Semiparametric topographical mixture models with symmetric errors

Cristina Butucea, Rodrigue Nguyep, Pierre Vandekerkhove

▶ **To cite this version:**

Cristina Butucea, Rodrigue Nguyep, Pierre Vandekerkhove. Semiparametric topographical mixture models with symmetric errors. 2013. hal-00912964v3

HAL Id: hal-00912964
https://hal.science/hal-00912964v3

Preprint submitted on 7 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semiparametric topographical mixture models with symmetric errors

Cristina Butucea†, Rodrigue Ngueyep Tzoumpe* and Pierre Vandekerkhove †‡
† *Université Paris-Est*
LAMA (UMR 8050), UPEMLV
F-77454, Marne-la-Vallée, France
* *H. Milton Stewart School of Industrial Systems and Engineering*
Georgia Institute of Technology
‡ *UMI Georgia Tech - CNRS 2958,*
School of aerospace
Georgia Institute of Technology

February 6, 2014

Abstract

Motivated by the analysis of a Positron Emission Tomography (PET) imaging data considered in Bowen et al. (2012), we introduce a semiparametric topographical mixture model able to capture the characteristics of dichotomous shifted response-type experiments. We propose a local estimation procedure, based on the symmetry of the local noise, for the proportion and locations functions involved in the proposed model. We establish under mild conditions the minimax properties and asymptotic normality of our estimators when Monte Carlo simulations are conducted to examine their finite sample performance. Finally a statistical analysis of the PET imaging data in Bowen et al. (2012) is illustrated for the proposed method.

AMS 2000 subject classifications. Primary 62G05, 62G20; secondary 62E10.
Key words and phrases. Asymptotic normality, consistency, contrast estimators, Fourier transform, identifiability, inverse problem, semiparametric, mixture model, symmetric errors, finite mixture of regressions.

1 Introduction

The model we propose to investigate in this paper is a semiparametric topographical mixture model able to capture the characteristics of dichotomous shifted response-type experiments such as the tumor data in Bowen et al. (2012, Fig. 4). Let suppose that we visit at random the space $\mathbb{R}^d$ ($d \geq 1$) by sampling a sequence of i.i.d. random variables $\mathbf{X}_i$, $i = 1, ..., n$, having common probability distribution function (p.d.f.) $\ell : \mathbb{R}^d \to \mathbb{R}_+$. For each $\mathbf{X}_i$ we observe an output response Y_i whose distribution is a mixture model with probability parameters depending on the design $\mathbf{X}_i$. For simplicity, let us consider first a mixture of two nonlinear regression model:

$$Y_i = W(\mathbf{X}_i)(a(\mathbf{X}_i) + \tilde{\varepsilon}_{1,i}) + (1 - W(\mathbf{X}_i))(b(\mathbf{X}_i) + \tilde{\varepsilon}_{2,i}), \tag{1}$$

where locations are a, $b : \mathbb{R}^d \to \mathbb{R}$, the errors $\{\tilde{\varepsilon}_{1,i}, \tilde{\varepsilon}_{2,i}\}_{i=1,...,n}$ are supposed to be i.i.d with zero-symmetric common p.d.f. f. The mixture in model (1) occurs according to the random variable $W(\mathbf{x})$ at point $\mathbf{x}$, with probability $\pi : \mathbb{R}^d \to (0, 1)$,

$$W(\mathbf{x}) \quad = \quad \begin{cases} 1 & \text{with probability} \quad \pi(\mathbf{x}), \\ 0 & \text{with probability} \quad 1 - \pi(\mathbf{x}). \end{cases}$$

Moreover we assume that, conditionally on the $\mathbf{X}_i$'s, the $\{\tilde{\varepsilon}_{1,i}, \tilde{\varepsilon}_{2,i}\}_i$'s and the $W(\mathbf{X}_i)$'s are independent. Such a model is linked to the class of Finite Mixtures of Regression (FMR), see Grün and Leisch (2006) for a good overview. Briefly, statistical inference for the class of parametric FMR model was first considered by Quandt and Ramsey (1978) who proposed a moment generating function based estimation method. An EM estimating approach was proposed by De Veaux (1989) in the two-component case. Variations of the latter approach were also considered in Jones and McLachlan (1992) and Turner (2000). Hawkins et al. (2001) studied the estimation problem of the number of components in the parametric FMR model using approaches derived from the likelihood equation. In Hurn et al. (2003), the authors investigated a Bayesian approach to estimate the regression coefficients and also proposed an extension of the model in which the number of components is unknown. Zhu and Zhang (2004) established the asymptotic theory for maximum likelihood estimators in parametric FMR models. More recently, Städler et al. (2010) proposed an ℓ_1-penalized method based on a Lasso-type estimator for a high-dimensional FMR model with $d \geq n$. As an alternative to parametric approaches to the estimation of a FMR model, some authors suggested the use of more flexible semiparametric approaches. These approaches can actually be classified into two groups: semiparametric FMR (SFMR) of type I and type II. The study of SFMR of type I comes from the seminal

work of Hall and Zhou (2003) in which d-variate semiparametric mixture models of random vectors with independent components were considered. These authors proved in particular that, for $d \geq 3$, we can identify a two-component mixture model without parametrizing the distributions of the component random vectors (Type I definition). To the best of our knowledge, Leung and Qin (2006) were the first in estimating a FMR model semiparametrically in that sense. In the two-component case, they studied the case where the components are related by Anderson (1979)'s exponential tilt model. Hunter and Young (2012) studied the identifiability of an m-component type I SFMR model and numerically investigated a Expectation-Maximization (EM) type algorithm for estimating its parameters. Vandekerkhove (2013) proposed an M-estimation method for a two-component semiparametric mixture of linear regressions with symmetric errors (type I) in which one component is known. Bordes et al. (2013) revisited the same model by establishing new moment-based identifiability results from which they derived explicit $\sqrt{n}$-convergent estimators. The study of type II SFMR models started with Huang and Yao (2012) who considered a semiparametric linear FMR model with Gaussian noise in which the mixing proportions are possibly covariates-dependent (Type II definition: parametric noises with mixing proportion and/or noises' parameters functionally depending on covariates). They established also the asymptotic normality of their local maximum likelihood estimator and investigated a modified EM-type algorithm. Huang et al. (2013) generalized the latter work to nonlinear FMR with possibly covariates-dependent noises. Toshiya (2013) considered a Gaussian FMR model where the joint distribution of the response and the covariate (possibly functional) is itself modeled as a mixture. More recently Montuelle et al. (2013) considered a penalized maximum likelihood approach for Gaussian FMR models with logistic weights.

To improve the flexibility of our FMR model (1) and address the study of models involving design-dependent noises, see radiotherapy application described in Section 5, we will consider a slightly more general model:

$$Y_i = W(\mathbf{X}_i)(a(\mathbf{X}_i) + \varepsilon_{1,i}(\mathbf{X}_i)) + (1 - W(\mathbf{X}_i))(b(\mathbf{X}_i) + \varepsilon_{2,i}(\mathbf{X}_i)), \tag{2}$$

such that, given $\{\mathbf{X} = \mathbf{x}\}$, the common p.d.f. of the $\varepsilon_{j,i}(\mathbf{x})$, $j = 1, 2$, denoted $f_{\mathbf{x}}$, is zero-symmetric. Note that the above model combines type I and type II properties since no parametric assumption is made about the noise and the mixing proportion, along with the location parameters, are possibly design dependent. Our model is still said *semiparametric* because, given $\{\mathbf{X} = \mathbf{x}\}$, the vector $\theta(\mathbf{x}) = (\pi(\mathbf{x},)a(\mathbf{x}), b(\mathbf{x}))$ will be viewed as an Euclidean parameter to be estimated.

Examples of design-point noise dependency.

i) (Topographical scaling) The most natural transformation is probably when considering a topographical scaling of the errors, with $\sigma : \mathbb{R}^d \to \mathbb{R}_+^*$, such that $\varepsilon_{j,i}(\mathbf{X}_i) = \sigma(\mathbf{X}_i)\tilde{\varepsilon}_{j,i}$, $j = 1, 2$, where the $\tilde{\varepsilon}_{j,i}$'s are similar to those involved in (1). The conditional p.d.f given $\{\mathbf{X} = \mathbf{x}\}$ is defined by

$$f_{\mathbf{x}}(y) = \frac{1}{\sigma(\mathbf{x})} f\left(\frac{y}{\sigma(\mathbf{x})}\right), \quad y \in \mathbb{R}. \tag{3}$$

Indeed, if f is zero-symmetric then the errors' distribution inherits trivially the same symmetry property.

ii) (Zero-symmetric varying mixture) Another useful example could be the varying mixing proportion mixture model of r zero-symmetric distributions. For $k = 1, \ldots, r$, we consider proportion functions $\lambda_k : \mathbb{R}^d \to (0, 1)$ with $\sum_{k=1}^r \lambda_k(\mathbf{x}) = 1$ for all $\mathbf{x} \in \mathbb{R}^d$. The conditional p.d.f given $\{\mathbf{X} = \mathbf{x}\}$ is defined by

$$f_{\mathbf{x}}(y) = \sum_{k=1}^{r} \lambda_k(\mathbf{x}) f_k(y), \quad y \in \mathbb{R},$$

where the f_k functions are zero-symmetric p.d.f.'s.

iii) (Antithetic location model) Consider a location function $\mu : \mathbb{R}^d \to \mathbb{R}$ and f any arbitrary p.d.f. The conditional p.d.f given $\{\mathbf{X} = \mathbf{x}\}$ is defined by

$$f_{\mathbf{x}}(y) = \frac{1}{2} f(y - \mu(\mathbf{x})) + \frac{1}{2} f(-y + \mu(\mathbf{x})), \quad y \in \mathbb{R},$$

and also results into a zero-symmetric p.d.f.

Note that any combination of the above situations could be considered in model (2) free from specifying any parametric family (provided the resulting zero-symmetry hold). This last remark reveals, according to us, the main strength of our model in the sense that it could prove to be a very flexible exploratory tool for the analysis of shifted response-type experiments. Our paper is organized as follows. Section 2 is devoted to a detailed description of our estimation method, while Section 3 is concerned with its asymptotic properties. The finite-sample performance of the proposed estimation method is studied for various scenarios through Monte Carlo experiments in Section 4. In Section 5 we propose to analyze the Positron Emission Tomography (PET) imaging data considered in Bowen et al. (2012). Finally Section 6 is devoted to auxiliary results and main proofs.

2 Estimation method

Let us define the joint density of couples $(Y_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, designed from model (2):

$$g(y, \mathbf{x}) = [\pi(\mathbf{x}) f_{\mathbf{x}}(y - a(\mathbf{x})) + (1 - \pi(\mathbf{x})) f_{\mathbf{x}}(y - b(\mathbf{x}))]\ell(\mathbf{x}), \quad (y, \mathbf{x}) \in \mathbb{R}^{d+1}, \tag{4}$$

while the conditional density of Y given $\{\mathbf{X} = \mathbf{x}\}$ (denoted for simplicity $Y/\mathbf{X} = \mathbf{x}$) is

$$g_{\mathbf{x}}(y) = g(y, \mathbf{x})/\ell(\mathbf{x}) = \pi(\mathbf{x}) f_{\mathbf{x}}(y - a(\mathbf{x})) + (1 - \pi(\mathbf{x})) f_{\mathbf{x}}(y - b(\mathbf{x})). \tag{5}$$

We are interested in estimating the parameter $\theta_0 = \theta(\mathbf{x}_0) = (\pi(\mathbf{x}_0), a(\mathbf{x}_0), b(\mathbf{x}_0))$ at some fixed point $\mathbf{x}_0$ belonging to the interior of the support of ℓ ($\ell(\mathbf{x}_0) > 0$), denoted $supp(\ell)$. For simplicity and identifiability matters, we will suppose that θ_0 belongs to the interior of the parametric space $\Theta = [p, P] \times \Delta$, where $0 < p \leq P < 1/2$ and Δ denotes a compact set of $\mathbb{R}^2 \backslash \{(x, x) : x \in \mathbb{R}\}$.

2.1 Mixture of regression functions as an inverse problem

We see in formula (5), that the conditional density of Y given $\{\mathbf{X} = \mathbf{x}\}$ can be viewed as a mixture of the errors distribution $f_{\mathbf{x}}$ given $\{\mathbf{X} = \mathbf{x}\}$ with locations $(a(\mathbf{x}), b(\mathbf{x}))$ and mixing proportion $\pi(\mathbf{x})$. Mixture of populations with different locations is a well known inverse problem. Our inversion procedure is here based on the Fourier transform of the conditional density $g_{\mathbf{x}}(y)$ of $Y/\mathbf{X} = \mathbf{x}$. If the p.d.f. $g_{\mathbf{x}}$ belongs to $\mathbb{L}_1 \cap \mathbb{L}_2$, define $g^*_{\mathbf{x}}(u) = \int \exp[iuy] g_{\mathbf{x}}(y) dy$ for all $u \in \mathbb{R}$, and observe that

$$g^*_{\mathbf{x}}(u) \quad = \quad \left(\pi(\mathbf{x}) e^{iua(\mathbf{x})} + (1 - \pi(\mathbf{x})) e^{iub(\mathbf{x})}\right) f^*_{\mathbf{x}}(u), \quad u \in \mathbb{R}.$$

Let us denote, for all $(t, u) = (\pi, a, b, u) \in \Theta \times \mathbb{R}$,

$$M(t, u) := \pi e^{iua} + (1 - \pi) e^{iub}. \tag{6}$$

Note that $|M(t, u)| \in [1 - 2P, 1]$ for all $(t, u) \in \Theta \times \mathbb{R}$. Then, we have

$$g^*_{\mathbf{x}}(u) \quad = \quad M(\theta(\mathbf{x}), u) f^*_{\mathbf{x}_0}(u).$$

Let us fix $\mathbf{x}_0 \in supp(\ell)$ such that $\theta(\mathbf{x}_0)$ belongs to the interior of Θ, denoted $\overset{\circ}{\Theta}$. Noticing that the p.d.f. $f_{\mathbf{x}_0}$ is zero-symmetric we therefore have that $f^*_{\mathbf{x}_0}(u) \in \mathbb{R}$, for all $u \in \mathbb{R}$. If t belongs to Θ, we prove in the next theorem the *picking* property

$$\Im\left(\frac{g^*_{\mathbf{x}_0}(u)}{M(t, u)}\right) = 0 \text{ for all } u \in \mathbb{R}, \text{ if and only if } t = \theta(\mathbf{x}_0),$$

where $\Im : \mathbb{C} \to \mathbb{R}$ denotes the imaginary part of a complex number. This result allows us to build a *contrast* function for the parameter $t \in \Theta$:

$$S(t) := \int \Im^2 \left(\frac{g^*_{\mathbf{x}_0}(u)}{M(t,u)} \right) \ell^2(\mathbf{x}_0) w(u) du. \tag{7}$$

The function $w : \mathbb{R}^d \to \mathbb{R}_+$ is a bounded p.d.f. which helps in computing the integral via Monte-Carlo method and solves integrability issues.

