Electronic Journal of Statistics
Vol. 17 (2023) 4011–4048
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2186

Adaptive warped kernel estimation for nonparametric regression with circular responses*

Tien Dat Nguyen[1], **Thanh Mai Pham Ngoc**[2], **and Vincent Rivoirard**[3]

[1]*Faculty of Mathematics and Computer Science, University of Science, Vietnam National University Ho Chi Minh city, Vietnam*
e-mail: ndat@hcmus.edu.vn

[2]*LAGA, Institut Galilée, CNRS, UMR 7539, Université Sorbonne Paris Nord, 93430, Villetaneuse, France*
e-mail: phamngoc@math.univ-paris13.fr

[3]*CEREMADE, CNRS, Université Paris-Dauphine, Université PSL, 75016 Paris, France*
e-mail: Vincent.Rivoirard@dauphine.fr

Abstract: In this paper, we deal with nonparametric regression for circular data, meaning that observations are represented by points lying on the unit circle. We propose a kernel estimation procedure with data-driven selection of the bandwidth parameter. For this purpose, we use a warping strategy combined with a Goldenshluger-Lepski type estimator. To study optimality of our methodology, we consider the minimax setting and prove, by establishing upper and lower bounds, that our procedure is nearly optimal on anisotropic Hölder classes of functions for pointwise estimation. The obtained rates also reveal the specific nature of regression for circular responses. Finally, a numerical study is conducted, illustrating the good performances of our approach.

MSC2020 subject classifications: Primary 62G08, 62H11.
Keywords and phrases: Circular data, nonparametric regression, warping method, kernel rule, adaptive minimax estimation, Goldenshluger-Lepski procedure.

Received April 2022.

Contents

*T.D.N. was supported by a public grant as part of the Investissement d'avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH.

1. Introduction

Directional statistics is the branch of statistics which deals with observations that are directions. In this paper, we will consider more specifically *circular data* which arises whenever using a periodic scale to measure observations. These data are represented by points lying on the unit circle of $\mathbb{R}^2$ denoted in the sequel by $\mathbb{S}^1$. Circular data are collected in many research fields, for example in ecology (animal orientations), earth sciences (wind, ocean current directions, cross-bed orientations to name a few), medicine (circadian rhythm), forensics (crime incidence) or social science (clock or calendar effects). Various comprehensive surveys on statistical methods for circular data can be found in Mardia and Jupp [19], Jammalamadaka and SenGupta [13], Ley and Verdebout [18] and recent advances are collected in Pewsey and García-Portugués [23]. Note that the term *circular data* is also used to distinguish them from data supported on the real line $\mathbb{R}$ (or some subset of it), which henceforth are referred to as *linear data*.

In the present work, we focus on a nonparametric regression model with a circular response and linear predictor. A real example of such a statistical problem is given for instance by data provided by Table 1 of Fisher and Lee [9] and which contains distances and directions moved by small blue periwinkles after relocation. The objective is to predict the angles given the distance moved. To model such a problem, it is fundamental to take into account the circular geometry of the directions taken by the snails. From a formal point of view, we then assume that we have an independent identically distributed (i.i.d. in the sequel) sample $\{(X_i, \Theta_i)\}_{i=1}^n$ distributed as (X, Θ), where Θ is a circular random variable, i.e. $\Theta \in \mathbb{S}^1$, and X is a random variable with density f_X supported on $\mathbb{R}$. We assume that the cumulative distribution function of X, denoted F_X, is known. We also assume that F_X is invertible on $\mathbb{R}$ meaning that f_X is positive on $\mathbb{R}$ and $F_X(\mathbb{R}) = (0,1)$. We aim at estimating a function m which contains the dependence structure between the predictors X_i and the observations Θ_i. For our setting, described in Section 2.1, the regression function m is derived in Equation (3). Note that in practice, F_X, which is usually unknown, is replaced by its empirical counterpart. See Remark 3.7 for more details about such a plug-in.

Because of the circular response, it is natural to wonder what statistical challenges are posed by the circular regression problem compared to classical regression with a real response which has been widely studied in the literature. At first, to what extent does circular regression require suitable methodologies adapted to the underlying geometry? Secondly, in order to measure the difficulty of this specific statistical problem, can we derive associated rates of convergence in the minimax setting and can we propose optimal estimators? These devised estimators should be data driven to be implemented. Accordingly, we strive to respond to these various statistical difficulties in this work: in a minimax point of view, we propose an original nearly optimal data-driven kernel estimator adapted to the circular geometry based on the Goldenshluger-Lepski bandwidth selection rule.

Regression with circular response and linear covariates has been first and mostly explored from a parametric point of view. Pioneered contributions are due to Gould [12], Johnson and Wehlry [14] or Fisher and Lee [9]. The latter proposed the most popular link-based function (namely the function $2\arctan$) to model the conditional mean. Major difficulties, among others of such link-based models involve computational drawbacks to estimate parameters as identified by Presnell et al. [26]. Presnell et al. [26] in turn suggested alternatively a spherically projected multivariate linear model. Since then, numerous parametric approaches have been proposed, we refer the reader to all the references in Pewsey and García-Portugués [23]. In order to get a more flexible approach, nonparametric paradigm has been considered, first in the pioneering work by Di Marzio et al. [20] and more recently in Meilán-Vila et al. [22] for the multivariate setting. Surprisingly enough, the nonparametric point of view has only been considered in very few papers. Note that contrary to all works aforementioned which classically focus on the conditional mean (which is our goal as well) Alonso-Pena and Crujeiras [1] proposed a nonparametric multimodal regression method for estimating the conditional density when for instance the latter is highly skewed or multimodal. Estimation procedures developed in [20] or [22] consist in estimating the arctangent function of the ratio of the trigonometric moments of Θ (more details about this approach are given in the next section as it is the starting point of our procedure). More precisely, in the case of pointwise estimation and covariates supported on $[0,1]$, Di Marzio et al. [20] investigated the performances of a Nadaraya-Watson and a local linear polynomial estimators. Theoretically, for regression functions being twice continuously differentiable, they obtained expressions for asymptotic bias and variance. Their proofs are based on linearization of the function arctangent by using Taylor expansions, but no sharp controls of the remainder terms in the expansions are obtained. Actually obtaining such controls would be very tedious with such an approach based on Taylor expansions. As for the more recent work of Meilán-Vila et al. [22], they studied the multivariate setting $[0,1]^d$ with the same estimators and proofs technics. In both papers, neither rates of convergence nor adaptation are obtained and cross-validation is used to select the kernel bandwidth in practice. By adaptation, we mean that the estimators do not require the specification of the regularity of the regression function which is crucial from a practical point of view. In view of this, we were motivated to fill the gap in the literature. Our goal is twofold: obtaining optimal rates of convergence for predictors supported on $\mathbb{R}$ and adaptation for estimating m the regression function. To achieve this, we propose a new strategy based on concentration inequalities along with warping methods.

Our contributions Under the assumption that the cumulative distribution function (c.d.f.) of the design X is known and invertible, warping methods used in this paper consist in introducing the auxiliary function $g := m \circ F_X^{<-1>}$, with $F_X^{<-1>}$ the inverse of F_X. We then use classical kernel rules to estimate the function g in the specific framework of circular data. Our procedure needs to select two bandwidths. Fully data-driven selection of bandwidths is performed by using a Goldenshluger-Lepski type procedure [11]. Then, theoretical performances are studied. We consider the minimax setting and prove by establishing upper and lower bounds that our procedure is nearly optimal on anisotropic Hölder classes of functions for pointwise estimation. These results are stated in Theorems 3.6 and 3.11 respectively. Then, we conduct a numerical study whose goal is twofold. We first investigate the best tuning parameters of our procedure. Once tuned, our estimates are used on artificial data and compared to other classical methods. The numerical study reveals the good performances of our methodology.

Plan In section 2, we explain how to take into account the circular nature of the response and then propose our data-driven kernel estimator of the regression function m based on warping strategy and the Goldenshluger-Lepski bandwidth selection rule. Section 3 contains the theoretical results. Section 4 presents numerical results including simulations. Finally, all the proofs are deferred to Section 5.

Notations It is necessary to equip the reader with some notations. In the sequel, a point on $\mathbb{S}^1$ will not be represented as a two-dimensional vector $\mathbf{w} = (w_2, w_1)^\top$ with unit Euclidean norm but as an angle $\theta = atan2(w_1, w_2)$ defined as follows:

Definition 1.1. The function atan2 : $\mathbb{R}^2 \setminus (0,0) \mapsto [-\pi;\pi]$ is defined for any $(w_1, w_2) \in \mathbb{R}^2 \setminus (0,0)$ by

$$\operatorname{atan2}(w_1, w_2) := \begin{cases} \arctan\left(\frac{w_1}{w_2}\right) & \text{if} \quad w_2 \geq 0, w_1 \neq 0 \\ 0 & \text{if} \quad w_2 > 0, w_1 = 0 \\ \arctan\left(\frac{w_1}{w_2}\right) + \pi & \text{if} \quad w_2 < 0, w_1 > 0 \\ \arctan\left(\frac{w_1}{w_2}\right) - \pi & \text{if} \quad w_2 < 0, w_1 \leq 0, \end{cases}$$

with arctan taking values in $[-\pi/2, \pi/2]$. In particular, for $w_1 > 0$, $\operatorname{atan2}(w_1, 0) = \arctan(+\infty) = \pi/2$ and $\operatorname{atan2}(-w_1, 0) = \arctan(-\infty) = -\pi/2$.

In this definition, one has arbitrarily fixed the origin of $\mathbb{S}^1$ at $(1, 0)^\top$ and uses the anti-clockwise direction as positive. Thus, a circular random variable can be represented as angle over $[-\pi, \pi)$. Observe that $\operatorname{atan2}(0, 0)$ is not defined. Hereafter, $\|\cdot\|_{\mathbb{L}^1(\mathbb{R})}$ and $\|\cdot\|_{\mathbb{L}^2(\mathbb{R})}$ respectively denote the $\mathbb{L}^1$ and $\mathbb{L}^2$ norm on $\mathbb{R}$ with respect to the Lebesgue measure:

$$\|f\|_{\mathbb{L}^1(\mathbb{R})} = \int_{\mathbb{R}} |f(y)| dy, \quad \|f\|_{\mathbb{L}^2(\mathbb{R})} = \Big(\int_{\mathbb{R}} |f(y)|^2 dy \Big)^{1/2}.$$

The $\mathbb{L}^\infty$ norm is defined by $\|f\|_\infty = \sup_{y \in \mathbb{R}} |f(y)|$. Moreover, we denote $*$ the classical convolution product defined for functions f, g by $f * g(x) := \int_{\mathbb{R}} f(x - y).g(y)dy$, for $x \in \mathbb{R}$. Finally, for $\alpha \in \mathbb{R}$, $[\alpha]_+ := \max\{\alpha; 0\}$, and for $\beta > 0$, $\lfloor \beta \rfloor$ denotes the largest integer strictly smaller than β.

2. The estimation procedure

After recalling the framework of circular data in Section 2.1, Section 2.2 is devoted to the construction of an estimator for $m(x)$, at a given point $x \in \mathbb{R}$ which will be fixed along this paper, using warped kernel methods. Then, Section 2.3 presents a data-driven procedure for bandwidth selection by using the Goldenshluger-Lepski methodology.

2.1. The framework of circular data

There is no doubt that, due to their periodic nature, circular data are fundamentally different from linear ones, and thus need specific tools. To measure the closeness between two angles θ_1 and θ_2, we do not consider the natural distance

$$d(\theta_1, \theta_2) := \min\big\{ \big|\theta_1 - \theta_2 + 2k\pi\big| : k \in \mathbb{Z} \big\}, \quad \theta_1, \theta_2 \in [-\pi, \pi),$$

but we focus on d_c with

$$d_c(\theta_1, \theta_2) := 1 - \cos(\theta_1 - \theta_2), \quad \theta_1, \theta_2 \in [-\pi, \pi),$$

which is extensively used in the literature of directional statistics (see for instance Section 2 in the seminal monograph by Mardia and Jupp [19], Section 3.2.1 of [18], [20] or [22]). Note that the divergence d_c corresponds to the usual squared Euclidean norm in $\mathbb{R}^2$. Indeed, the angles θ_1 and θ_2 determine the corresponding points $(\cos\theta_1, \sin\theta_1)$ and $(\cos\theta_2, \sin\theta_2)$ respectively on the unit circle $\mathbb{S}^1$. Then, the usual squared Euclidean norm in $\mathbb{R}^2$ reads

$$\big(\cos\theta_1 - \cos\theta_2\big)^2 + \big(\sin\theta_1 - \sin\theta_2\big)^2 = 2.\big[1 - \cos(\theta_1 - \theta_2)\big] = 2.d_c(\theta_1, \theta_2).$$

Hence, $\sqrt{d_c}$ is a distance on $[-\pi, \pi)$ and we naturally look for a measurable function m such that:

$$\mathbb{E}\big[d_c(\Theta, m(X))\big] = \min_{f\colon \mathbb{R}\to[-\pi,\pi)} \mathbb{E}\big[d_c(\Theta, f(X))\big], \tag{1}$$

where the minimum is taken over $[-\pi, \pi)$-valued functions f that are measurable with respect to the σ-algebra generated by X. It is interesting to notice that the minimization problem (1) is directly linked to the definition of the Frechet mean on the circle (see Charlier [6]). Furthermore, in the literature of directional statistics, the problem of finding such a regression function $m(X)$ as defined in (1) has been already considered to solve the circular regression problem (see [10] and [22]).

