, \boldsymbol{\alpha})$ is defined to be

$$d(m_1, m_2) = \left[\int_{S_\gamma} \int_{\mathbb{R}} \left\{ m_1^{1/2}(x, \boldsymbol{\alpha}) - m_2^{1/2}(x, \boldsymbol{\alpha}) \right\}^2 dx d\boldsymbol{\alpha} \right]^{1/2},$$

with S_γ being the support of the random vector $\boldsymbol{\alpha}$.

To facilitate our technical development, we assume that all the bandwidths, $h_1, \ldots, h_M$, are of the same order as $n \to \infty$. That is:

> Condition 0: There exists a common bandwidth $h > 0$ such that $C_1 \le \inf_{1 \le j \le M, n \ge 1} h_j/h \le \sup_{1 \le j \le M, n \ge 1} h_j/h \le C_2$ for some fixed constants $C_2 \ge C_1 > 0$.

Furthermore, we need the following conditions for deriving the theoretical results in this section.

> Condition 1: $h \to 0$ and $n^{1-\vartheta} h \to \infty$ when $n \to \infty$, where $\vartheta > 0$ is an arbitrarily small value.
> Condition 2: The kernel function $K(x)$ is symmetric about 0 and supported and continuous on $[-L, L]$ for some $L > 0$, and $\inf_{x \in [-L,L]} K(x) > 0$. The ath-order derivative $K^{(a)}(x)$ of $K(x)$ exists for every $a = 1, 2, \ldots$ and $x \in (-L, L)$. Further, $\sup_{a,x} |K^{(a)}(x)|$ is bounded.
> Condition 3: The true component pdfs $f_{0,j}(x)$, $j = 1, \ldots, M$ are bounded, supported on $S_x = [c_1, c_2]$, and twice continuously differentiable in (c_1, c_2) with bounded second-order derivatives. Furthermore, $\inf_{x \in S_x} f_{0,j}(x) > 0$.
> Condition 4: There exist $M \times 1$ vectors $\boldsymbol{\alpha}_{0,1}, \ldots, \boldsymbol{\alpha}_{0,M}$ in the support S_γ of $\gamma(\boldsymbol{\alpha})$ satisfying (i) and (ii) below.

(i). The M vectors $\boldsymbol{\alpha}_{0,1}, \ldots, \boldsymbol{\alpha}_{0,M}$ are linearly independent.

(ii). There exist balls $\mathcal{O}_j \subset S_\gamma, j = 1, \ldots, M$, where $\boldsymbol{\alpha}_{0,j} \in \mathcal{O}_j$, the $\mathcal{O}_j$'s are disjoint, and $\gamma(\boldsymbol{\alpha}) > 0$ for every $\boldsymbol{\alpha} \in \mathcal{O}_j$.

Note that Condition 1 requires that the M bandwidths satisfy $h_j \to 0$ and $n^{1-\vartheta} h_j \to \infty$. Condition 2 requires that the kernel function $K(x)$ is symmetric and sufficiently smooth. Condition 3 requires that the component pdfs are sufficiently smooth and positive on the support of X. Condition 4 is an identifiability condition, which is satisfied when $\boldsymbol{\alpha}$ is a continuous random vector, or a discrete random vector with at least M supports.

Theorem 4. *Assume Conditions 0–3. For any arbitrarily small $\vartheta > 0$, we have*

$$d(\gamma\widehat{p}, \gamma\widetilde{p}_0) = O_p(h^{0.5}) + O_p(n^{-0.5+\vartheta} h^{-0.5}),$$

where $\gamma(\boldsymbol{\alpha})$ is the marginal density of $\boldsymbol{\alpha}$, $\widetilde{p}_0(x, \boldsymbol{\alpha}) = \sum_{j=1}^M \alpha_j f_{0,j}(x)$ is the conditional density of X given $\boldsymbol{\alpha}$, and $f_{0,j}(x)$, $j = 1, \ldots, M$, are the true values of $f_j(x)$.

Next we establish the asymptotic convergence rate for $\mathcal{N}_{h_j} \widehat{f}_j$, $j = 1, \ldots, M$ under the L_1-distance. The proof of this theorem relies heavily on the results given in Theorem 4.

Theorem 5. *Assume Conditions 0–4. For any arbitrarily small $\vartheta > 0$ and $j = 1, \ldots, M$, we have*

$$\int_{\mathbb{R}} |\mathcal{N}_{h_j} \widehat{f}_j(x) - f_{0,j}(x)| dx = O_p(h^{1/2}) + O_p(n^{-0.5+\vartheta} h^{-0.5}).$$

Finally, we establish the L_1 convergence of $\widehat{f}_j(x)$. We observe that Theorems 2 and 5 play key roles in the proof.

Theorem 6. *Assume Conditions 0–4. For any arbitrarily small $\vartheta > 0$, we have*

$$\int_{\mathbb{R}} |\widehat{f}_j(x) - f_{0,j}(x)| dx = O_p(h^{1/2}) + O_p(n^{-0.5+\vartheta} h^{-0.5}), \quad j = 1, \ldots, M.$$

For presentational continuity, we have placed the long proofs of Theorems 4–6 in the Appendix. As observed in the Appendix, these proofs are technically challenging. The main obstacles are the following undesirable properties of $\mathcal{N}_h f(x)$ with $f(x)$ being an arbitrary pdf. First, $\mathcal{N}_h f(x)$ is neither a density nor necessarily sufficiently close to the corresponding $f(x)$. Therefore, the well-developed empirical process theory and techniques for M-estimators in density estimation (see for example Section 3.4.1 in van der Vaart and Wellner 1996) are not directly applicable. Secondly, $\mathcal{N}_h f(x)$ introduces significant bias on the boundary of the support of $f(x)$. For example, if $f(x)$ is supported on $[c_1, c_2]$, then $\mathcal{N}_h f(x)$ is supported on $[c_1 + Lh, c_2 - Lh]$, i.e. $\mathcal{N}_h f(x) = 0$ when $x \in [c_1, c_1 + Lh) \cup (c_2 - Lh, c_2]$. Here $[-L, L]$ is the support for the kernel function $K(x)$.

These two properties of $\mathcal{N}_h f(x)$ significantly challenge our technical development. To date, we can show only the asymptotic behaviour of $\widehat{p}(x)$, $\mathcal{N}_{h_j}\widehat{f}_j(x)$, and $\widehat{f}_j(x)$ as given in Theorems 4, 5, and 6. The convergence rate given in Theorems 5 and 6 may not be optimal; there is some room for improvement. However, because of these two properties of "$\mathcal{N}_h$", we conjecture that $O_p(h^{0.5})$ is the best rate achievable by $d(\gamma\widehat{p}, \gamma\widetilde{p}_0)$ under the assumption that the $f_{0,j}(x)$'s are supported on a compact support. The intuition is as follows. One can show that even in the extreme case where the $\widehat{f}_j(x)$'s are estimated ideally well, $\widehat{f}_j(x) = f_{0,j}(x)$ say, the convergence rate for $d(\gamma\widehat{p}, \gamma\widetilde{p}_0)$ can not be better than $O_p(h^{0.5})$. Consequently, based on this and the convergence rates shown in our theorems, we can only show that the best L_1 convergence rate of our estimators can be arbitrarily close to $n^{-0.25}$.

5. Bandwidth selection

The maximum smoothed likelihood estimates $\widehat{f}_1, \ldots, \widehat{f}_M$ depend on the choice of the bandwidths $h_1, \ldots, h_M$. In this section, we propose a mean integrated square error (MISE) based method to select them numerically. Recall that $\widehat{f}_j(x)$, for $j = 1, \ldots, M$, has the form:

$$\widehat{f}_j(x) = \frac{\sum_{i=1}^n \widehat{w}_{i,j}(X_i) K_{h_j}(x - X_i)}{\sum_{i=1}^n \widehat{w}_{i,j}(X_i)},$$

where $\widehat{w}_{i,j}(X_i) = \frac{\alpha_{i,j}\mathcal{N}_{h_j}\widehat{f}_j(X_i)}{\sum_{k=1}^M \alpha_{i,k}\mathcal{N}_{h_k}\widehat{f}_k(X_i)}$. We derive the bias of $\widehat{f}_j(x)$ first. In the proof of Theorem 6, we have derived that

$$\frac{1}{n}\sum_{i=1}^n \widehat{w}_{i,j}(X_i) \to \int_{\boldsymbol{\alpha}\in S_\gamma} \alpha_j \gamma(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \tag{5.1}$$

in probability. Therefore

$$\begin{aligned}
E\left\{\widehat{f}_j(x)\right\} &\approx \frac{\sum_{i=1}^n E\left\{\widehat{w}_{i,j}(X_i) K_{h_j}(x - X_i)\right\}}{\sum_{i=1}^n \widehat{w}_{i,j}(X_i)} \\
&\approx \frac{\sum_{i=1}^n \int \widehat{w}_{i,j}(u) K_{h_j}(x - u) p(u, \boldsymbol{\alpha}_i) du}{\sum_{i=1}^n \widehat{w}_{i,j}(X_i)} \\
&= \int K_{h_j}(x - u)\breve{f}_j(u) du \\
&= f_j(x) + \int K_{h_j}(x - u)\left\{\breve{f}(u) - f_j(u)\right\} du \\
&\quad + \int K_{h_j}(x - u)\left\{f_j(u) - f_j(x)\right\} du \\
&\approx f_j(x) + \int K_{h_j}(x - u)\left\{\breve{f}(u) - f_j(u)\right\} du + \frac{h_j^2 f_j''(x)}{2}\int t^2 K(t) dt,
\end{aligned}$$

where $p(u,\boldsymbol{\alpha}_i)=\sum_{k=1}^M \alpha_{i,k} f_k(u)$ is the conditional density of X_i given $\boldsymbol{\alpha}_i$; and

$$\begin{aligned}\breve{f}_j(u) &= \frac{1}{\sum_{i=1}^n \widehat{w}_{i,j}(X_i)} \sum_{i=1}^n \alpha_{i,j} \left\{ \frac{\sum_{k=1}^M \alpha_{i,k} f_k(u)}{\sum_{k=1}^M \alpha_{i,k} \mathcal{N}_{h_k} \widehat{f}_k(u)} \right\} \mathcal{N}_{h_j} \widehat{f}_j(u) \\ &= \frac{1}{\sum_{i=1}^n \widehat{w}_{i,j}(X_i)} \sum_{i=1}^n \alpha_{i,j} \widehat{w}_{i,j}(u) p(u,\boldsymbol{\alpha}_i).\end{aligned}$$

Next, we derive the variance of $\widehat{f}_j(x)$ as follows. Because of (5.1), we have

$$\begin{aligned}\operatorname{var}\{\widehat{f}_j(x)\} &\approx \left\{\frac{1}{\sum_{i=1}^n \widehat{w}_{i,j}(X_i)}\right\}^2 \sum_{i=1}^n \operatorname{var}\left\{\widehat{w}_{i,j}(X_i) K_{h_j}(x-X_i)\right\} \\ &= \left\{\frac{1}{\sum_{i=1}^n \widehat{w}_{i,j}(X_i)}\right\}^2 \sum_{i=1}^n \Big(E\left[\left\{\widehat{w}_{i,j}(X_i) K_{h_j}(x-X_i)\right\}^2\right] \\ &\qquad - \left[E\left\{\widehat{w}_{i,j}(X_i) K_{h_j}(x-X_i)\right\}\right]^2 \Big).\end{aligned}$$

Note that

$$\begin{aligned}&\sum_{i=1}^n E\left\{\widehat{w}_{i,j}(X_i) K_{h_j}(x-X_i)\right\}^2 \\ \approx\ & \sum_{i=1}^n \int \widehat{w}_{i,j}^2(u) K_{h_j}^2(x-u) p(u,\boldsymbol{\alpha}_i) du \\ =\ & \frac{1}{h_j} \sum_{i=1}^n \widehat{w}_{i,j}^2(x) p(x,\boldsymbol{\alpha}_i) \int K^2(t) dt + O_P\left(h_j \sum_{i=1}^n \{\widehat{w}_{i,j}(x) p(x,\boldsymbol{\alpha}_i)\}''\right) \\ \approx\ & \frac{1}{h_j} \sum_{i=1}^n \widehat{w}_{i,j}^2(x) p(x,\boldsymbol{\alpha}_i) \int K^2(t) dt, \qquad (5.2)\end{aligned}$$

and similarly

$$\begin{aligned}&\sum_{i=1}^n \left[E\left\{\widehat{w}_{i,j}(X_i) K_{h_j}(x-X_i)\right\}\right]^2 \\ \approx\ & \sum_{i=1}^n \left\{\int \widehat{w}_{i,j}(u) K_{h_j}(x-u) p(u,\boldsymbol{\alpha}_i) du\right\}^2 \\ \approx\ & \sum_{i=1}^n \left\{\widehat{w}_{i,j}(x) p(x,\boldsymbol{\alpha}_i)\right\},\end{aligned}$$

which is ignorable comparing to (5.2) under appropriate regularity conditions. Hence

$$\operatorname{var}\{\widehat{f}_j(x)\} \approx \left\{\frac{1}{\sum_{i=1}^n \widehat{w}_{i,j}^2(X_i)}\right\}^2 \frac{1}{h_j} \sum_{i=1}^n \widehat{w}_{i,j}(x) p(x,\boldsymbol{\alpha}_i) \int K^2(u) du.$$

Then the MISE of of $\widehat{f}_j(x)$ can be approximated by

$$\text{MISE}(\widehat{f}_j) \approx \int \left[\int K_{h_j}(x-u) \left\{ \breve{f}(u) - f_j(u) \right\} du + \frac{h_j^2 f_j''(x)}{2} \int t^2 K(t) dt \right]^2 dx \\ + \frac{1}{\left\{ \sum_{i=1}^n \widehat{w}_{i,j}^2(X_i) \right\}^2 h_j} \int \sum_{i=1}^n \widehat{w}_{i,j}(x) p(x, \boldsymbol{\alpha}_i) dx \int K^2(u) du,$$

which relies on $f_j(x)$ and $f_j''(x)$ that can be approximated by estimators from other approaches. In our numerical implementation, we assume $f_j(x) \sim N(\mu_j, \sigma_j^2)$, and estimate them by the classical EM algorithm.

Note that directly finding $h_j, j = 1, \ldots, M$, which minimize the MISE, may not be computationally feasible, partially because that $\widehat{w}_{i,j}(x)$ depends on h_j and need to be evaluated based on iterations, and their relation to h_j has no explicit formula. Based on this MISE, we propose an iterative algorithm to select h_j; in each iteration, we replace $\widehat{w}_{i,j}(x)$ with their estimates from the last iteration.

Given initial bandwidths $(h_1^{(0)}, \ldots, h_M^{(0)})$, we update $(h_1^{(t)}, \ldots, h_M^{(t)})$ for $t = 0, 1, 2, \cdots$ as follows.

Step 1. For every $i = 1, \ldots, n$ and $j = 1, \ldots, M$, update $w_{i,j}^{(t)}(\cdot)$ by the majorization-minimization algorithm given in Section 3 with $h_j^{(t)}$.

Step 2. Update $h_{j,1}^{(t)}$, $j = 1, \ldots, M$ to be the minimizer of

$$\int \left[\int K_{h_j^{(t)}}(x-u) \left\{ \breve{f}(u) - f_j(u) \right\} du + \frac{h_j^2 f_j''(x)}{2} \int t^2 K(t) dt \right]^2 dx \\ + \frac{1}{\left\{ \sum_{i=1}^n \widehat{w}_{i,j}^2(X_i) \right\}^2 h_j} \int \sum_{i=1}^n \widehat{w}_{i,j}(x) p(x, \boldsymbol{\alpha}_i) dx \int K^2(u) du,$$

with $\widehat{w}_{i,j}(\cdot)$ replaced with $w_{i,j}^{(t)}(\cdot)$ from Step 1, and $f_j(\cdot)$ and $f_j''(\cdot)$ replaced with the estimates from other approaches.

Step 3. Let n_j be the positive integer closest to $\sum_{i=1}^n \alpha_{i,j}$, which serves as an estimate of the average number of observations from the j the population. For each $j = 1, \ldots, M$, sort $w_{i,j}^{(t)}$: $w_{(1),j}^{(t)} \geq w_{(2),j}^{(t)} \geq \ldots \geq w_{(n),j}^{(t)}$. Let $\mathcal{S}_j^t = \{X_i : w_{i,j}^{(t)} \geq w_{(n_j),j}^{(t)}\}$. Treating the observations in $\mathcal{S}_j^t$ as if they are from a single population, we apply the available bandwidth-selection method for the classical kernel density estimate to choose h_j. Denote by $h_{j,2}^{(t)}$ the resulting bandwidth; we use it as an upper bound for our selected bandwidth to hold back the potential over-smoothing in Step 2.

Step 4. Based on $h_{j,1}^{(t)}$ from Step 2 and $h_{j,2}^{(t)}$ from Step 3, we update $h_j^{(t+1)} = \min\left\{ h_{j,1}^{(t)}, h_{j,2}^{(t)} \right\}$.

We update Steps 1–4 until convergence to obtain $w_{i,j}^{(\infty)}$ and $h_j^{(\infty)}$; and let

$$\widehat{f}_j(x) = \frac{\sum_{i=1}^n w_{i,j}^{(\infty)} K_{h_j^{(\infty)}}(x - X_i)}{\sum_{i=1}^n w_{i,j}^{(\infty)}}.$$

The philosophy of the selection method in Step 3 is as follows. S_j^t collects the n_j observations that are most likely to come from the jth population based on the preceding iteration. We use these observations to obtain an upper bound of the bandwidth for the corresponding density estimates in the current iteration.