Remark. The idea of using Fourier transform in order to solve the inverse mixture problem was introduced in Butucea and Vandekerkhove (2013) for density models. In the regression models we deal with the conditional density of $Y/\mathbf{X} = \mathbf{x}_0$. This has no incidence on the identifiability of the model but changes dramatically the behavior of the estimators as we shall see later on.

We prove in the following theorem that our model is identifiable and that $S(t)$ defines a contrast on the parametric space Θ.

Theorem 1 *(Identifiability and contrast property) Consider model (2) provided with $f_{\mathbf{x}}(\cdot) \in \mathbb{L}_1$ for all $\mathbf{x} \in \mathbb{R}^d$. For a fixed point $\mathbf{x}_0$ in the interior of the support of ℓ, we assume that $f_{\mathbf{x}_0}(\cdot)$ is zero-symmetric and that $\theta_0 = \theta(\mathbf{x}_0)$ is an interior point of Θ. Then we have the following properties:*

i) The collection of scalar parameters $\theta_0 = (\pi(\mathbf{x}_0), a(\mathbf{x}_0), b(\mathbf{x}_0))$ and the function $f_{\mathbf{x}_0}(\cdot)$ are identifiable.

ii) The function S in (7) is a contrast function, i.e. for all $t \in \Theta$, $S(t) \geq 0$ and $S(t) = 0$ if and only if $t = \theta_0$.

Proof. The proofs of i) and ii) are respectively similar to the proof of Theorem 1 and Proposition 1 in Butucea and Vandekerkhove (2013), replacing $f^*(\cdot)$ and $g^*(\cdot)$ by $f^*_{\mathbf{x}_0}(\cdot)$ and $g^*_{\mathbf{x}_0}(\cdot)$, and noticing that $\ell(\mathbf{x}_0)$ is bounded away from zero. Follows also Theorem 2.1 in Bordes et al. (2006). ■

Remark. For mixture models with higher number of components, i.e.

$$Y_i = \sum_{j=1}^{J} W_j(\mathbf{X}_i)(\gamma_j(\mathbf{X}_i) + \varepsilon_{j,i}(\mathbf{X}_i)), \quad i = 1, \ldots, n,$$

where $(W_1(\mathbf{x}), ..., W_J(\mathbf{x}))$ are distributed according to a J-components ($J > 2$) multinomial distribution with parameters $(\pi_1(\mathbf{x}), ..., \pi_J(\mathbf{x}))$, and noises $(\varepsilon_{j,i})$, $j = 1, \ldots, J$, i.i.d.

according to $f_{\mathbf{x}}$, we assume that there exists a compact set $\Psi \subset]0,1[^{J-1}\times\mathbb{R}^J$ of parameters $(\pi_1(\mathbf{x}),...,\pi_{J-1}(\mathbf{x}),\gamma_1(\mathbf{x}),...,\gamma_J(\mathbf{x}))$ where the model is *identifiable*, see Hunter et al. (2007, Section 2). Note that the 3-components mixture model has been studied closely in Bordes et al. (2006) and Hunter et al. (2007) where sufficient identifiability conditions were given. The case where $d > 3$ is more involved for full description and it is still an open question. In this setup, the estimation procedure described hereafter can be adapted over the parameter space Ψ with analogous results.

2.2 Estimation procedure

In order to build an estimator of the contrast $S(t)$ defined in (7), a local smoothing has to be performed in order to extract the information that the random design $X_1,...,X_n$ brings to the knowledge of the conditional law of $Y/\mathbf{X} = \mathbf{x}_0$. We use a kernel smoothing approach, but local polynomials or wavelet methods could also be employed. This smoothing is a major difference with respect to the density model considered in Butucea and Vandekerkhove (2013) and all the rates will depend on the smoothing parameter applied to the kernel function.

We choose a kernel function $K : \mathbb{R}^d \to \mathbb{R}$ belonging to $\mathbb{L}_1$ and to $\mathbb{L}_4$ and some bandwidth parameter $h > 0$ to be described later on. For $\mathbf{x}_0 \in supp(\ell)$ fixed, we denote

$$Z_k(t,u,h) := \left(\frac{e^{iuY_k}}{M(t,u)} - \frac{e^{-iuY_k}}{M(t,-u)}\right) K_h(\mathbf{X}_k - \mathbf{x}_0), \text{ where } K_h(\mathbf{x}) := \frac{1}{h^d} K\left(\frac{\mathbf{x}}{h}\right). \quad (8)$$

The empirical contrast of $S(t)$ is defined by

$$S_n(t) = -\frac{1}{4n(n-1)} \sum_{j\neq k, j,k=1}^{n} \int Z_k(t,u,h) Z_j(t,u,h) w(u) du, \quad (9)$$

where $w : \mathbb{R} \to \mathbb{R}_+^*$ is a bounded p.d.f., having a finite moment of order 4, i.e. $\int u^4 w(u)du < \infty$. From this empirical contrast we then define the estimator

$$\hat{\theta}_n = \arg\inf_{t\in\Theta} S_n(t), \quad (10)$$

of $\theta_0 = \theta(\mathbf{x}_0)$. We shall study successively the properties of $S_n(t)$ as an estimator of $S(t)$ and deduce consistency and asymptotic normality of $\hat{\theta}_n$ as an estimator of θ_0.

Estimation methodology for $f_{\mathbf{x_0}}$. For the estimation of the local noise density $f_{\mathbf{x}_0}$ we suggest to consider the natural smoothed version of the plug-in density estimate given in Butucea and Vandekerkhove (2013, Section 2.2).

Let us denote by $\varphi(\mathbf{x}, y) = \ell(\mathbf{x}) f_{\mathbf{x}}(y)$. We plug $\hat{\theta}_n$ in the natural smoothed nonparametric kernel estimator of $\varphi(\mathbf{x}, y)$ deduced from (6), whenever the unknown parameter θ_0 is required. For $\mathbf{x}_0$ fixed, we consider the Fourier transform of the resulting estimator of $\varphi(\mathbf{x}_0, y)$. This procedure gives, in Fourier domain,

$$\varphi^*_{\mathbf{x}_0,n}(u) = \frac{1}{n}\sum_{k=1}^{n} \frac{Q^*(h_{1,n}u)e^{iuY_k}}{M(\hat{\theta}_n, u)} K_{h_{2,n}}(\mathbf{X_k} - \mathbf{x}_0),$$

where Q is a univariate kernel ($\int Q = 1$ and $Q \in \mathbb{L}_2$) and $(h_{1,n}, h_{2,n})$ are bandwidth parameters properly chosen. Note that $G^*_n(u) := Q^*(h_{1,n}u)/M(\hat{\theta}_n, u)$ is in $\mathbb{L}_1$ and $\mathbb{L}_2$ and has an inverse Fourier transform which we denote by $G_n(u/h_{1,n})/h_{1,n}$. Therefore, the estimator of $\varphi(\mathbf{x_0}, y)$ is

$$\varphi_n(\mathbf{x}_0, y) = \frac{1}{nh_{1,n}} \sum_{k=1}^{n} G_n\left(\frac{x - X_k}{h_{1,n}}\right) K_{h_{2,n}}(\mathbf{X_k} - \mathbf{x}_0).$$

Finally the estimator of $f_{\mathbf{x}_0}$ is obtained by considering

$$\hat{f}_{\mathbf{x}_0}(y) = \frac{f_n(y|\mathbf{x}_0)\mathbb{I}_{f_n(y|\mathbf{x}_0)\geq 0}}{\int_{\mathbb{R}} f_n(y|\mathbf{x}_0)\mathbb{I}_{f_n(y|\mathbf{x}_0)\geq 0} dy}, \quad \text{where} \quad f_n(y|\mathbf{x}_0) = \frac{\varphi_n(\mathbf{x}_0, y)}{\ell_n(\mathbf{x}_0)}. \tag{11}$$

where $\ell_n(\mathbf{x}_0) = \frac{1}{n}\sum_{k=1}^{n} K_{h_{2,n}}(\mathbf{X_k} - \mathbf{x}_0)$. The asymptotic properties of this local density estimator are not established yet but we strongly guess that the bandwidth conditions required to prove its convergence and classical convergence rate are similar to those found in the conditional density estimation literature, see Brunel et al. (2010) or Cohen and Le Pennec (2012).

3 Performance of the method

We give upper bounds for the mean squared error of $S_n(t)$. We are interested in consistency and asymptotic normality of $\hat{\theta}_n$ and this requires some small amount of smoothness $\alpha \in (0, 1]$ for the p.d.f. of the errors and for the functions π, a and b. From now on, $\|v\|$ denotes the Euclidean norm of vector v. Recall that a function F is Lipschitz α-smooth if it belongs to the following class

$$L(\alpha, M) = \left\{ F : \mathbb{R}^d \to \mathbb{R}, |F(\mathbf{x}) - F(\mathbf{y})| \leq M\|\mathbf{x} - \mathbf{y}\|^\alpha, \ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d \right\},$$

for $\alpha \in (0, 1]$ and $M > 0$.

A1. We assume that the functions π, a, b, ℓ are Lipschitz α-smooth with constant $M > 0$.

Remark. We may actually suppose that the functions appearing in our model have different smoothness parameters, but the rate will be governed by the smallest smoothness parameter.

An important consequence of this assumption is that the density ℓ is uniformly bounded by some constant depending only on α and M, i.e. $\sup_{\ell \in L(\alpha,M)} \|\ell\|_\infty < \infty$.

A2. Assume that $f_{\mathbf{x}}(\cdot) \in \mathbb{L}_1 \cap \mathbb{L}_2$ for all $\mathbf{x} \in \mathbb{R}^d$. In addition, we require that there exists a w-integrable function φ such that

$$|f^*_{\mathbf{x}}(u) - f^*_{\mathbf{x}'}(u)| \leq \varphi(u)\|\mathbf{x} - \mathbf{x}'\|^\alpha, \quad (\mathbf{x}, \mathbf{x}') \in \mathbb{R}^d \times \mathbb{R}^d, \ u \in \mathbb{R}.$$

Remark. Note that for the scaling model (3), if f is the $\mathcal{N}(0,1)$ p.d.f. and $\sigma(\cdot)$ is bounded and Lipschitz α-smooth, we have:

$$|f^*_{\mathbf{x}}(u) - f^*_{\mathbf{x}'}(u)| \leq \frac{u^2}{2}|\sigma^2(\mathbf{x}) - \sigma^2(\mathbf{x}')| \leq \frac{u^2}{2}\|\mathbf{x} - \mathbf{x}'\|^\alpha.$$

A3. We assume that the kernel K is such that $\int |K| < \infty$, $\int K^4 < \infty$ and that it satisfies also the moment condition

$$\int \|\mathbf{x}\|^\alpha |K(\mathbf{x})| d\mathbf{x} < \infty.$$

A4. The weight function w is a p.d.f. such that

$$\int (u^4 + \varphi(u)) w(u) du < \infty.$$

Remark. We may suppose that the smoothness $\alpha > 1$. In that case, the class $L(\alpha, M)$ consists of all functions F with bounded derivatives up to order k, where $\alpha = k + \beta$, $k \in \mathbb{N}$ and $\beta \in (0, 1]$. Moreover, for all multi-index $j = (j_1, ..., j_d) \in \mathbb{N}^d$ such that $|j| = k$ where $|j| = j_1 + ... + j_d$, we have

$$|F^{(j)}(\mathbf{x}) - F^{(j)}(\mathbf{y})| \leq M\|\mathbf{x} - \mathbf{y}\|^\beta, \quad (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d.$$

The following results will hold true under the additional assumption on the kernel (see **A3**): $\int \mathbf{x}^j K(\mathbf{x}) d\mathbf{x} = 0$, for all j such that $|j| \leq k$.

Proposition 1 *For each $t \in \Theta$ and $\mathbf{x}_0 \in supp(\ell)$ fixed, suppose $\theta_0 \in \overset{\circ}{\Theta}$ and that assumptions* **A1-A4** *hold. Then, the empirical contrast function $S_n(\cdot)$ defined in (9) satisfies*

$$E\left[(S_n(t) - S(t))^2\right] \leq C_1 h^{2\alpha} + C_2 \frac{1}{nh^d},$$

if $h \to 0$ and $nh^d \to \infty$ as $n \to \infty$, where constants C_1, C_2 depend on Θ, K, w, α and M but are free from n, h, t and $\mathbf{x}_0$.

Theorem 2 (Consistency) *Let suppose that assumptions of Proposition 1 hold and consider model (2) with likelihood given by (4). If the p.d.f $f_{\mathbf{x}_0}$ is zero-symmetric, then the estimator $\hat{\theta}_n$ defined in (9-10) converges in probability to $\theta(\mathbf{x}_0) = \theta_0$ if $h \to 0$ and $nh^d \to \infty$ as $n \to \infty$.*

The following theorem establishes the asymptotic normality of the estimator $\hat{\theta}_n$ of θ_0. Recall that $\theta_0 = \theta(\mathbf{x}_0)$ belongs to Θ and that there exists $l > 0$ such that $\ell(\mathbf{x}_0) \geq l$. We see that the local smoothing with bandwidth $h > 0$ deteriorates the rate of convergence to $\sqrt{nh^d}$ instead of $\sqrt{n}$ for the density model. In the asymptotic variance we will use the following notation:

$$\dot{J}(\theta_0, u) := \Im\left(-\frac{\dot{M}(\theta_0, u)}{M(\theta_0, u)}\right) f^*_{\mathbf{x}_0}(u)\ell(\mathbf{x}_0), \tag{12}$$

and

$$V(\theta_0, u_1, u_2) := \int \left(\frac{e^{iu_1 y}}{M(\theta_0, u_1)} - \frac{e^{-iu_1 y}}{M(\theta_0, -u_1)}\right)\left(\frac{e^{iu_2 y}}{M(\theta_0, u_2)} - \frac{e^{-iu_2 y}}{M(\theta_0, -u_2)}\right) g_{\mathbf{x}_0}(y)dy, \tag{13}$$

where the function $M(\cdot, \cdot)$ is defined in (6). Note that $\dot{J}(\theta_0, \cdot)$ is uniformly bounded by some constant and that V is well defined for all $(u_1, u_2) \in \mathbb{R} \times \mathbb{R}$ and also uniformly bounded by some constant.