Now let us work conditionally to X. For $x \in \mathbb{R}$ let

$$m_1(x) := \mathbb{E}\big(\sin(\Theta)|X = x\big) \quad \text{and} \quad m_2(x) := \mathbb{E}\big(\cos(\Theta)|X = x\big). \tag{2}$$

Moreover, write for an arbitrary function $f : \mathbb{R} \to [-\pi, \pi)$

$$\begin{aligned}\mathbb{E}\big[\cos(\Theta - f(X))|X\big] &= \cos(f(X)).m_2(X) + \sin(f(X)).m_1(X)\\ &= \sqrt{(m_2(X))^2 + (m_1(X))^2}.\cos(f(X) - \gamma(X)),\end{aligned}$$

where $\gamma : \mathbb{R} \to [-\pi, \pi)$ is defined for $x \in \mathbb{R}$ by

$$\cos(\gamma(x)) := \frac{m_2(x)}{\sqrt{(m_2(x))^2 + (m_1(x))^2}}, \text{ and } \sin(\gamma(x)) := \frac{m_1(x)}{\sqrt{(m_2(x))^2 + (m_1(x))^2}}.$$

Observe that

$$\gamma(x) = \operatorname{atan2}(m_1(x), m_2(x)).$$

Thus, we have

$$\begin{aligned}&\min_{f\colon \mathbb{R}\to[-\pi,\pi)} \mathbb{E}\big[d_c(\Theta, f(X))\big]\\ &= 1 - \max_{f\colon \mathbb{R}\to[-\pi,\pi)} \mathbb{E}\Big[\mathbb{E}\big[\cos(\Theta - f(X))|X\big]\Big]\\ &= 1 - \max_{f\colon \mathbb{R}\to[-\pi,\pi)} \mathbb{E}\Big[\sqrt{m_1^2(X) + m_2^2(X)}\cos(f(X) - \gamma(X))\Big].\end{aligned}$$

Finally the minimizer of the minimization problem (1) is achieved for

$$f(x) = \gamma(x) = \operatorname{atan2}\big(m_1(x), m_2(x)\big).$$

In conclusion, the circular nature of the response is taken into account by the arctangent of the ratio of the conditional expectation of sine and cosine components of Θ given X and we tackle the problem by estimating the function

$$m(x) = \operatorname{atan2}(m_1(x), m_2(x)), \quad x \in \mathbb{R}, \tag{3}$$

with m_1 and m_2 defined in (2).

Remark 2.1. Observe that if $m_1(x) = m_2(x) = 0$, then $m(x)$ is not defined. This occurs if and only if

$$\phi_1(f(\cdot|x)) := \int_{-\pi}^{\pi} e^{i\theta} f(\theta|x)d\theta = 0,$$

where $f(\cdot|x)$ denotes the conditional density of $\Theta|X = x$. Note that $\phi_1(f(\cdot|x))$ plays a specific role in the literature of directional statistics. See for instance Section 3.4.2 of [19].

In the sequel, we estimate the circular regression function m as defined in (3) under the condition

$$\phi_1(f(\cdot|x)) \neq 0. \tag{4}$$

We set $\zeta := (\zeta_i)_{i=1,\dots,n}$ the vector of errors so that

$$\Theta_i = m(X_i) + \zeta_i \pmod{2\pi}, \quad i = 1, \dots, n. \tag{5}$$

Our estimation methodology is based on a warping strategy.

2.2. Warping strategy

The popular Nadaraya-Watson (NW) methodology provides a natural estimator of m of the form

$$\widehat{m}_h^{NW} : x \longmapsto \frac{\frac{1}{n}\sum_{j=1}^n \Theta_j.K_h(x - X_j)}{\frac{1}{n}\sum_{j=1}^n K_h(x - X_j)},$$

with $K : \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} K(y)dy = 1$ and $K_h(\cdot) := \frac{1}{h}K(\frac{\cdot}{h})$, for some bandwidth $h > 0$. However, on the one hand, the denominator which can be small may lead to some instability. On the other hand, as adaptive estimation requires the data-driven selection of the bandwidth, the ratio form of the NW estimate indicates that we should select two bandwidths: one for the numerator and one for the denominator. Consequently, considering NW estimators for m_1 and m_2 involve four bandwidths. This makes the study of these estimators quite intricate.

Recalling that $g = m \circ F_X^{<-1>}$ with $F_X^{<-1>}$ the inverse of F_X, warping methods then boil down to first estimating g by say $\widehat{g}$ and then estimating the regression function of interest m by $\widehat{g} \circ F_X$. To deal with regression with random design, the warping strategy has been applied for instance by Kerkyacharian and Picard [15], Pham Ngoc [24], Chagny [3] and Chagny et al. [5]. Among the advantages of this method, let us mention that a warped kernel estimator does not involve a ratio, which strengthens its stability whatever the design distribution, even when the design is inhomogeneous. In our framework, in order to construct an estimator for the regression function m, we first estimate m_1 and m_2 (see (3)). Consequently, we introduce two auxiliary functions $g_1, g_2 : (0, 1) \longmapsto \mathbb{R}$ defined by

$$g_1 := m_1 \circ F_X^{<-1>}, \quad \text{and} \quad g_2 := m_2 \circ F_X^{<-1>},$$

so that $m_1 = g_1 \circ F_X$ and $m_2 = g_2 \circ F_X$; we then have for $u \in (0,1)$

$$g(u) = \operatorname{atan2}\big(g_1(u), g_2(u)\big).$$

Our fully data-driven approach is based on the selection of two bandwidths that adapt automatically to the unknown smoothness of functions g_1 and g_2.

Now, we propose to adapt the strategy developed in the linear case by Chagny et al. in [5]. The warping device is based on the transformation $F_X(X_i)$ of the data X_i, $i = 1, \dots, n$. We first define kernels considered in our framework as follows.

Definition 2.2. Let $K : \mathbb{R} \to \mathbb{R}$ be an integrable function such that K is compactly supported, $K \in \mathbb{L}^\infty(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. We say that K is a kernel if it satisfies $\int_{\mathbb{R}} K(y)dy = 1$.

Then, for $u \in (0,1)$, we estimate $g_1(u)$ and $g_2(u)$ by

$$\begin{aligned} \widehat{g}_{1,h_1}(u) &:= \frac{1}{n}\sum_{i=1}^n \sin(\Theta_i).K_{h_1}(u - F_X(X_i)), \\ \widehat{g}_{2,h_2}(u) &:= \frac{1}{n}\sum_{i=1}^n \cos(\Theta_i).K_{h_2}(u - F_X(X_i)) \end{aligned} \tag{6}$$

respectively, where $h_1, h_2 > 0$ are bandwidths of kernels $K_{h_1}(\cdot)$ and $K_{h_2}(\cdot)$ respectively.

Thus, we estimate g by

$$\widehat{g}_h(u) := \operatorname{atan2}\big(\widehat{g}_{1,h_1}(u), \widehat{g}_{2,h_2}(u)\big), \quad u \in (0,1), \tag{7}$$

where we denote $h := (h_1, h_2)$. Moreover, as a consequence, for $x \in \mathbb{R}$, the estimators for m_1 and m_2 are

$$\widehat{m}_{1,h_1}(x) := \widehat{g}_{1,h_1}\big(F_X(x)\big) = \frac{1}{n}\sum_{i=1}^n \sin(\Theta_i).K_{h_1}\big(F_X(x) - F_X(X_i)\big), \tag{8}$$

and

$$\widehat{m}_{2,h_2}(x) := \widehat{g}_{2,h_2}\big(F_X(x)\big) = \frac{1}{n}\sum_{i=1}^n \cos(\Theta_i).K_{h_2}\big(F_X(x) - F_X(X_i)\big). \tag{9}$$

Using $m = g \circ F_X = \operatorname{atan2}\big(m_1, m_2\big)$, we then obtain an estimator of $m(x)$ at $x \in \mathbb{R}$ by setting

$$\widehat{m}_h(x) := \operatorname{atan2}\big(\widehat{m}_{1,h_1}(x), \widehat{m}_{2,h_2}(x)\big) = \widehat{g}_h\big(F_X(x)\big).$$

2.3. Bandwidth selection

We study the pointwise risk of the estimator $\widehat{m}_h(x)$ associated to the divergence d_c. The expression of the risk is then

$$\mathbb{E}\Big[d_c(\widehat{m}_h(x), m(x))\Big] = \mathbb{E}\Big[d_c\big(\widehat{g}_h(F_X(x)), g(F_X(x))\big)\Big].$$

We first focus on the estimator $\widehat{g}_h$ of g by studying the adaptive choice of bandwidths belonging to a convenient grid $\mathcal{H}_n$. To define the latter, we assume that the kernel K satisfies $\text{supp}(K) \subseteq [-A, A]$ for some $A > 0$ and we take $h_{\max}$ a constant such that $F_X(x) - A.h_{\max} > 0$ and $F_X(x) + A.h_{\max} < 1$. Then, we set

$$\mathcal{H}_n := \left\{ h = k^{-1} : \ k \in \mathbb{N}^*, \ h \leq h_{\max}, \ n.h > \max\Big(\frac{\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{\|K\|^2_\infty}; 1\Big). \log(n) \right\}. \tag{10}$$

Remark 2.3. Observe that the condition

$$F_X(x) - A.h_{\max} > 0, \quad F_X(x) + A.h_{\max} < 1$$

is satisfied for n large enough if $h_{\max}$ depends on n and goes to 0 (even slowly) when $n \to +\infty$.

We have $\text{Card}\big(\mathcal{H}_n\big) \lesssim n/\log n$. In the sequel, we apply the method proposed by Goldenshluger and Lepski in [11] to select an optimal value for bandwidths h_1 and h_2 automatically. Let $j \in \{1, 2\}$. For $h_j \in \mathcal{H}_n$ and $v \in (0, 1)$ we set

$$A_j(h_j, v) := \sup_{h'_j \in \mathcal{H}_n} \left\{ \big|\widehat{g}_{j,h_j,h'_j}(v) - \widehat{g}_{j,h'_j}(v)\big| - \sqrt{\widetilde{V}_j(n, h'_j)} \right\}_+, \tag{11}$$

with $\widetilde{V}_j(n, h'_j) := c_{0,j}.\frac{\log(n).\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h'_j}$, $c_{0,j} > 0$ a tuning parameter and

$$\widehat{g}_{j,h_j,h'_j}(v) := \big(K_{h'_j} * \widehat{g}_{j,h_j}\big)(v),$$

so that $\widehat{g}_{j,h_j,h'_j}(v) = \widehat{g}_{j,h'_j,h_j}(v)$. Then, a data-driven choice of bandwidth h_j is performed as follows:

$$\widehat{h}_j = \underset{h_j \in \mathcal{H}_n}{\text{argmin}} \Big\{ A_j(h_j, v) + \sqrt{\widetilde{V}_j(n, h_j)} \Big\}. \tag{12}$$

Observe that our bandwidth selection rule depends on x. The criterion (12) is inspired from [11], in order to mimic the optimal "bias-variance" trade-off in the pointwise quadratic decomposition:

$$\begin{aligned}\mathbb{E}[|\widehat{g}_{j,h_j}(v) - g_j(v)|^2] &= |\mathbb{E}[\widehat{g}_{j,h_j}(v)] - g_j(v)|^2 + \mathbb{E}[|\widehat{g}_{j,h_j}(v) - \mathbb{E}[\widehat{g}_{j,h_j}(v)]|^2] \\ &=: b^2(h_j, v) + V(h_j, v).\end{aligned}$$

It is common to use $\widetilde{V}_j(n, h'_j)$ to provide an upper bound for the variance term $V(h_j, v)$ (see Section 5.2), whereas the more involved task of the Goldenshluger-Lepski method is to provide an estimate for the bias term by comparing pair-by-pair several estimators. In our framework, the bias term corresponds to

$$b(h_j, v) = |\mathbb{E}[\widehat{g}_{j,h_j}(v)] - g_j(v)| = \big|\big(K_{h_j} * g_j\big)(v) - g_j(v)\big|,$$

(see (23)), so it is natural to estimate it by an estimator of the form $\left|\left(K_{h_j} * \widehat{g}_{j,h'_j}\right)(v) - \widehat{g}_{j,h'_j}(v)\right|$. Thus, the estimator of the bias term is $A_j(h_j, v)$, defined in (11), where the second term $\sqrt{\widetilde{V}_j(n, h'_j)}$ controls the fluctuations of the first term. Now, we define the kernel estimator of $g(v)$ with data-driven bandwidths as follows:

$$\widehat{g}_{\widehat{h}}(v) := \operatorname{atan2}\left(\widehat{g}_{1,\widehat{h}_1}(v), \widehat{g}_{2,\widehat{h}_2}(v)\right), \tag{13}$$

where we denote $\widehat{h} := (\widehat{h}_1, \widehat{h}_2)$. We finally define the adaptive estimator for $m(x)$ by

$$\widehat{m}_{\widehat{h}}(x) := \operatorname{atan2}\left(\widehat{m}_{1,\widehat{h}_1}(x), \widehat{m}_{2,\widehat{h}_2}(x)\right). \tag{14}$$

3. Theoretical results

3.1. Minimax rates of convergence

The minimax approach is a framework that shows the optimality of an estimate among all possible estimates. The minimax pointwise quadratic risk for the estimator $\widehat{g}_{\widehat{h}} = \operatorname{atan2}\left(\widehat{g}_{1,\widehat{h}_1}, \widehat{g}_{2,\widehat{h}_2}\right)$ will be derived from the following control of the pointwise quadratic risks of $\widehat{g}_{1,\widehat{h}_1}$ and $\widehat{g}_{2,\widehat{h}_2}$.