When implementing this algorithm in our numerical studies, we use the quartic kernel, which was also used by Ma et al. (2011). In Step 3, once $\mathcal{S}_j^t$ is obtained, we use R function `dpik()` to obtain $h_{j,2}^{(t)}$, $j = 1, \dots, M$. `dpik()` in the R package `KernSmooth` is implemented by Wand and Matt (publicly available at http://CRAN.R-project.org/package=KernSmooth). This package is based on the kernel methods in Wand and Jones (1995). Furthermore, the initial bandwidths are set to $h_j^{(0)} = h^{(0)}$ for every $j = 1, \dots, M$, where $h^{(0)}$ is the output of `dpik()` based on all the observations $X_1, \dots, X_n$. We iterate Steps 1–4 until

$$\sum_{j=1}^{M} \left(h_j^{(t+1)} - h_j^{(t)} \right)^2 \leq 0.005^2.$$

6. Simulation study

We use the following two simulation studies to examine the numerical performance of our density estimates.

In Study I, we generate data using two populations, i.e. $M = 2$. The first population has a standard normal distribution, so that $f_{0,1} = \phi_0$, where ϕ_0 denotes the pdf of the standard normal distribution. The second population has a mixture normal distribution: $f_{0,2}(x) = \lambda\phi_0(x) + (1 - \lambda)\phi(x - \mu)$; we consider different values of λ and μ so that $f_{0,2}(x)$ has different mixture structures. For every value of (λ, μ), we generate $X_1, \dots, X_n$ with $n = 400$. For every X_i, we set $\boldsymbol{\alpha}_i = (\alpha_{i,1}, \alpha_{i,2})^\tau$ with $\alpha_{i,1} = u_{i,1}/(u_{i,1} + u_{i,2})$, where $u_{i,1}, u_{i,2}$ are generated independently from the uniform distribution over $[0, 1]$. Therefore, approximately 200 observations will come from each of the populations.

In Study II, we simulate densities that mimic the shape of those estimated from the real-data example in Section 7. The data are generated via:

$$\begin{aligned} X_i|\alpha_i &\sim \alpha f_1 + (1 - \alpha) f_2(x) \qquad \text{when} \quad i > n_1 \\ X_i|\alpha_i &\sim 0.677 f_1(x) + 0.323 f_2(x) \qquad \text{when} \quad i \leq n_1, \end{aligned}$$

where $n_1 = 211$, $n_2 = 81$, and $f_1(x)$ and $f_2(x)$ are the pdfs of $N(10.77, 1.19)$ and $0.48N(5.68, 1.04)+0.52N(9.17, 0.78)$ respectively. Here, $N(\mu, \sigma)$ denotes the normal distribution with mean μ and variance σ^2. $f_1(x)$ and $f_2(x)$ are similar to the densities estimated from the real-data example in Section 7. We consider different values of α. When $\alpha = 0$, the simulated data has the same $\boldsymbol{\alpha}$ values as those in the real data.

For every combination of (λ, μ) in Study I and every α in Study II, we repeat the simulation 1000 times and therefore obtain 1000 replicated simulation data sets, $\{X_i, \boldsymbol{\alpha}_i\}_{i=1}^n$; here $n = 400$ for study I and $n = 292$ for Study II.

For both studies, we apply the algorithm in Section 5 to obtain $\widehat{f}_1$ and $\widehat{f}_2$. The 5%, 50%, and 95% point-wise quantiles for $\widehat{f}_1$ (left panel) and $\widehat{f}_2$ (right panel) over 1000 replications are given in Figure 1 (top panels: Study I, $\lambda = 0$, $\mu = 0$; middle panels: Study I, $\lambda = 0.5$, $\mu = 2$; bottom panels: Study II, $\alpha = 0$). We observe that the 90% confidence bands of f_1 and f_2 cover the corresponding true density.

For study II, the number of elements in the support of $\boldsymbol{\alpha}$ is equal to that of the mixing components. A referee pointed out that we can use the following alternative method to estimate f_1 and f_2. Denote by $g_1(x) = \alpha f_1(x) + (1 - \alpha) f_2(x)$ and $g_2(x) = 0.667 f_1(x) + 0.323 f_2(x)$, which are respectively the pdfs of the observations $\{X_{n_1+1}, \ldots, X_n\}$ and $\{X_1, \ldots, X_{n_1}\}$. Therefore $g_1(\cdot)$ and $g_2(\cdot)$ can be estimated separately by classical kernel density estimates based on their corresponding observations. We denote these estimates by $\widehat{g}_{1,A}(\cdot)$ and $\widehat{g}_{2,A}(\cdot)$. As a consequence, the estimates for $f_1(\cdot)$ and $f_2(\cdot)$, denoted by $\widehat{f}_{1,A}(\cdot)$ and $\widehat{f}_{2,A}(\cdot)$, can be obtained by solving the aforementioned linear equations. We call this method the "alternative method" and compare it with other methods in study II.

We compare our method with the methods proposed by Ma et al. (2011) in both studies, and with the alternative method in study II. We compute the average values of the L_1 errors for $\widehat{f}_1$ and $\widehat{f}_2$ over 1000 replications for both studies; for study I, we consider different combinations of λ and μ; for study II, we consider $\alpha = 0, 0.5$, and 1. We give the results together with those of the alternative method and Ma et al. (2011), "OLS, ICV", named the Ma et al. method hereafter, in Table 1. As observed in that paper, the other methods of Ma et al. (2011) give results that are similar or not as good. Here L_1 error is defined to be

$$L_1(\widehat{f}_j) = \int \left| \widehat{f}_j(x) - f_{0,j}(x) \right| dx.$$

Table 1 clearly shows that our method gives smaller or comparable average values of the L_1 errors to those of the alternative and the Ma et al. methods. The improvement is significant, particularly when $f_{0,1}$ and $f_{0,2}$ are simulated similarly (i.e., $\lambda = 0$ and $\mu = 0$). Furthermore, we observe that both the Ma et al. and the alternative methods do not inherit the nonnegativity property of a regular density function.

7. Real-data example

We consider the malaria data described by Vounatsou et al. (1998). The data come from a cross-sectional survey of parasitemia and fever of children less than

TABLE 1

Average values of the L_1 errors for our method and the Ma et al. method. Each value in the table was computed from 1000 replications.

Study I		$L_1(\widehat{f}_1)$			$L_1(\widehat{f}_2)$		
λ	μ	our	Ma et al.		our	Ma et al.	
0	0	0.137	0.165		0.138	0.166	
0	1	0.155	0.169		0.157	0.169	
0.5	1	0.144	0.168		0.139	0.167	
0	2	0.182	0.177		0.181	0.178	
0.5	2	0.158	0.175		0.134	0.163	
0	3	0.191	0.208		0.192	0.205	
0.5	3	0.164	0.186		0.155	0.182	
0	4	0.175	0.220		0.175	0.223	
0.5	4	0.168	0.204		0.192	0.238	
Study II		$L_1(\widehat{f}_1)$			$L_1(\widehat{f}_2)$		
α		our	Ma et al.	alternative	our	Ma et al.	alternative
0		0.186	0.218	0.209	0.189	0.247	0.215
0.5		0.339	0.406	0.449	0.554	0.639	0.761
1		0.140	0.157	0.160	0.406	0.443	0.512

a year old in a village in the Kilombero district of Tanzania (Kitua et al. 1996). They considered a subset of this data for children of between six and nine months collected in two seasons: (1) January–June, the wet season, when malaria prevalence is high; (2) July–December, the dry season, when malaria prevalence is low. We use one of these data sets, which has also been analyzed by Qin and Leung (2005) with other statistical methods.

The measurements are the parasite levels (per μl), ranging from 0 to 399952.1. There are $n_1 = 211$ observations with positive parasite levels from the mixture sample and $n_2 = 81$ observations with positive parasite levels for nonmalaria cases in the community. If we denote these parasite levels (after log transformation) as $X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_n$ with $n = n_1 + n_2$, then

$$X_i|\alpha_i \sim \alpha_i f_1(x) + (1-\alpha_i) f_2(x),$$

where $f_1(x)$ and $f_2(x)$ are the pdfs of the log parasite levels for the malaria and nonmalaria subjects respectively; α_i is the probability that the ith subject is a malaria patient. Clearly, when $i > n_1$, $\alpha_i = 0$ since it is known that all the subjects in this group are nonmalaria patients. When $i \le n_1$, $\alpha_i \approx 0.677$, which is estimated from the ratio of malaria patients to fevered patients in the endemicity and the community (Qin and Leung 2005). Therefore,

$$\begin{aligned} X_i|\alpha_i &\sim f_2(x) \qquad \text{when} \quad i > n_1 \\ X_i|\alpha_i &\sim 0.677 f_1(x) + 0.323 f_2(x) \qquad \text{when} \quad i \le n_1. \end{aligned}$$

We apply our method and the Ma et al. method to $\{X_i, \boldsymbol{\alpha}_i\}_{i=1}^n$ above, where $\boldsymbol{\alpha}_i = (\alpha_i, 1-\alpha_i)^\tau$. The density estimates from our method, named $\widehat{f}_1(x)$ and $\widehat{f}_2(x)$, and Ma et al., name $\widetilde{f}_1(x)$ and $\widetilde{f}_2(x)$, are displayed in Figure 2. The "hat" and "tilde" esitmates for f_1 (and f_2) are similar in shape, but $\widetilde{f}_1(x)$ is not always

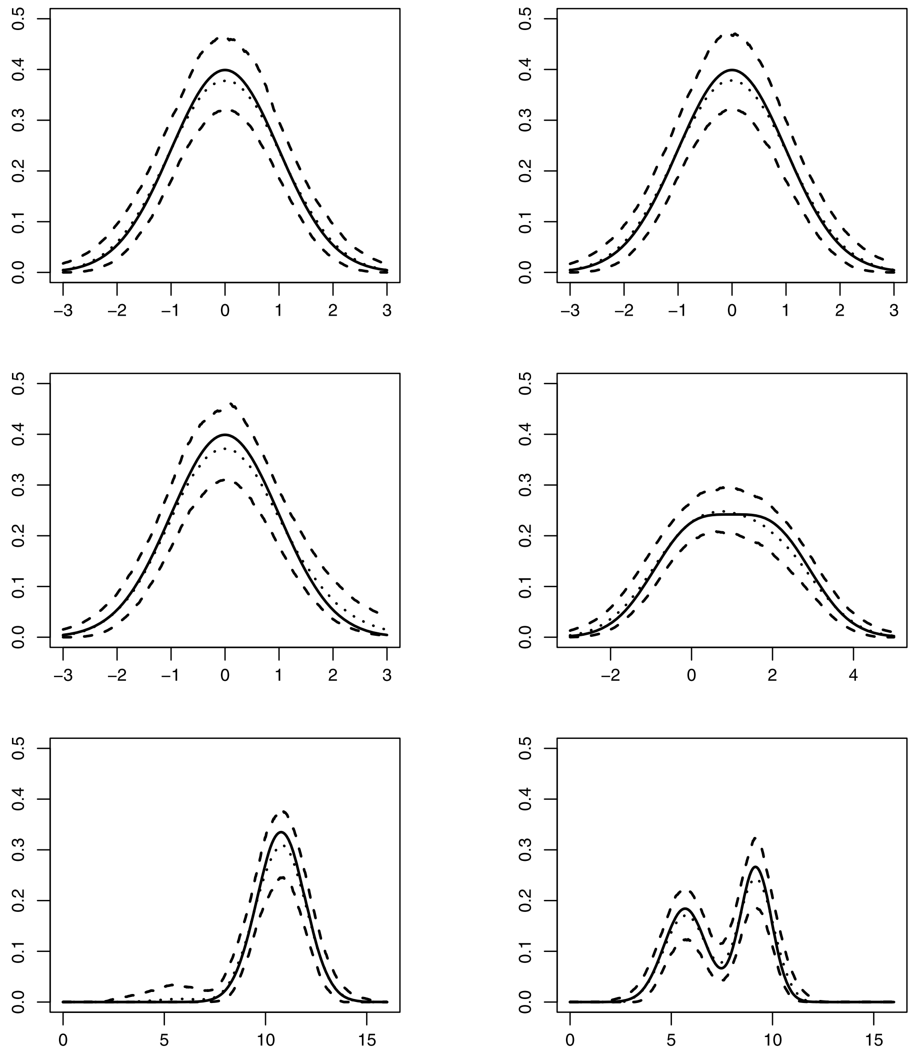

Fig 1. *Point-wise quantile density estimates for Study I (top panels: $\lambda = 0, \mu = 0$; middle panel: $\lambda = 0.5, \mu = 2$), and Study II (bottom panels). In each plot, the solid line is the true density and the other three curves are the point-wise quantiles for density estimates over 1000 replicates: median (dotted), 5% (dashed), and 95% (dashed).*

nonnegative. Considering these estimates together with the observations in our simulation studies, we expect that $\widehat{f}_1(x)$ and $\widehat{f}_2(x)$ are more accurate than $\widetilde{f}_1(x)$ and $\widetilde{f}_2(x)$. Figure 3 presents histograms for the nonmalaria sample (i.e. that for $f_2(x)$) and the mixture sample (i.e. that for $0.677f_1(x) + 0.323f_2(x)$) with the corresponding density estimates from our method. From this figure, we observe that our density estimates agree well with the observed data (see the histogram of the observations from the relevant sample).

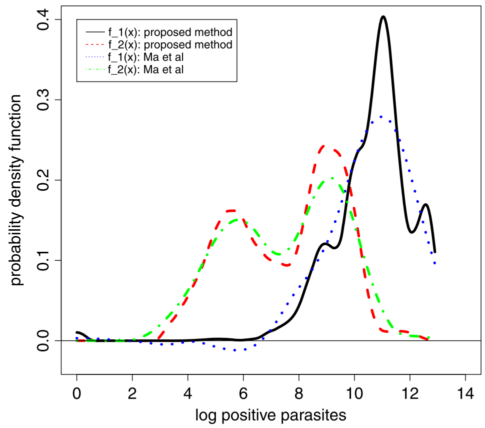

FIG 2. *Component density estimates for malaria data based on our method and Ma et al.*

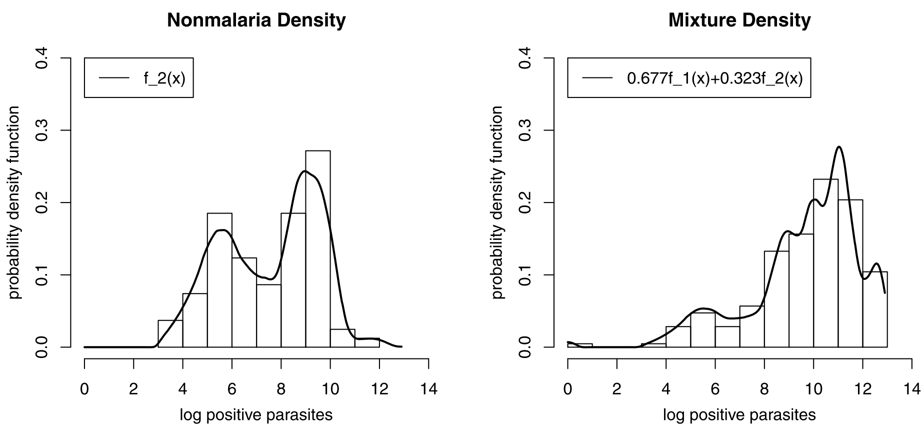

FIG 3. *Histograms for the nonmalaria sample (i.e. that for $f_2(x)$) and the mixture sample (i.e. that for $0.677f_1(x) + 0.323f_2(x)$) along with the corresponding density estimates based on our method.*

Furthermore, from Figure 2, we observe that the density estimate for the log parasite levels of the malaria patients (the black solid line) has a clearer peak and more concentrated curve (centred around 11) than that for the nonmalaria sample (the red dashed line), which has a bimodal feature. From a practical point of view, we argue that this observation is not surprising: the log parasite levels for the nonmalaria sample may result from more than one cause; these causes may lead to different parasite levels and therefore the corresponding density is in fact a mixture of a number of subpopulations. In contrast, the cause for the malaria sample is clear, i.e. the malaria disease; therefore, the density is concentrated and has a clear peak.

8. Discussion

In this paper, we consider the density estimation for several subpopulations, where every observation in the data is composed of a measurement and the probability of every subpopulation that this measurement comes from. With the smoothed likelihood principal, we have proposed density estimators and a majorization-minimization algorithm that numerically computes these density estimates. In theory, we have shown the convergence of the proposed majorization-minimization algorithm, and established the asymptotic L_1 convergence rate of our estimates when the sample size goes to infinity. However, because of the features of the nonlinear operator "$\mathcal{N}_h$", the theoretical development for the asymptotic behaviors of our proposed estimators is technically challenging, and so far, we can only obtain the convergence rate presented in Theorems 5 and 6. We conjecture there is some room to improve this convergence rate. We leave it for future research. We have conducted numerical studies to illustrate the effectiveness of our method and compared our proposed method with the existing methods in the literature. We have observed that our method does lead to comparable or smaller L_1 errors.