Theorem 3 (Asymptotic normality) *Suppose that assumptions of Theorem 2 hold. The estimator $\hat{\theta}_n$ of θ_0 defined by (9-10), with $h \to 0$ such that $nh^d \to \infty$ and such that $h^{2\alpha+d} = o(n^{-1})$, as $n \to \infty$, is asymptotically normally distributed:*

$$\sqrt{nh^d}(\hat{\theta}_n - \theta_0) \to N(0, \mathcal{S}) \quad \textit{in distribution},$$

where $\mathcal{S} = \frac{1}{4}\mathcal{I}^{-1}\Sigma\mathcal{I}$, with

$$\mathcal{I} = -\frac{1}{2}\int \dot{J}(\theta_0, u)\dot{J}(\theta_0, u)^\top dw(u),$$

and

$$\Sigma := \int\int \dot{J}(\theta_0, u_1)\dot{J}^\top(\theta_0, u_2)V(\theta_0, u_1, u_2)w(u_1)w(u_2)du_1 du_2,$$

for $\dot{J}$ defined in (12) and V in (13).

The above results show that our estimator of θ_0 behaves like any nonparametric pointwize estimator. This is indeed the case and we provide in the next theorem the best achievable convergence rates uniformly over the large set of functions involved in our model, see assumptions **A1**-**A2**. For length matters, we will just provide some hints of proof of the next theorem.

Theorem 4 (Minimax rates) *Suppose* **A1**-**A4** *and consider* $\mathbf{x}_0 \in supp(\ell)$ *fixed such that* $\ell(\mathbf{x}_0) \geq L_* > 0$ *for all* $\ell \in L(\alpha, M)$ *and* $\theta_0 = \theta(\mathbf{x}_0) \in \overset{\circ}{\Theta}$. *The estimator* $\hat{\theta}_n$ *of* θ_0 *defined by (9-10), with* $h \asymp n^{-1/(2\alpha+d)}$, *as* $n \to \infty$, *is such that*

$$\sup E[\|\hat{\theta}_n - \theta_0\|^2] \leq C n^{-\frac{2\alpha}{2\alpha+d}},$$

where the supremum is taken over all the functions π, a, b, ℓ *and* f^* *checking assumptions* **A1**-**A2**. *Moreover,*

$$\inf_{T_n} \sup E[\|T_n - \theta_0\|^2] \geq c n^{-\frac{2\alpha}{2\alpha+d}},$$

where C, $c > 0$ *depend only on* α, M, Θ, K *and* w, *and the infimum is taken over the set of all the estimators* T_n *(measurable function of the observations* $(X_1, \ldots, X_n)$*) of* θ_0.

Proof hints. Throughout the proofs of the previous results we learn that the estimator $\hat{\theta}_n$ of θ_0, behaves asymptotically as $\dot{S}_n(\theta_0)$ which is a U-statistic with a dominant term whose bias is of order $h^{2\alpha}$ and whose variance is smaller than $C_2(nh^d)^{-1}$. The bias-variance compromise will produce an optimal choice of the bandwidth h of order $n^{-1/(2\alpha+d)}$ and a rate $n^{-\frac{2\alpha}{2\alpha+d}}$. It is the optimal rate for estimating a Lipschitz α-smooth regression function at a fixed point and the optimality results in the previous theorem are a consequence of the general nonparametric problem, see Stone (1977), Ibragimov and Has'minski (1981) and Tsybakov (2009).

4 Practical behaviour

4.1 Algorithm

We describe below the initialization scheme and the optimization method used to determine the estimates of the locations $a(\mathbf{x}_k)$, $b(\mathbf{x}_k)$ and the weight functions $\pi(\mathbf{x}_k)$ for a fixed sequence of testing points $\{\mathbf{x}_k,\ k = 1, \ldots, K\}$. To simply differentiate these testing points from the design data points we will allocate specifically the index k for the numbering of the testing points and the index i for the numbering of the dataset points, i.e. $\{(\mathbf{x}_i, y_i),\ i = 1, \ldots, n\}$.

Initialization

1. For each design data point $\mathbf{x}_i$, $i = 1, \ldots, n$, fit a kernel regression smoothing $\bar{m}(\mathbf{x}_i)$ with local bandwidth $\bar{h}_{\mathbf{x}_i}$. The `R` package `lokerns`, see Herrmann (2013), can be used.

2. Classify each data point $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, n$ according to: if $y_i > \bar{m}(\mathbf{x}_i)$ classify $(\mathbf{x}_i, y_i)$ in group 1 associated with location $a(\cdot)$, otherwise classify it in group 2 associated with $b(\cdot)$.

3. For each $\mathbf{x}_k$, $k = 1, \ldots, K$, obtain initial value $\bar{a}(\mathbf{x}_k)$, respectively $\bar{b}(\mathbf{x}_k)$, by fitting a kernel regression smoothing based on the observations $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, n$, previously classified in group 1 with local bandwidth $\bar{h}_{1,\mathbf{x}_k}$, respectively in group 2 with local bandwidth $\bar{h}_{2,\mathbf{x}_k}$.

4. Compute the local bandwidth $h_{\mathbf{x}_k} = \min(\bar{h}_{1,\mathbf{x}_k}, \bar{h}_{2,\mathbf{x}_k})$.

5. Fix an arbitrary single value $\bar{\pi}$ for all the $\pi(\mathbf{x}_k)$'s.

Estimation

1. Generate one w-distributed i.i.d sample (U_r), $r = 1, \ldots, N$ dedicated to the pointwize Monte Carlo estimation of $S_n(t)$ defined by:

$$S_n^{MC}(t) = -\frac{1}{4n(n-1)N} \sum_{j \neq k, j,k=1}^{n} \sum_{r=1}^{N} Z_k(t, U_r, h) Z_j(t, U_r, h).$$

   In the Sections 4.2 and 5, we will consider $N = n$ and w the p.d.f. corresponding to the mixture $0.1 * \mathcal{N}(0,1) + 0.9 * \mathcal{U}_{[-2,2]}$.

2. Compute the minimizer $\hat{\theta}(\mathbf{x}_k) = (\hat{\pi}(\mathbf{x}_k), \hat{a}(\mathbf{x}_k), \hat{b}(\mathbf{x}_k))$ of $S_n^{MC}(\cdot)$ evaluated at each point $\mathbf{x}_0 = \mathbf{x}_k$, by using the starting values $(\bar{\pi}, \bar{a}(\mathbf{x}_k), \bar{b}(\mathbf{x}_k))$ and the local bandwidth $h_{\mathbf{x}_k}$.

In our simulations, the above minimization will be, contrarily to the theoretical requirements, deliberately done over a non-constrained space, i.e. generically $\theta(\cdot) \in [0.05, 0.95] \times [A, B]^2$, with $A < B$. Our goal is to analyze experimentally if a performant initialization procedure is able to prevent from spurious phenomenons like the label switching or component merging occurring when $\pi(\mathbf{x}_0)$ is close to 0.5. This kind of information is actually very relevant to interpret correctly some cross-over effects as the one we will observe in Fig. 6 (a). Note that other initialization methods can be figured out. We can for instance use, similarly to Huang et al. (2013), a mixture of polynomial regressions with constant proportions and variances to pick initial values $\bar{a}(x)$ and $\bar{b}(x)$, or the `R` package `flexmix`, see Gruen et al. (2013), that implements a general framework for finite mixture of regression models based on EM-type algorithms (we selected this latter approach for the analysis of radiotherapy application in Section 5).

4.2 Simulations

In this section, we propose to measure the performances of our estimator $\hat{\theta}_n(\cdot)$ over a testing sequence $\{\mathbf{x}_k = k/K\}$, $k = 1, \ldots, K = 20$. Given that in the simulation setting the true function $\theta(\cdot)$ is known, we can compute, similarly to Huang et al. (2013), the Root Average Squared Errors (RASE) of our estimator. To this end we generate $M = 100$ datasets $(\mathbf{X}_i^{[z]}, Y_i^{[z]})_{1 \leq i \leq n}$, $z = 1, \ldots, M$ of sizes n= 400, 800, 1200, for each of the scenario described below and, for each scalar parameter $s = a,\ b,\ \pi$, denote by $RASE_s^{[z]}$ the RASE performance associated to the z-th dataset, defined by $RASE_s^{[z]} = (1/K \sum_{k=1}^{K} R_s^{[z]}(k))^{1/2}$, where $R_s^{[z]}(k) = \left(\hat{s}^{[z]}(\mathbf{x}_k) - s(\mathbf{x}_k)\right)^2$, and the empirical RASE by

$$RASE_s = \frac{1}{M} \sum_{z=1}^{M} RASE_s^{[z]}. \tag{14}$$

Let us also define the empirical squared deviation at point $\mathbf{x}_k$ by $\nu_k = \frac{1}{M} \sum_{z=1}^{M} R_s^{[z]}(k)$, and empirical variance of the squared deviation at $\mathbf{x}_k$ by $\sigma_s^2(k) = \frac{1}{M-1} \sum_{z=1}^{M} \left(R_s^{[z]}(k) - \nu_k\right)^2$. From these quantities we deduce the averaged variance of the squared deviations defined by

$$\sigma_s^2 = \frac{1}{K} \sum_{k=1}^{K} \sigma_s^2(k). \tag{15}$$

In all the simulation setups, we use the same mixing proportion function $\pi(\cdot)$:

$$\pi(\mathbf{x}) \quad = \quad \frac{\sin(3\pi x) - 1}{15} + 0.4, \quad \mathbf{x} \in [0, 1].$$

Gaussian setup (G). The errors $\varepsilon_{j,i}(\mathbf{x})$'s are distributed according to a Gaussian topographical scaling model corresponding to (3), i.e. f is the $\mathcal{N}(0, 1)$ p.d.f. when the location and scaling functions are

$$a(\mathbf{x}) = 4 - 2\sin(2\pi\mathbf{x}), \quad b(\mathbf{x}) = 1.5\cos(3\pi\mathbf{x}) - 3, \quad \sigma(\mathbf{x}) = 0.9\exp(\mathbf{x}), \quad \mathbf{x} \in [0, 1].$$

Student setup (T). The errors $\varepsilon_{j,i}(\mathbf{x})$'s are distributed according to a Student distribution with continuous degrees of freedom function denoted $df(\mathbf{x})$. The locations and degrees of freedom functions are

$$a(\mathbf{x}) = 3 - 2\sin(2\pi x), \quad b(\mathbf{x}) = 1.5\cos(3\pi x) - 2, \quad df(\mathbf{x}) = -5x + 8, \quad \mathbf{x} \in [0, 1].$$

Laplace setup (L). The errors $\varepsilon_{j,i}(\mathbf{x})$'s are distributed according to a Laplace distribution with scaling function $\nu(\mathbf{x})$. The locations and scaling functions are

$$a(\mathbf{x}) = 5 - 3\sin(2\pi\mathbf{x}), \quad b(\mathbf{x}) = 2\cos(3\pi\mathbf{x}) - 4, \quad \nu(\mathbf{x}) = \mathbf{x} + 1, \quad \mathbf{x} \in [0, 1].$$

Comments on Tables 1-3. We report for the simulation setups **(G)**, **(T)** and **(L)** the quantities $RASE_s$ defined in (14), and between parenthesis σ_s^2 defined in (15), for $s = \pi,\ a,\ b$. In these tables, we label our method as NMR-SE (Nonparametric Mixture of Regression with Symmetric Errors). To illustrate the contribution of our method, we compare our results with the RASE obtained by using the local EM-type algorithm proposed by Huang et al. (2013) for Nonparametric Mixture of Regression models with Gaussian noises (method labeled for simplicity NMRG). When the errors of the simulated model are Gaussian, the NMRG estimation should outperform our method, since the NMRG method assumes correctly that the errors are normally distributed, while our method does not make any parametric assumption on the distribution of the errors. When the sample size $n = 400$, the NMRG is more precise than our method, since the $RASE_s$'s and σ_s^2's are both smaller for the NMRG . When we increase the sample size of the simulated datasets to $n = 800,\ 1200$, our method becomes more competitive and yields $RASE_s$'s and σ_s^2's that are lower than those obtained by NMRG . This surprising behavior is probably due to the fact that in model (2) we impose the equality in law of the noises up to a shift parameter, when in the NMRG approach possibly different variances are fitted to each kind of noise, increasing by the way drastically the degrees of freedom of the model to be addressed. In Tables 2 and 3 we observe that our method has globally smaller $RASE_s$'s

Sample size	Method	$RASE_\pi$ (σ_π^2)	$RASE_a$ (σ_a^2)	$RASE_b$ (σ_b^2)
$n = 400$	NMRG	0.011 (0.015)	0.523 (0.952)	0.237 (0.415)
	NMR-SE	0.018 (0.034)	0.661 (1.485)	0.304 (0.833)
$n = 800$	NMRG	0.010 (0.012)	0.436 (0.767)	0.206 (0.368)
	NMR-SE	0.006 (0.013)	0.311 (0.696)	0.145 (0.370)
$n = 1200$	NMRG	0.009 (0.013)	0.469 (0.896)	0.197 (0.340)
	NMR-SE	0.003 (0.008)	0.209 (0.439)	0.094 (0.230)

Table 1: $RASE_z$'s and σ_z^2's for data with Gaussian Errors

and σ_s^2's. This result is not surprising, given that in the estimation methodology of Huang et al. (2013), the distribution of the noise are then completely misspecified under the simulation setups **(T)** and **(L)**. Note however, that when the sample size is small $n = 400$,

the NMRG displays better results, which can be explained by the fact that when we generate small size datasets, the points that are supposed to be in the tails of the non-normal distributions are less likely to appear in the dataset. So in that case it can be reasonable to assume that the Gaussian distribution approximates the errors distribution well.