Proposition 3.1. *Consider the collection of bandwidths $\mathcal{H}_n$ defined in (10). Let $j \in \{1, 2\}$ and $q \geq 1$ and assume that $\min\{c_{0,1}; c_{0,2}\} \geq 16\left(2+q\right)^2.\left(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})}\right)^2$. Then, with probability larger than $1 - 4.n^{-q}$,*

$$\left|\widehat{g}_{j,\widehat{h}_j}(F_X(x)) - g_j(F_X(x))\right| \leq \inf_{h_j \in \mathcal{H}_n} \Big\{\left(1 + 2.\|K\|_{\mathbb{L}^1(\mathbb{R})}\right).\left\|g_j - K_{h_j} * g_j\right\|_\infty + 3.\sqrt{\widetilde{V}_j(n, h_j)}\Big\}.$$

The proof of Proposition 3.1 is given in Section 5.4.1. Roughly speaking, in view of results of Section 5.1, the right hand side of the inequality stated in Proposition 3.1 may be viewed as the bias-variance decomposition of the pointwise quadratic-risk of the best warped-kernel estimate, up to a logarithmic term.

Remark 3.2. Examining the proof of Proposition 3.1, the "uniform bias" comes from Inequality (33). This control can be refined and the term $\left\|g_j - K_{h_j} * g_j\right\|_\infty$ can be replaced by

$$\sup_{t \in V(u_x)} \left|g_j(t) - \left(K_{h_j} * g_j\right)(t)\right|,$$

with $V(u_x) := \{t : |t - u_x| \leq A h_{\max}\}$. Observe that the size of this neighborhood of u_x goes to 0 if $h_{\max} \to 0$.

Since the function $\operatorname{atan2}(w_1, w_2)$ is undefined when $w_1 = w_2 = 0$, it is reasonable to consider the following assumption:

Assumption 3.3. *We have*

$$m_1(x) \neq 0 \quad \textit{or} \quad m_2(x) \neq 0.$$

Then, we define $\delta > 0$ *such that*

$$\delta = \begin{cases} \min\big(|m_1(x)|, |m_2(x)|\big) & \textit{if} \quad m_1(x) \neq 0 \textit{ and } m_2(x) \neq 0, \\ |m_1(x)| & \textit{if} \quad m_2(x) = 0, \\ |m_2(x)| & \textit{if} \quad m_1(x) = 0. \end{cases} \tag{15}$$

In the minimax setting, we need some assumptions on the regularity of g_1 and g_2. Thus, we introduce the following Hölder classes that are adapted to local estimation.

Definition 3.4. Let $\beta > 0$ and $L > 0$. The Hölder class $\mathcal{H}(\beta, L)$ is the set of functions $f : (0,1) \longmapsto \mathbb{R}$, such that f admits derivatives up to the order $\lfloor\beta\rfloor$, and for any $(y, \widetilde{y}) \in (0,1)^2$,

$$\left|\frac{d^{\lfloor\beta\rfloor}f}{(dy)^{\lfloor\beta\rfloor}}(\widetilde{y}) - \frac{d^{\lfloor\beta\rfloor}f}{(dy)^{\lfloor\beta\rfloor}}(y)\right| \leq L.\big|\widetilde{y} - y\big|^{\beta - \lfloor\beta\rfloor}.$$

We also consider the following assumption on the kernel K:

Assumption 3.5. *The kernel K is of order $\mathcal{L} \in \mathbb{R}_+$, i.e.*

(i) $C_{K,\mathcal{L}} := \int_{\mathbb{R}} (1 + |y|)^{\mathcal{L}}.|K(y)|dy < \infty$*;*
(ii) $\forall k \in \{1, \ldots, \lfloor\mathcal{L}\rfloor\}$, $\int_{\mathbb{R}} y^k.K(y)dy = 0$.

Now, we obtain an upper bound for the pointwise risk of our final estimator $\widehat{m}_{\widehat{h}}$ at x defined in (14):

Theorem 3.6. *Let $\beta_1, \beta_2, L_1, L_2 > 0$. Suppose that g_1 belongs to $\mathcal{H}(\beta_1, L_1)$, g_2 belongs to $\mathcal{H}(\beta_2, L_2)$, the kernel K satisfies Assumption 3.5 with an index $\mathcal{L} \in \mathbb{R}_+$ such that $\mathcal{L} \geq \max(\beta_1, \beta_2)$. Let $q \geq 1$, and suppose that* $\min\{c_{0,1}; c_{0,2}\} \geq 16\big(2 + q\big)^2.\big(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})}\big)^2$*. Then, by taking $h_{\max} = (\log n)^{-1}$, under Assumption 3.3, for n sufficiently large,*

$$\mathbb{E}\Big[d_c(\widehat{m}_{\widehat{h}}(x), m(x))\Big] \leq \frac{C}{\delta^2}.\max\left\{\psi_n^2(\beta_1), \psi_n^2(\beta_2)\right\},$$

where

$$\psi_n(\beta_1) = \big(\log(n)/n\big)^{\frac{\beta_1}{2\beta_1+1}}, \quad \psi_n(\beta_2) = \big(\log(n)/n\big)^{\frac{\beta_2}{2\beta_2+1}},$$

δ is defined in (15) and C is a constant depending on $\beta_1, \beta_2, L_1, L_2, c_{0,1}, c_{0,2}$ and K.

A proof of Theorem 3.6 is given in Section 5.4.2. Observe that if $\beta_1 = \beta_2 = \beta$, then we obtain the rate $\psi_n(\beta) = (\log n/n)^{\beta/(2\beta+1)}$, which is the optimal rate for adaptive univariate regression function estimation and pointwise risk (see e.g. Section 2 in [2]). Note that the logarithmic term appearing in the rate of convergence is expected since we deal with pointwise adaptive estimation. For further details, we refer the reader to Lepski [16] and Lepski and Spokoiny [17] who have highlighted and discussed this fact for the Gaussian white noise model.

Remark 3.7. Eventually, to obtain a fully computable estimator, we replace the c.d.f. F_X by its natural estimate $y \in \mathbb{R} \longmapsto \widehat{F}_n(y) := n^{-1} \sum_{i=1}^{n} 1_{\{X_i \leq y\}}$. The theoretical justification of this plug-in lies on the use of the Dvoretzky-Kiefer-Wolfowitz inequality [21] which states that, roughly speaking, the function $\widehat{F}_n$ converges uniformly to F_X at the parametric rate $\sqrt{n}$. This has been proposed and fully detailed in [15] and [4] in which tedious calculations show that rates are not deteriorated by this replacement.

3.2. Minimax lower bounds

To establish minimax lower bounds, we assume that the ζ_i's are i.i.d. random angles with zero mean direction, finite concentration and independent of the X_i's. We also assume that Model (5) satisfies the following assumption.

Assumption 3.8. *The design points X_i's are i.i.d. random variables with density $f_X(.)$ on $[0,1]$ such that there exists $\mu_0 < \infty$ and $f_X(t) \leq \mu_0$ $\forall t \in [0,1]$ and the errors ζ_i have common density $p_\zeta(.)$ on $\mathbb{S}^1$ with respect to the Lebesgue measure on $\mathbb{S}^1$, verifying*

$$\exists\, p_* > 0,\ \exists\, \theta_0 > 0 : \int p_\zeta(t) \log \frac{p_\zeta(t)}{p_\zeta(t+\theta)} dt \leq p_* \theta^2,\ \forall |\theta| \leq \theta_0. \tag{16}$$

The subsequent minimax lower bound is based on a reduction scheme based on some well-chosen probability distributions. The closeness between the associated models is measured by using the Kullback-Leibler divergence and is controlled by using Assumption 3.8. In the sequel, the function m belongs to the class $\tilde{\Sigma}(\beta, L)$ defined as the set of functions $f : [0,1] \longmapsto \mathbb{S}^1$ such that the derivative $f^{(l)}$, $l = \lfloor \beta \rfloor$ exists and verifies

$$\sqrt{d_c(f^{(l)}(t), f^{(l)}(t'))} \leq L|t - t'|^{\beta - l}, \quad \forall t, t' \in [0,1].$$

Remark 3.9. For two classes of functions $\mathcal{D}$ and $\mathcal{D}'$ such that $\mathcal{D} \subset \mathcal{D}'$, a lower bound for the minimax rate of convergence for $\mathcal{D}$ will also be a lower bound for the minimax rate for $\mathcal{D}'$. Hence, this justifies the restriction of the study of the lower bound to circular functions m defined on $[0,1]$.

Remark 3.10. The classical von Mises distribution with location parameter $\mu \in [-\pi, \pi)$ and concentration parameter $\kappa > 0$ whose density $f_{vM(\mu,\kappa)}$ is defined for any $\theta \in [-\pi, \pi)$ by

$$f_{vM(\mu,\kappa)}(\theta) = c(\kappa). \exp\big(\kappa. \cos(\theta - \mu)\big), \tag{17}$$

with $c(\kappa)$ the normalizing constant, satisfies condition (16). This is proved in Lemma 5.7. Note that apart from the most popular von Mises distribution, two other classical circular distributions namely the cardioid and the wrapped Cauchy distributions, respectively defined by (see [18])

$$\theta \in [-\pi, \pi) \longmapsto \frac{1}{2\pi}\big(1 + 2\rho \cos(\theta - \mu)\big), \quad \rho \in \left[0, \frac{1}{2}\right),\ \mu \in [-\pi, \pi)$$

and

$$\theta \in [-\pi, \pi) \longmapsto \frac{1}{2\pi} \frac{1-\ell^2}{1+\ell^2-2\ell\cos(\theta-\mu)}, \quad \ell \in [0,1),\ \mu \in [-\pi, \pi)$$

also satisfy (16). Proofs are very similar to the von Mises case.

We obtain the following lower bound:

Theorem 3.11. *Let $\beta > 0$ and $L > 0$. Under Assumptions 3.8, we have*

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{m \in \tilde{\Sigma}(\beta,L)} \mathbb{E}\big[n^{\frac{2\beta}{2\beta+1}} d_c(T_n(x), m(x))\big] \geq \tilde{c},$$

where $\tilde{c}$ depends only on β, L, p_* and μ_0 and the infimum is taken over all possible estimates based on observations $(\Theta_i, X_i)_{i=1,\dots,n}$.

According to Remark 3.9, Theorem 3.11 entails that the lower bound for the minimax risk for functions $m : \mathbb{R} \longmapsto \mathbb{S}^1$ such that $m \in \tilde{\Sigma}(\beta, L)$ is $n^{-\frac{2\beta}{2\beta+1}}$. Now let us connect this result to the upper bound obtained in Theorem 3.6. As the function atan2 is infinitely differentiable on $\mathbb{R}^* \times \mathbb{R}^*$, and if F_X is smoother than g_1 and g_2, then if one writes $m(x) = \text{atan2}(g_1(F_X(x)), g_2(F_X(x))$, the smoothness β of m will be the minimum of the smoothness of g_1 and the smoothness of g_2. Hence, the result of Theorem 3.6 guarantees the near optimal rate of our adaptive estimator provided that F_X is known.

4. Numerical simulations

In this section, we implement some simulations to study the numerical performances of our procedure. We consider three different regression models:

$$\text{M1.} \quad \Theta = \text{atan2}\big(2X-1, X^2+2\big) + \zeta \pmod{2\pi}, \tag{18}$$

$$\text{M2.} \quad \Theta = \text{atan2}\big(-2X+1, X^2-1\big) + \zeta \pmod{2\pi}, \tag{19}$$

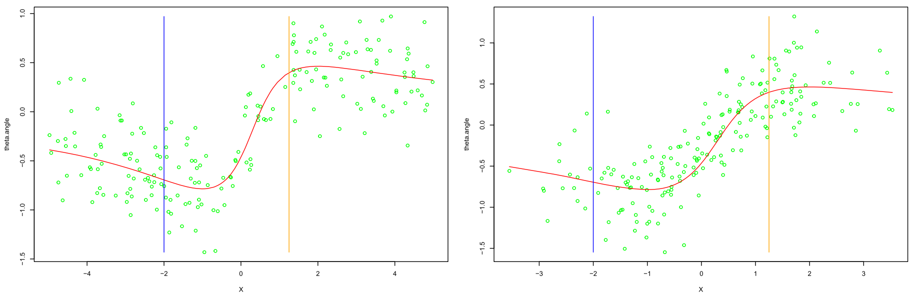
$$\text{M3.} \quad \Theta = \arccos\big(X^5-1\big) + 3.\arcsin\big(X^3-X+1\big) + \zeta \pmod{2\pi}, \tag{20}$$

where the circular error, ζ, is distributed according to a von Mises distribution $f_{vM(0,10)}$ (see (17) and is independent from X.

In the sequel, for models M1 and M2, we consider two cases: $X \sim U([-5,5])$ and $X \sim \mathcal{N}(0, 1.5)$. For model M3, we consider $X \sim U([0,1])$. Then, for different values of n, we draw a sample $\big(\Theta_i, X_i\big)_{i=1,\dots,n}$ with the same distribution as (Θ, X). To implement the Goldenshluger-Lepski methodology, we shall consider either the Gaussian kernel defined by $y \longmapsto K(y) = \frac{1}{\sqrt{2\pi}}.e^{-\frac{y^2}{2}}$, or the Epanechnikov kernel K defined by $y \longmapsto K(y) = \frac{3}{4}.(1-y^2).1_{|y|\leq 1}$. Moreover, we consider the following collection of bandwidths $\mathcal{H}_n$ defined as

$$\mathcal{H}_n := \left\{k^{-1} : k \in \mathbb{N}, 1 \leq k \leq \frac{n}{\log(n)}\right\}.$$

Finally, as explained in Remark 3.7, the final estimators are computed using $y \in \mathbb{R} \longmapsto \widehat{F}_n(y) := \frac{1}{n}\sum_{j=1}^{n} 1_{X_j \leq y}$ instead of F_X.