As far as we are aware, there are a number of interesting future research topics that can be very closely related to our works in this paper. One is to develop better convergence rate and establish the asymptotic normality of our density estimates. We may also extend the current method and theory to the censored data, as considered by Wang et al. (2012) and Qin et al. (2014). We can further consider imposing various constraints on the densities under the framework of this paper based on practical and scientific interests. For example, in the real data example given in Section 7, it could be reasonable to assume that the posterior probability of a subject having malaria given the log parasite level x is a nondecreasing function of x; this is equivalent to assume that $f_1(x)/f_2(x)$ is a nondecreasing function of x; see Yu et al. (2017) and the references therein for more details. We plan to incorporate such a constraint condition into our method under the framework of this paper and study all the relevant theoretical and numerical properties in the near future. We also expect that the technical tools in this paper may benefit the theoretical development for the smoothed likelihood density estimates for mixture data of other kinds.

Appendix A: Proof of Theorems 1–3

A.1. Proof of Theorem 1

The proof of this theorem uses a strategy similar to that in Levine et al. (2011). Recall that for $(f_1, \dots, f_M) \in \mathcal{C}$, $w_{i,j} = \frac{\alpha_{i,j}\mathcal{N}_{h_j} f_j(X_i)}{\sum_{k=1}^M \alpha_{i,k}\mathcal{N}_{h_k} f_k(X_i)}$. Then for every $i = 1, \dots, n$, $\sum_{j=1}^M w_{i,j} = 1$. By the concavity of the logarithm function, we have for every $(g_1, \dots, g_M) \in \mathcal{C}$,

$$l_n(g_1, \dots, g_M) - l_n(f_1, \dots, f_M)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \log \frac{\sum_{j=1}^{M} \alpha_{ij} \mathcal{N}_{h_j} g_j(X_i)}{\sum_{j=1}^{M} \alpha_{ij} \mathcal{N}_{h_j} f_j(X_i)} \\
&= \sum_{i=1}^{n} \log \sum_{j=1}^{M} w_{i,j} \frac{\mathcal{N}_{h_j} g_j(X_i)}{\mathcal{N}_{h_j} f_j(X_i)} \\
&\geq \sum_{i=1}^{n} \sum_{j=1}^{M} w_{i,j} \left\{ \log \mathcal{N}_{h_j} g_j(X_i) - \log \mathcal{N}_{h_j} f_j(X_i) \right\} \\
&= \sum_{j=1}^{M} \left\{ b_j(g_1, \ldots, g_M) - b_j(f_1, \ldots, f_M) \right\}, \qquad \text{(A.1)}
\end{aligned}
$$

where

$$
\begin{aligned}
b_j(g_1, \ldots, g_M) &= \sum_{i=1}^{n} w_{i,j} \log \mathcal{N}_{h_j} g_j(X_i) \\
&= \int \sum_{i=1}^{n} w_{i,j} K_h(u - X_i) \log g_j(u) du, \qquad \text{(A.2)}
\end{aligned}
$$

which is maximized when $g_j(x) = \frac{\sum_{i=1}^{n} w_{i,j} K_h(x - X_i)}{\sum_{i=1}^{n} w_{i,j}} = f_j^{\mathcal{G}}(x)$. This together with (A.1) completes the proof of this theorem. ∎

A.2. Proof of Theorem 2

We first show necessity. Assume $l_n(\widehat{f}_1, \ldots, \widehat{f}_M) = \sup_{(f_1,\ldots,f_M)\in\mathcal{C}} l_n(f_1, \ldots, f_M)$. Based on Theorem 1, we immediately have $l(\widehat{f}_1, \ldots, \widehat{f}_M) = l(\mathcal{G}(\widehat{f}_1, \ldots, \widehat{f}_M))$. Next we show that $(\widehat{f}_1, \ldots, \widehat{f}_M) = \mathcal{G}(\widehat{f}_1, \ldots, \widehat{f}_M)$ almost surely under the Lebesgue measure.

With exactly the same calculation as for (A.1) and (A.2), we have

$$
\begin{aligned}
0 &= l(\mathcal{G}(\widehat{f}_1, \ldots, \widehat{f}_M)) - l(\widehat{f}_1, \ldots, \widehat{f}_M) \\
&\geq \sum_{j=1}^{M} \left\{ \left(\sum_{i=1}^{n} \widehat{w}_{i,j} \right) \int \widehat{f}_j^{\mathcal{G}}(x) \log \frac{\widehat{f}_j^{\mathcal{G}}(x)}{\widehat{f}_j(x)} dx \right\},
\end{aligned}
$$

where $\widehat{f}_j^{\mathcal{G}}$ denotes the jth component of $\mathcal{G}(\widehat{f}_1, \ldots, \widehat{f}_M)$, $\widehat{w}_{i,j} = \frac{\alpha_{i,j} \mathcal{N}_{h_j} \widehat{f}_j(X_i)}{\sum_{k=1}^{M} \alpha_{i,k} \mathcal{N}_{h_k} \widehat{f}_k(X_i)}$. On the other hand, since $\widehat{f}_j^{\mathcal{G}}$ and $\widehat{f}_j$ are pdfs, we have

$$
\int \widehat{f}_j^{\mathcal{G}}(x) \log \frac{\widehat{f}_j^{\mathcal{G}}(x)}{\widehat{f}_j(x)} dx \geq 0.
$$

Furthermore, for every $j = 1, \ldots, M$, since $\sum_{i=1}^{n} \alpha_{i,j} > 0$ and $(\widehat{f}_1, \ldots, \widehat{f}_M) \in \mathcal{C}$, we have $\sum_{i=1}^{n} \widehat{w}_{i,j} > 0$. Therefore,

$$
\int \widehat{f}_j^{\mathcal{G}}(x) \log \frac{\widehat{f}_j^{\mathcal{G}}(x)}{\widehat{f}_j(x)} dx = 0,
$$

which together with the fact that $\log(\cdot)$ is strictly concave leads to $\widehat{f}_j^{\mathcal{G}}(x) = \widehat{f}_j(x)$ almost surely under the Lebesgue measure. That is, $(\widehat{f}_1, \ldots, \widehat{f}_M) = \mathcal{G}(\widehat{f}_1, \ldots, \widehat{f}_M)$ almost surely under the Lebesgue measure as claimed before.

We proceed to show sufficiency. Assume $(\widehat{f}_1, \ldots, \widehat{f}_M) = \mathcal{G}(\widehat{f}_1, \ldots, \widehat{f}_M)$. Let $\widehat{\boldsymbol{f}} = (\widehat{f}_1, \ldots, \widehat{f}_M)$. For an arbitrary $\boldsymbol{f} = (f_1, \ldots, f_M) \in \mathcal{F}_n$, we need to show that $l_n(\boldsymbol{f}) \le l_n(\widehat{\boldsymbol{f}})$.

Define

$$H(t) = l_n(\widehat{\boldsymbol{f}} + t(\boldsymbol{f} - \widehat{\boldsymbol{f}})), \tag{A.3}$$

with $t \in [0, 1]$. Next, we verify that $H(\cdot)$ has the following properties:

(P1). $H(t)$ is a concave function in $[0, 1]$.
(P2). $H(t)$ is continuously differentiable in $(0, 1)$, $H'(0+)$ exists, and $H'(0+) = 0$.

We first show (P1) above. Note that l_n is concave in $\mathcal{C}$, so we immediately have for every $t_1, t_2 \in [0, 1]$,

$$\begin{aligned} H\left(\frac{t_1 + t_2}{2}\right) &= l_n\left(\frac{\left\{\widehat{\boldsymbol{f}} + t_1(\boldsymbol{f} - \widehat{\boldsymbol{f}})\right\} + \left\{\widehat{\boldsymbol{f}} + t_2(\boldsymbol{f} - \widehat{\boldsymbol{f}})\right\}}{2}\right) \\ &\ge \frac{1}{2} l_n\left(\widehat{\boldsymbol{f}} + t_1(\boldsymbol{f} - \widehat{\boldsymbol{f}})\right) + \frac{1}{2} l_n\left(\widehat{\boldsymbol{f}} + t_2(\boldsymbol{f} - \widehat{\boldsymbol{f}})\right) \\ &= \frac{1}{2} H(t_1) + \frac{1}{2} H(t_2), \end{aligned}$$

leading to (P1).

We proceed to show (P2). First, to verify that $H(t)$ is continuously differentiable in $(0, 1)$ and the existence of $H'(0+)$, it suffices to verify that for every $x \in S_x$ and $j = 1, \ldots, M$, $\int K_h(u - x) \log\left[\widehat{f}_j(u) + t\left\{f_j(u) - \widehat{f}_j(u)\right\}\right] du$ is continuously differentiable when $t \in (0, 1)$ and right differentiable at $t = 0$, and that the derivative can be exchanged with the integration. This is valid because of the definition of $\mathcal{F}_n$ and the dominant convergence theorem. Therefore, it remains to verify $H'(0+) = 0$. For notational convenience, we write $\boldsymbol{f}_t = \widehat{\boldsymbol{f}} + t(\boldsymbol{f} - \widehat{\boldsymbol{f}}) = (f_{1,t}, \ldots, f_{M,t})$ and let $(f_{1,t}^{\mathcal{G}}, \ldots, f_{M,t}^{\mathcal{G}}) = \mathcal{G}(f_{1,t}, \ldots, f_{M,t})$. Using the chain rule for derivatives, we have for every $t \in (0, 1)$,

$$\begin{aligned} H'(t) &= \sum_{i=1}^{n} \sum_{j=1}^{M} \frac{\alpha_{i,j} \mathcal{N}_{h_j} f_{j,t}}{\sum_{k=1}^{M} \alpha_{i,k} \mathcal{N}_{h_k} f_{k,t}(X_i)} \int \frac{K_{h_j}(u - X_i)}{f_{j,t}(u)} \left\{f_j(u) - \widehat{f}_j(u)\right\} du \\ &= \sum_{j=1}^{M} \int \frac{f_{j,t}^{\mathcal{G}}(u)}{f_{j,t}(u)} \left\{f_j(u) - \widehat{f}_j(u)\right\} du. \end{aligned}$$

Noting that $f_{j,0} = \widehat{f}_j$ and $\widehat{f}_j^{\mathcal{G}} = \widehat{f}_j$ almost surely under the Lebesgue measure because of our assumption, we immediately have

$$\begin{aligned} H'(0+) &= \sum_{j=1}^{M} \int \frac{f_{j,0}^{\mathcal{G}}(u)}{f_{j,0}(u)} \left\{ f_j(u) - \widehat{f}_j(u) \right\} du \\ &= \sum_{j=1}^{M} \int \left\{ f_j(u) - \widehat{f}_j(u) \right\} du = 0, \end{aligned}$$

which completes our proof of (P2) above. Now, from (P1) and (P2) and the property of concave functions, we immediately have

$$H(1) \leq H(0) + H'(0+)(1-0),$$

which is

$$l_n(\boldsymbol{f}) \leq l_n(\widehat{\boldsymbol{f}}).$$

This completes the proof of the theorem. ∎

A.3. Proof of Theorem 3

Since $(f_1^s, \ldots, f_M^s) \in \mathcal{F}_n$, for every $j = 1, \ldots, M$, we can write

$$f_j^s(x) = \frac{\sum_{i=1}^n w_{i,j}^s K_{h_j}(x - X_i)}{\sum_{i=1}^n w_{i,j}^s}.$$

Clearly, for every s, the set of coefficients $w^s = \{w_{i,j}^s : i = 1, \ldots, n; j = 1, \ldots, M\}$ belongs to

$$\Omega_w = \left\{ \{w_{i,j} : i = 1, \ldots, n; j = 1, \ldots, M\} : 0 \leq w_{i,j} \leq 1 \right\},$$

which is a closed subset of $\mathbb{R}^{nM}$. Therefore, there exists a subsequence of w^s, namely w^{s_l}, and $w^\infty = \{w_{i,j}^\infty : i = 1, \ldots, n; j = 1, \ldots, M\} \in \Omega_w$, such that

$$\lim_{l \to \infty} w^{s_l} = w^\infty. \tag{A.4}$$

Let

$$f_j^\infty(x) = \frac{\sum_{i=1}^n w_{i,j}^\infty K_{h_j}(x - X_i)}{\sum_{i=1}^n w_{i,j}^\infty}.$$

We can readily check that

$$\lim_{l \to \infty} f_j^{s_l}(x) = f_j^\infty(x) \tag{A.5}$$

for all $x \in S_x$ and hence

$$\lim_{l \to \infty} l_n(f_1^{s_l}, \ldots, f_M^{s_l}) = l_n(f_1^\infty, \ldots, f_M^\infty),$$

which together with Theorem 1 ensures that

$$\lim_{s\to\infty} l_n(f_1^s,\dots,f_M^s) = l_n(f_1^\infty,\dots,f_M^\infty)$$

almost surely under the Lebesgue measure. It remains to show that

$$\mathcal{G}(f_1^\infty,\dots,f_M^\infty) = (f_1^\infty,\dots,f_M^\infty). \tag{A.6}$$

Then based on Theorem 2, we have

$$l_n(f_1^\infty,\dots,f_M^\infty) = l_n(\widehat{f}_1,\dots,\widehat{f}_M),$$

which completes our proof of this theorem.

In fact, along the subsequence s_l defined above, using the same derivations as for (A.1) and (A.2), we have

$$\begin{aligned} 0 &= \lim_{l\to\infty}\left\{l_n(f_1^{s_l+1},\dots,f_M^{s_l+1}) - l_n(f_1^{s_l},\dots,f_M^{s_l})\right\} \\ &\geq \lim_{l\to\infty}\sum_{j=1}^M\left\{\left(\sum_{i=1}^n w_{i,j}^{s_l}\right)\int f_j^{s_l+1}(x)\log\frac{f_j^{s_l+1}(x)}{f_j^{s_l}(x)}dx\right\}\geq 0. \end{aligned}$$

Hence

$$\lim_{l\to\infty}\sum_{j=1}^M\left\{\left(\sum_{i=1}^n w_{i,j}^{s_l}\right)\int f_j^{s_l+1}(x)\log\frac{f_j^{s_l+1}(x)}{f_j^{s_l}(x)}dx\right\} = 0. \tag{A.7}$$

On the other hand, (A.5) implies $\lim_{l\to\infty}\mathcal{G}(f_1^{s_l},\dots,f_M^{s_l}) = \mathcal{G}(f_1^\infty,\dots,f_M^\infty)$, or equivalently,

$$\lim_{l\to\infty}(f_1^{s_l+1},\dots,f_M^{s_l+1}) = (f_1^{\infty,\mathcal{G}},\dots,f_M^{\infty,\mathcal{G}}), \tag{A.8}$$

where $(f_1^{\infty,\mathcal{G}},\dots,f_M^{\infty,\mathcal{G}}) = \mathcal{G}(f_1^\infty,\dots,f_M^\infty)$. Combining (A.4), (A.5), (A.7), and (A.8), we have

$$\sum_{j=1}^M\left\{\left(\sum_{i=1}^n w_{i,j}^{\infty}\right)\int f_j^{\infty,\mathcal{G}}(x)\log\frac{f_j^{\infty,\mathcal{G}}(x)}{f_j^{\infty}(x)}dx\right\} = 0,$$

which indicates that for every $j = 1,\dots,M$,

$$\int f_j^{\infty,\mathcal{G}}(x)\log\frac{f_j^{\infty,\mathcal{G}}(x)}{f_j^{\infty}(x)}dx = 0. \tag{A.9}$$

Since $\log(\cdot)$ is strictly concave, (A.9) implies $f_j^\infty(x) = f_j^{\infty,\mathcal{G}}(x)$. That is,

$$\mathcal{G}(f_1^\infty,\dots,f_M^\infty) = (f_1^\infty,\dots,f_M^\infty)$$

almost surely under the Lebesgue measure, which proves (A.6), and therefore completes the proof of this theorem. ■

Appendix B: Proof of Theorems 4–6

B.1. Preliminaries

The proofs of Theorems 4–6 rely heavily on well-developed results for M-estimation in empirical processes. We use van der Vaart and Wellner (1996) (VM) as the main reference and adapt the commonly used notation of this book. In this section, we introduce some necessary notation and review two important results.