Sample size	Method	$RASE_\pi$ (σ_π^2)	$RASE_a$ (σ_a^2)	$RASE_b$ (σ_b^2)
$n = 400$	NMRG	0.013 (0.018)	0.342 (0.631)	0.126 (0.205)
	NMR-SE	0.012 (0.025)	0.294 (0.664)	0.117 (0.249)
$n = 800$	NMRG	0.011 (0.014)	0.236 (0.377)	0.110 (0.189)
	NMR-SE	0.004 (0.008)	0.108(0.238)	0.047 (0.093)
$n = 1200$	NMRG	0.010 (0.013)	0.216 (0.352)	0.099 (0.153)
	NMR-SE	0.003 (0.006)	0.067 (0.125)	0.035 (0.072)

Table 2: $RASE_z$'s and σ_z^2's for data with Student Errors

Sample size	Method	$RASE_\pi$ (σ_π^2)	$RASE_a$ (σ_a^2)	$RASE_b$ (σ_b^2)
$n = 400$	NMRG	0.012 (0.004)	0.250 (0.156)	0.108 (0.036)
	NMR-SE	0.022 (0.012)	0.462 (0.623)	0.105 (0.088)
$n = 800$	NMRG	0.009 (0.003)	0.202 (0.100)	0.091 (0.036)
	NMR-SE	0.004 (0.002)	0.109 (0.010)	0.039 (0.014)
$n = 1200$	NMRG	0.009 (0.003)	0.192 (0.082)	0.091 (0.035)
	NMR-SE	0.002 (0.001)	0.064 (0.025)	0.027 (0.010)

Table 3: $RASE_z$'s and σ_z^2's for data with Laplace Errors

Comments on Figures 1-5. To illustrate the sensitivity of our method and compare it graphically to the NMRG approach we plot in Fig. 1 different samples coming from the setups **(G)**, **(T)**, and **(L)** for $n = 1200$, and in blue lines the corresponding true location functions $a(\cdot)$ and $b(\cdot)$. In Fig. 2, respectively Fig. 3, we plot in grey the $M = 100$ segment-line interpolation curves obtained by connecting the points $(\mathbf{x}_k, \hat{s}^{[z]}(\mathbf{x}_k))$, $k = 1, \ldots, K$ where $s(\cdot) = a(\cdot)$, $b(\cdot)$ for the NMRG method, respectively our NMR-SE method. In Fig. 4 and 5 we do the same for $s(\cdot) = \pi(\cdot)$. In Fig. 2-5 the dashed red lines represent the mean curves obtained by connecting the points $(\mathbf{x}_k, \bar{s}(\mathbf{x}_k))$, $k = 1, \ldots, K$ with $\bar{s}(\mathbf{x}_k) = 1/M \sum_{z=1}^{M} \hat{s}^{[z]}(\mathbf{x}_k)$ and $s(\cdot) = a(\cdot)$, $b(\cdot)$ and $\pi(\cdot)$. Let us observe first that the good behavior of the NMR-SE method is confirmed by the small variability of the

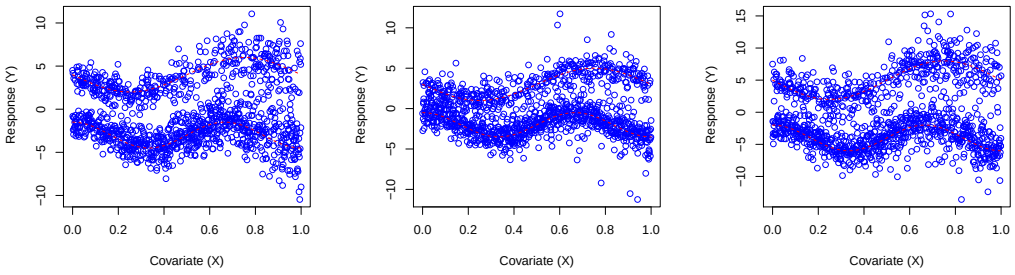
curves in Fig. 3 and 5 compared to those in Fig. 2 and 4 corresponding to the NMRG method. Secondly it is important to notice that sometime, since we did not constrained or method to have $\pi \in [p, P]$ with $0 < p < P < 1/2$, we run into some spurious estimation due to label switching or component merging phenomenon.

Label switching. This well known phenomenon, due to the lack of identifiability when the parametric space is not lexicographically ordered, translate into our case by a double-representation of the mixture model (5), i.e.

$$\pi(\mathbf{x})f_{\mathbf{x}}(y - a(\mathbf{x})) + (1 - \pi(\mathbf{x}))f_{\mathbf{x}}(y - b(\mathbf{x})) = \pi'(\mathbf{x})f_{\mathbf{x}}(y - a'(\mathbf{x})) + (1 - \pi'(\mathbf{x}))f_{\mathbf{x}}(y - b'(\mathbf{x}))$$

where $a'(\cdot) = b(\cdot)$, $b'(\cdot) = a(\cdot)$, and $\pi'(\cdot) = 1 - \pi(\cdot)$. This switching phenomenon is observable on the interval $[0, 0.2]$ of Fig. 3 (b) where the two populations of the mixture strongly overlap, see Fig. 1 (b).

Component merging. When $\pi(\cdot)$ is close to 0.5 it is actually hard to decide if we have only one shifted symmetric distribution, i.e. $g_{\mathbf{x}}(y) = 1 * f(y - c(\mathbf{x})) + 0$ where $c(\mathbf{x}) = (b + a)(\mathbf{x})/2$ and $f(y) = 1/2f(y + (b - a)(\mathbf{x})/2) + 1/2f(y - (b - a)(\mathbf{x})/2)$ or a balanced two-component mixture $g_{\mathbf{x}}(y) = 1/2f(y - a(\mathbf{x})) + 1/2f(y - b(\mathbf{x}))$. This phenomenon happens clearly when $\hat{\pi}^{[z]}(\cdot)$ is unexpectedly attracted by the single values 0 or 1, as it occurs sometimes on the intervals $[0, 0.2]$ or $[0.8, 1]$, see Fig. 5 (a-c).

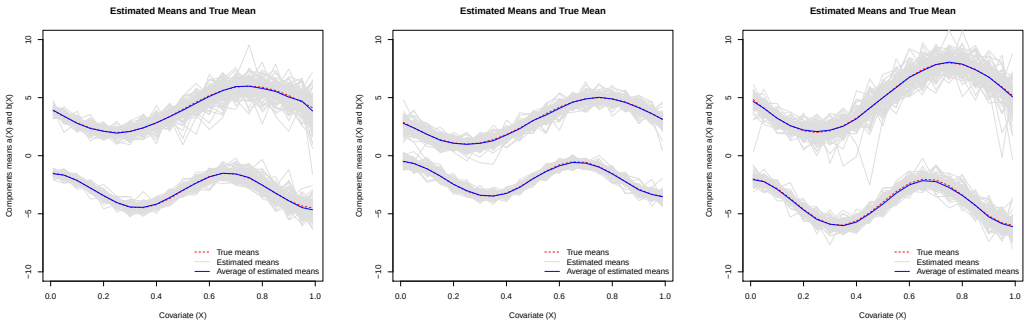


(a) Gaussian distribution (b) Student distribution (c) Laplace distribution

Figure 1: Examples of simulated datasets with different distribution errors

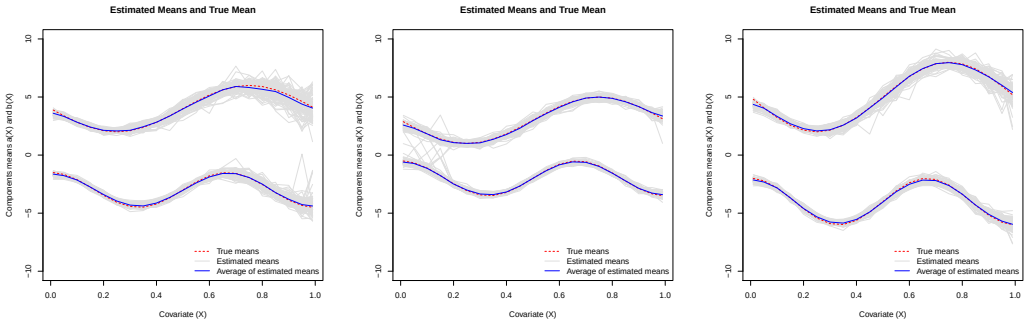
5 Application in radiotherapy

In this section, we implement the proposed methodology to a dataset obtained from applying radiation therapy to a canine patient with locally advanced Sinonasal Neoplasia.



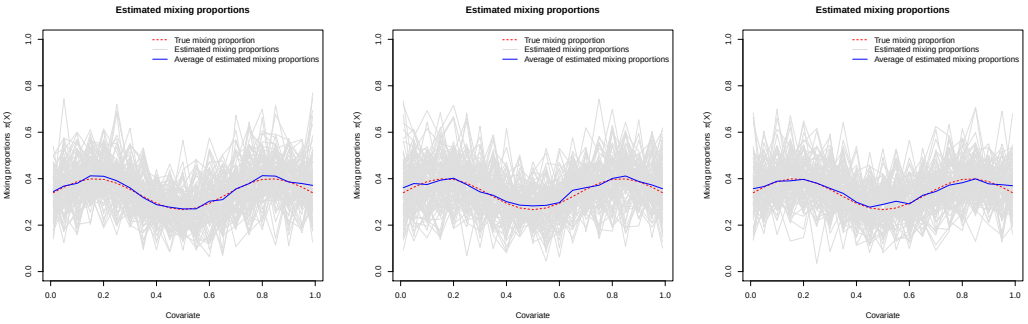
(a) Gaussian distribution (b) Student distribution (c) Laplace distribution

Figure 2: Mean Curves estimated with NMRG



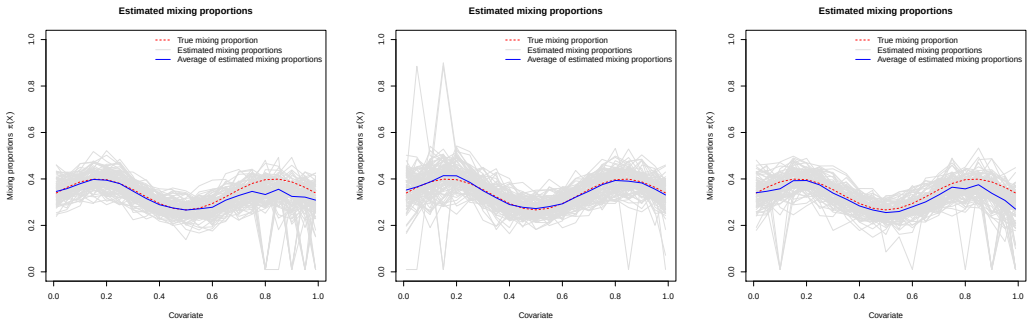
(a) Gaussian distribution (b) Student distribution (c) Laplace distribution

Figure 3: Mean Curves estimated with NMR-SE



(a) Gaussian distribution (b) Student distribution (c) Laplace distribution

Figure 4: Mixing proportions estimated with NMRG



(a) Gaussian distribution (b) Student distribution (c) Laplace distribution

Figure 5: Mixing proportions curves estimated with NMR-SE

These data were provided by Bowen et al. (2012, Fig. 4) who used them to quantify the associations between pre-radiotherapy and post-radiotherapy PET parameters via spatially resolved mixture of linear regressions. Intensity Modulated Radiotherapy is an advanced radiotherapy method that uses computer controlled device to deliver radiation of varying intensities to tumor or smaller areas within the tumor. There is evidence showing that the tumor is not homogeneous in its response to the radiation, and that some regions are more resistant than others. Functional imaging techniques (such as Positron Emission Tomography) can be used to identify the radiotherapy resistant regions within the tumor. For instance, an uptake in PET imaging of follow-up 2-deoxy-2-[^{18}F]fluoro-D-glucose (FDG) is empirically linked to a local recurrence of the disease. Bowen et al. (2012), use this approach to construct a prescription function that maps the image intensity values into a local radiation dose that will maximize the probability of a desired clinical outcome. In their manuscript they validate the use of molecular imaging based prescription function against clinical outcome by establishing an association between imaging biomarkers (PET imaging pre-radiotherapy) and regional imaging response to known dosage of therapy (PET imaging post-radiotherapy). The regional imaging response captures the change in imaging signal over an individual image volume element (called a voxel). In our model of interest (2), the pre-radiotherapy PET imaging intensities correspond to the input $\mathbf{X}_i$'s, and the post-radiotherapy PET imaging levels are the outputs Y_i's. For many patients, the empirical link between post-treatment PET of FDG (regional imaging response) and pre-treatment PET of FDG (imaging biomarker at baseline) is well captured by a mixture regression model with two components. For a set of voxels with similar pre-treatment PET intensities, the nature of the response to the radiotherapy leads to two groups of voxels. The first group corresponds to voxels that respond well to the radiotherapy, and

the second group contains the non-responding voxels. In our model of interest (2), the non-responding voxel group corresponds to the case where $W(\mathbf{X}_i) = 1$. The location parameters of each group appears to change as the pre-radiotherapy imaging intensity $\mathbf{X}_i$ varies. These changes in location are captured in our model by the location functions $a(\cdot)$ or $b(\cdot)$, where $a(\cdot)$, respectively $b(\cdot)$, is the component mean function for the completely responding (CR), respectively non-responding (NR), voxel. Additionally, the proportion of voxels $\pi(\mathbf{X}_i)$ that respond well to treatment depends on the pre-treatment level of the PET, so the mixture model should also account for a mixing proportion that depends on the input $\mathbf{X}_i$. For a given input $\mathbf{x}$, we assume that the intensity level of the completely responding and the non-responding voxel have approximately the same p.d.f. $f_{\mathbf{x}}$ up to a shift parameter, with the topographical scaling structure (3) presented in the Introduction. The variance of the distribution also changes with the level of the covariate (pre-treatment PET FDG). In many cases the variance increases as the intensity of a voxel's PET pre-radiotherapy increases, this is simply due to the fact the responding voxels will have a low post-treatment PET intensity, while the non-responding voxels will not. The aforementioned topographical scaling property, will allow to model this behavior. To obtain initial values for the location curves $a(\cdot)$ and $b(\cdot)$, we first use the `R` package `flexmix`, see Gruen et. al (2013), which allows us to fit defined parametric functions to the mixture. For the mixing proportion function we set a fixed constant value $\bar{\pi}(\mathbf{x}) = 0.4$. The bandwidths are computed according to the methodology described in Section 4.1, except that the groups are now determined as an output of the `flexmix` package. To stress the fact that the identification of the topographical model (2) his highly hazardous in the neighborhood of the design value 2.5 due to a component crossing (local non-identifiability), we plot in dashed line the behavior of our method over the interval $[2, 3]$ and will rule out this domain from the following discussion.