(a) Model M1 : $X_i \sim U([-5,5])$ (b) Model M1 : $X_i \sim \mathcal{N}(0,1.5)$

Fig 1. *Illustration of model M1 with $n = 200$ for two different density functions of the design. Simulated data $(\Theta_i)_{i=1}^n$ are displayed in green points. The red curve represents the regression function m, while the blue vertical line displays the point $x = -2$ and the orange vertical line displays the point $x = 1.25$ where we aim at estimating $m(x)$.*

4.1. Practical calibration of tuning parameters

In the bandwidth selection procedure described in Section 2.3, we need to tune two parameters $c_{0,1}$ and $c_{0,2}$ in order to find an optimal value of the pointwise risk

$$\mathcal{R} := 1 - \cos\big(\widehat{m}_{\widehat{h}}(x), m(x)\big), \tag{21}$$

with $\widehat{m}_{\widehat{h}}(x) = \operatorname{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(\widehat{F}_n(x)), \widehat{g}_{2,\widehat{h}_2}(\widehat{F}_n(x))\big)$. To do this, we implement preliminary simulations to calibrate $c_{0,1}$ and $c_{0,2}$ by only considering model M1. Figure 1 displays an illustration of our setting.

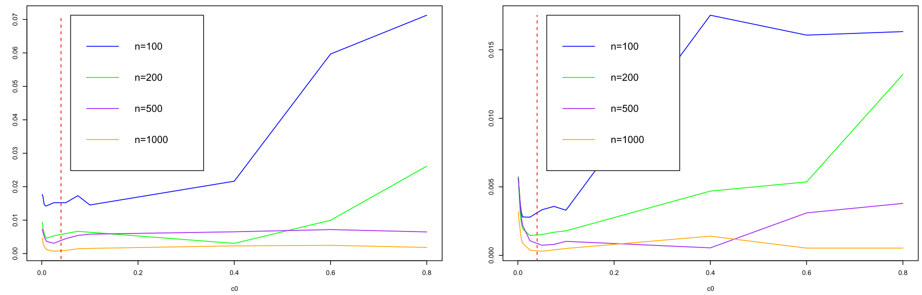
4.1.1. The case $c_{0,1} = c_{0,2}$

To select $\widehat{h}_1$ and $\widehat{h}_2$, we first consider the case $c_{0,1} = c_{0,2} = c_0$. For different sample sizes $n \in \{100; 200; 500; 1000\}$, we compute the risk $\mathcal{R}$ defined in (21) as a function of c_0 on the following discretization grid

$$G_{c_0} := \{0.001;\, 0.0025;\, 0.005;\, 0.0075;\, 0.01;\, 0.025;\, 0.05;\, 0.075;\, 0.1;\, 0.2;\, 0.3;\, 0.4\}\,.$$

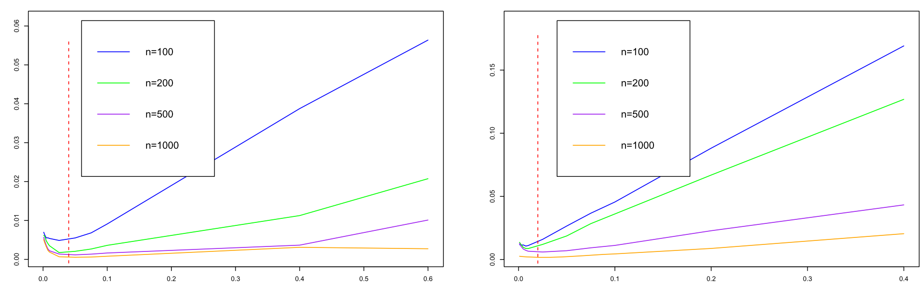
We denote $\mathcal{R} \equiv \mathcal{R}(c_0)$. The numerical illustrations are displayed in Figure 2 for $x = -2$ and in Figure 3 for $x = 1.25$, respectively.

To further study an influence of the kernel rule, we consider the Gaussian kernel. The associated numerical illustrations are provided in Figure 4 for $X \sim U([-5,5])$ and for $X \sim \mathcal{N}(0,1.5)$. This brief numerical study shows that the choice $c_{0,1} = c_{0,2} = 0.04$ is convenient for each numerical scheme.



(a) $X_i \sim \mathcal{N}(0, 1.5)$ (b) $X_i \sim U([-5, 5])$

FIG 2. *Model M1. Plot of the Monte Carlo estimation of the function* $c_0 \in G_{c_0} \longmapsto \mathcal{R}(c_0)$*, based on 50 runs, for* $x = -2$*, the Gaussian and the uniform designs and by using the Epanechnikov kernel for* $n \in \{100; 200; 500; 1000\}$ *corresponding to the line in blue, green, violet and orange, respectively. The red vertical line displays the point* $c_0 = 0.04$*.*



(a) $X_i \sim \mathcal{N}(0, 1.5)$ (b) $X_i \sim U([-5, 5])$

FIG 3. *Model M1. Plot of the Monte Carlo estimation of the function* $c_0 \in G_{c_0} \longmapsto \mathcal{R}(c_0)$*, based on 50 runs, for* $x = 1.25$*, the Gaussian and the uniform designs and by using the Epanechnikov kernel for* $n \in \{100; 200; 500; 1000\}$ *corresponding to the line in blue, green, violet and orange, respectively. The red vertical line displays the point* $c_0 = 0.04$*.*

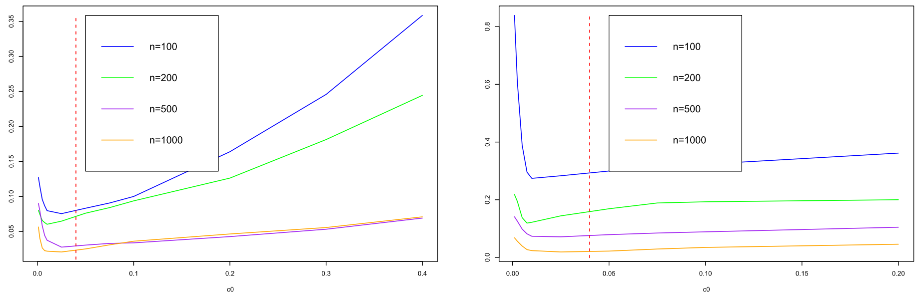
4.1.2. The case $c_{0,1} \neq c_{0,2}$

We do no longer assume that $c_{0,1} = c_{0,2}$. For $n = 200$, we compute the risk $\mathcal{R}$ defined in (21) as a function of $(c_{0,1}, c_{0,2})$ on the following discretization grid

$$G_{c_0} := \{0.001;\ 0.005;\ 0.01;\ 0.025;\ 0.05;\ 0.075;\ 0.1;\ 0.2;\ 0.3;\ 0.4\}.$$

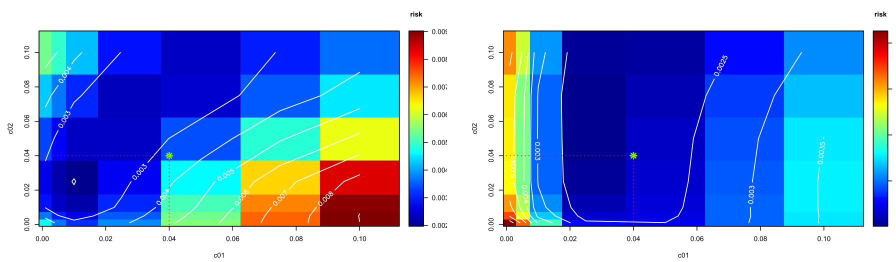
We denote $\mathcal{R} \equiv \mathcal{R}(c_{0,1}, c_{0,2})$. The associated numerical illustrations are provided in Figure 5 and Figure 6 for the case $X \sim \mathcal{N}(0, 1.5)$ and $X \sim U([-5, 5])$, respectively.

Even if it is not the best one, the choice of $c_{0,1} = c_{0,2} = 0.04$ is reasonable. For sake of simplicity, we fix $c_{0,1} = c_{0,2} = 0.04$ for subsequent numerical simulations.



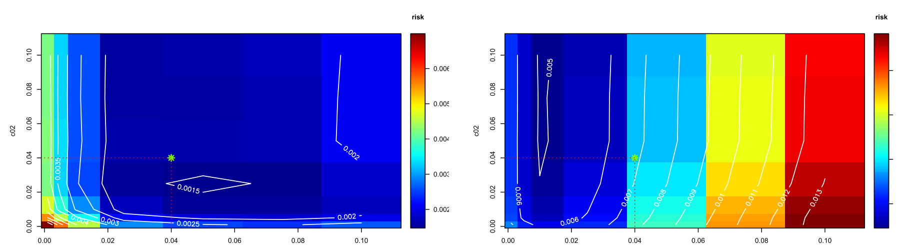
(a) $X_i \sim U([-5, 5])$. (b) $X_i \sim \mathcal{N}(0, 1.5)$.

Fig 4. *Model M1. Plot of the Monte Carlo estimation of the function* $c_0 \in G_{c_0} \longmapsto \mathcal{R}(c_0)$, *based on 50 runs, for* $x = -2$, *the Gaussian and the uniform designs and by using the Gaussian kernel for* $n \in \{100; 200; 500; 1000\}$ *corresponding to the line in color of blue, green, violet and orange, respectively.*



(a) estimation at $x = -2$ (b) estimation at $x = 1.25$

Fig 5. *Model M1. 2D-representation of the Monte Carlo estimation of the function* $(c_{0,1}, c_{0,2}) \in G_{c_0} \times G_{c_0} \longmapsto \mathcal{R}(c_{0,1}, c_{0,2})$, *based on 50 runs, with* $X_i \sim \mathcal{N}(0, 1.5)$, $n = 200$ *at* $x = -2$ *and* $x = 1.25$ *by using the Epanechnikov kernel. We display the specific point* $c_{0,1} = c_{0,2} = 0.04$.

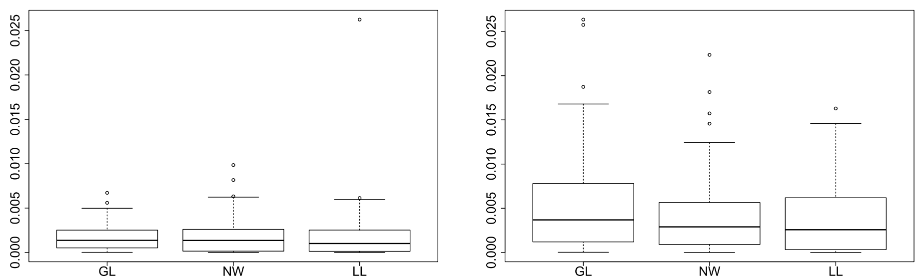


(a) estimation at $x = -2$ (b) estimation at $x = 1.25$

Fig 6. *Model M1. 2D-representation of the Monte Carlo estimation of the function* $(c_{0,1}, c_{0,2}) \in G_{c_0} \times G_{c_0} \longmapsto \mathcal{R}(c_{0,1}, c_{0,2})$, *based on 50 runs with* $X_i \sim U([-5, 5])$, $n = 200$ *at* $x = -2$ *and* $x = 1.25$ *by using the Epanechnikov kernel. We display the specific point* $c_{0,1} = c_{0,2} = 0.04$.

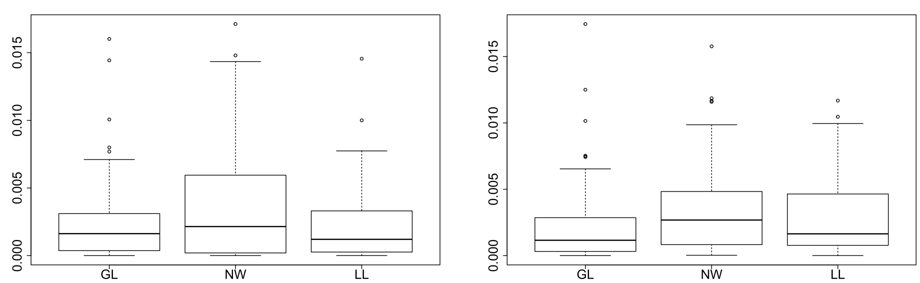
4.2. Numerical results

We now illustrate the numerical performances obtained by our methodology for models M1, M2 and M3 by using the Epanechnikov kernel. They are also compared to other approaches. A similar scheme is conducted by using the Gaussian kernel. Remember that in the following numerical experiments, our estimate is tuned with $c_{0,1} = c_{0,2} = 0.04$. We first display several graphs to illustrate numerical performances obtained by our methodology, denoted GL, by using the Epanechnikov kernel. More precisely, we display boxplots in Figures 7 and 8 summarizing our numerical results for model M1 in the case $X \sim U([-5, 5])$ and in the case $X \sim \mathcal{N}(0, 1.5)$, respectively. In both cases, for model M1, we estimate $m(x)$ at $x = -2$ and at $x = 1.25$. Figure 9 shows simulations for model M2 with $X \sim \mathcal{N}(0, 1.5)$ and we estimate $m(x)$ at $x = 1.05$. Figure 10 shows simulations for model M3 with $X \sim U([0, 1])$ and we estimate $m(x)$ at $x = 0.95$.