We first review some notation necessary for introducing the result for the M-estimation. Let "$\lesssim$" ("$\gtrsim$") denote smaller (greater) than, up to a universal constant. Throughout, we will use C to denote a sufficiently large universal constant. For a function $m(x, \boldsymbol{\alpha})$, we define

$$\begin{aligned} \mathbb{P}_n\{m(X, \boldsymbol{\alpha})\} &= \frac{1}{n}\sum_{i=1}^{n} m(X_i, \boldsymbol{\alpha}_i); \\ \mathbb{P}\{m(X, \boldsymbol{\alpha})\} &= \int_{S_\gamma}\int_{\mathbb{R}} m(x, \boldsymbol{\alpha})\gamma(\boldsymbol{\alpha})\widetilde{p}_0(x, \boldsymbol{\alpha})dxd\boldsymbol{\alpha}. \end{aligned}$$

When $m(x, \boldsymbol{\alpha})$ is a nonrandom function, $\mathbb{P}\{m(X, \boldsymbol{\alpha})\} = E_0\{m(X, \boldsymbol{\alpha})\}$, where E_0 means that the expectation is taken under $\gamma(\boldsymbol{\alpha})\widetilde{p}_0(x, \boldsymbol{\alpha})$. This convention will be used throughout the proofs. For a set $\mathcal{M}$ of functions of $(x, \boldsymbol{\alpha})$, we define

$$\mathbb{G}_n m = \sqrt{n}\left[\mathbb{P}_n\{m(X, \boldsymbol{\alpha})\} - \mathbb{P}\{m(X, \boldsymbol{\alpha})\}\right] \text{ for } m \in \mathcal{M}; \tag{B.1}$$

$$\|\mathbb{G}_n\|_{\mathcal{M}} = \sup_{m\in\mathcal{M}} |\mathbb{G}_n m|. \tag{B.2}$$

Let $\mathcal{P}_n$ denote the class of functions:

$$\mathcal{P}_n = \left\{ p(x, \boldsymbol{\alpha}) = \sum_{j=1}^{M} \alpha_j \mathcal{N}_{h_j} f_j(x) : (f_1, \dots, f_M) \in \mathcal{F}_n \right\}, \tag{B.3}$$

where $\mathcal{F}_n$ is defined by (3.3). For any nonnegative functions $p(x, \boldsymbol{\alpha})$ and $p_1(x, \boldsymbol{\alpha})$, we define

$$\begin{aligned} m_{p,p_1}(X, \boldsymbol{\alpha}) &= \log \frac{p(X, \boldsymbol{\alpha}) + p_1(X, \boldsymbol{\alpha})}{2p_1(X, \boldsymbol{\alpha})}; \\ \mathbb{M}_n(p, p_1) &= \mathbb{P}_n\left\{m_{p,p_1}(X, \boldsymbol{\alpha})\right\} = \frac{1}{n}\sum_{i=1}^{n} m_{p,p_1}(X_i, \boldsymbol{\alpha}_i); \\ M_n(p, p_1) &= \mathbb{P}\left\{m_{p,p_1}(X, \boldsymbol{\alpha})\right\} = \int_{S_\gamma}\int_{\mathbb{R}} m_{p,p_1}(x, \boldsymbol{\alpha})\gamma(\boldsymbol{\alpha})\widetilde{p}_0(x, \boldsymbol{\alpha})dxd\boldsymbol{\alpha}; \\ \mathcal{M}_{n,\delta,p,p_1} &= \left\{m_{p,p_1} - m_{p_1,p_1} : p \in \mathcal{P}_n, d(\gamma p, \gamma p_1) < \delta\right\}. \end{aligned}$$

With the above preparation, we present an important lemma, which is an application of Theorem 3.4.1 of van der Vaart and Wellner (1996) to our setup. It serves as the basis for the proof of Theorem 4.

Lemma 1. *Suppose $\mathbb{M}_n$, M_n, and $\|\mathbb{G}_n\|_{\mathcal{M}_{n,\delta,p,\widetilde{p}_0}}$ are as defined above, $\widetilde{p}_0(x,\boldsymbol{\alpha}) = \sum_{j=1}^M \alpha_j f_{0,j}(x)$ is the true conditional density of X given $\boldsymbol{\alpha}$, and $\gamma(\cdot)$ is the marginal density of $\boldsymbol{\alpha}$. Suppose further that the following three conditions are satisfied:*

(a) for every n and $p \in \mathcal{F}_n$, $M_n(p,\widetilde{p}_0) - M_n(\widetilde{p}_0,\widetilde{p}_0) \lesssim -d^2(\gamma p, \gamma\widetilde{p}_0)$;
(b) for every n and $\delta > 0$, $E_0\|\mathbb{G}_n\|_{\mathcal{M}_{n,\delta,p,\widetilde{p}_0}} \lesssim \phi_n(\delta)$ for functions $\phi_n(\cdot)$ such that $\phi_n(\delta)/\delta^\alpha$ is decreasing on $(0,\infty)$ for some $\alpha < 2$;
(c) $\mathbb{M}_n(\widehat{p},\widetilde{p}_0) \geq \mathbb{M}_n(\widetilde{p}_0,\widetilde{p}_0) - O_p(r_n^{-2})$, where $\widehat{p}(x,\boldsymbol{\alpha}) = \sum_{j=1}^M \alpha_j \mathcal{N}_{h_j}\widehat{f}_j(x)$ and r_n satisfies $r_n^2\phi(1/r_n) \leq \sqrt{n}$, for every n.

Then we have

$$r_n d(\gamma\widehat{p}, \gamma\widetilde{p}_0) = O_p(1).$$

A difficult step in the application of the above lemma is to verify Condition (b). A useful technique is to establish a connection between $E_0\|\mathbb{G}_n\|_{\mathcal{M}_{n,\delta,p,\widetilde{p}_0}}$ and the bracketing integral of the class $\gamma\mathcal{P}_n$. For convenience of presentation in the next subsections, we introduce some necessary notation and review an important lemma.

We first introduce the concept of bracketing numbers, which will be used to define the bracketing integral. Consider a set $\mathcal{M}$ of functions and the norm $\|\cdot\|$ defined on the set $\mathcal{M}$. For any $\epsilon > 0$, the bracketing number $N_{[]}(\epsilon, \mathcal{M}, \|\cdot\|)$ is the minimum number of N for which there exists a set of pairs of functions or brackets $\{[l_j, u_j], j = 1, \ldots, N\}$ such that (i) $\|u_j - l_j\| < \epsilon$ and (ii) for any $m \in \mathcal{M}$, there exists a $j = j(m)$ such that $l_j \leq m \leq u_j$. The bracketing integral of the class $\mathcal{M}$ is then defined to be

$$\widetilde{J}_{[]}(\delta, \mathcal{M}, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[]}(\epsilon, \mathcal{M}, \|\cdot\|)}\, d\epsilon. \tag{B.4}$$

Next, we review a result about the bracketing number of a class of continuous functions, which will be useful to calculate the bracketing number of $\gamma\mathcal{P}_n$ and the bracketing integral of $\gamma\mathcal{P}_n$. For every function f defined on $\mathcal{A} \subset \mathbb{R}$ and a positive integer a, define the norm

$$\|f\|_a = \max_{k:k\leq a} \sup_{x\in\mathcal{A}} |f^{(k)}(x)|,$$

where $f^{(k)}(x)$ denotes the kth-order derivative of f; $f^{(0)} = f$. Let $C_W^a(\mathcal{A})$ be the set of all continuous functions $f : \mathcal{A} \mapsto \mathbb{R}$ with $\|f\|_a \leq W$.

Lemma 2. *Let $\mathcal{A}$ be an interval with finite length in $\mathbb{R}$. Then*

$$\log N_{[]}\Big(\epsilon, C_1^a(A), L_r(Q)\Big) \lesssim 1/\epsilon^{1/a},$$

for every $r \geq 1$, $\epsilon > 0$, and any probability measure Q on $\mathbb{R}$, where the universal constant in "$\lesssim$" depends only on a and the length of $\mathcal{A}$. Here $L_r(Q)$ is the L_r-norm under the probability measure Q.

This lemma is a special case of Corollary 2.7.2 of VW; see p. 157.

B.2. Proof of Theorem 4

In this section, we show Theorem 4, which establishes the consistency of $d(\gamma\widehat{p}, \gamma\widetilde{p}_0)$ and plays a key role in the proofs of Theorems 5 and 6. Recall that we need to show

$$d(\gamma\widehat{p}, \gamma\widetilde{p}_0) = O_p(h^{0.5}) + O_p(n^{-0.5+\vartheta}h^{-0.5}).$$

This proof contains three steps. At each step, we verify one condition in Lemma 1.

In *Step 1*, we verify that Condition (a) in Lemma 1 is satisfied. We need the following lemma giving a property of the smoothing operator $\mathcal{N}_h$.

Lemma 3. *Given $\mathcal{N}_h f(x)$ defined by (2.2), for any density function $f(x)$, we have*

$$\int_{\mathbb{R}} \mathcal{N}_h f(x)dx \le 1.$$

Proof. By the concavity of the logarithm and Jensen's inequality, the result follows. ∎

We now verify Condition (a). For any $p \in \mathcal{F}_n$, let $q = (p + \widetilde{p}_0)/2$. Since $\log x \le 2(\sqrt{x} - 1)$ for every $x > 0$, we have

$$\begin{aligned} M_n(p, \widetilde{p}_0) - M_n(\widetilde{p}_0, \widetilde{p}_0) &= E_0\left(\log \frac{q}{\widetilde{p}_0}\right) \le 2E_0\left(\frac{q^{1/2}}{\widetilde{p}_0^{1/2}} - 1\right) \\ &= -d^2(\gamma\widetilde{p}_0, \gamma q) + \int_{\mathbb{R}} \gamma(q - \widetilde{p}_0)dxd\boldsymbol{\alpha} \\ &= -d^2(\gamma\widetilde{p}_0, \gamma q) + 0.5\int_{\mathbb{R}} \gamma\left\{\int_{\mathbb{R}} pdx - 1\right\}d\boldsymbol{\alpha} \\ &\le -d^2(\gamma\widetilde{p}_0, \gamma q), \end{aligned}$$

where, to achieve the last "$\le$", we have applied Lemma 3. Note that

$$\left|\sqrt{\gamma p} - \sqrt{\gamma\widetilde{p}_0}\right| = 2\frac{\sqrt{\gamma q} + \sqrt{\gamma\widetilde{p}_0}}{\sqrt{\gamma p} + \sqrt{\gamma\widetilde{p}_0}}\left|\sqrt{\gamma q} - \sqrt{\gamma\widetilde{p}_0}\right| \le 4\left|\sqrt{\gamma q} - \sqrt{\gamma\widetilde{p}_0}\right|,$$

which implies that

$$-d^2(\gamma\widetilde{p}_0, \gamma q) \le -\frac{1}{16}d^2(\gamma\widetilde{p}_0, \gamma p).$$

Therefore

$$M_n(p, \widetilde{p}_0) - M_n(\widetilde{p}_0, \widetilde{p}_0) \le -\frac{1}{16}d^2(\gamma\widetilde{p}_0, \gamma p).$$

Hence Condition (a) of Lemma 1 is satisfied.

In *Step 2*, we establish the upper bound for $E_0\|\mathbb{G}_n\|_{\mathcal{M}_{n,\delta,p,\widetilde{p}_0}}$. Following exactly the same process as for Theorem 3.4.4 of VM, we get

$$E_0\|\mathbb{G}_n\|_{\mathcal{M}_{n,\delta,p,\widetilde{p}_0}} \lesssim \widetilde{J}_{[]}(\delta, \gamma\mathcal{P}_n, d)\left\{1 + \frac{\widetilde{J}_{[]}(\delta, \gamma\mathcal{P}_n, d)}{\delta^2\sqrt{n}}\right\}, \tag{B.5}$$

where the bracketing integral $\widetilde{J}_{[]}$ is defined in (B.4). Lemma 4 below gives the upper bound for $\widetilde{J}_{[]}(\delta, \gamma\mathcal{P}_n, d)$, which, combined with (B.5), immediately leads to $\phi_n(\cdot)$ in Condition (b) of Lemma 1.

Lemma 4. *Let a be an arbitrary positive integer. Then*

$$\widetilde{J}_{[]}(\delta, \gamma\mathcal{P}_n, d) \lesssim \delta^{1-1/(2a)}\sum_{j=1}^{M}|\log h_j|^{0.5}h_j^{-0.5-0.25/a}. \tag{B.6}$$

Proof. Consider

$$\mathcal{P}_{n,j} = \left\{\mathcal{N}_{h_j}f : f = \frac{\sum_{i=1}^n w_{i,j}K_{h_j}(x - X_i)}{\sum_{i=1}^n w_{i,j}}; 0 \le w_{i,j} \le 1\right\}.$$

Let $S_x^* = [c_1 - \Delta, c_2 + \Delta]$, where $\Delta > 0$ is an arbitrarily small constant. Note that for any $g \in \mathcal{P}_{n,j}$, $g(x) = 0$ when $x \notin S_x^*$. In the following proof, we focus on the function class defined on S_x^*.

With Condition 2, we first check that for any arbitrary $a > 0$, we have

$$\left(\frac{h_j}{|\log h_j|}\right)^a\sqrt{h_j}C_3\sqrt{\mathcal{P}_{n,j}} \subset C_1^a(S_x^*) \tag{B.7}$$

for some universal constant $C_3 > 0$. For presentational brevity, we show only the case $a = 1$; the cases $a = 2, 3, \ldots$, can be proved similarly. For any $\sqrt{N_{h_j}f} \in \sqrt{\mathcal{P}_{n,j}}$, using the conditions that $K(t)$ is bounded below and $|K'(t)|$ is bounded in Condition (b), and by straightforward calculus, we have

$$\begin{aligned}
&\left|\left(\sqrt{N_{h_j}f}\right)'\right| \\
&\lesssim \frac{1}{h_j}\exp\left\{0.5\int_{\mathbb{R}}K(t)\log f(x+th_j)dt\right\}\int_{\mathbb{R}}K(t)\left|\log f(x+th_j)\right|dt \\
&\le \frac{1}{h_j}\exp\left[0.5\int_{\mathbb{R}}K(t)\left\{\log f(x+th_j)\right\}^+dt - 0.5\int_{\mathbb{R}}K(t)\left\{\log f(x+th_j)\right\}^-dt\right] \\
&\quad\times\left[\int_{\mathbb{R}}K(t)\left\{\log f(x+th_j)\right\}^+dt + \int_{\mathbb{R}}K(t)\left\{\log f(x+th_j)\right\}^-dt\right] \\
&\lesssim \frac{1}{h_j^{1.5}}\exp\left[-0.5\int_{\mathbb{R}}K(t)\left\{\log f(x+th_j)\right\}^-dt\right] \\
&\quad\times\left[\log(1/h_j) + \int_{\mathbb{R}}K(t)\left\{\log f(x+th_j)\right\}^-dt\right]
\end{aligned}$$

$$\lesssim \frac{1}{h_j^{1.5}} \log(1/h_j) + \frac{1}{h_j^{1.5}}$$
$$\lesssim \frac{1}{h_j^{1.5}} \log(1/h_j),$$

where we have used the facts that $\int_{\mathbb{R}} K(t) \{\log f(x + th_j)\}^+ \, dt \lesssim \log(1/h_j)$ and that for any $x \geq 0$, $x \exp(-0.5x) < 1$. Therefore, by Lemma 2 and viewing d on $\frac{h_j^{2a+1}}{|\log h_j|^{2a}} C_3^2 \mathcal{P}_{n,j}$ as the L_2-distance on $\frac{h_j^{a+0.5}}{|\log h_j|^{a}} C_3 \sqrt{\mathcal{P}_{n,j}}$, we have

$$\log N_{[]} \left(\epsilon, \frac{h_j^{2a+1}}{|\log h_j|^{2a}} C_3^2 \mathcal{P}_{n,j}, d \right)$$
$$= \log N_{[]} \left(\epsilon, \frac{h_j^{a+0.5}}{|\log h_j|^{a}} C_3 \sqrt{\mathcal{P}_{n,j}}, L_2 \right) \lesssim 1/\epsilon^{1/a}.$$

On the other hand, under d, every ϵ-length bracket of $\frac{h_j^{2a+1}}{|\log h_j|^{2a}} C_3^2 \mathcal{P}_{n,j}$ is a length $\epsilon |\log h_j|^a / (h_j^{a+0.5} C_3)$ bracket in $\mathcal{P}_{n,j}$. Therefore,

$$\log N_{[]} \Big(\epsilon |\log h_j|^a / (h_j^{a+0.5} C_3), \mathcal{P}_{n,j}, d \Big)$$
$$= \log N_{[]} \left(\epsilon, \frac{h_j^{2a+1}}{|\log h_j|^{2a}} C_3^2 \mathcal{P}_{n,j}, d \right) \lesssim 1/\epsilon^{1/a},$$

which immediately implies that

$$\log N_{[]} \left(\epsilon, \mathcal{P}_{n,j}, d \right) \lesssim |\log h_j| / \{\epsilon (h_j^{a+0.5})\}^{1/a}. \tag{B.8}$$

For notational simplicity, we write $N_j = N_{[]}(\epsilon, \mathcal{P}_{n,j}, d)$. Then for every j, there exists a set of ϵ-brackets $\mathcal{B}_j = \{[u_{i_j,j}, v_{i_j,j}] : i_j = 1, \ldots, N_j\}$ that covers $\mathcal{P}_{n,j}$. Let

$$\mathcal{B} = \left\{ \Big[p_L(x, \boldsymbol{\alpha}), p_U(x, \boldsymbol{\alpha}) \Big] : \begin{array}{l} p_L = \sum_{j=1}^M \alpha_j u_{i_j,j}(x), p_U = \sum_{j=1}^M \alpha_j v_{i_j,j}(x) \\ \text{for every } j, i_j \in \{1, \ldots, N_j\} \end{array} \right\}.$$

Clearly, $\mathcal{B}$ covers $\gamma \mathcal{P}_n$ with $\Pi_{j=1}^M N_j$ brackets.