In Fig. 6(a), we show the PET imaging response to radiotherapy at 3 months, measured by FDG PET uptake, versus the pre-treatment FDG PET uptake and the fitted location functions of the two groups of voxels. For this canine patient, the fitted location curve $a(\mathbf{x})$ of the non-responding voxels increase with the pre-treatment FDG PET uptake, showing a positive relationship between the imaging response and the pre-treatment FDG PET. The location function $b(\mathbf{x})$ corresponding to the completely responding voxels, shows little variation across the range of values of pre-treatment FDG PET and remains relatively flat. This findings are in line with the results obtained by Bowen et al. (2012), however our model is able to capture more than the linear variation in the location curves. Our model also yields the mixing proportions function $\pi(\mathbf{x})$ that can be used to determine the optimal local radiation dose. As illustrated in Fig. 6(b), for this patient voxels

tend to be completely-responding when the pre-treatment FDG PET uptake is between 6.5 and 7.5 SUVs (Standardized Uptake Values), the proportion of non-responding voxels at that level decreases to 0.25. This suggests that the current radiation dose could be appropriate for voxels that have pre-treatment FDG PET uptake close to the range aforementioned. In figure 7, we show the estimator $\hat{f}_{\mathbf{x}}$ of $f_{\mathbf{x}}$, defined in (11), for different values of pre-treatment FDG PET uptake $\mathbf{x}$. We see that these conditional distributions are about zero-symmetric with reasonably small trimming effect due to $\mathbb{I}_{f_n(y|\mathbf{x}_0)\geq 0}$ in (11) (tiny wave effect on both sides of the main mode). This is a good model validation tool since we are actually able to recover, after local Fourier inversion, the basic symmetry assumption technically made on the distributions of the errors; see for quality comparison other existing (nonconditional) semiparametric inversion density estimates performed on real datasets: Fig. 1-2 (a) in Bordes et al. (2006), Fig. 3 in Butucea and Vandekerkhove (2013), Fig. 5 in Vandekerkhove (2013), or Fig. 2-3 in Bordes et al. (2013).

6 Auxiliary results and main proofs

Let us denote by $\|\cdot\|$ the Euclidean norm of a vector and by $\|\cdot\|_2$ the Frobenius norm of any squared matrix. Recall the definition of Z_k in (8) and let $J(t,u,h) := E[Z_1(t,u,h)]$. Let $\dot{Z}_k$ and $\dot{J}$ denote respectively the gradient of Z_k and J with respect to their first argument t.

Lemma 1 *Under assumption* **A1** *we have:*

i) For all $(u,h) \in \mathbb{R}\times\mathbb{R}_+^*$ *and any* $k = 1,...,n$,

$$\sup_{t\in\Theta}|Z_k(t,u,h)| \leq \frac{2}{1-2P}\frac{\|K\|_\infty}{h^d}, \quad \sup_{t\in\Theta}|J(t,u,h)| \leq \frac{2}{1-2P}\|\ell\|_\infty \cdot \int |K|.$$

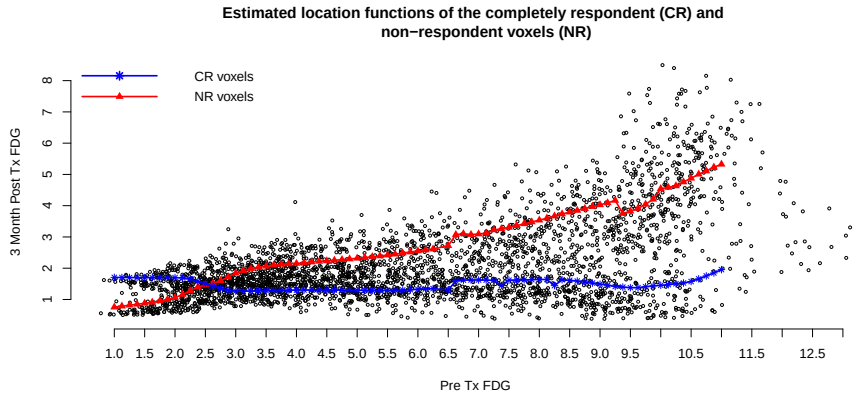
ii) For all $(u,h) \in \mathbb{R}\times\mathbb{R}_+^*$ *and any* $k = 1,...,n$,

$$\sup_{t\in\Theta}\|\dot{Z}_k(t,u,h)\| \leq \frac{4(1+|u|)}{(1-2P)^2}\frac{\|K\|_\infty}{h^d}, \quad \sup_{t\in\Theta}\|\dot{J}(t,u,h)\| \leq \frac{4(1+|u|)}{(1-2P)^2}\|\ell\|_\infty \cdot \int |K|.$$

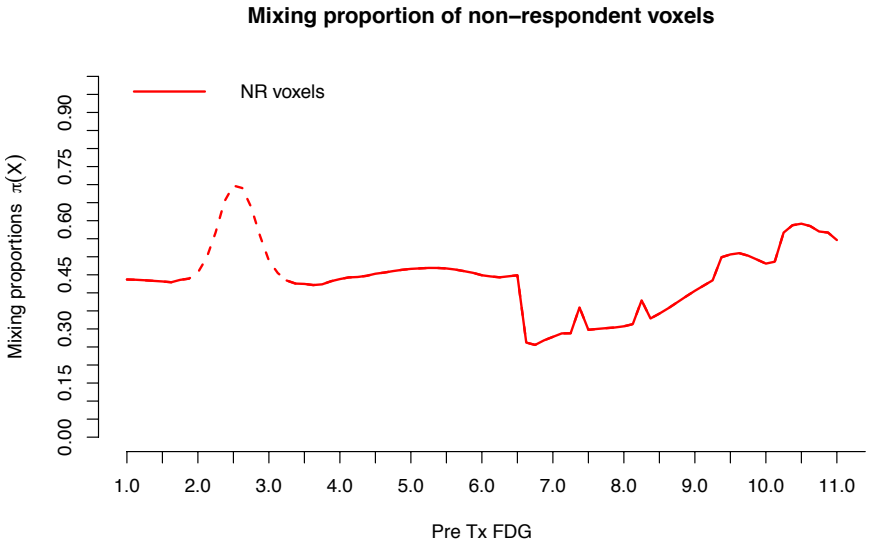
iii) For all $(u,h) \in \mathbb{R}\times\mathbb{R}_+^*$ *and any* $k = 1,...,n$,

$$\begin{aligned}\sup_{t\in\Theta}\|\ddot{Z}_k(t,u,h)\|_2 &\leq \frac{C(1+|u|+u^2)}{(1-2P)^3}\frac{\|K\|_\infty}{h^d},\\ \sup_{t\in\Theta}\|\ddot{J}_k(t,u,h)\|_2 &\leq \frac{C(1+|u|+u^2)}{(1-2P)^3}\|\ell\|_\infty \cdot \int |K|,\end{aligned}$$

for some constant $C > 0$.



(a) Scatter of plots of pre-treatment FDG PET vs. post-treatment FDG PET and estimated location functions for the completely respondent and non-respondent voxel subpopulations



(b) Estimated mixing proportions for the completely (CR) and non-respondent (NR) voxel subpopulation

Figure 6

Proof of Lemma 1. i) It is easy to see, from $1-2P \le |M(t,u)| \le 1$, that

$$|Z_k(t,u,h)| \le \frac{2}{|M(t,u)|} K_h(\mathbf{X}_k - \mathbf{x}_0) \le \frac{2}{(1-2P)} \frac{\|K\|_\infty}{h^d},$$

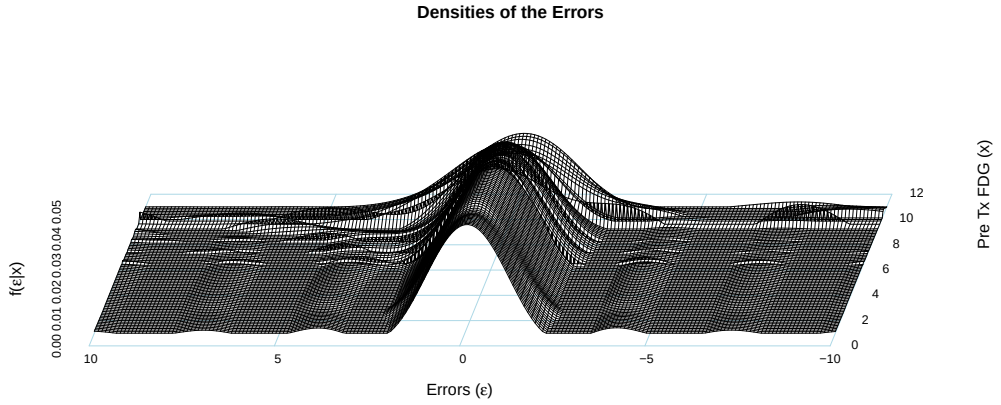


Figure 7: Density Estimates of the errors for the different levels of PET Tx FDG values

and that

$$|J(t,u)| \leq 2\left|\int \Im\left(\frac{g^*_{\mathbf{x}}(u)}{M(t,u)}\right) K_h(\mathbf{x}-\mathbf{x}_0)\ell(\mathbf{x})d\mathbf{x})\right| \leq \frac{2}{(1-2P)}\|\ell\|_\infty . \int |K|.$$

ii) We note that

$$\begin{aligned}\dot{Z}_k(t,u,h) &= -\left\{\frac{e^{iuY_k}}{M^2(t,u)}\begin{pmatrix} e^{iu\alpha}-e^{iu\beta} \\ iupe^{iu\alpha} \\ iu(1-p)e^{iu\beta}\end{pmatrix}\right. \\ &\quad \left. +\frac{e^{-iuY_k}}{M^2(t,-u)}\begin{pmatrix} e^{-iu\alpha}-e^{-iu\beta} \\ -iupe^{-iu\alpha} \\ -iu(1-p)e^{-iu\beta}\end{pmatrix}\right\} K_h(\mathbf{X}_k-\mathbf{x}_0),\end{aligned}$$

and that

$$\begin{aligned}E[\dot{Z}_k(t,u,h)] = \dot{J}_k(t,u,h) &= -\int\left\{\frac{g_{\mathbf{x}_0}(u)}{M^2(t,u)}\begin{pmatrix} e^{iu\alpha}-e^{iu\beta} \\ iupe^{iu\alpha} \\ iu(1-p)e^{iu\beta}\end{pmatrix}\right. \\ &\quad \left. +\frac{g_{\mathbf{x}_0}(-u)}{M^2(t,-u)}\begin{pmatrix} e^{-iu\alpha}-e^{-iu\beta} \\ -iupe^{-iu\alpha} \\ -iu(1-p)e^{-iu\beta}\end{pmatrix}\right\} K_h(\mathbf{x}-\mathbf{x}_0)\ell(\mathbf{x})d\mathbf{x}.\end{aligned}$$

We thus have

$$\begin{aligned}\|\dot{Z}_k(t,u,h)\| &= \left\| \frac{e^{iuY_k}}{M^2(t,u)}\dot{M}(t,u) + \frac{e^{-iuY_k}}{M^2(t,-u)}\dot{M}(t,-u) \right\| K_h(\mathbf{X}_k - \mathbf{x}_0) \\ &\leq \frac{1}{(1-2P)^2}\left(2\left(2^2 + p^2u^2 + (1-p)^2u^2\right)\right)^{1/2} K_h(\mathbf{X}_k - \mathbf{x}_0) \\ &\leq \frac{4(1+|u|)}{(1-2P)^2}\frac{\|K\|_\infty}{h^d},\end{aligned}$$

and

$$\begin{aligned}\|\dot{J}_k(t,u,h)\| &= \int \left\| \frac{g^*_{\mathbf{x}}(u)}{M^2(t,u)}\dot{M}(t,u) + \frac{g^*_{\mathbf{x}}(-u)}{M^2(t,-u)}\dot{M}(t,-u) \right\| K_h(\mathbf{X}_k - \mathbf{x}_0)\ell(\mathbf{x})d\mathbf{x} \\ &\leq \frac{1}{(1-2P)^2}\left(2\left(2^2 + p^2u^2 + (1-p)^2u^2\right)\right)^{1/2} \int |K_h(\mathbf{X}_k - \mathbf{x}_0)\ell(\mathbf{x})|d\mathbf{x} \\ &\leq \frac{4(1+|u|)}{(1-2P)^2}\|\ell\|_\infty . \int |K|.\end{aligned}$$

iii) Formula of $\ddot{M}(t,u)$ being tedious, we shortly write that

$$\begin{aligned}\ddot{Z}_k(t,u,h) &= \left\{ -\frac{e^{iuY_k}}{M^2(t,u)}\ddot{M}(t,u) + \frac{e^{-iuY_k}}{M^2(t,-u)}\ddot{M}(t,-u) \right. \\ &\left. +2\frac{e^{iuY_k}}{M^3(t,u)}\dot{M}(t,u)\dot{M}(t,u)^\top - 2\frac{e^{-iuY_k}}{\dot{M}^3(t,-u)}\dot{M}(t,-u)\dot{M}(t,-u)^\top \right\} K_h(\mathbf{X}_k - \mathbf{x}_0),\end{aligned}$$

and deduce our bound from the above expression using arguments similar to i) and ii). ■

Lemma 2 *i) For all $(t,t') \in \Theta^2$, there exists a constant $C_1 > 0$ such that*

$$|S_n(t) - S_n(t')| \leq C_1\|t - t'\| \sum_{j\neq k,j,k=1}^{n} \frac{K_h(\mathbf{X}_k - \mathbf{x}_0)K_h(\mathbf{X}_j - \mathbf{x}_0)}{n(n-1)}.$$

ii) For all $(t,t') \in \Theta^2$, there exists a constant $C_2 > 0$ such that

$$\|\ddot{S}_n(t) - \ddot{S}_n(t')\|_2 \leq C_2\|t - t'\| \sum_{j\neq k,j,k=1}^{n} \frac{K_h(\mathbf{X}_k - \mathbf{x}_0)K_h(\mathbf{X}_j - \mathbf{x}_0)}{n(n-1)}.$$

iii) There exists some constants C_1, $C_2 > 0$ depending on Θ, α, M, K such that

$$E\left[\left(\sum_{j\neq k,j,k=1}^{n} \frac{K_h(\mathbf{X}_k - \mathbf{x}_0)K_h(\mathbf{X}_j - \mathbf{x}_0)}{n(n-1)} - \ell^2(\mathbf{x}_0)\right)^2\right] \leq C_1h^{2\alpha} + \frac{C_2}{nh^d},$$

as $h \to 0$ and $nh^d \to \infty$.