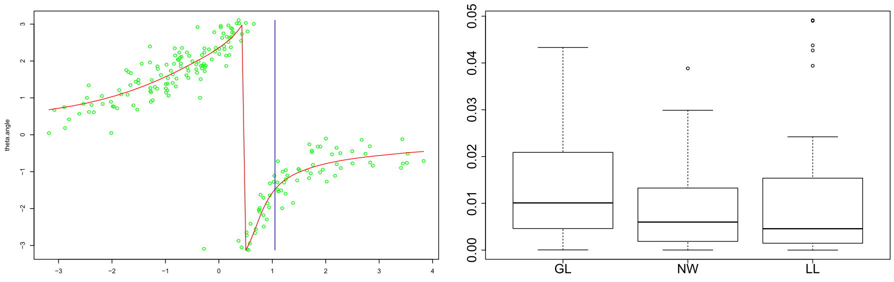
(a) estimation at $x = -2$ (b) estimation at $x = 1.25$

FIG 7. *Model M1. Boxplots of the estimated risk with* 50 *runs for the GL, NW and LL methodologies (from left-hand side to right-hand side on each plot (a) and (b) for* $n = 200$ *and* $X \sim U([-5, 5])$ *by using the Epanechnikov kernel.*



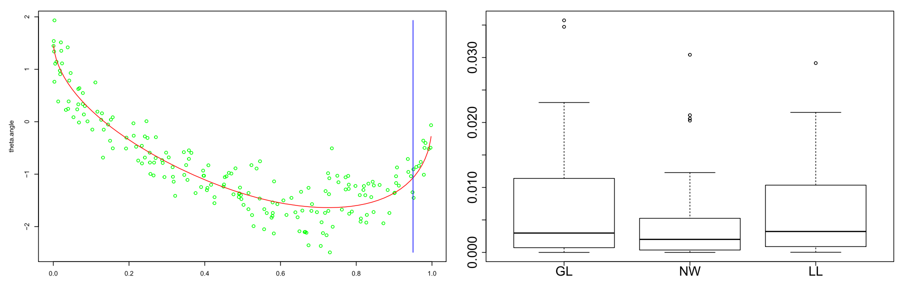
(a) estimation at $x = -2$ (b) estimation at $x = 1.25$

FIG 8. *Model M1. Boxplots of the estimated risk with* 50 *runs for the GL, NW and LL methodologies (from left-hand side to right-hand side on each plot (a) and (b) for* $n = 200$ *and* $X \sim \mathcal{N}(0, 1.5)$ *by using the Epanechnikov kernel.*



(a) $X_i \sim \mathcal{N}(0, 1.5)$ (b) Boxplots for GL, NW and LL methodologies

Fig 9. *(a): Model M2. Simulated data $(\Theta_i)_{i=1}^n$ of model M2 (green points) with $n = 200$ and $X_i \sim \mathcal{N}(0, 1.5)$. The red curve represents the true regression function m, while the blue vertical line displays the point $x = 1.05$; (b): Boxplots of the estimated risk with 50 runs for the GL, NW and LL methodologies (from left-hand side to right-hand side) by using the Epanechnikov kernel.*

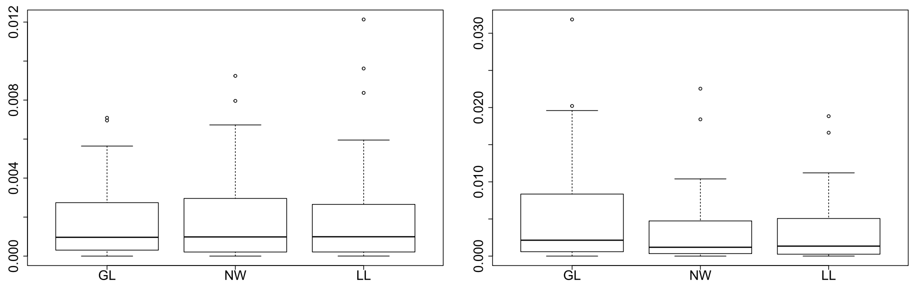


(a) $X_i \sim U([0, 1])$ (b) Boxplots for GL, NW and LL methodologies

Fig 10. *(a): Model M3. Simulated data $(\Theta_i)_{i=1}^n$ of Model M3 (green points) with $n = 200$ and $X_i \sim U([0, 1])$. The red curve represents the true regression function m, while the blue vertical line displays the point $x = 0.95$; (b): Boxplots of the estimated risk with 50 runs for the GL, NW and LL methodologies (from left-hand side to right-hand side) by using the Epanechnikov kernel.*

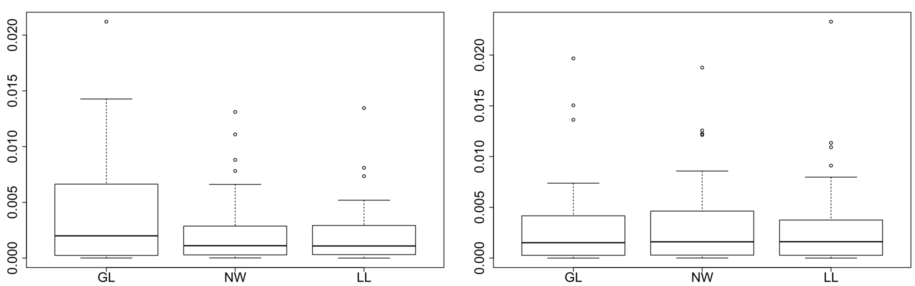
Moreover, to make a comparison with our adaptive estimator, as proposed in [20], we also compute the Nadaraya-Watson (NW) estimator $\widehat{m}_h^{NW}$ and the version of the local linear (LL) estimator proposed by [20, Section 4.2]) denoted by $\widehat{m}_h^{LL}$. Cross-Validation is used to select the bandwidth parameter for $\widehat{m}_h^{NW}$ and $\widehat{m}_h^{LL}$. Boxplots in Figures 7, 8, 9 and 10 show that the performances of our estimator are quite satisfying.

We finally repeat the previous numerical experiments but with the use of the Gaussian kernel: Figures 11 and 12 are the analogs of Figures 7 and 8. Figure 13a shows the numerical simulation for model M2 with $X \sim \mathcal{N}(0, 1.5)$ and we estimate $m(x)$ at $x = 1.05$. Figure 13b shows the numerical simulation for model M3 with $X \sim U([0, 1])$ and we estimate $m(x)$ at $x = 0.95$. These graphs show that the performances of our adaptive estimator associated with the Gaussian kernel are quite satisfying as well.



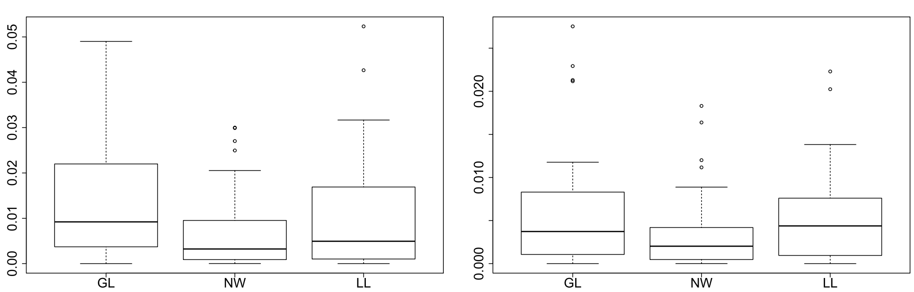
(a) estimation at $x = -2$ (b) estimation at $x = 1.25$

Fig 11. *Model M1. Boxplots of the estimated risk with* 50 *runs for the GL, NW and LL methodologies (from left-hand side to right-hand side on each plot (a) and (b)) for* $n = 200$ *and* $X \sim U([-5, 5])$ *by using the Gaussian kernel.*



(a) estimation at $x = -2$ (b) estimation at $x = 1.25$

Fig 12. *Model M1. Boxplots of the estimated risk with* 50 *runs for the GL, NW and LL methodologies (from left-hand side to right-hand side on each plot (a) and (b)) for* $n = 200$ *and* $X \sim \mathcal{N}(0, 1.5)$ *by using the Gaussian kernel.*



(a) Model M2 with $X \sim \mathcal{N}(0, 1.5)$: estimation at $x = 1.05$ (b) Model M3 with $X \sim U([0, 1])$: estimation at $x = 0.95$

Fig 13. *Model M2 (left), Model M3 (right). Boxplots of the estimated risk with* 50 *runs for the GL, NW and LL methodologies (from left-hand side to right-hand side on each plot (a) and (b)) for* $n = 200$ *by using the Gaussian kernel.*

5. Proofs

Along this section, we fix x in $\mathbb{R}$ and we set $u_x := F_X(x)$.

5.1. Preliminary results

In this section, we study several preliminary results for $\widehat{g}_{1,h_1}$ and $\widehat{g}_{2,h_2}$ defined in (6). First of all, via the warping method, we observe

$$\begin{aligned}\mathbb{E}\big(\widehat{g}_{1,h_1}(u_x)\big) &= \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n}\sin(\Theta_k).K_{h_1}\big(u_x - F_X(X_k)\big)\right]\\ &= \mathbb{E}\Big[\mathbb{E}\big[\sin(\Theta)|X\big].K_{h_1}\big(u_x - F_X(X)\big)\Big]\\ &= \int_{\mathbb{R}}\mathbb{E}[\sin(\Theta)|X=y].K_{h_1}\big(u_x - F_X(y)\big).f_X(y)dy\\ &= \int_{\mathbb{R}} m_1(y).K_{h_1}\big(u_x - F_X(y)\big).f_X(y)dy\\ &= \int_{F_X(\mathbb{R})} g_1(w).K_{h_1}\big(u_x - w\big)dw.\end{aligned}$$

Then, using the choice of $h_{\max}$, since $u_x = F_X(x) \in (0,1) = F_X(\mathbb{R})$ is fixed and $K_{h_1}\big(u_x - w\big) = 0$ for $w \in (0,1)$ such that $\big|u_x - w\big| > A.h_{\max}$, we have

$$\begin{aligned}\int_{F_X(\mathbb{R})} g_1(w).K_{h_1}\big(u_x - w\big)dw &= \int_{u_x - A.h_{\max}}^{u_x + A.h_{\max}} g_1(w).K_{h_1}\big(u_x - w\big)dw\\ &= \big(K_{h_1} * g_1\big)(u_x). \end{aligned} \tag{22}$$

Thus, we obtain

$$\mathbb{E}\big(\widehat{g}_{1,h_1}(u_x)\big) = \big(K_{h_1} * g_1\big)(u_x) \tag{23}$$

and similarly,

$$\mathbb{E}\big(\widehat{g}_{2,h_2}(u_x)\big) = \big(K_{h_2} * g_2\big)(u_x). \tag{24}$$

We obtain upper bounds for the bias and variance terms.

Lemma 5.1. *Let $j \in \{1,2\}$. Suppose that g_j belongs to $\mathcal{H}(\beta_j, L_j)$, with $L_j, \beta_j \in \mathbb{R}_+^*$. Assume that the kernel K satisfies Assumption 3.5 with an index $\mathcal{L} \in \mathbb{R}_+$ such that $\mathcal{L} \geq \beta_j$. Then, for any $h_j \in \mathcal{H}_n$,*

$$\Big|\mathbb{E}\big(\widehat{g}_{j,h_j}(u_x)\big) - g_j(u_x)\Big| \leq C_{K,\mathcal{L}}.L_j.h_j^{\beta_j}, \quad \text{and} \quad Var\big(\widehat{g}_{j,h_j}(u_x)\big) \leq \frac{\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_j},$$

with $C_{K,\mathcal{L}}$ the constant defined in Assumption 3.5.

The proof of Lemma 5.1 is given in Section 5.2.

We introduce in the sequel several events on which we will establish some concentration results for $\widehat{g}_{j,h_j}$, $j \in \{1,2\}$.

Definition 5.2. For $n \in \mathbb{N}^*$, $p \geq 1$, and $h_j > 0$, we define for an arbitrary $v \in F_X(\mathbb{R})$ the following event

$$\Omega_{j,n}(v, h_j) := \left\{ \left| \widehat{g}_{j,h_j}(v) - \mathbb{E}[\widehat{g}_{j,h_j}(v)] \right| \leq c_1(p).\sqrt{\widetilde{V}_j(n, h_j)} \right\}$$

with $c_1(p)$ satisfying $c_1(p).\sqrt{\min\{c_{0,1}; c_{0,2}\}} \geq 4p$.

Furthermore, for $L_j, \beta_j > 0$, we also introduce

$$E_{j,n}(v, h_j) := \left\{ \left| \widehat{g}_{j,h_j}(v) - g_j(v) \right| \leq \Phi_j(n, h_j) \right\}$$

where

$$\Phi_j(n, h_j) := c_1(p).\sqrt{\widetilde{V}_j(n, h_j)} + C_{K,\mathcal{L}}.L_j.h_j^{\beta_j}.$$

Then, the following proposition gives a concentration inequality for $\widehat{g}_{j,h_j}(u_x)$.

Proposition 5.3. *For $p \geq 1$ and $h_j \in \mathcal{H}_n$ (defined in (10)), we have:*

$$\mathbb{P}\left(\left(\Omega_{j,n}(u_x, h_j) \right)^c \right) \leq 2.n^{-p}.$$

Consequently, suppose further that g_j belongs to $\mathcal{H}(\beta_j, L_j)$ with $L_j, \beta_j \in \mathbb{R}_+^*$ and the kernel K satisfies Assumption 3.5 with an index $\mathcal{L} \in \mathbb{R}_+$ such that $\mathcal{L} \geq \beta_j$, then we get:

$$\mathbb{P}\left(\left(E_{j,n}(u_x, h_j) \right)^c \right) \leq 2.n^{-p}.$$

The proof of Proposition 5.3 is given in Section 5.3.