Next we consider the minimum bracket length. Note that for any $x, x', y, y' \geq 0$, we have

$$\{(x + y)^{1/2} - (x' + y')^{1/2}\}^2 \leq (x^{1/2} - x'^{1/2})^2 + (y^{1/2} - y'^{1/2})^2.$$

Hence for any $[p_L(x, \boldsymbol{\alpha}), p_U(x, \boldsymbol{\alpha})] \in \mathcal{B}$,

$$d^2(p_L, p_U) \leq \sum_{j=1}^M d^2(\alpha_j u_{i_j,j}, \alpha_j v_{i_j,j}) \leq \sum_{j=1}^M d^2(u_{i_j,j}, v_{i_j,j}) \leq M \epsilon^2.$$

This indicates that for every $\epsilon > 0$,

$$\log N_{[]}(\epsilon, \gamma \mathcal{P}_n, d) \lesssim \log N_{[]}(\sqrt{M} \epsilon, \gamma \mathcal{P}_n, d) \leq \sum_{j=1}^M \log N_j \lesssim \sum_{j=1}^M \frac{|\log h_j|}{\epsilon^{1/a} h_j^{1+0.5/a}}. \quad \blacksquare$$

With the help of Lemma 4, we set

$$\begin{aligned}\phi_n(\delta) &= \delta^{1-1/(2a)} \sum_{j=1}^{M} |\log h_j|^{0.5} h_j^{-0.5-0.25/a} \\ &\quad \times \left(1 + \frac{1}{\sqrt{n}\delta^{1+1/(2a)}} \sum_{j=1}^{M} |\log h_j|^{0.5} h_j^{-0.5-0.25/a}\right).\end{aligned}$$

Obviously, $\phi_n(\delta)/\delta^{\alpha}$ with $\alpha = 1$ is a decreasing function of δ. This verifies Condition (b) of Lemma 1.

In *Step 3*, we check

$$\mathbb{M}_n(\widehat{p}, \widetilde{p}_0) \geq \mathbb{M}_n(\widetilde{p}_0, \widetilde{p}_0) + O_p(h). \tag{B.9}$$

Let $p_n(x, \boldsymbol{\alpha}) = \sum_{j=1}^{M} \alpha_j \mathcal{N}_{h_j} \mathcal{S}_{h_j} f_{0,j}(x)$, where for $j = 1, \ldots, M$,

$$\mathcal{S}_{h_j} f_{0,j}(x) = \begin{cases} c_{h_j,j} f_{0,j}(c_2), & x \in [c_2, c_2 + Lh_j] \\ c_{h_j,j} f_{0,j}(x), & x \in [c_1, c_2] \\ c_{h_j,j} f_{0,j}(c_1), & x \in [c_1 - Lh_j, c_1] \\ 0, & \text{otherwise} \end{cases}, \tag{B.10}$$

where $c_{h_j,j}$ is a constant such that $\int_{\mathbb{R}} \mathcal{S}_{h_j} f_{0,j}(x) dx = 1$.

Note that $\mathbb{M}_n(\widetilde{p}_0, \widetilde{p}_0) = 0$ and $\log(x)$ is concave. We have

$$\begin{aligned}
&\mathbb{M}_n(\widehat{p}, \widetilde{p}_0) - \mathbb{M}_n(\widetilde{p}_0, \widetilde{p}_0) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{\widehat{p}(X_i, \boldsymbol{\alpha}_i) + \widetilde{p}_0(X_i, \boldsymbol{\alpha}_i)}{2\widetilde{p}_0(X_i, \boldsymbol{\alpha}_i)} \\
\geq\ & \frac{1}{2n} \sum_{i=1}^{n} \{\log \widehat{p}(X_i, \boldsymbol{\alpha}_i) - \log \widetilde{p}_0(X_i, \boldsymbol{\alpha}_i)\} \\
=\ & \frac{1}{2n} \sum_{i=1}^{n} \{\log \widehat{p}(X_i, \boldsymbol{\alpha}_i) - \log p_n(X_i, \boldsymbol{\alpha}_i)\} \\
& + \frac{1}{2n} \sum_{i=1}^{n} \{\log p_n(X_i, \boldsymbol{\alpha}_i) - \log \widetilde{p}_0(X_i, \boldsymbol{\alpha}_i)\} \\
\geq\ & \frac{1}{2n} \sum_{i=1}^{n} \{\log p_n(X_i, \boldsymbol{\alpha}_i) - \log \widetilde{p}_0(X_i, \boldsymbol{\alpha}_i)\} \equiv I_n,
\end{aligned}$$

where the last "$\geq$" follows from the fact that

$$\begin{aligned}
& \sum_{i=1}^{n} \{\log \widehat{p}(X_i, \boldsymbol{\alpha}_i) - \log p_n(X_i, \boldsymbol{\alpha}_i)\} \\
=\ & l_n(\widehat{f}_1, \ldots, \widehat{f}_M) - l_n(\mathcal{S}_{h_1} f_{0,1}, \ldots, \mathcal{S}_{h_M} f_{0,M}) \geq 0.
\end{aligned}$$

Therefore, to show (B.9), we need only to verify that $I_n = O_p(h)$, which is valid because of Lemma 5 below and Chebyshev's inequality.

Lemma 5. *Assume Conditions 1–3. We have*

$$E_0\left\{\log\frac{p_n(X,\boldsymbol{\alpha})}{\widetilde{p}_0(X,\boldsymbol{\alpha})}\right\} = O(h), \tag{B.11}$$

$$Var_0\left\{\log\frac{p_n(X,\boldsymbol{\alpha})}{\widetilde{p}_0(X,\boldsymbol{\alpha})}\right\} = O(h^2), \tag{B.12}$$

where Var_0 means that the variance is taken under $\gamma(\boldsymbol{\alpha})\tilde{p}_0(x,\boldsymbol{\alpha})$.

Proof. In the proof, we need the approximation of $\log(p_n/\widetilde{p}_0)$. Note that

$$\log(p_n/\widetilde{p}_0) = \log\left(\frac{p_n-\widetilde{p}_0}{\widetilde{p}_0}+1\right).$$

By Condition 3, we have that for $x \in [c_1, c_2]$ and $\boldsymbol{\alpha} \in S_\gamma$,

$$\begin{aligned}\left|\frac{p_n(x,\boldsymbol{\alpha})-\widetilde{p}_0(x,\boldsymbol{\alpha})}{\widetilde{p}_0(x,\boldsymbol{\alpha})}\right| &\lesssim \left|\sum_{j=1}^M \alpha_{i,j}\{\mathcal{N}_{h_j}\mathcal{S}_{h_j}f_{0,j}(x) - f_{0j}(x)\}\right| \\ &\le \sum_{j=1}^M \left|\mathcal{N}_{h_j}\mathcal{S}_{h_j}f_{0,j}(x) - f_{0,j}(x)\right|. \end{aligned} \tag{B.13}$$

Applying Condition 3 again, we further note that

$$\sup_{x\in[c_1,c_2]} |\mathcal{N}_{h_j}\mathcal{S}_{h_j}f_{0,j}(x) - f_{0,j}(x)| = O(h_j) = O(h), \tag{B.14}$$

where the last step follows from Condition 0. Hence

$$\sup_{x\in[c_1,c_2],\ \boldsymbol{\alpha}}\left|\frac{p_n(x,\boldsymbol{\alpha})-\widetilde{p}_0(x,\boldsymbol{\alpha})}{\widetilde{p}_0(x,\boldsymbol{\alpha})}\right| = O(h). \tag{B.15}$$

Applying the second-order Taylor expansion and using (B.15), we get

$$\log(p_n/\widetilde{p}_0) = \frac{p_n(x,\boldsymbol{\alpha})-\widetilde{p}_0(x,\boldsymbol{\alpha})}{\widetilde{p}_0(x,\boldsymbol{\alpha})} + R(x,\boldsymbol{\alpha}), \tag{B.16}$$

where the remaining term $R(x,\alpha)$ satisfies

$$\sup_{x\in[c_1,c_2],\ \boldsymbol{\alpha}} |R(x,\boldsymbol{\alpha})| = O(h^2). \tag{B.17}$$

We now prove (B.11). Combining (B.16) and (B.17), we have

$$\begin{aligned}\left|E_0\left\{\log\frac{p_n(X,\boldsymbol{\alpha})}{\widetilde{p}_0(X,\boldsymbol{\alpha})}\right\}\right| &\le E_0\left|\frac{p_n(X,\boldsymbol{\alpha})-\widetilde{p}_0(X,\boldsymbol{\alpha})}{\widetilde{p}_0(X,\boldsymbol{\alpha})}\right| + O(h^2) \\ &\lesssim \sum_{j=1}^M E_0|\mathcal{N}_{h_j}\mathcal{S}_{h_j}f_{0,j}(X) - f_{0,j}(X)| + O(h^2) = O(h),\end{aligned}$$

where we have used (B.13) in the second step and (B.14)–(B.17) in the third step.

Finally, we show (B.12). Note that

$$Var_0\left(\log\frac{p_n(X,\boldsymbol{\alpha})}{\widetilde{p}_0(X,\boldsymbol{\alpha})}\right)\le E_0\log^2\left\{\frac{p_n(X,\boldsymbol{\alpha})}{\widetilde{p}_0(X,\boldsymbol{\alpha})}\right\}. \tag{B.18}$$

Combining (B.15)–(B.17) and (B.18), we further get that

$$Var_0\left(\log\frac{p_n(X,\boldsymbol{\alpha})}{\widetilde{p}_0(X,\boldsymbol{\alpha})}\right)\le E_0\left[\left\{\frac{p_n(X,\boldsymbol{\alpha})-\widetilde{p}_0(X,\boldsymbol{\alpha})}{\widetilde{p}_0(X,\boldsymbol{\alpha})}\right\}^2\right]+O(h^3)=O(h^2). \qquad \blacksquare$$

We have finished verifying Conditions (a)–(b) in Lemma 1. Recall that

$$\begin{aligned}\phi_n(\delta) &= \delta^{1-1/(2a)}\sum_{j=1}^M|\log h_j|^{0.5}h_j^{-0.5-0.25/a}\\ &\quad\times\left(1+\frac{1}{\sqrt{n}\delta^{1+1/(2a)}}\sum_{j=1}^M|\log h_j|^{0.5}h_j^{-0.5-0.25/a}\right)\end{aligned}$$

and $\mathbb{M}_n(\widehat{p},\widetilde{p}_0)\ge\mathbb{M}_n(\widetilde{p}_0,\widetilde{p}_0)+O_p(h)$. Applying Lemma 1, we have $d(\gamma\widehat{p},\gamma\widetilde{p}_0)=O_p(r_n^{-1})$ with r_n satisfying $r_n^2\phi_n(1/r_n)\le\sqrt{n}$ and $r_n^{-2}=O_p(h)$ for every $a>0$. Note that $r_n^2\phi_n(1/r_n)\le\sqrt{n}$ is equivalent to

$$r_n^2(r_n^{-1})^{1-1/(2a)}\sum_{j=1}^M|\log h_j|^{0.5}h_j^{-0.5-0.25/a}\lesssim\sqrt{n},$$

which implies that

$$r_n\lesssim\left(n^{0.5}\sum_{j=1}^M|\log h_j|^{-0.5}h_j^{0.5+0.25/a}\right)^{1/(1+1/(2a))}.$$

Set

$$r_n^{-1}=O_p(h^{0.5})+O_p\left(\sum_{j=1}^M\frac{|\log h_j|^{0.5/(1+1/(2a))}}{n^{0.5/(1+1/(2a))}h_j^{0.5}}\right).$$

With Condition 0, we get

$$d(\gamma\widehat{p},\gamma\widetilde{p}_0)=O_P(r_n^{-1})=O_p(h^{0.5})+O_p\left(\frac{|\log h|^{0.5/(1+1/(2a))}}{n^{0.5/(1+1/(2a))}h^{0.5}}\right).$$

Note that

$$O_p\left(\frac{|\log h|^{0.5/(1+1/(2a))}}{n^{0.5/(1+1/(2a))}h^{0.5}}\right)=O_p\left(\frac{h^{\frac{1}{8a^2+4a}}|\log h|^{\frac{a}{2a+1}}}{(nh)^{\frac{1}{8a^2+4a}}}\right)\cdot O_p\left(\frac{1}{n^{0.5-0.5/(2a)}h^{0.5}}\right).$$

With Condition 1, for any arbitrarily small $\vartheta > 0$, we can find a sufficiently large a such that

$$O_p\left(\frac{|\log h|^{0.5/(1+1/(2a))}}{n^{0.5/(1+1/(2a))}h^{0.5}}\right) = O_p(n^{-0.5+\vartheta}h^{-0.5})$$

and hence

$$d(\gamma\widehat{p}, \gamma\widetilde{p}_0) = O_p(h^{0.5}) + O_p(n^{-0.5+\vartheta}h^{-0.5}),$$

which completes the proof of this theorem. ∎

B.3. Proof of Theorem 5

In this section, we mainly establish the consistency of $\int_{\mathbb{R}} |\mathcal{N}_{h_j}\widehat{f}_j(x) - f_{0,j}(x)|dx$ as claimed in Theorem 5 by using the consistency result for $d(\gamma\widehat{p}, \gamma\widetilde{p}_0)$ in Theorem 4. We need the following lemma.

Lemma 6. *Assume Condition 4. For any $p(x, \boldsymbol{\alpha}) = \sum_{j=1}^M \alpha_j \mathcal{N}_{h_j} f_j(x) \in \mathcal{P}_n$, we have*

$$\int_{\mathbb{R}} |\mathcal{N}_{h_j} f_j(x) - f_{0,j}(x)|dx \lesssim d(\gamma p, \gamma\widetilde{p}_0).$$

Proof. With $\mathcal{O}_j, j = 1, \ldots, M$ and $\boldsymbol{\alpha}_{0,j}$ given in Condition 4, we have

$$\begin{aligned}
& \sum_{j=1}^M \int_{\mathbb{R}} \left\{\sqrt{p(x, \boldsymbol{\alpha}_{0,j})} - \sqrt{\widetilde{p}_0(x, \boldsymbol{\alpha}_{0,j})}\right\}^2 dx \\
\lesssim \quad & \sum_{j=1}^M \int\int_{\boldsymbol{\alpha}\in\mathcal{O}_j; x\in\mathbb{R}} \left(\sqrt{p(x, \boldsymbol{\alpha})} - \sqrt{\widetilde{p}_0(x, \boldsymbol{\alpha})}\right)^2 \gamma(\boldsymbol{\alpha}) dx d\boldsymbol{\alpha} \\
\leq \quad & d^2(\gamma p, \gamma\widetilde{p}_0),
\end{aligned}$$

which indicates that for every $j = 1, \ldots, M$,

$$\int_{\mathbb{R}} \left\{\sqrt{p(x, \boldsymbol{\alpha}_{0,j})} - \sqrt{\widetilde{p}_0(x, \boldsymbol{\alpha}_{0,j})}\right\}^2 dx \lesssim d^2(\gamma p, \gamma\widetilde{p}_0). \tag{B.19}$$

Next we show that $\int_{\mathbb{R}} |\mathcal{N}_{h_j} f_j(x) - f_{0,j}(x)|dx$ can be bounded by a linear combination of the left-hand side of (B.19). We need some notation. Let $A = (\boldsymbol{\alpha}_{0,1}, \ldots, \boldsymbol{\alpha}_{0,M})$ be an $M \times M$ invertible matrix and write

$$A^{-1} = \left(a_{j,k}\right)_{j=1,\ldots,M;\ k=1,\ldots,M}.$$

Then $\mathcal{N}_{h_j} f_j(x) = \sum_{k=1}^M a_{j,k} p(x, \boldsymbol{\alpha}_{0,k})$, $f_{0,j}(x) = \sum_{k=1}^M a_{j,k}\widetilde{p}_0(x, \boldsymbol{\alpha}_{0,k})$. Therefore,

$$\int_{\mathbb{R}} |\mathcal{N}_{h_j} f_j(x) - f_{0,j}(x)|dx \leq \sum_{j=1}^M |a_{j,k}| \int_{\mathbb{R}} |p(x, \boldsymbol{\alpha}_{0,k}) - \widetilde{p}_0(x, \boldsymbol{\alpha}_{0,k})|dx$$

$$
\begin{aligned}
&\le \sum_{j=1}^{M} |a_{j,k}| \sqrt{\int_{\mathbb{R}} \left\{ \sqrt{p(x,\boldsymbol{\alpha}_{0,j})} - \sqrt{\widetilde{p}_0(x,\boldsymbol{\alpha}_{0,j})} \right\}^2 dx} \\
&\quad \times \sqrt{\int_{\mathbb{R}} \left\{ \sqrt{p(x,\boldsymbol{\alpha}_{0,j})} + \sqrt{\widetilde{p}_0(x,\boldsymbol{\alpha}_{0,j})} \right\}^2 dx} \qquad \text{(B.20)} \\
&\le \sum_{j=1}^{M} |a_{j,k}| d(\gamma p, \gamma \widetilde{p}_0) \sqrt{2 \int_{\mathbb{R}} \{ p(x,\boldsymbol{\alpha}_{0,j}) + \widetilde{p}_0(x,\boldsymbol{\alpha}_{0,j}) \} dx} \lesssim d(\gamma p, \gamma \widetilde{p}_0), \qquad \text{(B.21)}
\end{aligned}
$$

where from (B.20) to (B.21), we use (B.19) and the fact that $(a+b)^2 \le 2(a^2+b^2)$; to derive the last "$\lesssim$", we have applied Lemma 3; specifically,

$$
\int_{\mathbb{R}} p(x,\boldsymbol{\alpha}_{0,j}) = \sum_{j=1}^{M} \alpha_{0,j} \int_{\mathbb{R}} \mathcal{N}_{h_j} f_j(x) dx \le \sum_{j=1}^{M} \alpha_{0,j} = 1,
$$

and likewise $\int_{\mathbb{R}} \widetilde{p}_0(x,\boldsymbol{\alpha}_{0,j}) \le 1$. ∎

Combining Theorem 4 and Lemma 6, we can immediately conclude the consistency of $\int_{\mathbb{R}} |\mathcal{N}_{h_j} \widehat{f}_j(x) - f_{0,j}(x)| dx$. That is, for any $\vartheta > 0$, we have

$$
\int_{\mathbb{R}} |\mathcal{N}_{h_j} \widehat{f}_j(x) - f_{0,j}(x)| dx = O_p(h^{0.5}) + O_p(n^{-0.5+\vartheta} h^{-0.5}), \qquad \text{(B.22)}
$$

which completes our proof of Theorem 5. ∎

B.4. Proof of Theorem 6

In this section, we prove Theorem 6, which establishes the L_1 consistency of $\widehat{f}_j(x)$, $j = 1, \ldots, M$. Recall that

$$
\widehat{f}_j(x) = \frac{\sum_{i=1}^{n} \widehat{w}_{i,j} K_{h_j}(x - X_i)}{\sum_{i=1}^{n} \widehat{w}_{i,j}} \qquad \text{(B.23)}
$$

with $\widehat{w}_{i,j} = \frac{\alpha_{i,j} \mathcal{N}_{h_j} \widehat{f}_j(X_i)}{\widehat{p}(X_i, \boldsymbol{\alpha}_i)}$ and $\widehat{p}(x,\boldsymbol{\alpha}) = \sum_{k=1}^{M} \alpha_k \mathcal{N}_{h_k} \widehat{f}_k(x)$. We investigate the asymptotic properties of the numerator and denominator of (B.23) separately, and then establish the consistency of $\widehat{f}_j(x)$.