Proof. i) By a first order Taylor expansion we have

$$\begin{aligned} & S_n(t) - S_n(t'))K_h(\mathbf{X}_j - \mathbf{x}_0) \\ = & -\frac{1}{2n(n-1)} \int (t-t')^\top \sum_{j\neq k,j,k=1}^{n} \dot{Z}_k(t_u,u,h) Z_j(t_u,u,h) dw(u), \end{aligned}$$

where for all $u \in \mathbb{R}$, t_u lies in the line segment with extremities t and t'. Therefore, according to calculations made in the proofs of Lemma 1 i) and ii), we obtain

$$|S_n(t) - S_n(t')| \leq \frac{4}{(1-2P)^3} \|t-t'\| \int_{\mathbb{R}} (1+|u|) w(u) du \left| \sum_{j\neq k,j,k=1}^{n} \frac{K_h(\mathbf{X}_k - \mathbf{x}_0) K_h(\mathbf{X}_j - \mathbf{x}_0)}{n(n-1)} \right|,$$

which ends the proof of i) by using assumption **A4**.

ii) Let recall first that

$$\ddot{S}_n(t) = \frac{-1}{2n(n-1)} \sum_{k\neq j} \int \left[\ddot{Z}_k(t,u,h) Z_j(t,u,h) + \dot{Z}_k(t,u,h) \dot{Z}_j(t,u)^\top \right] dw(u).$$

We shall bound from above as follows

$$\begin{aligned} \|\ddot{S}_n(t,u) - \ddot{S}_n(t',u)\|_2 \quad \leq \quad & \frac{1}{2n(n-1)} \sum_{k\neq j} \left\{ \left\| \int (\ddot{Z}_k(t,u,h) - \ddot{Z}_k(t',u,h)) Z_j(t,u) dw(u) \right\|_2 \right. \\ & + \left\| \int \ddot{Z}_k(t',u,h)(Z_j(t,u,h) - Z_j(t',u,h)) dw(u) \right\|_2 \\ & + \left\| \int \dot{Z}_k(t,u,h)(\dot{Z}_j(t,u,h) - \dot{Z}_j(t',u,h))^\top dw(u) \right\|_2 \\ & \left. + \left\| \int (\dot{Z}_k(t,u,h) - \dot{Z}_k(t',u,h)) \dot{Z}_j(t',u,h)^\top dw(u) \right\|_2 \right\}. \end{aligned}$$

For each term in the previous sum, we use Taylor expansion and upper-bounds similar to those developed in the proof of Lemma 1, and get

$$\begin{aligned} & \left\| \ddot{S}_n(t,u) - \ddot{S}_n(t',u) \right\|_2 \\ \leq \quad & \|t-t'\| \frac{C \int (1+|u|+u^2+|u|^3) dw(u)}{(1-2P)^5} \left| \sum_{j\neq k,j,k=1}^{n} \frac{K_h(\mathbf{X}_k - \mathbf{x}_0) K_h(\mathbf{X}_j - \mathbf{x}_0)}{n(n-1)} \right|, \end{aligned}$$

for some constant $C > 0$, which finishes the proof by using assumption **A4**.

iii) The proof is a consequence of Proposition 1 hereafter. ■

Proof of Proposition 1. We shall bound from above the mean square error by the usual decomposition into squared bias plus variance.

Note that

$$E[S_n(t)] = -\frac{1}{4}\int (E[Z_1(t,u,h)])^2 w(u)du$$

as $(Y_i, \mathbf{X}_i)$, $i = 1, ..., n$ are independent. Moreover,

$$\begin{aligned} E[Z_1(t,u,h)] &= \int\int \left(\frac{e^{iuy}}{M(t,u)} - \frac{e^{-iuy}}{M(t,-u)}\right) K_h(\mathbf{x}-\mathbf{x}_0) g(y,\mathbf{x}) dy d\mathbf{x} \\ &= \int \left(\int \left(\frac{e^{iuy}}{M(t,u)} - \frac{e^{-iuy}}{M(t,-u)}\right) g_{\mathbf{x}}(y) dy\right) \ell(\mathbf{x}) K_h(\mathbf{x}-\mathbf{x}_0) d\mathbf{x} \\ &= \int \left(\frac{g^*_{\mathbf{x}}(u)}{M(t,u)} - \frac{g^*_{\mathbf{x}}(-u)}{M(t,-u)}\right) \ell(\mathbf{x}) K_h(\mathbf{x}-\mathbf{x}_0) d\mathbf{x}. \end{aligned}$$

Let us denote by $L(\mathbf{x},t,u) := \frac{g^*_{\mathbf{x}}(u)}{M(t,u)} - \frac{g^*_{\mathbf{x}}(-u)}{M(t,-u)}$, which is further equal to

$$L(\mathbf{x},t,u) = 2i \cdot \Im\left(\frac{g^*_{\mathbf{x}}(u)}{M(t,u)}\right) = 2i \cdot \Im\left(\frac{M(\theta(\mathbf{x}),u)}{M(t,u)}\right) f^*_{\mathbf{x}}(u).$$

We can write $E[Z_1(t,u,h)] = [(L(\cdot,t,u)\ell) \star K_h](\mathbf{x}_0)$, where $\star$ denotes the convolution product. The bias of $S_n(t)$ is bounded from above as follows:

$$\begin{aligned} |E[S_n(t)] - S(t)| &= \frac{1}{4}\left|\int \left([(L(\cdot,t,u)\ell) \star K_h]^2(\mathbf{x}_0) - L^2(\mathbf{x}_0,t,u)\ell^2(\mathbf{x}_0)\right) w(u)du\right| \\ &\leq \frac{1}{4}\int |[(L(\cdot,t,u)\ell) \star K_h](\mathbf{x}_0) - L(\mathbf{x}_0,t,u)\ell(\mathbf{x}_0)| \\ &\quad \cdot |[(L(\cdot,t,u)\ell) \star K_h](\mathbf{x}_0) + L(\mathbf{x}_0,t,u)\ell(\mathbf{x}_0)|\, w(u)du. \end{aligned}$$

Now

$$|L(\mathbf{x}_0,t,u)\ell(\mathbf{x}_0)| \leq \frac{4\|\ell\|_\infty}{1-2P} \leq \frac{4C}{1-2P},$$

as $\|\ell\|_\infty$ is further bounded by a constant $C = C(\alpha, M)$ depending only on α, $M > 0$, uniformly over $\ell \in L(\alpha, M)$ (see remark following condition **A1**). We also have

$$\begin{aligned} E[Z_1(t,u,h)] = |[(L(\cdot,t,u)\ell) \star K_h](\mathbf{x}_0)| &\leq \int |L(\mathbf{x},t,u)| l(\mathbf{x}) |K|_h(\mathbf{x}-\mathbf{x}_0) d\mathbf{x} \\ &\leq \frac{4C}{1-2P}\int |K|. \end{aligned} \tag{16}$$

Moreover, for all $u \in \mathbb{R}$,

$$\begin{aligned} & |[(L(\cdot,t,u)\ell) \star K_h](\mathbf{x}_0) - L(\mathbf{x}_0,t,u)\ell(\mathbf{x}_0)| \\ \leq & \int |L(\mathbf{x}+\mathbf{x}_0,t,u)\ell(\mathbf{x}+\mathbf{x}_0) - L(\mathbf{x}_0,t,u)\ell(\mathbf{x}_0)| \cdot |K|_h(\mathbf{x}) d\mathbf{x} \\ \leq & c(|u| + \varphi(u)) \int \|\mathbf{x}\|^\alpha \cdot |K|_h(\mathbf{x}) d\mathbf{x} \leq c \cdot h^\alpha (|u| + \varphi(u)) \int \|\mathbf{x}\|^\alpha \cdot |K|(\mathbf{x}) d\mathbf{x}, \end{aligned}$$

under our assumptions **A1-A4**. Indeed, that implies that $L(\cdot,t,u)\ell(\cdot)$ is Lipschitz α-smooth for all $(t,u)\in\Theta\times\mathbb{R}$, with some constant $c>0$, see Lemma 3. Therefore we get

$$|E[S_n(t)]-S(t)|\leq\frac{4C(1+\int|K|)}{1-2P}\,c\left(\int\|\mathbf{x}\|^\alpha\cdot|K|(\mathbf{x})d\mathbf{x}\right)\cdot\left(\int|u|w(u)du\right)\cdot h^\alpha.$$

Similarly to $S_n(t)$ variance decomposition, we write

$$\begin{aligned}
& S_n(t)-E[S_n(t)]\\
= & \frac{-1}{4n(n-1)}\sum_{j\neq k}\left(\int(Z_j(t,u,h)Z_k(t,u,h)-E^2[Z_1(t,u,h)])w(u)du\right)\\
= & \frac{-1}{2n}\sum_j\int(Z_j(t,u,h)-E[Z_1(t,u,h)])E[Z_1(t,u,h)]w(u)du\\
& +\frac{-1}{4n(n-1)}\sum_{j\neq k}\left(\int(Z_j(t,u,h)-E[Z_1(t,u,h)])(Z_k(t,u,h)-E[Z_1(t,u,h)])w(u)du\right)\\
= & T_1+T_2,\ \text{say.}
\end{aligned}$$

Terms in T_1 and T_2 are uncorrelated and thus $Var(S_n(t))=Var(T_1)+Var(T_2)$.
On the one hand,

$$\begin{aligned}
Var(T_1) &= \frac{1}{4n}Var\left(\int(Z_1(t,u,h)-E[Z_1(t,u,h)])E[Z_1(t,u,h)]w(u)du\right)\\
&= \frac{1}{4n}E\left[\left(\int(Z_1(t,u,h)-E[Z_1(t,u,h)])E[Z_1(t,u,h)]w(u)du\right)^2\right]\\
&\leq \frac{1}{4n}E\left[\int(Z_1(t,u,h)-E[Z_1(t,u,h)])^2w(u)du\right]\int E^2[Z_1(t,u,h)]w(u)du,
\end{aligned}$$

according to Cauchy-Schwarz inequality. Now we use (16) and obtain

$$Var(T_2)\ \leq\ \frac{1}{4n}\left(\frac{4C\int|K|}{1-2P}\right)^2\int E[Z_1(t,u,h)^2]w(u)du.$$

We have,

$$\begin{aligned}
E[Z_1(t,u,h)^2] &= E\left[E\left[\left(2i\cdot\Im\left(\frac{e^{iuY}}{M(t,u)}\right)\right)^2\middle|\mathbf{X}\right](K_h(\mathbf{X}-\mathbf{x}_0))^2\right]\\
&= -4E\left[\left(\Im\left(\frac{g^*_{\mathbf{X}}(u)}{M(t,u)}\right)\right)^2(K_h(\mathbf{X}-\mathbf{x}_0))^2\right]\\
&\leq \frac{4}{(1-2P)^2}\int\frac{1}{h^{2d}}K^2\left(\frac{\mathbf{x}-\mathbf{x}_0}{h}\right)\ell(\mathbf{x})d\mathbf{x}\\
&\leq \frac{4C\int K^2}{(1-2P)^2h^d}.
\end{aligned}$$

Therefore,

$$Var(T_1) \leq \frac{16C^3(\int |K|)^2 \int K^2}{(1-2P)^4 nh^d}, \tag{17}$$

for all $t \in \Theta$, $h > 0$.

On the other hand,

$$\begin{aligned}
Var(T_2) &= \frac{1}{16n(n-1)} E\left[\left(\int (Z_1(t,u,h) - E[Z_1(t,u,h)])(Z_2(t,u,h) - E[Z_1(t,u,h)])w(u)du\right)^2\right] \\
&\leq \frac{1}{16n(n-1)} E\left[\int (Z_1(t,u,h) - E[Z_1(t,u,h)])^2 (Z_2(t,u,h) - E[Z_1(t,u,h)])^2 w(u)du\right] \\
&\leq \frac{1}{16n(n-1)} \int E^2[Z_1(t,u,h)^2] w(u)du \\
&\leq \frac{1}{16n(n-1)} \left(\frac{4C\int K^2}{(1-2P)^2 h^d}\right)^2 \\
&= \frac{C^2(\int K^2)^2}{n(n-1)(1-2P)^4 h^{2d}},
\end{aligned}$$

which is clearly a $o((nh^d)^{-1})$ and concludes the proof. ■

Lemma 3 (Smoothness of $L(\mathbf{x},t,u)\ell(\mathbf{x})$) *Assume* **A1**-**A4**. *There exists a constant $C > 0$, such that for all $(\mathbf{x},\mathbf{x}') \in \mathbb{R}^d \times \mathbb{R}^d$ and all $(t,u) \in \Theta \times \mathbb{R}$:*

$$|L(\mathbf{x},t,u)\ell(\mathbf{x}) - L(\mathbf{x}',t,u)\ell(\mathbf{x}')| \leq C(|u| + \varphi(u))\|\mathbf{x} - \mathbf{x}'\|^\alpha.$$

Proof. For $t = (\pi, a, b) \in \Theta$, and $(\mathbf{x}, u) \in \mathbb{R}^d \times \mathbb{R}$ we write

$$L(\mathbf{x},t,u)\ell(\mathbf{x}) = f_{\mathbf{x}_0}(u)\ell(\mathbf{x})\mathcal{T}(\mathbf{x},t,u), \text{ and } \mathcal{T}(\mathbf{x},t,u) := \frac{\sum_{i=1}^4 \mathcal{T}_i(\mathbf{x},t,u)}{1 - 2\pi(1-\pi)\cos[u(a-b)]}$$

where

$$\begin{aligned}
\mathcal{T}_1(\mathbf{x},t,u) &= \pi(\mathbf{x})\pi \sin[u(a(\mathbf{x}) - a)], \quad \mathcal{T}_2(\mathbf{x},t,u) = \pi(\mathbf{x})(1-\pi)\sin[u(a(\mathbf{x}) - b)], \\
\mathcal{T}_3(\mathbf{x},t,u) &= (1-\pi(\mathbf{x}))\pi \sin[u(b(\mathbf{x}) - a)], \quad \mathcal{T}_4(\mathbf{x},t,u) = (1 - \pi(\mathbf{x})(1-\pi)\sin[u(b(\mathbf{x}) - b)].
\end{aligned}$$