5.2. *Proof of Lemma 5.1*

Proof. First, for the bias of $\widehat{g}_{j,h}(u_x)$ at $u_x = F_X(x)$, using (22), we can write

$$\begin{aligned} \mathbb{E}\big(\widehat{g}_{j,h_j}(u_x)\big) - g_j(u_x) &= \frac{1}{h_j}.\int_{u_x - A.h_{\max}}^{u_x + A.h_{\max}} K\Big(\frac{u_x - z}{h_j}\Big).g_j(z)dz - g_j(u_x) \\ &= \int_{-A}^{A} K(w).\big(g_j(u_x - h_j.w) - g_j(u_x)\big)dw. \end{aligned} \tag{25}$$

Since g_j belongs to $\mathcal{H}(\beta_j, L_j)$, using a Taylor expansion for g_j, we get for $w \in [-A, A]$, $0 \leq \tau \leq 1$,

$$\begin{aligned} g_j(u_x - h_j.w) = g_j(u_x) &+ g_j'(u_x).(-h_j).w \\ &+ \cdots + \frac{(-h_j.w)^{\lfloor \beta_j \rfloor}}{(\lfloor \beta_j \rfloor)!}.g_j^{(\lfloor \beta_j \rfloor)}(u_x - \tau.h_j.w). \end{aligned}$$

Then, under Assumption 3.5 with an index $\mathcal{L} \in \mathbb{R}_+$ satisfying $\mathcal{L} \geq \beta_j$, from (25) one gets

$$\int_{-A}^{A} K(w).\big(g_j(u_x - h_j.w) - g_j(u_x)\big)dw$$

$$= \int_{-A}^{A} K(w).\frac{(-h_j.w)^{\lfloor\beta_j\rfloor}}{(\lfloor\beta_j\rfloor)!}.g_j^{(\lfloor\beta_j\rfloor)}(u_x - \tau.h_j.w)dw$$
$$= \int_{-A}^{A} K(w).\frac{(-h_j.w)^{\lfloor\beta_j\rfloor}}{(\lfloor\beta_j\rfloor)!}.\Big(g_j^{(\lfloor\beta_j\rfloor)}(u_x - \tau.h_j.w) - g_j^{(\lfloor\beta_j\rfloor)}(u_x)\Big)dw.$$

This implies that with $0 \le \tau \le 1$, with $C_{K,\mathcal{L}}$ the constant defined in Assumption 3.5, since $g_j \in \mathcal{H}(\beta_j, L_j)$,

$$\begin{aligned}
&\Big|\mathbb{E}\big(\widehat{g}_{j,h_j}(u_x)\big) - g_j(u_x)\Big| \\
&\quad\le \int_{-A}^{A} |K(w)|.\frac{|h_j.w|^{\lfloor\beta_j\rfloor}}{(\lfloor\beta_j\rfloor)!}.\Big|g_j^{(\lfloor\beta_j\rfloor)}(u_x - \tau.h_j.w) - g_j^{(\lfloor\beta_j\rfloor)}(u_x)\Big|dw \\
&\quad\le \int_{-A}^{A} |K(w)|.\frac{|h_j.w|^{\lfloor\beta_j\rfloor}}{(\lfloor\beta_j\rfloor)!}.L_j.\big|\tau.h_j.w\big|^{\beta_j - \lfloor\beta_j\rfloor}dw \\
&\quad\le L_j.h_j^{\beta_j}.\int_{-A}^{A} |K(w)|.|w|^{\beta_j}dw \\
&\quad\le L_j.h_j^{\beta_j}.\int_{-A}^{A} |K(w)|.(1+|w|)^{\beta_j}dw \le L_j.h_j^{\beta_j}.C_{K,\mathcal{L}}.
\end{aligned}$$

For the variance of $\widehat{g}_{j,h_j}(u_x)$, one gets, with $a_1(\Theta_k) = \sin(\Theta_k)$ and $a_2(\Theta_k) = \cos(\Theta_k)$,

$$\begin{aligned}
\mathrm{Var}\big(\widehat{g}_{j,h_j}(u_x)\big) &= \mathbb{E}\Big[\Big(\widehat{g}_{j,h_j}(u_x) - \mathbb{E}\big(\widehat{g}_{j,h_j}(u_x)\big)\Big)^2\Big] \\
&= \mathrm{Var}\left(\frac{1}{n}.\sum_{k=1}^{n} a_j(\Theta_k).K_{h_j}\big(u_x - F_X(X_k)\big)\right) \\
&\le \frac{1}{n}.\mathbb{E}\left[\big(a_j(\Theta)\big)^2.\big[K_{h_j}\big(u_x - F_X(X)\big)\big]^2\right] \\
&\le \frac{1}{n}.\mathbb{E}\left(\big[K_{h_j}\big(u_x - F_X(X)\big)\big]^2\right) \le \frac{\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_j}.
\end{aligned}$$

This concludes the proof of Lemma 5.1. ☐

5.3. Proof of Proposition 5.3

We shall use the following version of Bernstein inequality (see [8, Lemma 2]).

Lemma 5.4 (Bernstein inequality). *Let $T_1, \ldots, T_n$ be i.i.d. random variables and $S_n = \sum_{j=1}^{n}\big[T_j - \mathbb{E}(T_j)\big]$. Then, for any $\eta > 0$,*

$$\mathbb{P}\Big(\big|S_n\big| \ge n.\eta\Big) \le 2.\max\left(\exp\Big(-\frac{n.\eta^2}{4.V}\Big), \exp\Big(-\frac{n.\eta}{4.b}\Big)\right),$$

with $Var(T_1) \le V$ and $|T_1| \le b$, where V and b are two positive deterministic constants.

Now, we can start to prove Proposition 5.3.

Proof of Proposition 5.3. We follow the procedure proposed in [8, Proposition 6]. First of all, we define random variables $Z_k(u_x) := a_j(\Theta_k).K_{h_j}\big(u_x - F_X(X_k)\big)$, for $1 \leq k \leq n$, with $a_1(\Theta_k) = \sin(\Theta_k)$ and $a_2(\Theta_k) = \cos(\Theta_k)$. Hence, $\widehat{g}_{j,h_j}(u_x) = \frac{1}{n}\sum_{k=1}^{n} Z_k(u_x)$. Notice that $\mathbb{E}(Z_k(u_x)) = \big(K_{h_j} * g_j\big)(u_x)$ (see (23) and (24)). Since $\|K\|_\infty < +\infty$, we then have for any $v \in F_X(\mathbb{R})$:

$$|Z_k(v)| = \big|a_j(\Theta_k).K_{h_j}\big(v - F_X(X_k)\big)\big| \leq \frac{\|K\|_\infty}{h_j} =: b(h_j), \tag{26}$$

and $\mathrm{Var}\big(Z_k(v)\big) = n.\mathrm{Var}\big(\widehat{g}_{j,h_j}(v)\big) \leq n.\frac{\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_j} =: n.V_0(n,h_j)$. Now, applying Lemma 5.4 to the $Z_k(v)$'s, we obtain for $\eta(h_j) > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left|\widehat{g}_{j,h_j}(v) - \mathbb{E}\big(\widehat{g}_{j,h_j}(v)\big)\right| \geq \eta(h_j)\right) \\
&\qquad = \mathbb{P}\left(\Big|\sum_{k=1}^{n} Z_k(v) - \mathbb{E}\Big(Z_k(v)\Big)\Big| \geq n.\eta(h_j)\right) \\
&\qquad \leq 2.\max\left\{\exp\left(-\frac{n.\big(\eta(h_j)\big)^2}{4n.V_0(n,h_j)}\right); \exp\left(-\frac{n.\eta(h_j)}{4b(h_j)}\right)\right\}.
\end{aligned}$$

For $p \geq 1$, choose $\eta(h_j) = c_1(p).\sqrt{\widetilde{V}_j(n,h_j)}$, with

$$\widetilde{V}_j(n,h_j) = c_{0,j}.\log(n).V_0(n,h_j). \tag{27}$$

Then,

$$\begin{aligned}
&\mathbb{P}\left(\left|\widehat{g}_{j,h_j}(v) - \mathbb{E}\big(\widehat{g}_{j,h_j}(v)\big)\right| \geq c_1(p).\sqrt{\widetilde{V}_j(n,h_j)}\right) \\
&\leq 2.\max\left\{\exp\left(-\frac{n.c_1(p)^2.\widetilde{V}_j(n,h_j)}{4n.V_0(n,h_j)}\right); \exp\left(-\frac{n.c_1(p).\sqrt{\widetilde{V}_j(n,h_j)}}{4b(h_j)}\right)\right\}.
\end{aligned} \tag{28}$$

We then choose $c_1(p)$.

\+ First, $c_1(p)$ is chosen such that

$$\frac{n.c_1(p)^2.\widetilde{V}_j(n,h_j)}{4n.V_0(n,h_j)} = \frac{n.c_1(p)^2.c_{0,j}.\log(n).V_0(n,h_j)}{4n.V_0(n,h_j)} \geq p.\log(n), \tag{29}$$

that is $c_1(p)$ satisfies $c_1(p)^2.c_{0,j} \geq 4p$.

\+ Secondly, we can write

$$\frac{n.c_1(p).\sqrt{\widetilde{V}_j(n,h_j)}}{4b(h_j)} = \frac{c_1(p).\sqrt{c_{0,j}}}{4}.\sqrt{\log(n)}.\frac{n.\sqrt{V_0(n,h_j)}}{b(h_j)}$$

and for $h_j \in \mathcal{H}_n$,

$$n.\frac{\sqrt{V_0(n,h_j)}}{b(h_j)} = n.\frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}}{\sqrt{n.h_j}}.\frac{h_j}{\|K\|_\infty} = \sqrt{n.h_j}.\frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}}{\|K\|_\infty} > \sqrt{\log(n)},$$

then, we have

$$\begin{aligned}\frac{n.c_1(p).\sqrt{\widetilde{V}_j(n,h_j)}}{4b(h_j)} &= \frac{c_1(p).\sqrt{c_{0,j}}}{4}.\sqrt{\log(n)}.\frac{n.\sqrt{V_0(n,h_j)}}{b(h_j)} \geq \frac{c_1(p).\sqrt{c_{0,j}}}{4}.\log(n) \\ &\geq p.\log(n), \qquad (30)\end{aligned}$$

provided that $c_1(p).\sqrt{c_{0,j}} \geq 4p$. Note that this condition also ensures the constraint $c_1(p)^2.c_{0,j} \geq 4p$.

Now, combining (29) and (30), we get from (28) for any $p \geq 1$:

$$\begin{aligned}\mathbb{P}\left(\left(\Omega_{j,n}(u_x,h_j)\right)^c\right) &= \mathbb{P}\left(\left|\widehat{g}_{j,h_j}(u_x) - \mathbb{E}\big(\widehat{g}_{j,h_j}(u_x)\big)\right| > c_1(p).\sqrt{\widetilde{V}_j(n,h_j)}\right) \\ &\leq 2.n^{-p}.\end{aligned}$$

This implies that with probability larger than $1-2.n^{-p}$, we have:

$$\left|\widehat{g}_{j,h_j}(u_x) - \mathbb{E}\big(\widehat{g}_{j,h_j}(u_x)\big)\right| \leq c_1(p).\sqrt{\widetilde{V}_j(n,h_j)}.$$

Then, with probability larger than $1-2.n^{-p}$, we obtain

$$\begin{aligned}\left|\widehat{g}_{j,h_j}(u_x) - g_j(u_x)\right| &\leq \left|\widehat{g}_{j,h_j}(u_x) - \mathbb{E}\big(\widehat{g}_{j,h_j}(u_x)\big)\right| + \left|\mathbb{E}\big(\widehat{g}_{j,h_j}(u_x)\big) - g_j(u_x)\right| \\ &\leq c_1(p).\sqrt{\widetilde{V}_j(n,h_j)} \quad + \quad L_j.h_j^{\beta_j}.C_{K,\mathcal{L}}.\end{aligned}$$

Recall that $\Phi_j(n,h_j) = c_1(p).\sqrt{\widetilde{V}_j(n,h_j)} + L_j.h_j^{\beta_j}.C_{K,\mathcal{L}}$, therefore, we finally obtain for $p \geq 1$ that

$$\mathbb{P}\left(\left(E_{j,n}(u_x,h_j)\right)^c\right) = \mathbb{P}\left(\left|\widehat{g}_{j,h_j}(u_x) - g_j(u_x)\right| > \Phi_j(n,h_j)\right) \leq 2.n^{-p}. \qquad (31)$$

This concludes the proof of Proposition 5.3. $\square$

5.4. Proofs of main results

5.4.1. Proof of Proposition 3.1

We first have the following concentration result.

Corollary 5.5. *Let $j \in \{1,2\}$. Under the Assumptions of Proposition 3.1, for all $h_j, h'_j \in \mathcal{H}_n$, for all $p \geq 1$, for $u_x = F_X(x)$,*

$$\mathbb{P}\left(\left|\widehat{g}_{j,h_j,h'_j}(u_x) - \mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big]\right| > c_1(p).\,\|K\|_{\mathbb{L}^1(\mathbb{R})}.\sqrt{\widetilde{V}_j(n,h'_j)}\right) \leq 2.n^{-p},$$

with $c_1(p)$ satisfying $c_1(p).\sqrt{\min\{c_{0,1};c_{0,2}\}} \geq 4p$.

Proof of Corollary 5.5. We define random variables $\widetilde{Z}_k(u_x) := a_j(\Theta_k).\big(K_{h'_j} * K_{h_j}\big)\big(u_x - F_X(X_k)\big)$, for $1 \leq k \leq n$, with $a_1(\Theta_k) = \sin(\Theta_k)$ and $a_2(\Theta_k) = \cos(\Theta_k)$. Hence, $\widehat{g}_{j,h_j,h'_j}(u_x) = \frac{1}{n}\sum_{k=1}^{n} \widetilde{Z}_k(u_x)$. Since $\|K\|_\infty < +\infty$, we then have:

$$\left|\widetilde{Z}_k(u_x)\right| = \left|a_j(\Theta_k).\big(K_{h'_j} * K_{h_j}\big)\big(u_x - F_X(X_k)\big)\right| \leq \frac{\|K\|_\infty \cdot \|K\|_{\mathbb{L}^1(\mathbb{R})}}{h_j} =: \widetilde{b}(h_j),$$

and $\mathrm{Var}\big(\widetilde{Z}_k(u_x)\big) = n.\mathrm{Var}\big(\widehat{g}_{j,h_j,h'_j}(u_x)\big) \leq n.\frac{\|K\|^2_{\mathbb{L}^1(\mathbb{R})} \cdot \|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_j} =: n.\widetilde{V}_0(n, h_j)$.