Based on Condition 3, we can find a $c > 0$, such that $\inf_{x \in S_x, \boldsymbol{\alpha}} \widetilde{p}_0(x,\boldsymbol{\alpha}) > 2c$. Denote $I_1(x) = I_{1,1}(x) - I_{1,2}(x) + I_{1,3}(x)$, and $I_2 = \frac{1}{n} \sum_{i=1}^{n} \widehat{w}_{i,j}$, where

$$
\begin{aligned}
I_{1,1}(x) &= \frac{1}{n} \sum_{i=1}^{n} K_{h_j}(x - X_i) \frac{\alpha_{i,j} \mathcal{N}_{h_j} \widehat{f}_j(X_i)}{\widehat{p}(X_i, \boldsymbol{\alpha}_i)} I\{\widehat{p}(X_i, \alpha_i) \le c\}; \qquad \text{(B.24)} \\
I_{1,2}(x) &= \frac{1}{n} \sum_{i=1}^{n} K_{h_j}(x - X_i) \frac{\alpha_{i,j} \mathcal{N}_{h_j} \widehat{f}_j(X_i)}{c} I\{\widehat{p}(X_i, \boldsymbol{\alpha}_i) \le c\}; \qquad \text{(B.25)}
\end{aligned}
$$

$$
\begin{aligned}
I_{1,3}(x) &= \frac{1}{n}\sum_{i=1}^{n} K_{h_j}(x-X_i)\frac{\alpha_{i,j}\mathcal{N}_{h_j}\widehat{f}_j(X_i)}{\widehat{p}(X_i,\boldsymbol{\alpha}_i)} I\{\widehat{p}(X_i,\boldsymbol{\alpha}_i)>c\} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} K_{h_j}(x-X_i)\frac{\alpha_{i,j}\mathcal{N}_{h_j}\widehat{f}_j(X_i)}{c} I\{\widehat{p}(X_i,\boldsymbol{\alpha}_i)\le c\} \\
&= \frac{1}{n}\sum_{i=1}^{n} K_{h_j}(x-X_i)\frac{\alpha_{i,j}\mathcal{N}_{h_j}\widehat{f}_j(X_i)}{\widehat{p}(X_i,\boldsymbol{\alpha}_i)I\{\widehat{p}(X_i,\boldsymbol{\alpha}_i)>c\}+cI\{\widehat{p}(X_i,\boldsymbol{\alpha}_i)\le c\}};
\end{aligned}
\tag{B.26}
$$

With straightforward manipulation, we note that $\widehat{f}_j(x)$ given in (B.23) can be decomposed as follows:

$$
\widehat{f}_j(x) = \frac{I_1(x)}{I_2} = \frac{I_{1,1}(x) - I_{1,2}(x) + I_{1,3}(x)}{I_2}. \tag{B.27}
$$

Next we study the asymptotic behaviour of $I_{1,1}(x), I_{1,2}(x)$, and $I_{1,3}(x)$, separately. Studying I_2 is similar but easier. We first consider $I_{1,3}(x)$.

B.4.1. Asymptotic property of $I_{1,3}(x)$

We need some preparation, and we first define some notation. Let

$$
g_{j,0}(y,\boldsymbol{\alpha}) = \frac{\alpha_j f_{0,j}(y)}{\widetilde{p}_0(y,\boldsymbol{\alpha})}, \quad \widehat{g}_{j,c}(y,\boldsymbol{\alpha}) = \frac{\alpha_j \mathcal{N}_{h_j}\widehat{f}_j(y)}{\widehat{p}(y,\boldsymbol{\alpha})I\{\widehat{p}(y,\boldsymbol{\alpha})>c\}+cI\{\widehat{p}(y,\boldsymbol{\alpha})\le c\}}. \tag{B.28}
$$

Then we can write

$$
I_{1,3}(x) = I_{1,3,1}(x) + I_{1,3,2}(x), \tag{B.29}
$$

where

$$
\begin{aligned}
I_{1,3,1}(x) &= \mathbb{P}_n\left[K_{h_j}(X-x)\cdot\{\widehat{g}_{j,c}(X,\boldsymbol{\alpha}) - g_{j,0}(X,\boldsymbol{\alpha})\}\right], \\
I_{1,3,2}(x) &= \mathbb{P}_n\left\{K_{h_j}(X-x)\cdot g_{j,0}(X,\boldsymbol{\alpha})\right\},
\end{aligned}
$$

with "$\mathbb{P}_n$" operated on $(X,\boldsymbol{\alpha})$. Next we define two classes of functions:

- $\mathcal{F}_{c,j} = \left\{\frac{\alpha_j f_j(y)}{p(y,\boldsymbol{\alpha})I\{p(y,\boldsymbol{\alpha})>c\}+cI\{p(y,\boldsymbol{\alpha})\le c\}} : p(y,\boldsymbol{\alpha}) = \sum_{k=1}^{M}\alpha_k f_k(y), f_k\in\mathcal{P}_{n,k}\right\}$, where for $k=1,\dots,M$
$$
\mathcal{P}_{n,k} = \left\{\mathcal{N}_{h_k} f : f = \frac{\sum_{i=1}^n w_{i,k}K_{h_k}(y-X_i)}{\sum_{i=1}^n w_{i,k}}; 0\le w_{i,k}\le 1\right\};
$$
- $\widetilde{\mathcal{F}}_{c,j} = \{|g_{j,c} - g_{j,0}| : g_{j,c}\in\mathcal{F}_{c,j}\}$.

Clearly, $\widehat{g}_{j,c}(y,\boldsymbol{\alpha})\in\mathcal{F}_{c,j}$, $|\widehat{g}_{j,c}(y,\boldsymbol{\alpha}) - g_{j,0}(y,\boldsymbol{\alpha})|\in\widetilde{\mathcal{F}}_{c,j}$.

The following lemma calculates the bracketing numbers of these function classes, and will be helpful to establish the asymptotic properties for $I_{1,3}(x)$.

Lemma 7. *Let P_0 denote the probability measure under the true joint distribution $\gamma(\boldsymbol{\alpha})\tilde{p}_0(x,\boldsymbol{\alpha})$ of $(X,\boldsymbol{\alpha})$. For every $\epsilon > 0$,*

(a) for an arbitrary positive integer a, $\log N_{[]}\Big(\epsilon, \mathcal{F}_{c,j}, L_2(P_0)\Big) \lesssim \sum_{k=1}^{M} \frac{|\log h_k|}{\epsilon^{1/a} h_k^{1+1/a}}$;

(b) for an arbitrary positive integer a, $\log N_{[]}\Big(\epsilon, \widetilde{\mathcal{F}}_{c,j}, L_2(P_0)\Big) \lesssim \sum_{k=1}^{M} \frac{|\log h_k|}{\epsilon^{1/a} h_k^{1+1/a}}$.

In the above, "$\lesssim$" are up to universal constants depending on the upper bound of $K(\cdot)$, a, c, and M.

Proof. For part (a), using the same strategy as in the proof of (B.8) in Lemma 4, we can verify that for $k = 1,\ldots,M$,

$$\log N_{[]}\Big(\epsilon, \mathcal{P}_{n,k}, L_2(P_0)\Big) \lesssim \frac{|\log h_k|}{\epsilon^{1/a} h_k^{1+1/a}}.$$

For notational convenience, we write $N_k = N_{[]}\Big(\epsilon, \mathcal{P}_{n,k}, L_2(P_0)\Big)$. Then for every $k = 1,\ldots,M$, there exists a set of ϵ-brackets $\mathcal{B}_k = \{[u_{i_k,k}, v_{i_k,k}] : i_k = 1,\ldots,N_k\}$ that covers $\mathcal{P}_{n,k}$. We consider

$$\widetilde{\mathcal{B}}_j = \left\{ [g_L(y,\boldsymbol{\alpha}), g_U(y,\boldsymbol{\alpha})] : \begin{array}{l} g_L(y,\boldsymbol{\alpha}) = \frac{\alpha_j u_{i_j,j}}{p_U}; \quad g_U(y,\boldsymbol{\alpha}) = \frac{\alpha_j v_{i_j,j}}{p_L}; \\ p_U = \widetilde{p}_U I\{\widetilde{p}_U > c\} + cI\{\widetilde{p}_U \le c\}; \\ \widetilde{p}_U = \sum_{k=1}^{M} \alpha_k v_{i_k,k}; \\ p_L = \widetilde{p}_L I\{\widetilde{p}_L > c\} + cI\{\widetilde{p}_L \le c\}; \\ \widetilde{p}_L = \sum_{k=1}^{M} \alpha_k u_{i_k,k}; \\ \text{for every } i_k = 1,\ldots,N_k \quad \text{and} \quad k = 1,\ldots,M \end{array} \right\},$$

which contains $\Pi_{k=1}^{M} N_k$ pairs of functions.

We now verify that $\widetilde{\mathcal{B}}_j$ covers $\mathcal{F}_{c,j}$. Recall that for every $k = 1,\ldots,M$, $\mathcal{B}_k$ covers $\mathcal{P}_{n,k}$. Then for every

$$g_{j,c}(y,\boldsymbol{\alpha}) = \frac{\alpha_j f_j(y)}{p(y,\boldsymbol{\alpha}) I\{p(y,\boldsymbol{\alpha}) > c\} + cI\{p(y,\boldsymbol{\alpha}) \le c\}} \in \mathcal{F}_{c,j},$$

there exist $(i_1,\ldots,i_M)$, where $1 \le i_k \le N_k$ for every $k = 1,\ldots,M$, such that $u_{i_k,k} \le f_k \le v_{i_k,k}$, which implies that

(i) $\alpha_j u_{i_j,j} \le \alpha_j f_j \le \alpha_j v_{i_j,j}$; and further
(ii) $\widetilde{p}_L \le p \le \widetilde{p}_U$, where $\widetilde{p}_L = \sum_{k=1}^{M} \alpha_k u_{i_k,k}$ and $\widetilde{p}_U = \sum_{k=1}^{M} \alpha_k v_{i_k,k}$.

With the fact that for any two functions g_1 and g_2, $g_1 \le g_2$ implies $g_1 I\{g_1 > c\} + cI\{g_1 \le c\} \le g_2 I\{g_2 > c\} + cI\{g_2 \le c\}$, (i) and (ii) above lead to

(iii) $p_L \le pI\{p > c\} + cI\{p \le c\} \le p_U$, where $p_L = \widetilde{p}_L I\{\widetilde{p}_L > c\} + cI\{\widetilde{p}_L > c\}$ and $p_U = \widetilde{p}_U I\{\widetilde{p}_U > c\} + cI\{\widetilde{p}_U \le c\}$.

(i) and (iii) imply that $g_L \le g_{j,c} \le g_U$, where $g_L = \frac{\alpha_j u_{i_j,j}}{p_U}$, $g_U = \frac{\alpha_j v_{i_j,j}}{p_L}$. Hence $[g_L, g_U]$ is a bracket in $\widetilde{\mathcal{B}}_j$ and we have verified that $\widetilde{\mathcal{B}}_j$ covers $\mathcal{F}_{c,j}$.

We need to calculate the sizes of the brackets in $\widetilde{\mathcal{B}}_j$ under $L_2(P_0)$. To this end, we consider an arbitrary $[g_L, g_U] \in \widetilde{\mathcal{B}}_j$. Noting that $|p_U - p_L| \leq |\widetilde{p}_U - \widetilde{p}_L|$, $0 \leq \alpha_j \leq 1$, $0 \leq \alpha_j u_{i_j,j} \leq p_L$, and $p_U \geq p_L \geq c > 0$, we have

$$\begin{aligned} |g_U - g_L| &\leq \frac{\alpha_j}{p_L}\Big|v_{i_j,j} - u_{i_j,j}\Big| + \frac{\alpha_j u_{i_j,j}}{p_U p_L}\Big|p_U - p_L\Big| \leq \frac{|v_{i_j,j} - u_{i_j,j}|}{c} + \frac{|p_U - p_L|}{c} \\ &\leq \frac{|v_{i_j,j} - u_{i_j,j}|}{c} + \frac{|\widetilde{p}_U - \widetilde{p}_L|}{c} \leq \frac{1}{c}\sum_{k=1}^{M} |v_{i_k,k} - u_{i_k,k}|, \end{aligned}$$

which immediately leads to

$$\begin{aligned} &\int_{S_\gamma}\int_{\mathbb{R}} |g_U(x,\boldsymbol{\alpha}) - g_L(x,\boldsymbol{\alpha})|^2 \gamma(\boldsymbol{\alpha})\widetilde{p}_0(x,\boldsymbol{\alpha})dxd\boldsymbol{\alpha} \\ \lesssim\ & \sum_{k=1}^{M}\int_{S_\gamma}\int_{\mathbb{R}} |u_{i_k,k} - v_{i_k,k}|^2 \gamma(\boldsymbol{\alpha})\widetilde{p}_0(x,\boldsymbol{\alpha})dxd\boldsymbol{\alpha} \lesssim \epsilon^2, \end{aligned}$$

where the last "$\lesssim$" is because for every $k = 1, \ldots, M$, $[u_{i_k,k}, v_{i_k,k}]$ is an ϵ-bracket in $\mathcal{B}_k$ under $L_2(P_0)$. This together with the facts that $\widetilde{\mathcal{B}}_j$ covers $\mathcal{F}_{c,j}$ and $\widetilde{\mathcal{B}}_j$ contains $\Pi_{k=1}^{M} N_k$ brackets completes our proof of part (a) in this Lemma.