For all $(\mathbf{x},\mathbf{x}') \in \mathbb{R}^d \times \mathbb{R}^d$ we have

$$\begin{aligned}
&|L(\mathbf{x},t,u)\ell(\mathbf{x}) - L(\mathbf{x}',t,u)\ell(\mathbf{x}')| \\
&\leq |f_{\mathbf{x}_0}(u)\ell(\mathbf{x})||\mathcal{T}(\mathbf{x},t,u) - \mathcal{T}(\mathbf{x}',t,u)| + |\mathcal{T}(\mathbf{x},t,u)||f_{\mathbf{x}_0}(u)\ell(\mathbf{x}) - f^*_{\mathbf{x}'}(u)\ell(\mathbf{x}')| \\
&\leq \|\ell\|_\infty |\mathcal{T}(\mathbf{x},t,u) - \mathcal{T}(\mathbf{x}',t,u)| + (1-2P)^{-1}|f^*_{\mathbf{x}}(u)\ell(\mathbf{x}) - f^*_{\mathbf{x}'}(u)\ell(\mathbf{x}')|.
\end{aligned}$$

Let us now show the α-smooth Lipschitz property of $\mathcal{T}_1$, the proof for the other $\mathcal{T}_i$'s being completely similar. For all $(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^d \times \mathbb{R}^d$

$$\begin{aligned}
|\mathcal{T}_1(\mathbf{x}, t, u) - \mathcal{T}_1(\mathbf{x}', t, u)| &\leq |\sin[u(a(\mathbf{x}) - a)] - \sin[u(a(\mathbf{x}') - a)]| + |\pi(\mathbf{x}) - \pi(\mathbf{x}')| \\
&\leq |u||(a(\mathbf{x}) - a(\mathbf{x}')]| + |\pi(\mathbf{x}) - \pi(\mathbf{x}')| \\
&\leq M|u|\|\mathbf{x} - \mathbf{x}'\|^\alpha + M\|\mathbf{x} - \mathbf{x}'\|^\alpha.
\end{aligned}$$

On the other hand we have

$$\begin{aligned}
|f^*(u|\mathbf{x})\ell(\mathbf{x}) - f^*(u|\mathbf{x}')\ell(\mathbf{x}')| &\leq |\ell(\mathbf{x}) - \ell(\mathbf{x}')| + \|\ell\|_\infty |f_{\mathbf{x}_0}(u) - f^*_{\mathbf{x}'}(u)|, \\
&\leq (M + \|\ell\|_\infty \varphi(u))\|\mathbf{x} - \mathbf{x}'\|^\alpha,
\end{aligned}$$

which concludes the proof. ■

Proof of Theorem 2. Our method is based on a consistency proof for mininum contrast estimators by Dacunha-Castelle and Duflo (1993, pp.94–96). Let us consider a countable dense set D in Θ, then $\inf_{t\in\Theta} S_n(t) = \inf_{t\in D} S_n(t)$, is a measurable random variable. We define in addition the random variable

$$W(n, \xi) = \sup\left\{|S_n(t) - S_n(t')|;\ (t, t') \in D^2,\ \|t - t'\| \leq \xi\right\},$$

and recall that $S(\theta_0) = 0$. Let us consider a non-empty open ball B_* centered on θ_0 such that S is bounded from below by a positive real number 2ε on $\Theta\backslash B_*$. Let us consider a sequence $(\xi_p)_{p\geq 1}$ decreasing to zero, and take p such that there exists a covering of $\Theta\backslash B_*$ by a finite number κ of balls $(B_i)_{1\leq i\leq\kappa}$ with centers $t_i \in \Theta$, $i = 1, \ldots, \kappa$, and radius less than ξ_p. Then, for all $t \in B_i$, we have

$$S_n(t) \geq S_n(t_i) - |S_n(t) - S_n(t_i)| \geq S_n(t_i) - \sup_{t\in B_i} |S_n(t) - S_n(t_i)|,$$

which leads to

$$\inf_{t\in\Theta\backslash B_*} S_n(t) \geq \inf_{1\leq i\leq\kappa} S_n(t_i) - W(n, \xi_p).$$

As a consequence we have the following events inclusions

$$\begin{aligned}
\left\{\hat{\theta}_n \notin B_*\right\} &\subseteq \left\{\inf_{t\in\Theta\backslash B_*} S_n(t) < \inf_{t\in B_*} S_n(t) < S_n(\theta_0)\right\} \\
&\subseteq \left\{\inf_{1\leq i\leq\kappa} S_n(t_i) - W(n, \xi_p) < S_n(\theta_0)\right\} \\
&\subseteq \{W(n, \xi_p) > \varepsilon\} \cup \left\{\inf_{1\leq i\leq\kappa} (S_n(t_i) - S_n(\theta_0)) \leq \varepsilon\right\}.
\end{aligned}$$

In addition we have

$$P\left(\inf_{1\leq i\leq\kappa}(S_n(t_i)-S_n(\theta_0))\leq\varepsilon\right)$$
$$\leq 1-\prod_{i=1}^{\kappa}(1-[P(|S_n(t_i)-S(t_i)|\geq\varepsilon)+P(|S_n(\theta_0)-S(\theta_0)|\geq\varepsilon)]),$$

where, according to Proposition 1, the last two terms in the right hand side of the above inequality vanish to zero if $h^d n\to\infty$ and $h\to 0$ as $n\to\infty$. To conclude we use Lemma 2 and notice that, for all $(t,t')\in\Theta^2$, we have

$$\begin{aligned}
&|S_n(t)-S_n(t')|\\
&\leq\frac{C\|t-t'\|}{n(n-1)}\left|\sum_{j\neq k,j,k=1}^{n}K_h(\mathbf{X}_k-\mathbf{x}_0)K_h(\mathbf{X}_j-\mathbf{x}_0)\right|\\
&\leq C\|t-t'\|\ell^2(\mathbf{x}_0)+C\|t-t'\|\left|\sum_{j\neq k,j,k=1}^{n}\frac{K_h(\mathbf{X}_k-\mathbf{x}_0)K_h(\mathbf{X}_j-\mathbf{x}_0)}{n(n-1)}-\ell^2(\mathbf{x}_0)\right|. \quad (18)
\end{aligned}$$

We deduce from above that

$$\begin{aligned}
P(W(n,\xi_p)>\varepsilon) \;\leq\; & P\left(C\xi_p\ell^2(\mathbf{x}_0)>\frac{\varepsilon}{2}\right)\\
&+\left(\frac{2C\xi_p}{\varepsilon}\right)^2 E\left[\left(\sum_{j\neq k,j,k=1}^{n}\frac{K_h(\mathbf{X}_k-\mathbf{x}_0)K_h(\mathbf{X}_j-\mathbf{x}_0)}{n(n-1)}-\ell^2(\mathbf{x}_0)\right)^2\right],
\end{aligned}$$

where the last term in the right hand side is of order $(nh^d)^{-1}+h^{2\alpha}$ and tends to 0 by our assumption on h. Since for p sufficiently large we have $C\xi_p\ell^2(\mathbf{x}_0)<\varepsilon/2$ and thus $P\left(C\xi_p\ell^2(\mathbf{x}_0)>\varepsilon/2\right)=0$, this concludes the proof of the consistency in probability of $\hat{\theta}_n$ when $nh^d\to\infty$ and $h\to 0$ as $n\to\infty$. ■

Proof of Theorem 3. By a Taylor expansion of $\dot{S}_n$ around θ_0, we have

$$0=\dot{S}_n(\hat{\theta}_n)=\dot{S}_n(\theta_0)+\ddot{S}_n(\bar{\theta}_n)(\hat{\theta}_n-\theta_0),$$

where $\bar{\theta}_n$ lies in the line segment with extremities $\hat{\theta}_n$ and θ_0.

Let us study the behaviour of

$$\dot{S}_n(\theta_0)=\frac{-1}{2n(n-1)}\sum_{j\neq k}\int\dot{Z}_k(\theta_0,u,h)Z_j(\theta_0,u,h)w(u)du,$$

where $\dot{Z}_k$ denotes the gradient of Z_k with respect to the first argument. Recall that $\theta_0 = \theta(\mathbf{x}_0) = (\pi(\mathbf{x}_0), a(\mathbf{x}_0), b(\mathbf{x}_0))$ and therefore

$$J(t,u,h) = E[Z_1(t,u,h)] = 2i \int \Im\left(\frac{M(\theta(\mathbf{x}_0),u)}{M(t,u)}\right) f^*_{\mathbf{x}}(u)\ell(\mathbf{x})K_h(\mathbf{x}-\mathbf{x}_0)d\mathbf{x},$$

satisfies $J(\theta_0,u,h) \to 0$ as $h \to 0$. Indeed, the last integral may be equal to 0 if the set $\{\mathbf{x} : \theta(\mathbf{x}) = \theta(\mathbf{x}_0)\}$ has Lebesgue measure 0, or tends (by uniform continuity in $\mathbf{x}$ of the integrand) to

$$2i\Im\left(\frac{M(\theta(\mathbf{x}_0),u)}{M(\theta(\mathbf{x}_0),u)}\right) f^*_{\mathbf{x}_0}(u)\ell(\mathbf{x}_0) = 0.$$

Moreover,

$$\dot{Z}_k(t,u,h) = \Im\left(-\dot{M}(t,u)\frac{e^{iuY_k}}{M^2(t,u)}\right) K_h(\mathbf{X}_k - \mathbf{x}_0).$$

Denote $\dot{J}(t,u,h) = E[\dot{Z}_k(t,u,h)]$ and observe that

$$\dot{J}(t,u,h) = \int \Im\left(-\dot{M}(t,u)\frac{M(\theta(\mathbf{x}),u)f^*_{\mathbf{x}}(u)}{M^2(t,u)}\right) K_h(\mathbf{x}-\mathbf{x}_0)\ell(\mathbf{x})d\mathbf{x}.$$

Then, we decompose $\dot{S}_n(\theta_0)$ as follows

$$\begin{aligned}
&\dot{S}_n(\theta_0)\\
= \;& \frac{-1}{2n(n-1)}\sum_{j\neq k}\int \left(\dot{Z}_k(\theta_0,u,h) - \dot{J}(\theta_0,u,h)\right)\left(Z_j(\theta_0,u,h) - E[Z_j(\theta_0,u,h)]\right) w(u)du\\
& -\frac{1}{2n}\sum_{j=1}^{n}\int \dot{J}(\theta_0,u,h)(Z_j(\theta_0,u,h) - E[Z_j(\theta_0,u,h)])w(u)du\\
:= \;& -\frac{1}{2}(A_n(h) + B_n(h)), \qquad (19)
\end{aligned}$$

where terms in $A_n(h)$ and $B_n(h)$ are uncorrelated. On the one hand, we use a multivariate Central Limit Theorem for independent random variables taking values in a Hilbert space, following Kandelaki and Sozanov (1964) or Gikhman and Skorokhod (2004, Theorem 4, page 396). This will give us the limit behavior of the term

$$B_n(h) = \frac{1}{n}\sum_{j=1}^{n} U_j(h), \quad U_j(h) := \int \dot{J}(\theta_0,u,h)(Z_j(\theta_0,u,h) - E[Z_j(\theta_0,u,h)])w(u)du.$$

The random variables $U_j(h)$, $j = 1, ..., n$ are independent, centered, but their common law depend on n via h. Our goal is to show that

$$nh^d Var(B_n(h)) = \sum_{j=1}^{n} Var\left(\sqrt{\frac{h^d}{n}}U_j(h)\right) \to \Sigma, \quad \text{as } n \to \infty \qquad (20)$$

and that

$$\sum_{j=1}^{n} E\left[\left\|\sqrt{\frac{h^d}{n}}U_j(h)\right\|^4\right] = \frac{h^{2d}}{n}E[\|U_1(h)\|^4] \to 0, \quad \text{as } n \to \infty. \tag{21}$$

Indeed, (21) implies the Lindeberg's condition in Kandelaki and Sozanov (1964):

$$\sum_{j=1}^{n} E\left[\left\|\sqrt{\frac{h^d}{n}}U_j(h)\right|^2 \cdot \mathbb{I}_{\left\|\sqrt{h^d/n}U_j(h)\right\| \geq \varepsilon}\right] \to 0, \quad \text{as } n \to \infty, \text{ for any } \varepsilon > 0.$$

On the other hand, we prove that

$$\sqrt{nh^d}A_n(h) \to 0, \text{ in probability, as } n \to \infty, \tag{22}$$

stating that $\sqrt{nh^d}A_n(h)$ is a negligible term and that, as a consequence, the limiting behavior of $\sqrt{nh^d}\dot{S}_n(\theta_0)$ is only driven by $\sqrt{nh^d}B_n(h)$. This will end the proof of the theorem.

Let us prove (20) and (21). Note that $nh^d Var(B_n(h)) = h^d Var(U_1(h))$ and that

$$\begin{aligned} & Var(U_1(h)) \\ = & \int\int \dot{J}(\theta_0, u_1, h)\dot{J}^\top(\theta_0, u_2, h)Cov(Z_1(\theta_0, u_1, h), Z_1(\theta_0, u_2, h))w(u_1)w(u_2)du_1du_2. \end{aligned}$$

Similarly to Proposition 1, by uniform continuity in $\mathbf{x}$ of the integrand in $\dot{J}$, we get

$$\lim_{h\to 0} \dot{J}(\theta_0, u, h) = \dot{J}(\theta_0, u).$$

See that $\|\dot{J}(\theta_0, u)\| \leq 2(1+|u|)\|\ell\|_\infty/(1-2P)$ and that the latter upper bound is integrable with respect to the measure $w(u)du$ by assumption on w. It remains to study:

$$\begin{aligned} & Cov(Z_1(\theta_0, u_1, h), Z_1(\theta_0, u_2, h)) \\ = & E\left[Z_1(\theta_0, u_1, h)Z_1(\theta_0, u_2, h)\right] - E\left[Z_1(\theta_0, u_1, h)\right]E\left[Z_1(\theta_0, u_2, h)\right]. \end{aligned}$$