Using similar arguments of the proof of Proposition 5.3, we obtain with a probability greater than $1 - 2.n^{-p}$ that

$$\left|\widehat{g}_{j,h_j,h'_j}(u_x) - \mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big]\right| \leq \|K\|_{\mathbb{L}^1(\mathbb{R})} .c_1(p).\sqrt{\widetilde{V}_j(n, h_j)}.$$

This concludes the proof of Corollary 5.5. ☐

Now, we can start to prove Proposition 3.1.

Proof of Proposition 3.1. We follow the strategy proposed in [8, Theorem 2]. The target is to find an upper bound for $\big|\widehat{g}_{j,\widehat{h}_j}(u_x) - g_j(u_x)\big|$. Let $h_j \in \mathcal{H}_n$ be fixed. We consider the following decomposition:

$$\begin{aligned}\big|\widehat{g}_{j,\widehat{h}_j}(u_x) - g_j(u_x)\big| \leq \underbrace{\big|\widehat{g}_{j,\widehat{h}_j}(u_x) - \widehat{g}_{j,h_j,\widehat{h}_j}(u_x)\big|}_{=:I_{g_j,1}} &+ \underbrace{\big|\widehat{g}_{j,h_j,\widehat{h}_j}(u_x) - \widehat{g}_{j,h_j}(u_x)\big|}_{=:I_{g_j,2}} \\ &+ \big|\widehat{g}_{j,h_j}(u_x) - g_j(u_x)\big|.\end{aligned}$$

By the definition of $A_2(h_j, u_x)$, we have

$$\begin{aligned}I_{g_j,1} &= \big|\widehat{g}_{j,\widehat{h}_j}(u_x) - \widehat{g}_{j,h_j,\widehat{h}_j}(u_x)\big| \\ &= \big|\widehat{g}_{j,\widehat{h}_j}(u_x) - \widehat{g}_{j,h_j,\widehat{h}_j}(u_x)\big| - \sqrt{\widetilde{V}_j(n, \widehat{h}_j)} + \sqrt{\widetilde{V}_j(n, \widehat{h}_j)} \\ &\leq \sup_{h'_j \in \mathcal{H}_n} \left\{\big|\widehat{g}_{j,h'_j}(u_x) - \widehat{g}_{j,h_j,h'_j}(u_x)\big| - \sqrt{\widetilde{V}_j(n, h'_j)}\right\}_+ + \sqrt{\widetilde{V}_j(n, \widehat{h}_j)} \\ &= A_2(h_j, u_x) + \sqrt{\widetilde{V}_j(n, \widehat{h}_j)}.\end{aligned}$$

And similarly, by the definition of $A_2(\widehat{h}_j, u_x)$,

$$\begin{aligned}I_{g_j,2} &= \big|\widehat{g}_{j,h_j,\widehat{h}_j}(u_x) - \widehat{g}_{j,h_j}(u_x)\big| \\ &\leq \sup_{h'_j \in \mathcal{H}_n} \left\{\big|\widehat{g}_{j,h'_j,\widehat{h}_j}(u_x) - \widehat{g}_{j,h'_j}(u_x)\big| - \sqrt{\widetilde{V}_j(n, h'_j)}\right\}_+ + \sqrt{\widetilde{V}_j(n, h_j)} \\ &= A_2(\widehat{h}_j, u_x) + \sqrt{\widetilde{V}_j(n, h_j)}.\end{aligned}$$

Therefore, by using the definition of $\widehat{h}_j$, we get

$$\begin{aligned} I_{g_j,1} + I_{g_j,2} &\leq A_2(h_j, u_x) + \sqrt{\widetilde{V}_j(n, \widehat{h}_j)} + A_2(\widehat{h}_j, u_x) + \sqrt{\widetilde{V}_j(n, h_j)} \\ &\leq 2.\Big[A_2(h_j, u_x) + \sqrt{\widetilde{V}_j(n, h_j)}\Big]. \end{aligned}$$

Hence, we obtain

$$\big|\widehat{g}_{j,\widehat{h}_j}(u_x) - g_j(u_x)\big| \leq 2.A_2(h_j, u_x) + 2.\sqrt{\widetilde{V}_j(n, h_j)} + \big|\widehat{g}_{j,h_j}(u_x) - g_j(u_x)\big|. \tag{32}$$

Now, to study $A_2(h_j, u_x)$, we can write:

$$\begin{aligned} \widehat{g}_{j,h'_j}(u_x) - \widehat{g}_{j,h_j,h'_j}(u_x) = \widehat{g}_{j,h'_j}(u_x) - \mathbb{E}\big[\widehat{g}_{j,h'_j}(u_x)\big] - \big(\widehat{g}_{j,h_j,h'_j}(u_x) \\ - \mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big]\big) + \mathbb{E}\big[\widehat{g}_{j,h'_j}(u_x)\big] - \mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big], \end{aligned}$$

and, we have $\mathbb{E}\big[\widehat{g}_{j,h'_j}(u_x)\big] = \big(K_{h'_j} * g_j\big)(u_x)$ as well as $\mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big] = \mathbb{E}\big[K_{h'_j} * \widehat{g}_{j,h_j}(u_x)\big] = \big(K_{h'_j} * K_{h_j} * g_j\big)(u_x)$.

Thus,

$$\begin{aligned} \big|\widehat{g}_{j,h'_j}(u_x) - \widehat{g}_{j,h_j,h'_j}(u_x)\big| - \sqrt{\widetilde{V}_j(n, h'_j)} \leq{}& \big|\widehat{g}_{j,h'_j}(u_x) - \mathbb{E}\big[\widehat{g}_{j,h'_j}(u_x)\big]\big| \\ &- \frac{\sqrt{\widetilde{V}_j(n, h'_j)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \\ &+ \big|\widehat{g}_{j,h_j,h'_j}(u_x) - \mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big]\big| \\ &- \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \frac{\sqrt{\widetilde{V}_j(n, h'_j)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \\ &+ \big|\mathbb{E}\big[\widehat{g}_{j,h'_j}(u_x)\big] - \mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big]\big|. \end{aligned}$$

However, for any $h'_j \in \mathcal{H}_n$,

$$\begin{aligned} \big|\mathbb{E}\big[\widehat{g}_{j,h'_j}(u_x)\big] - \mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big]\big| &= \big|K_{h'_j} * \big(g_j - K_{h_j} * g_j\big)(u_x)\big| \\ &\leq \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \big\|g_j - K_{h_j} * g_j\big\|_\infty . \end{aligned} \tag{33}$$

Hence, incorporating this bound in the definition of $A_2(h_j, u_x)$, we obtain

$$A_2(h_j, u_x) \tag{34}$$

$$= \sup_{h'_j \in \mathcal{H}_n} \Big\{\big|\widehat{g}_{j,h'_j}(u_x) - \widehat{g}_{j,h_j,h'_j}(u_x)\big| - \sqrt{\widetilde{V}_j(n, h'_j)}\Big\}_+$$

$$\leq \sup_{h'_j \in \mathcal{H}_n} \left\{\big|\widehat{g}_{j,h'_j}(u_x) - \mathbb{E}\big[\widehat{g}_{j,h'_j}(u_x)\big]\big| - \frac{\sqrt{\widetilde{V}_j(n, h'_j)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})}\right\}_+ \tag{35}$$

$$+\sup_{h'_j\in\mathcal{H}_n}\left\{\left|\widehat{g}_{j,h_j,h'_j}(u_x)-\mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big]\right|-\|K\|_{\mathbb{L}^1(\mathbb{R})}.\frac{\sqrt{\widetilde{V}_j(n,h'_j)}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\right\}_+ + \|K\|_{\mathbb{L}^1(\mathbb{R})}.\left\|g_j-K_{h_j}*g_j\right\|_\infty.$$

From Corollary 5.5, for $h_j,h'_j\in\mathcal{H}_n$,

$$\mathbb{P}\left(\left|\widehat{g}_{j,h_j,h'_j}(u_x)-\mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big]\right|>c_1(p).\|K\|_{\mathbb{L}^1(\mathbb{R})}.\sqrt{\widetilde{V}_j(n,h'_j)}\right)\le 2.n^{-p}.$$

It implies that if we take $c_1(p)=\frac{1}{1+\|K\|_{\mathbb{L}^1(\mathbb{R})}}$ and if $c_{0,j}\ge 16p^2.\big(1+\|K\|_{\mathbb{L}^1(\mathbb{R})}\big)^2$, then

$$\begin{aligned}\mathbb{P}\left(\sup_{h'_j\in\mathcal{H}_n}\left\{\left|\widehat{g}_{j,h'_j}(u_x)-\mathbb{E}\big[\widehat{g}_{j,h'_j}(u_x)\big]\right|-\frac{\sqrt{\widetilde{V}_j(n,h'_j)}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\right\}_+>0\right)&\le 2.\sum_{h_j\in\mathcal{H}_n}n^{-p}\\&\le 2.n^{1-p},\end{aligned}$$

as $\mathrm{Card}(\mathcal{H}_n)\le n$. Consequently, the following set

$$\begin{aligned}\widetilde{A}_2:=&\left\{\sup_{h'_j\in\mathcal{H}_n}\left\{\left|\widehat{g}_{j,h'_j}(u_x)-\mathbb{E}\big[\widehat{g}_{j,h'_j}(u_x)\big]\right|-\frac{\sqrt{\widetilde{V}_j(n,h'_j)}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\right\}_+=0\right\}\\&\cap\left\{\forall\, h_j\in\mathcal{H}_n,\ \sup_{h'_j\in\mathcal{H}_n}\left\{\left|\widehat{g}_{j,h_j,h'_j}(u_x)-\mathbb{E}\big[\widehat{g}_{j,h_j,h'_j}(u_x)\big]\right|\right.\right.\\&\left.\left.\qquad-\|K\|_{\mathbb{L}^1(\mathbb{R})}.\frac{\sqrt{\widetilde{V}_j(n,h'_j)}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\right\}_+=0\right\}\end{aligned}$$

has probability greater than $(1-4.n^{2-p})$. Now, choose $p=2+q$ and then $c_{0,j}\ge 16\big(2+q\big)^2.\big(1+\|K\|_{\mathbb{L}^1(\mathbb{R})}\big)^2$. Thus, we obtain that $\mathbb{P}\big(\widetilde{A}_2\big)>1-4.n^{-q}$.

Combining inequalities (32) and (34), we have on $\widetilde{A}_2$:

$$\begin{aligned}\left|\widehat{g}_{j,\widehat{h}_j}(u_x)-g_j(u_x)\right|&\le 2.A_2(h_j,u_x)+2.\sqrt{\widetilde{V}_j(n,h_j)}+\left|\widehat{g}_{j,h_j}(u_x)-g_j(u_x)\right|\\&\le 2.\|K\|_{\mathbb{L}^1(\mathbb{R})}.\left\|g_j-K_{h_j}*g_j\right\|_\infty+2.\sqrt{\widetilde{V}_j(n,h_j)}\\&\qquad+\left|\widehat{g}_{j,h_j}(u_x)-g_j(u_x)\right|,\end{aligned}$$

but still on $\widetilde{A}_2$, one gets $\left|\widehat{g}_{j,h_j}(u_x)-\mathbb{E}\big[\widehat{g}_{j,h_j}(u_x)\big]\right|-\frac{\sqrt{\widetilde{V}_j(n,h_j)}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\le 0$, so

$$\begin{aligned}\left|\widehat{g}_{j,h_j}(u_x)-g_j(u_x)\right|&\le\left|\mathbb{E}\big[\widehat{g}_{j,h_j}(u_x)\big]-g_j(u_x)\right|+\left|\widehat{g}_{j,h_j}(u_x)-\mathbb{E}\big[\widehat{g}_{j,h_j}(u_x)\big]\right|\\&\qquad-\frac{\sqrt{\widetilde{V}_j(n,h_j)}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}+\frac{\sqrt{\widetilde{V}_j(n,h_j)}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\\&\le\left\|g_j-K_{h_j}*g_j\right\|_\infty+\sqrt{\widetilde{V}_j(n,h_j)}.\end{aligned}$$

Therefore, on $\widetilde{A}_2$, we finally obtain

$$\left|\widehat{g}_{j,\widehat{h}_j}(u_x) - g_j(u_x)\right| \le (1 + 2\|K\|_{\mathbb{L}^1(\mathbb{R})}).\left\|g_j - K_{h_j} * g_j\right\|_\infty + 3.\sqrt{\widetilde{V}_j(n,h_j)}.$$

This concludes the proof of Proposition 3.1. □

5.4.2. *Proof of Theorem 3.6*

First, we establish a concentration result for $\widehat{g}_{1,\widehat{h}_1}(u_x)$ and $\widehat{g}_{2,\widehat{h}_2}(u_x)$ as follows. In the sequel, we consider $j \in \{1,2\}$ and we set

$$\psi_n(\beta_j) = \left(\frac{\log(n)}{n}\right)^{\frac{\beta_j}{2\beta_j+1}}.$$

Corollary 5.6. *Suppose that g_j belongs to $\mathcal{H}(\beta_j, L_j)$ for $\beta_j, L_j > 0$. Then, under the assumptions of Proposition 3.1, for $q \ge 1$, there exists a constant C_j depending on $\beta_j, L_j, c_{0,j}$ and K such that, with*

$$\widetilde{E}_{j,n}(u_x, \widehat{h}_j)) := \left\{\left|\widehat{g}_{j,\widehat{h}_j}(u_x) - g_j(u_x)\right| \le C_j.\psi_n(\beta_j)\right\},$$

we have, for n large enough,

$$\mathbb{P}\Big(\big(\widetilde{E}_{j,n}(u_x, \widehat{h}_j)\big)^c\Big) \le 4.n^{-q}.$$

Proof of Corollary 5.6. Since g_j belongs to $\mathcal{H}(\beta_j, L_j)$, from Lemma 5.1, we have

$$\left\|g_j - K_{h_j} * g_j\right\|_\infty \le L_j.h_j^{\beta_j}.C_{K,\mathcal{L}}.$$

From Proposition 3.1, this implies that with a probability greater than $1-4.n^{-q}$, one gets for any $h_j \in \mathcal{H}_n$:

$$\left|\widehat{g}_{j,\widehat{h}_j}(u_x) - g_j(u_x)\right| \le \big(1 + 2\|K\|_{\mathbb{L}^1(\mathbb{R})}\big).L_j.h_j^{\beta_j}.C_{K,\mathcal{L}} + 3.\sqrt{\widetilde{V}_j(n,h_j)}. \quad (36)$$

In (36), we take h_j so that h_j^{-1} is an integer and h_j is of order $\left(\frac{\log(n)}{n}\right)^{\frac{1}{2\beta_j+1}}$. Since $h_{\max} = (\log n)^{-1}$ and $1/(2\beta_j+1) < 1$, $h_j \in \mathcal{H}_n$, for n large enough. As a result, we obtain with probability greater than $1 - 4.n^{-q}$, that

$$\left|\widehat{g}_{j,\widehat{h}_j}(u_x) - g_j(u_x)\right| \le C_j.\psi_n(\beta_j),$$

with a constant C_j (depending on $\beta_j, L_j, c_{0,j}$ and K). This concludes the proof of Corollary 5.6. □

Now, we start to prove Theorem 3.6.