For part (b), let $\mathcal{F}_{c,j,0} = \{g_{j,c} - g_{j,0} : g_{j,c} \in \mathcal{F}_{c,j}\}$. It is straightforward to check that

$$\log N_{[]}\Big(\epsilon, \mathcal{F}_{c,j,0}, L_2(P_0)\Big) \lesssim \sum_{k=1}^{M} \frac{|\log h_k|}{\epsilon^{1/a} h_k^{1+1/a}}. \tag{B.30}$$

On the other hand, let $|f|$ be an arbitrary function in $\widetilde{\mathcal{F}}_{c,j}$ with $f \in \mathcal{F}_{c,j,0}$. Let $[g_L, g_U]$ be the ϵ-bracket in $\mathcal{F}_{c,j,0}$ such that $g_L \leq f \leq g_U$. By noting that for any y and $\boldsymbol{\alpha}$, we must have

$$g_L^+ + g_U^- \leq |f| \leq g_U^+ + g_L^- \tag{B.31}$$

we get

$$\left|g_U^+ + g_L^- - g_L^+ - g_U^-\right| \leq |g_L^- - g_U^-| + |g_U^+ - g_L^+| \leq 2|g_U - g_L|. \tag{B.32}$$

(B.31) and (B.32) imply that every ϵ-bracket under $L_2(P_0)$ in $\mathcal{F}_{c,j,0}$ leads to a 2ϵ-bracket under $L_2(P_0)$ in $\widetilde{\mathcal{F}}_{c,j}$. This together with (B.30) completes our proof of part (b) in this lemma. ∎

With the lemma above, we study the asymptotic properties for $I_{1,3}$ given in (B.29). We will consider $I_{1,3,1}(x)$ and $I_{1,3,2}(x)$ separately. First, we show that

$$\int_{\mathbb{R}} |I_{1,3,1}(x)|dx = O_p\left(\sum_{k=1}^{M} \frac{\sqrt{|\log h_k|}}{n^{0.5} h_k^{0.5+0.5/a}}\right) + d(\gamma\widehat{p}, \gamma\widetilde{p}_0). \tag{B.33}$$

To this end, note that

$$\begin{aligned}\int_{\mathbb{R}} |I_{1,3,1}(x)|dx &\le \mathbb{P}_n\left\{\int_{\mathbb{R}} K_{h_j}(X-x)dx \cdot |\widehat{g}_{j,c}(X,\boldsymbol{\alpha}) - g_{j,0}(X,\boldsymbol{\alpha})|\right\} \\ &= \mathbb{P}_n\left\{|\widehat{g}_{j,c}(X,\boldsymbol{\alpha}) - g_{j,0}(X,\boldsymbol{\alpha})|\right\}, \qquad \text{(B.34)}\end{aligned}$$

where $\mathbb{P}_n$ is operated on $(X,\boldsymbol{\alpha})$. Note that $|\widehat{g}_{j,c}(y,\boldsymbol{\alpha}) - g_{j,0}(y,\boldsymbol{\alpha})| \in \widetilde{\mathcal{F}}_{c,j}$, and for any function $f \in \widetilde{\mathcal{F}}_{c,j}$, we have $\mathbb{P}\{f^2(X,\boldsymbol{\alpha})\} \le 4$, $\sup_{y,\boldsymbol{\alpha}} |f(y,\boldsymbol{\alpha})| \le 2$, which incorporated with Lemma 3.4.2 in VM leads to

$$E_0\|\mathbb{G}\|_{\widetilde{\mathcal{F}}_{c,j}} \lesssim \widetilde{J}_{[]}\left(2,\widetilde{\mathcal{F}}_{c,j},L_2(P_0)\right)\left\{1+\frac{\widetilde{J}_{[]}\left(2,\widetilde{\mathcal{F}}_{c,j},L_2(P_0)\right)}{\sqrt{n}\cdot 4}\cdot 2\right\}. \qquad \text{(B.35)}$$

By part (b) of Lemma 7, we have

$$\widetilde{J}_{[]}\left(2,\widetilde{\mathcal{F}}_{c,j},L_2(P_0)\right) \lesssim \int_0^2 \sqrt{1+\sum_{k=1}^M \frac{|\log h_k|}{\epsilon^{1/a} h_k^{1+1/a}}}d\epsilon \lesssim \sum_{k=1}^M \frac{\sqrt{|\log h_k|}}{h_k^{0.5+0.5/a}},$$

which together with (B.35) and Condition 1 leads to

$$E_0\|\mathbb{G}\|_{\widetilde{\mathcal{F}}_{c,j}} \lesssim \sum_{k=1}^M \frac{\sqrt{|\log h_k|}}{h_k^{0.5+0.5/a}}.$$

This together with Chebyshev's inequality implies

$$\|\mathbb{G}\|_{\widetilde{\mathcal{F}}_{c,j}} = O_p\left(\sum_{k=1}^M \frac{\sqrt{|\log h_k|}}{h_k^{0.5+0.5/a}}\right),$$

and hence

$$\begin{aligned}&\mathbb{P}_n\left\{|\widehat{g}_{j,c}(X,\boldsymbol{\alpha}) - g_{j,0}(X,\boldsymbol{\alpha})|\right\} - \mathbb{P}\left\{|\widehat{g}_{j,c}(X,\boldsymbol{\alpha}) - g_{j,0}(X,\boldsymbol{\alpha})|\right\} \\ &\qquad = O_p\left(\sum_{k=1}^M \frac{\sqrt{|\log h_k|}}{n^{0.5} h_k^{0.5+0.5/a}}\right), \qquad \text{(B.36)}\end{aligned}$$

which together with the convergence result of $\mathbb{P}\left\{|\widehat{g}_{j,c}(X,\boldsymbol{\alpha}) - g_{j,0}(X,\boldsymbol{\alpha})|\right\}$ in the following lemma implies the result in (B.33).

Lemma 8. *Recall $g_{j,0}(y,\boldsymbol{\alpha})$ and $\widehat{g}_{j,c}(y,\boldsymbol{\alpha})$ defined by (B.28). We have*

$$\mathbb{P}\left\{|\widehat{g}_{j,c}(y,\boldsymbol{\alpha}) - g_{j,0}(y,\boldsymbol{\alpha})|\right\} \lesssim d(\gamma\widehat{p},\gamma\widetilde{p}_0). \qquad \text{(B.37)}$$

Proof. With straightforward manipulation, we can write

$$\widehat{g}_{j,c}(y,\boldsymbol{\alpha}) = \frac{\alpha_j \mathcal{N}_{h_j}\widehat{f}_j(y)}{\widehat{p}(y,\boldsymbol{\alpha})} I\{\widehat{p}(y,\boldsymbol{\alpha}) > c\} + \frac{\alpha_j \mathcal{N}_{h_j}\widehat{f}_j(y)}{c} I\{\widehat{p}(y,\boldsymbol{\alpha}) \le c\}.$$

Therefore

$$\mathbb{P}\left\{\left|\widehat{g}_{j,c}(X,\boldsymbol{\alpha})-g_{j,0}(X,\boldsymbol{\alpha})\right|\right\} \le I_{1,3,1,1}+I_{1,3,1,2}+I_{1,3,1,3}, \tag{B.38}$$

where

$$\begin{aligned}
I_{1,3,1,1} &= \int_{\mathbb{R}}\int_{S_\gamma}\left|\frac{\alpha_j\mathcal{N}_{h_j}\widehat{f}_j(y)}{\widehat{p}(y,\boldsymbol{\alpha})}-\frac{\alpha_j f_{0,j}(y)}{\widetilde{p}_0(y,\boldsymbol{\alpha})}\right| I\{\widehat{p}(y,\boldsymbol{\alpha})>c\}\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha})d\boldsymbol{\alpha}dy;\\
I_{1,3,1,2} &= \int_{\mathbb{R}}\int_{S_\gamma}\frac{\alpha_j\mathcal{N}_{h_j}\widehat{f}_j(y)}{c} I\{\widehat{p}(y,\boldsymbol{\alpha})\le c\}\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha})d\boldsymbol{\alpha}dy;\\
I_{1,3,1,3} &= \int_{\mathbb{R}}\int_{S_\gamma}\frac{\alpha_j f_{0,j}(y)}{\widetilde{p}_0(y,\boldsymbol{\alpha})} I\{\widehat{p}(y,\boldsymbol{\alpha})\le c\}\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha})d\boldsymbol{\alpha}dy.
\end{aligned}$$

We first consider $I_{1,3,1,1}$:

$$\begin{aligned}
&I_{1,3,1,1}\\
&\le \int_{\mathbb{R}}\int_{S_\gamma}\frac{\alpha_j}{\widehat{p}(y,\boldsymbol{\alpha})}\left|\mathcal{N}_{h_j}\widehat{f}_j(y)-f_{0,j}(y)\right| I\{\widehat{p}(y,\boldsymbol{\alpha})>c\}\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha})d\boldsymbol{\alpha}dy\\
&\quad+\int_{\mathbb{R}}\int_{S_\gamma}\frac{\alpha_j f_{0,j}(y)}{\widehat{p}(y,\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha})}\left|\widehat{p}(y,\boldsymbol{\alpha})-\widetilde{p}_0(y,\boldsymbol{\alpha})\right| I\{\widehat{p}(y,\boldsymbol{\alpha})>c\}\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha})d\boldsymbol{\alpha}dy\\
&\lesssim \int_{\mathbb{R}}\left|\mathcal{N}_{h_j}\widehat{f}_j(y)-f_{0,j}(y)\right|dy+\int_{\mathbb{R}}\int_{S_\gamma}\left|\widehat{p}(y,\boldsymbol{\alpha})-\widetilde{p}_0(y,\boldsymbol{\alpha})\right|\gamma(\boldsymbol{\alpha})d\boldsymbol{\alpha}dy\\
&\lesssim \int_{\mathbb{R}}\left|\mathcal{N}_{h_j}\widehat{f}_j(y)-f_{0,j}(y)\right|dy+\sum_{k=1}^{M}\int_{\mathbb{R}}\left|\mathcal{N}_{h_k}\widehat{f}_k(y)-f_{0,k}(y)\right|dy \lesssim d(\gamma\widehat{p},\gamma\widetilde{p}_0),
\end{aligned} \tag{B.39}$$

where for the last "$\lesssim$", we have applied Lemma 6.

Next, we consider $I_{1,3,1,2}$ and $I_{1,3,1,3}$ together. It can be seen that

$$I_{1,3,1,2}\lesssim I_{1,3,1,4} \quad \text{and} \quad I_{1,3,1,3}\lesssim I_{1,3,1,4}, \tag{B.40}$$

where $I_{1,3,1,4} = \int_{\mathbb{R}}\int_{S_\gamma} I\{\widehat{p}(y,\boldsymbol{\alpha}) \le c\}\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha})d\boldsymbol{\alpha}dy$. Recalling that $\inf_{y\in S_x,\boldsymbol{\alpha}}\widetilde{p}_0(y,\boldsymbol{\alpha})>2c$, we have

$$\begin{aligned}
&I_{1,3,1,4}\\
&\le \int_{\mathbb{R}}\int_{S_\gamma} I\{|\widetilde{p}_0(y,\boldsymbol{\alpha})-\widehat{p}(y,\boldsymbol{\alpha})|>c\}\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha})d\boldsymbol{\alpha}dy\\
&\le \int_{\mathbb{R}}\int_{S_\gamma} I\{|\widetilde{p}_0(x,\boldsymbol{\alpha})-\widehat{p}(y,\boldsymbol{\alpha})|>c\}\frac{|\widetilde{p}_0(y,\boldsymbol{\alpha})-\widehat{p}(y,\boldsymbol{\alpha})|}{c}\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha})d\boldsymbol{\alpha}dy\\
&\lesssim \int_{\mathbb{R}}\int_{S_\gamma}|\widetilde{p}_0(y,\boldsymbol{\alpha})-\widehat{p}(y,\boldsymbol{\alpha})|\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha})d\boldsymbol{\alpha}dy\lesssim d(\gamma\widehat{p},\gamma\widetilde{p}_0).
\end{aligned} \tag{B.41}$$

Combining (B.38), (B.39), (B.40), and (B.41), we conclude (B.37). ∎

Second, we verify that

$$\begin{aligned} & \sup_{x\in S_x^*}\left|I_{1,3,2}(x) - \int_{\mathbb{R}} K_{h_j}(y-x) f_{0,j}(y)dy \int_{S_\gamma} \alpha_j \gamma(\boldsymbol{\alpha}) d\boldsymbol{\alpha}\right| \\ = \;& O_p\left(\sum_{k=1}^{M} \frac{1}{n^{0.5} h_k^{0.5+0.5/a}}\right) \end{aligned} \tag{B.42}$$

and

$$\sup_{x\notin S_x^*}\left|I_{1,3,2}(x) - \int_{\mathbb{R}} K_{h_j}(y-x) f_{0,j}(y)dy \int_{S_\gamma} \alpha_j \gamma(\boldsymbol{\alpha}) d\boldsymbol{\alpha}\right| = 0, \tag{B.43}$$

where $S_x^* = [c_1 - \Delta, c_2 + \Delta]$ with $\Delta > 0$ being an arbitrarily small constant.

Recall the definition of $I_{1,3,2}(x)$ in (B.29):

$$I_{1,3,2}(x) = \mathbb{P}_n\left\{K_{h_j}(X-x)\cdot g_{j,0}(X,\boldsymbol{\alpha})\right\} \tag{B.44}$$

where $g_{j,0}(y,\boldsymbol{\alpha}) = \frac{\alpha_j f_{0,j}(y)}{\widetilde{p}_0(y,\boldsymbol{\alpha})}$. (B.43) follows directly from the definition.

We now consider the proof of (B.42). For every n, we consider the class of functions of $(y,\boldsymbol{\alpha})$:

$$\mathcal{F}_{K,j} = \{K_{h_j}(y-x)\cdot g_{j,0}(y,\boldsymbol{\alpha}) : \text{indexed by } x \in S_x^*\}.$$

Then, we can readily check that for every $x \in S_x^*$,

$$\mathbb{P}\left[\{K_{h_j}(X-x)\cdot g_{j,0}(X,\boldsymbol{\alpha})\}^2\right] \lesssim 1/h_j \text{ and } \sup_{y,\boldsymbol{\alpha}} |K_{h_j}(y-x)\cdot g_{j,0}(y,\boldsymbol{\alpha})| \lesssim 1/h_j,$$

which incorporated with Lemma 3.4.2 in VM leads to

$$\begin{aligned} & E_{P_0}\|\mathbb{G}_n\|_{\mathbb{P}\mathcal{F}_{K,j}} \\ \lesssim \;& \widetilde{J}_{[]}\left(1/\sqrt{h_j}, \mathcal{F}_{K,j}, L_2(P_0)\right)\left\{1 + \frac{\widetilde{J}_{[]}\left(1/\sqrt{h_j}, \mathcal{F}_{K,j}, L_2(P_0)\right)}{\sqrt{n}/h_j}\frac{1}{h_j}\right\}. \end{aligned} \tag{B.45}$$

Applying Theorem 2.7.11 of VM, we can check that for every $\epsilon > 0$,

$$N_{[]}\left(\epsilon, \mathcal{F}_{K,j}, L_2(P_0)\right) \lesssim \frac{1}{h_j^2 \epsilon}.$$

Hence

$$\widetilde{J}_{[]}\left(1/\sqrt{h_j}, \mathcal{F}_{K,j}, L_2(P_0)\right) \lesssim h_j^{-b} h_j^{-0.5(1-b/2)},$$

for any arbitrary $0 < b < 1$. Setting $b = 2/(3a)$, we have

$$\widetilde{J}_{[]}\left(1/\sqrt{h_j}, \mathcal{F}_{K,j}, L_2(P_0)\right) \lesssim h_j^{-0.5-0.5/a},$$

which together with (B.45) leads to $E_0\|\mathbb{G}_n\|_{\mathcal{F}_{K,j}} \lesssim \frac{1}{h_j^{0.5+0.5/a}}$. This together with Chebyshev's inequality immediately implies

$$\begin{aligned}&\sup_{x\in S_x^*}\left|\mathbb{P}_n\left\{K_{h_j}(X-x)\cdot g_{j,0}(X,\boldsymbol{\alpha})\right\}-\mathbb{P}\left\{K_{h_j}(X-x)\cdot g_{j,0}(X,\boldsymbol{\alpha})\right\}\right|\\ =\ & O_p\left(\frac{1}{n^{0.5}h_j^{0.5+0.5/a}}\right)=O_p\left(\sum_{k=1}^{M}\frac{1}{n^{0.5}h_k^{0.5+0.5/a}}\right).\end{aligned} \tag{B.46}$$

Furthermore, it can be checked that

$$\mathbb{P}\left\{K_{h_j}(X-x)\cdot g_{j,0}(X,\boldsymbol{\alpha})\right\}=\int_{\mathbb{R}}K_{h_j}(y-x)f_{0,j}(y)dy\int_{S_\gamma}\alpha_j\gamma(\boldsymbol{\alpha})d\boldsymbol{\alpha},$$

which together with (B.44) and (B.46) leads to (B.42).

Combining (B.33), (B.42), and (B.43) with (B.29), we conclude that

$$\begin{aligned}\int_{\mathbb{R}}&\left|I_{1,3}(x)-\int_{\mathbb{R}}K_{h_j}(y-x)f_{0,j}(y)dy\int_{S_\gamma}\alpha_j\gamma(\boldsymbol{\alpha})d\boldsymbol{\alpha}\right|dx\\ &=O_p\left(\sum_{k=1}^{M}\frac{\sqrt{|\log h_k|}}{n^{0.5}h_k^{0.5+0.5/a}}\right)+d(\gamma\widehat{p},\gamma\widetilde{p}_0).\end{aligned} \tag{B.47}$$

B.4.2. Asymptotic properties of $I_{1,1}(x)$ and $I_{1,2}(x)$

We proceed to consider the consistency of $I_{1,1}(x)$ and $I_{1,2}(x)$. Note that they are respectively defined in (B.24) and (B.25). Recall that

$$I_{1,1}(x)=\frac{1}{n}\sum_{i=1}^{n}K_{h_j}(x-X_i)\frac{\alpha_{i,j}\mathcal{N}_{h_j}\widehat{f}_j(X_i)}{\widehat{p}(X_i,\boldsymbol{\alpha}_i)}I\{\widehat{p}(X_i,\alpha_i)\le c\}\le I_{1,4}(x), \tag{B.48}$$

$$I_{1,2}(x)=\frac{1}{n}\sum_{i=1}^{n}K_{h_j}(x-X_i)\frac{\alpha_{i,j}\mathcal{N}_{h_j}\widehat{f}_j(X_i)}{c}I\{\widehat{p}(X_i,\boldsymbol{\alpha}_i)\le c\}\le I_{1,4}(x), \tag{B.49}$$

where

$$0\le I_{1,4}(x)=\frac{1}{n}\sum_{i=1}^{n}K_{h_j}(x-X_i)I\{\widehat{p}(X_i,\alpha_i)\le c\}.$$

Recalling that $\inf_{x\in S_x,\boldsymbol{\alpha}}\widetilde{p}_0(x,\boldsymbol{\alpha})>2c$, we have

$$\begin{aligned}0&\le\int_{\mathbb{R}}I_{1,4}(x)dx=\frac{1}{n}\sum_{i=1}^{n}I\{\widehat{p}(X_i,\alpha_i)\le c\}\\ &\le\frac{1}{n}\sum_{i=1}^{n}I\{\widetilde{p}_0(X_i,\boldsymbol{\alpha})-\widehat{p}(X_i,\boldsymbol{\alpha})>c\}\\ &\le\frac{1}{nc}\sum_{i=1}^{n}I\{\widetilde{p}_0(X_i,\boldsymbol{\alpha})-\widehat{p}(X_i,\boldsymbol{\alpha})>c\}\left\{\widetilde{p}_0(X_i,\boldsymbol{\alpha})-\widehat{p}(X_i,\boldsymbol{\alpha})\right\}\end{aligned}$$

$$\lesssim \frac{1}{n}\sum_{i=1}^{n} I\{\widetilde{p}_0(X_i,\boldsymbol{\alpha}) - \widehat{p}(X_i,\boldsymbol{\alpha}) > 0\}\{\widetilde{p}_0(X_i,\boldsymbol{\alpha}) - \widehat{p}(X_i,\boldsymbol{\alpha})\}. \tag{B.50}$$

Let $\widehat{g}_p(y,\boldsymbol{\alpha}) = \widetilde{p}_0(y,\boldsymbol{\alpha}) - \widehat{p}(y,\boldsymbol{\alpha})$. Then

$$0 \le I_{1,4}(x) \le \mathbb{P}_n\left[\widehat{g}_p(X,\boldsymbol{\alpha}) I\{\widehat{g}_p(X,\boldsymbol{\alpha}) > 0\}\right]. \tag{B.51}$$

Define

$$\mathcal{F}_{p,I} = \Big\{ \{\widetilde{p}_0(y,\boldsymbol{\alpha}) - p(y,\boldsymbol{\alpha})\} I\{\widetilde{p}_0(y,\boldsymbol{\alpha}) - p(y,\boldsymbol{\alpha}) > 0\} : p \in \mathcal{P}_n \Big\},$$

where we refer to (B.3) for the definition of $\mathcal{P}_n$. Clearly

$$\widehat{g}_p(y,\boldsymbol{\alpha}) I\{\widehat{g}_p(y,\boldsymbol{\alpha}) > 0\} \in \mathcal{F}_{p,I}.$$

In the lemma below, we establish the ϵ-bracketing number of $\mathcal{F}_{p,I}$ under $L_2(P_0)$.