From (16) we deduce that

$$h^d|E\left[Z_1(\theta_0, u_1, h)\right]E\left[Z_1(\theta_0, u_2, h)\right]| \leq h^d\left(\frac{4C\int|K|}{1-2P}\right)^2 \to 0,$$

when $h \to 0$ as $n \to \infty$. We also have

$$\begin{aligned}
& h^d E\left[Z_1(\theta_0,u_1,h)Z_1(\theta_0,u_2,h)\right] \\
= & \int\int \left(\frac{e^{iu_1y}}{M(\theta_0,u_1)} - \frac{e^{-iu_1y}}{M(\theta_0,-u_1)}\right)\left(\frac{e^{iu_2y}}{M(\theta_0,u_2)} - \frac{e^{-iu_2y}}{M(\theta_0,-u_2)}\right)\frac{1}{h^d}K^2(\frac{\mathbf{x}-\mathbf{x}_0}{h})g(y,\mathbf{x})dyd\mathbf{x} \\
= & \int \left(\frac{e^{iu_1y}}{M(\theta_0,u_1)} - \frac{e^{-iu_1y}}{M(\theta_0,-u_1)}\right)\left(\frac{e^{iu_2y}}{M(\theta_0,u_2)} - \frac{e^{-iu_2y}}{M(\theta_0,-u_2)}\right)g(y,\mathbf{x}_0)dy(\int K^2)(1+o(1)) \\
= & \int \left(\frac{e^{iu_1y}}{M(\theta_0,u_1)} - \frac{e^{-iu_1y}}{M(\theta_0,-u_1)}\right)\left(\frac{e^{iu_2y}}{M(\theta_0,u_2)} - \frac{e^{-iu_2y}}{M(\theta_0,-u_2)}\right)g_{\mathbf{x}_0}(y)dy\cdot\ell(\mathbf{x}_0)(\int K^2)(1+o(1)),
\end{aligned}$$

as $h \to 0$. See also that we can write

$$\begin{aligned}
V(\theta_0,u_1,u_2) &:= \int \left(\frac{e^{iu_1y}}{M(\theta_0,u_1)} - \frac{e^{-iu_1y}}{M(\theta_0,-u_1)}\right)\left(\frac{e^{iu_2y}}{M(\theta_0,u_2)} - \frac{e^{-iu_2y}}{M(\theta_0,-u_2)}\right)g_{\mathbf{x}_0}(y)dy \\
&= \frac{M(\theta_0,u_1+u_2)}{M(\theta_0,u_1)M(\theta_0,u_2)}f_{\mathbf{x}_0}(u_1+u_2) - \frac{M(\theta_0,u_1-u_2)}{M(\theta_0,u_1)M(\theta_0,-u_2)}f_{\mathbf{x}_0}(u_1-u_2) \\
&\quad -\frac{M(\theta_0,-u_1+u_2)}{M(\theta_0,-u_1)M(\theta_0,u_2)}f_{\mathbf{x}_0}(-u_1+u_2) + \frac{M(\theta_0,-u_1-u_2)}{M(\theta_0,-u_1)M(\theta_0,-u_2)}f_{\mathbf{x}_0}(-u_1-u_2)
\end{aligned}$$

and this is a bounded function with respect to u_1 and u_2. Therefore

$$h^d Var(U_1(h)) \to \int\int \dot{J}(\theta_0,u_1)\dot{J}^\top(\theta_0,u_2)V(\theta_0,u_1,u_2)w(u_1)w(u_2)du_1du_2 =: \Sigma,$$

as $h \to 0$. This proves (20).

Now, denote by $v^{(k)}$ the k-th coordinate of a vector v and use Jensen inequality to see that

$$\begin{aligned}
E[\|U_1(h)\|^4] &\leq 3\left(E[(U_1^{(1)}(h))^4] + E[(U_1^{(2)}(h))^4] + E[(U_1^{(3)}(h))^4]\right) \\
&\leq 3\sum_{k=1}^{3} E\left[\left(\int \dot{J}^{(k)}(\theta_0,u,h)(Z_1(\theta_0,u,h) - E[Z_1(\theta_0,u,h)])w(u)du\right)^4\right] \\
&\leq 3\sum_{k=1}^{3}\int |\dot{J}^{(k)}(\theta_0,u,h)|^4 E\left[|Z_1(\theta_0,u,h)|^4\right]w(u)du.
\end{aligned}$$

We have $|\dot{J}^{(k)}(\theta_0,u,h)| \leq 4(1+|u|)(\int|K|)\|\ell\|_\infty/(1-2P)^2$ by Lemma 1 and

$$\begin{aligned}
E\left[|Z_1(\theta_0,u,h)|^4\right] &= \int\int 4\left|\Im\left(\frac{e^{iuy}}{M(\theta_0,u)}\right)\right|^4 \frac{1}{h^{4d}}K^4\left(\frac{\mathbf{x}-\mathbf{x}_0}{h}\right)g(y,\mathbf{x})dyd\mathbf{x} \\
&\leq \frac{4}{h^{3d}(1-2P)^4}\int \frac{1}{h^d}K^4\left(\frac{\mathbf{x}-\mathbf{x}_0}{h}\right)\ell(\mathbf{x})d\mathbf{x} \\
&\leq \frac{O(1)}{h^{3d}}\left(\int K^4\right)\|\ell\|_\infty,
\end{aligned}$$

as $h \to 0$. Therefore,

$$\frac{h^{2d}}{n} E[\|U_1(h)\|^4] \leq \frac{O(1)}{nh^d} \int |K| \cdot \int K^4 \cdot \int (1+|u|)^4 w(u) du = o(1),$$

as $n \to \infty$ and $h \to 0$ such that $nh^d \to \infty$. This proves (21).
To prove (22), we notice that $A_n(h)$ defined in (19) can be treated similarly to T_1 in (17). By this remark, we easily prove that $Var(A_n) = o\left((nh^d)^{-1}\right)$ which insure the wanted result.

Let us prove that

$$\ddot{S}_n(\theta_n) \longrightarrow \mathcal{I}(\theta_0), \text{ in probability, as } n \to \infty,$$

where $\mathcal{I} = \mathcal{I}(\theta_0) = -\frac{1}{2} \int \dot{J}(\theta_0, u) \dot{J}^\top(\theta_0, u) w(u) du$, and $\dot{J}(\theta_0, u)$ is defined in (12). We start by writing the triangular inequality

$$\|\ddot{S}_n(\theta_n) - \mathcal{I}\| \leq \|\ddot{S}_n(\theta_n) - \ddot{S}_n(\theta_0)\| + \|\ddot{S}_n(\theta_0) - E(\ddot{S}_n(\theta_0))\| + \|E(\ddot{S}_n(\theta_0)) - \mathcal{I}\|.$$

Then using upper bounds similar to (18) slighly adapted to $\ddot{S}_n$ instead of S_n and the convergence in probability of $\hat{\theta}_n$ towards θ_0 established in Theorem 2, we have that $\|\ddot{S}_n(\theta_n) - \ddot{S}_n(\theta_0)\| \to 0$ in probability as $n \to \infty$. By writting

$$E(\ddot{S}_n(\theta_0)) = -\frac{1}{2} \int \left(\ddot{J}(\theta_0, u, h) J(\theta_0, u, h) + \dot{J}(\theta_0, u, h) \dot{J}(\theta_0, u, h)^\top \right) w(u) du$$

and noticing, according to Bochner's Lemma, that $J(\theta_0, u, h) \to 0$ and $\dot{J}(\theta_0, u, h) \to \dot{J}(\theta_0, u)$ as $h \to 0$, we have, according to the Lebesgue's theorem, that $E[\ddot{S}_n(\theta_0)]$ tends to $\mathcal{I}$ as $h \to 0$. Finally we decompose $-2n(n-1)(\ddot{S}_n(\theta_0) - E[\ddot{S}_n(\theta_0)]) = \sum_{l=1}^{3} (D_{1,l} + D_{2,l})$ where

$$\begin{aligned}
D_{1,1} &= \sum_{k \neq j} \int (\ddot{Z}_k(\theta_0, u, h) - \ddot{J}(\theta_0, u, h))(Z_j(\theta, u, h) - J(\theta_0, u, h)) w(u) du \\
D_{1,2} &= (n-1) \sum_k \int (\ddot{Z}_k(\theta_0, u, h) - \ddot{J}(\theta_0, u, h)) J(\theta_0, u, h) w(u) du \\
D_{1,3} &= (n-1) \sum_j \int \ddot{J}(\theta_0, u, h)(Z_j(\theta, u, h) - J(\theta_0, u, h)) w(u) du,
\end{aligned}$$

and

$$
\begin{aligned}
D_{2,1} &= \sum_{k\neq j}\int (\dot{Z}_k(\theta_0,u,h)-\dot{J}(\theta_0,u,h))(\dot{Z}_j(\theta,u,h)-\dot{J}(\theta_0,u,h))^\top w(u)du \\
D_{2,2} &= (n-1)\sum_{k}\int (\dot{Z}_k(\theta_0,u,h)-\dot{J}(\theta_0,u,h))J(\theta_0,u,h)^\top w(u)du \\
D_{2,3} &= (n-1)\sum_{j}\int \dot{J}(\theta_0,u,h)(Z_j(\theta,u,h)-J(\theta_0,u,h))^\top w(u)du.
\end{aligned}
$$

Noticing that terms $D_{i,3}$, $i=1,2$, respectively $D_{i,j,}$, $i=1,2$ and $j=2,3$, can be treated as T_1 respectively T_2 in the proof of Proposition 1, we obtain

$$Var\left(\ddot{S}_n(\theta_0)\right)=O\left(\frac{1}{nh^d}\right),$$

which concludes the proof. ■

Aknowledgements. The authors thank warmly Dr.'s Bowen and Chappell for providing the Positron Emission Tomography dataset presented in Bowen et al. (2012, Fig. 4), as well as Dr. Wang for sharing the EM-type algorithm code developed in Huang et al. (2013).

References

[1] Anderson, J. A.. (1979). Multivariate logistic compounds. *Biometrika*, 17–26.

[2] Bordes, L., Kojadinovic, I. and Vandekerkhove, P. (2013) Semiparametric estimation of a two-component mixture of linear regressions in which one component is known. *Electr. J. Statist.*, 2603-2644.

[3] Bordes, L., Mottelet, S. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Statist.* **34** 1204–1232.

[4] Bowen, R. S., Chappell R. J., Bentzen S. M., Deveau, M. A., Forrest L. J., and Jeraj, R. (2012). Spatially resolved regression analysis of pre-treatment FDG, FLT and Cu-ATSM PET from post-treatment FDG PET: an exploratory study. *Radiother. Oncol.* **105**, 41–48.

[5] Brunel E., Comte F. and Lacour, C. (2010) Minimax estimation of the conditional cumulative distribution function under random censorship. *Sankhya Series A*, **72**, 293-330.

[6] Butucea, C. and Vandekerkhove, P. (2013). Semiparametric mixtures of symmetric distributions. *Scand. J. Statist.*, In press.

[7] Cohen, S. and Le Pennec, E. (2012). Conditional Density Estimation by Penalized Likelihood Model Selection and Applications. URL `http://arxiv.org/abs/1103.2021`.

[8] Dacunha-Castelle, D. and Duflo, M. (1983). *Probabilités et Statistique 2. Problèmes à temps mobile.* Masson, Paris.

[9] De Veaux, R. D. (1989). Mixtures of linear regressions. *Comput. Statist. Data Analyis*, **8**, 227–245.

[10] Gikhman, I. and Skorokhod, A. (2004). *The theory of stochastic processes. I* Springer-Verlag, Berlin.

[11] Gruen, B., Leisch, F., and Sarkar, D. (2013) *flexmix: Flexible Mixture Modeling.* URL http://CRAN.R-project.org/package=flexmix. R package version 2.3-11.

[12] Grün, B. and Leisch, F. (2006) *Fitting finite mixtures of linear regression models with varying and fixed effects in R.* In A. Rizzi and M. Vichi, editors, Compstat 2006, Proceedings in Computational Statistics, 853–860.

[13] Hall, P., and Zhou, X-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.* **31**, 201–224.

[14] Hawkins, D. S., Allen, D. M. and Stomber, A. J. (2001). Determining the number of components in mixtures of linear models. *Computational Statistics and Data Analysis*, **38**, 15–48.

[15] Herrmann E. (2013). *lokern: Kernel Regression Smoothing with Local or Global Plug-in Bandwidth,* 2013. URL http://CRAN.R-project.org/package=lokern. R package version 1.1-4.

[16] Huang, M., Li, R. and Wang, S. (2013). Nonparametric mixture of regression models. *J. Amer. Statist. Soc.* **108**, 229–241.

[17] Huang, M. and Yao, W. (2012). Mixture of Regression Models with Varying Mixing Proportions: A Semiparametric Approach. *J. Amer. Statist. Assoc.* **107**, 711-724.

[18] Hunter, D. R. and Young, D. S. (2012) Semiparametric mixtures of regressions. *J. Nonparam. Statist.* **24**, 19-38.

[19] Hunter, D. R., Wang, S. and Hettmanspeger, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.* **35** 224–251.

[20] Hurn, M., Justel, A. and Robert, C. P. (2003). Estimating mixtures of regressions. *J. Comput. Graph. Statist.* **12**, 1–25.

[21] Ibragimov, I. A. and Has'minski, R. Z. (1981). *Statistical estimation. Asymptotic theory.* Applications of Mathematics. Springer-Verlag, New York-Berlin.

[22] Jones, P. N. and McLachlan, G. J. (1992). Fitting finite mixture models in a regression context. *Australian J. Statist.* **34**, 233–240.

[23] Kandelaki, N. P., and Sozanov, V. V. (1964). On a central limit theorem for random elements with values in Hilbert space. *Theory Probab. Appl.* **71** 38–46.

[24] Montuelle, L., Le Pennec, E., and Cohen, S. (2013). Gaussian Mixture Regression model with logistic weights, a penalized maximum likelihood approach. URL `http://arxiv.org/pdf/1304.2696v1.pdf`.

[25] Leung, D. H-Y., and Qin, J. (2006). Semi-parametric inference in a bivariate (multivariate) mixture model. *Statistica Sinica*, **16**, 153–163.

[26] Quandt, R. and Ramsey, J. (1978). Estimating mixtures of normal distributions and switching regression. *J. Amer. Statist. Assoc.* **73**, 730–738.

[27] N. Städler, N., Bühlmann, P. — and van de Geer, S. (2010). ℓ_1-penalization for mixture of regression models. *Test*, **19**, 209–256.

[28] Stone, C. J. (1977) Consistent nonparametric regression. With discussion and a reply by the author. *Ann. Statist.* **5**, 595645.

[29] Toshiya, H. (2013). Mixture regression for observational data, with application to functional regression models. URL `http://arxiv.org/abs/1307.0170`.

[30] Tsybakov, A. B. (2009) *Introduction to nonparametric estimation.* Springer Series in Statistics. Springer, New York.

[31] Turner, R. (2000). Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions. *Applied Statistics.* **49**, 371–384.

[32] Turner, R. (2011). Mixreg: Functions to fit mixtures of regressions. http://CRAN.R-project.org/package=mixreg. R package version 0.0-4.

[33] Vandekerkhove, P. (2013). Estimation of a semiparametric mixture of regressions model. *J. Nonparam. Statist.*, 25, 181-208.

[34] Zhu, H. and Zhang, H. (2004). Hypothesis testing in mixture regression models. *J. Roy. Statist. Soc. Ser. B*, **66**, 3–16.