Lemma 9. *For arbitrary $\epsilon > 0$ and a positive integer a, we have*

$$\log N_{[]}\Big(\epsilon, \mathcal{F}_{p,I}, L_2(P_0)\Big) \lesssim \sum_{k=1}^{M} \frac{|\log h_k|}{\epsilon^{1/a} h_k^{1+1/a}}.$$

Proof. Using exactly the same procedure as in Lemma 4, we have

$$\log N_{[]}\Big(\epsilon, \mathcal{P}_n, L_2(P_0)\Big) \lesssim \sum_{k=1}^{M} \frac{|\log h_k|}{\epsilon^{1/a} h_k^{1+1/a}}.$$

Let

$$\mathcal{B}_{p,I} = \Big\{[g_{L,i}, g_{U,i}], i = 1,\dots, N_{[]}\Big(\epsilon, \mathcal{P}_{n,0}, L_2(P_0)\Big)\Big\}$$

be the set of ϵ-brackets for $\mathcal{P}_n$. We consider

$$\widetilde{\mathcal{B}}_{p,I} = \Big\{[\widetilde{g}_{L,i}, \widetilde{g}_{U,i}], i = 1,\dots, N_{[]}\Big(\epsilon, \mathcal{P}_{n,0}, L_2(P_0)\Big)\Big\},$$

with

$$\widetilde{g}_{L,i} = (\widetilde{p}_0 - g_{U,i}) I\{\widetilde{p}_0 - g_{U,i} > 0\} \text{ and } \widetilde{g}_{U,i} = (\widetilde{p}_0 - g_{L,i}) I\{g_{L,i} > 0\}.$$

For any arbitrary functions g_1, g_2, if $g_1 \le g_2$, then $g_1 I\{g_1 > 0\} \le g_2 I\{g_2 > 0\}$. Hence we conclude that $\widetilde{\mathcal{B}}_{p,I}$ covers $\mathcal{F}_{p,I}$. Furthermore, it is straightforward to check that $0 \le \widetilde{g}_{U,i} - \widetilde{g}_{L,i} \le (\widetilde{p}_0 - g_{L,i}) - (\widetilde{p}_0 - g_{U,i}) = g_{U,i} - g_{L,i}$. Therefore, we have

$$\log N_{[]}\Big(\epsilon, \mathcal{F}_{p,I}, L_2(P_0)\Big) \le \log N_{[]}\Big(\epsilon, \mathcal{P}_n, L_2(P_0)\Big) \lesssim \sum_{k=1}^{M} \frac{|\log h_k|}{\epsilon^{1/a} h_k^{1+1/a}},$$

which completes our proof of this lemma. ∎

We continue with our analysis of the asymptotic property for $I_{1,4}(x)$. Since $\widetilde{p}_0$ is bounded, $\mathcal{F}_{p,I}$ is uniformly bounded. For any function $f \in \mathcal{F}_{p,I}$, we have $\mathbb{P}\{f^2(X,\boldsymbol{\alpha})\} \lesssim 1$ and $\sup_{y,\boldsymbol{\alpha}} |f(y,\boldsymbol{\alpha})| \lesssim 1$, which incorporated with Lemma 3.4.2 in VM leads to

$$E_{P_0}\|\mathbb{G}_n\|_{\mathcal{F}_{p,I}} \lesssim \widetilde{J}_{[]}\Big(1, \mathcal{F}_{p,I}, L_2(P_0)\Big)\left\{1 + \frac{\widetilde{J}_{[]}\Big(1, \mathcal{F}_{p,I}, L_2(P_0)\Big)}{\sqrt{n}}\right\}. \tag{B.52}$$

Applying Lemma 9, we have

$$\widetilde{J}_{[]}\left(1, \mathcal{F}_{p,I}, L_2(P_0)\right) \quad \lesssim \quad \int_0^1 \sqrt{1 + \sum_{k=1}^M \frac{|\log h_k|}{\epsilon^{1/a} h_k^{1+1/a}}} d\epsilon \lesssim \sum_{k=1}^M \frac{\sqrt{|\log h_k|}}{h_k^{0.5+0.5/a}},$$

which together with (B.52) leads to

$$E_{P_0}\|\mathbb{G}_n\|_{\mathcal{F}_{p,I}} \lesssim \sum_{k=1}^M \frac{\sqrt{|\log h_k|}}{h_k^{0.5+0.5/a}}.$$

This together with Chebyshev's inequality immediately implies that

$$\begin{aligned} &\mathbb{P}_n\left[\widehat{g}_p(X,\boldsymbol{\alpha}) I\{\widehat{g}_p(X,\boldsymbol{\alpha}) > 0\}\right] - \mathbb{P}\left[\widehat{g}_p(X,\boldsymbol{\alpha}) I\{\widehat{g}_p(X,\boldsymbol{\alpha}) > 0\}\right] \\ &\qquad = O_p\left(\sum_{k=1}^M \frac{\sqrt{|\log h_k|}}{n^{0.5} h_k^{0.5+0.5/a}}\right). \end{aligned} \tag{B.53}$$

It remains to examine $\mathbb{P}\left[\widehat{g}_p(X,\boldsymbol{\alpha}) I\{\widehat{g}_p(X,\boldsymbol{\alpha}) > 0\}\right]$. In fact

$$\begin{aligned} &\mathbb{P}\left[\widehat{g}_p(X,\boldsymbol{\alpha}) I\{\widehat{g}_p(X,\boldsymbol{\alpha}) > 0\}\right] \\ &= \int_{\mathbb{R}}\int_{S_\gamma} I\{\widetilde{p}_0(y,\boldsymbol{\alpha}) - \widehat{p}(y,\boldsymbol{\alpha}) > 0\}\{\widetilde{p}_0(y,\boldsymbol{\alpha}) - \widehat{p}(y,\boldsymbol{\alpha})\}\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha}) d\boldsymbol{\alpha} dy \\ &\le \int_{\mathbb{R}}\int_{S_\gamma} |\widetilde{p}_0(y,\boldsymbol{\alpha}) - \widehat{p}(y,\boldsymbol{\alpha})|\gamma(\boldsymbol{\alpha})\widetilde{p}_0(y,\boldsymbol{\alpha}) d\boldsymbol{\alpha} dy \lesssim d(\gamma\widehat{p}, \gamma\widetilde{p}_0). \end{aligned} \tag{B.54}$$

Now, we combine (B.48), (B.49), (B.51), (B.53), and (B.54) to give

$$\int_{\mathbb{R}} I_{1,1}(x)dx \le \int_{\mathbb{R}} I_{1,4}(x)dx \lesssim d(\gamma\widehat{p}, \gamma\widetilde{p}_0) + O_p\left(\sum_{k=1}^M \frac{\sqrt{|\log h_k|}}{n^{0.5} h_k^{0.5+0.5/a}}\right); \tag{B.55}$$

$$\int_{\mathbb{R}} I_{1,2}(x)dx \le \int_{\mathbb{R}} I_{1,4}(x)dx \lesssim d(\gamma\widehat{p}, \gamma\widetilde{p}_0) + O_p\left(\sum_{k=1}^M \frac{\sqrt{|\log h_k|}}{n^{0.5} h_k^{0.5+0.5/a}}\right). \tag{B.56}$$

B.4.3. Asymptotic property of I_2

With procedures that are similar to but easier than the above, we can show that

$$\left| I_2 - \int_{\boldsymbol{\alpha}\in S_\gamma} \alpha_j \gamma(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \right| \lesssim d(\gamma\widehat{p}, \gamma\widetilde{p}_0) + O_p\left(\sum_{k=1}^M \frac{\sqrt{|\log h_k|}}{n^{0.5} h_k^{0.5+0.5/a}}\right). \tag{B.57}$$

B.4.4. Summary

We now prove Theorem 6. Recall the decomposition of $\widehat{f}_j(x)$ in (B.27). We then have

$$\begin{aligned}
&\int_{\mathbb{R}} |\widehat{f}_j(x) - f_{0,j}(x)| dx \\
&\leq \frac{1}{I_2}\left\{\int_{\mathbb{R}} I_{1,1}(x)dx + \int_{\mathbb{R}} I_{1,2}(x)dx + \int_{\mathbb{R}} |I_{1,3}(x) - I_2 f_{0,j}(x)|dx\right\} \\
&\leq d(\gamma\widehat{p}, \gamma\widetilde{p}_0) + O_p\left(\sum_{k=1}^{M} \frac{\sqrt{|\log h_k|}}{n^{0.5} h_k^{0.5+0.5/a}}\right) + O_P(1)\int_{\mathbb{R}} |I_{1,3}(x) - I_2 f_{0,j}(x)|dx \\
&\leq O_P(1)\int_{\mathbb{R}}\left|\int_{\mathbb{R}} K_{h_j}(u-x) f_{0,j}(u)du - f_{0,j}(x)\right| dx \int \alpha_j \gamma(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \\
&\quad + d(\gamma\widehat{p}, \gamma\widetilde{p}_0) + O_p\left(\sum_{k=1}^{M} \frac{\sqrt{|\log h_k|}}{n^{0.5} h_k^{0.5+0.5/a}}\right),
\end{aligned} \tag{B.58}$$

where in the second "≤" we have used (B.55)–(B.57) and in the third "≤" we have used (B.47) and (B.57).

Given Conditions 0, 2, and 3, it can be checked that

$$\int_{\mathbb{R}}\left|\int_{\mathbb{R}} K_{h_j}(u-x) f_{0,j}(u)du - f_{0,j}(x)\right| dx = O(h_j) = O(h), \tag{B.59}$$

and

$$\begin{aligned}
O_p\left(\sum_{k=1}^{M} \frac{\sqrt{|\log h_k|}}{n^{0.5} h_k^{0.5+0.5/a}}\right) &= O_p\left(\frac{\sqrt{|\log h|}}{n^{0.5} h^{0.5+0.5/a}}\right) \\
&= O_p\left(\frac{1}{n^{0.5-1/a} h^{0.5}}\right) \cdot O_p\left(\frac{h^{0.5/a}\sqrt{|\log h|}}{(nh)^{1/a}}\right).
\end{aligned}$$

With Condition 1, for any arbitrarily small $\vartheta > 0$, we can choose a large enough a such that

$$O_p\left(\sum_{k=1}^{M} \frac{\sqrt{|\log h_k|}}{n^{0.5} h_k^{0.5+0.5/a}}\right) = O_p(n^{-0.5+\vartheta} h^{-0.5}). \tag{B.60}$$

Combining (B.58)–(B.60) and Theorem 4, we complete the proof of Theorem 6. ■

Acknowledgements

The authors thank the editor, the associate editor, and the referee for constructive comments and suggestions that lead to a significant improvement over the article. The authors also thank Dr. Yanyuan Ma for sending them the code to implement the method in Ma et al. (2011).

References

Acar, E. F. and Sun, L. (2013). A generalized Kruskal-Wallis test incorporating group uncertainty with application to genetic association studies. *Biometrics*, **69**, 427–435. MR3071061

Carvalho, B. S., Louis, T. A., and Irizarry, R. A. (2010). Quantifying uncertainty in genotype calls. *Bioinformatics*, **26**, 242–249.

Eggermont, P. P. B. (1999). Nonlinear smoothing and the EM algorithm for positive integral equations of the first kind. *Applied Mathematics Optimization*, **39**, 75–91. MR1654562

Eggermont P. P. B. and LaRiccia V. N. (1995a). Maximum smoothed likelihood density estimation. *Nonparametric Statistics*, **4**, 211–222. MR1366769

Eggermont P. P. B. and LaRiccia V. N. (1995b). Maximum smoothed likelihood density estimation for inverse problems. *The Annals of Statistics,* **23**, 199–220. MR1331664

Eggermont, P. P. B. and LaRiccia, V. N. (2000). Maximum likelihood estimation of smooth monotone and unimodal densities. *The Annals of Statistics,* **28**, 922–947. MR1792794

Eggermont, P. P. B. and LaRiccia, V. N. (2001). *Maximum Penalized Likelihood Estimation.* New York: Springer. MR1837879

Groeneboom, P. (2014). Maximum smoothed likelihood estimators for the interval censoring model. *The Annals of Statistics*, **42**, 2092–2137. MR3262478

Groeneboom, P., Jongbloed, G., and Witte, B.I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *The Annals of Statistics*, **38**, 352–387. MR2589325

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

Kitua, A. Y., Smith, T., Alonso, P.L., Masanja, H., Urassa, H., Menendez, C., Kimario, J., and Tanner, M. (1996). Plasmodium falciparum malaria in the first year of life in an area of intense and perennial transmission. *Tropical Medicine and International Health*, **1**, 475–484.

Kosorok, M. R. (2008) *Introduction to Empirical Processes and Semiparametric Inference.* New York: Springer. MR2724368

Lander, E.S. and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 743–756.

Levine, M., Hunter, D. R., and Chauvead, D. (2011). Maximum smoothed likelihood for multivariate mixtures. *Biometrika*, **98**, 403–416. MR2806437

Li, Y., Willer, C. J., Sanna, S., and Abecasis, G.R. (2009). Genotype imputation. *Annual Review of Genomics and Human Genetics*, **10**, 387–406.

Ma, Y., Hart, J. D., and Carroll, R.J. (2011). Density estimation in several populations with uncertain population membership. *Journal of the American Statistical Association,* **106**, 1180–1192. MR2894773

Ma, Y. and Wang, Y. (2012). Efficient distribution estimation for data with unobserved sub-population identifiers. *Electronic Journal of Statistics*, **6**, 710–

737. MR2988426

Qin, J. and Leung, D. H.Y. (2005). A semiparametric two-component "compound" mixture model and its application to estimating malaria attributable fractions. *Biometrics*, **61**, 456–464. MR2140917

Qin, J., Garcia, T.P., Ma, Y., Tang, M., Marder, K., and Wang, Y. (2014). Combining isotonic regression and EM algorithm to predict genetic risk under monotonicity constraint. *Annals of Applied Statistics,* **8**, 1182–1208. MR3262550

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall. MR0848134

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* New York: Springer. MR1385671

Vounatsou, P., Smith, T., and Smith, A.F.M. (1998). Bayesian analysis of two-component mixture distributions applied to estimating malaria attributable fractions. *Applied Statistics*, **47**, 575–587.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* London: Chapman and Hall. MR1319818

Wang, Y., Clark, L.N., Louis, E.D., Mejia-Santana, H., Harris, J., Cote, L.J., Waters, C., Andrews, D., Ford, B., Frucht, S., Fahn, S., Ottman, R., Rabinowitz, D. and Marder, K. (2008). Risk of Parkinson's disease in carriers of Parkin mutations: Estimation using the kin-cohort method. *Archives of Neurology*, **65**, 467–474.

Wang, Y., Garcia, T.P., and Ma, Y. (2012). Nonparametric estimation for censored mixture data with application to the cooperative Huntington's observational research trial. *Journal of the American Statistical Association*, **107**, 1324–1338. MR3036398

Wu, R., Ma, C., and Casella, G. (2007). *Statistical Genetics of Quantitative Traits: Linkage, Maps, and QTL.* New York: Springer. MR2344949

Yu, T., Li, P., and Qin, J. (2017). Density estimation in the two-sample problem with likelihood ratio ordering. *Biometrika*, **104**, 141–152. MR3626481