Proof of Theorem 3.6. We have

$$\mathbb{E}\Big[d_c\big(\widehat{m}_{\widehat{h}}(x), m(x)\big)\Big] = \mathbb{E}\Big[d_c\big(\widehat{g}_{\widehat{h}}(u_x), g(u_x)\big)\Big].$$

We study

$$R_n := \mathbb{E}\left[d_c(\widehat{g}_{\widehat{h}}(u_x), g(u_x)).1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\right].$$

We have

$$\begin{aligned} R_n &= \mathbb{E}\Big[\Big(1-\cos\big(\operatorname{atan2}(\widehat{g}_{1,\widehat{h}_1}(u_x), \widehat{g}_{2,\widehat{h}_2}(u_x)) \\ &\qquad\qquad - \operatorname{atan2}(g_1(u_x), g_2(u_x))\big)\Big).1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\ &= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}\Big(\operatorname{atan2}(\widehat{g}_{1,\widehat{h}_1}(u_x), \widehat{g}_{2,\widehat{h}_2}(u_x)) \\ &\qquad\qquad - \operatorname{atan2}(g_1(u_x), g_2(u_x))\Big)\Big).1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]. \end{aligned}$$

We now distinguish 3 cases.

Case 1: $|g_1(u_x)| > 0$ ***and*** $|g_2(u_x)| > 0$

We denote

$$\delta_1 = |g_1(u_x)| \quad \text{and} \quad \delta_2 = |g_2(u_x)|,$$

meaning that $\delta = \min(\delta_1, \delta_2)$. First, on the event $\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)$, for n large enough satisfying

$$C_2.\psi_n(\beta_2) < \delta_2/2 \quad \text{and} \quad C_1.\psi_n(\beta_1) < \delta_1/2.$$

we have

$$\big|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)\big| \le C_2.\psi_n(\beta_2) < \frac{\big|g_2(u_x)\big|}{2}$$

and

$$\big|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)\big| \le C_1.\psi_n(\beta_1) < \frac{\big|g_1(u_x)\big|}{2}.$$

Thus, we get

$$\widehat{g}_{2,\widehat{h}_2}(u_x).g_2(u_x) > 0 \quad \text{and} \quad \widehat{g}_{1,\widehat{h}_1}(u_x).g_1(u_x) > 0.$$

Therefore,

$$\begin{aligned} R_n = 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}\Big(\arctan\Big(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big) - \arctan\Big(\frac{g_1(u_x)}{g_2(u_x)}\Big)\Big)\Big) \\ .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \end{aligned}$$

$$\leq \frac{1}{2}\mathbb{E}\left[\left|\arctan\left(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right) - \arctan\left(\frac{g_1(u_x)}{g_2(u_x)}\right)\right|^2 .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\right]$$
$$\leq \mathbb{E}\Big[\Big|\arctan\left(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right) - \arctan\left(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right)\Big|^2 .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]$$
$$+ \mathbb{E}\Big[\Big|\arctan\left(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right) - \arctan\left(\frac{g_1(u_x)}{g_2(u_x)}\right)\Big|^2 .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]. \tag{37}$$

For n sufficiently large, $\big|\widehat{g}_{2,\widehat{h}_2}(u_x)\big| \geq \big|g_2(u_x)\big| - \big|g_2(u_x) - \widehat{g}_{2,h}(u_x)\big| > \delta_2 - C_2.\psi_n(\beta_2) \geq \delta_2/2$ on the event $\widetilde{E}_{2,n}(u_x,\widehat{h}_2)$, and using the 1-Lipschitz continuity of arctan, we get for the first term in (37), since on $\widetilde{E}_{1,n}(u_x,\widehat{h}_1)$ one has $\big|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)\big| \leq C_1.\psi_n(\beta_1)$, that

$$\mathbb{E}\Big[\Big|\arctan\left(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right) - \arctan\left(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right)\Big|^2 .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]$$
$$\leq \frac{4}{\delta_2^2}.\mathbb{E}\Big[\Big|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)\Big|^2 .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]$$
$$\leq \frac{4}{\delta_2^2}.\mathbb{E}\Big[C_1^2.\psi_n(\beta_1)^2.1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]$$
$$\leq \frac{4}{\delta_2^2}.C_1^2.\psi_n(\beta_1)^2.\mathbb{P}\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big)$$
$$\leq \frac{4}{\delta_2^2}.C_1^2.\psi_n(\beta_1)^2 \leq \frac{4}{\delta^2}.C_1^2.\psi_n(\beta_1)^2.$$

Moreover, for the second term in (37), since $\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}.\frac{g_1(u_x)}{g_2(u_x)} > 0$ on $\widetilde{E}_{2,n}(u_x,\widehat{h}_2)$, we have

$$\mathbb{E}\Big[\Big|\arctan\left(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right) - \arctan\left(\frac{g_1(u_x)}{g_2(u_x)}\right)\Big|^2 .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]$$
$$= \mathbb{E}\Big[\Big|\arctan\left(\frac{\widehat{g}_{2,\widehat{h}_2}(u_x)}{g_1(u_x)}\right) - \arctan\left(\frac{g_2(u_x)}{g_1(u_x)}\right)\Big|^2 .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]$$
$$\leq \frac{1}{\big|g_1(u_x)\big|^2}.\mathbb{E}\Big[\Big|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)\Big|^2 .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]$$
$$\leq \frac{1}{\delta_1^2}.\mathbb{E}\Big[C_2^2.\psi_n(\beta_2)^2.1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]$$
$$\leq \frac{1}{\delta_1^2}.C_2^2.\psi_n(\beta_2)^2 \leq \frac{1}{\delta^2}.C_2^2.\psi_n(\beta_2)^2.$$

Therefore, on the event $\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)$, for n sufficiently large such that $C_2.\psi_n(\beta_2) \leq \delta_2/2$ and $C_1.\psi_n(\beta_1) \leq \delta_1/2$, we obtain

$$\mathbb{E}\Big[\big|\widehat{g}_{\widehat{h}}(u_x) - g(u_x)\big|^2 .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]$$

$$\leq \frac{4}{\delta^2}.C_1^2.\psi_n(\beta_1)^2 + \frac{1}{\delta^2}.C_2^2.\psi_n(\beta_2)^2.$$

On the other hand, on the complementary $\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\big)^c \cup \big(\widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big)^c$, using the fact that $\big|\mathrm{atan2}(w_1,w_2)\big| \leq \pi$, $\forall (w_1,w_2)$, we can simply obtain an upper-bound as follows:

$$\begin{aligned}
&\mathbb{E}\Big[d_c(\widehat{g}_{\widehat{h}}(u_x)-g(u_x)).1_{\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\big)^c\cup\big(\widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big)^c}\Big]\\
&\leq \frac{1}{2}.\mathbb{E}\Big[\Big|\mathrm{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(u_x),\widehat{g}_{2,\widehat{h}_2}(u_x)\big)-\mathrm{atan2}\big(g_1(u_x),g_2(u_x)\big)\Big|^2\\
&\qquad\qquad\qquad\qquad .1_{\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\big)^c\cup\big(\widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big)^c}\Big]\\
&\leq 2\pi^2.\mathbb{P}\left(\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\big)^c\right)+2\pi^2.\mathbb{P}\left(\big(\widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big)^c\right)\leq 4\pi^2.4.n^{-q},
\end{aligned}$$

by Corollary 5.6. For $q \geq 1$, we get that n^{-q} is negligible in comparison with $C_1^2.\psi_n(\beta_1)^2 = C_1^2.\Big(\frac{\log(n)}{n}\Big)^{\frac{2\beta_1}{2\beta_1+1}}$ and $C_2^2.\psi_n(\beta_2)^2 = C_2^2.\Big(\frac{\log(n)}{n}\Big)^{\frac{2\beta_2}{2\beta_2+1}}$.

Case 2: $g_1(u_x) = 0$ and $|g_2(u_x)| > 0$

In this case $\delta = |g_2(u_x)|$.

– If $g_2(u_x) > 0$, then, as previously, on the event $\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)$, for n large enough, $\widehat{g}_{2,\widehat{h}_2}(u_x) > 0$. Then,

$$\begin{aligned}
R_n &= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\mathrm{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(u_x),\widehat{g}_{2,\widehat{h}_2}(u_x)\big)-\mathrm{atan2}\big(g_1(u_x),g_2(u_x)\big)\Big)\Big)\\
&\qquad\qquad .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]\\
&= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\mathrm{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(u_x),\widehat{g}_{2,\widehat{h}_2}(u_x)\big)-0\Big)\Big)\\
&\qquad\qquad .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]\\
&= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\arctan\Big(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big)-\arctan\Big(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big)\Big)\Big)\\
&\qquad\qquad .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]\\
&\leq \frac{1}{2}.\mathbb{E}\Big[\Big|\arctan\Big(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big)-\arctan\Big(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big)\Big|^2\\
&\qquad\qquad .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big],
\end{aligned}$$

and we conclude as for the first case.

– If $g_2(u_x) < 0$, then, as previously, on the event $\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)$, for n large enough, $\widehat{g}_{2,\widehat{h}_2}(u_x) < 0$. Then,

$$
\begin{aligned}
R_n &= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}\Big(\text{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(u_x), \widehat{g}_{2,\widehat{h}_2}(u_x)\big) - \text{atan2}\big(g_1(u_x), g_2(u_x)\big)\Big)\Big) \\
&\qquad .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap \widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\text{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(u_x), \widehat{g}_{2,\widehat{h}_2}(u_x)\big) + \pi\Big)\Big) \\
&\qquad .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap \widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\arctan\Big(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big) + 2.\pi.1_{\{\widehat{g}_{1,\widehat{h}_1}(u_x)>0\}} \\
&\qquad - \arctan\Big(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big)\Big)\Big).1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap \widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\arctan\Big(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big) - \arctan\Big(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big)\Big)\Big) \\
&\qquad .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap \widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&\le \frac{1}{2}.\mathbb{E}\Big[\Big|\arctan\Big(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big) - \arctan\Big(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big)\Big|^2 \\
&\qquad .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap \widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big],
\end{aligned}
$$

and we conclude as for the first case.

Case 3: $|g_1(u_x)| > 0$ and $g_2(u_x) = 0$

In this case $\delta = |g_1(u_x)|$.

– If $g_1(u_x) > 0$, then, as previously, on the event $\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)$, for n large enough, $\widehat{g}_{1,\widehat{h}_1}(u_x) > 0$. Then,

$$
\begin{aligned}
R_n &= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\text{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(u_x), \widehat{g}_{2,\widehat{h}_2}(u_x)\big) - \text{atan2}\big(g_1(u_x), g_2(u_x)\big)\Big)\Big) \\
&\qquad .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap \widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\arctan\Big(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big) + \pi.1_{\{\widehat{g}_{2,\widehat{h}_2}(u_x)<0\}} - \frac{\pi}{2}\Big)\Big) \\
&\qquad .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap \widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]
\end{aligned}
$$

$$= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\arctan\Big(\frac{\widehat{g}_{2,\widehat{h}_2}(u_x)}{\widehat{g}_{1,\widehat{h}_1}(u_x)}\Big) - \arctan\Big(\frac{g_2(u_x)}{\widehat{g}_{1,\widehat{h}_1}(u_x)}\Big)\Big)\Big) .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big],$$

where the last equality is obtained by using for $x \in \mathbb{R}^*$,

$$\arctan(x) + \arctan(1/x) = \pi/2 \times sign(x)$$

and by distinguishing the cases according to the sign of $\widehat{g}_{2,\widehat{h}_2}(u_x)$.

– If $g_1(u_x) < 0$, then, as previously, on the event $\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)$, for n large enough, $\widehat{g}_{1,\widehat{h}_1}(u_x) < 0$. Then, similarly,

$$\begin{aligned} R_n &= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\text{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(u_x),\widehat{g}_{2,\widehat{h}_2}(u_x)\big) - \text{atan2}\big(g_1(u_x),g_2(u_x)\big)\Big)\Big) .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\ &= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\arctan\left(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right) - \pi.1_{\{\widehat{g}_{2,\widehat{h}_2}(u_x)<0\}} + \frac{\pi}{2}\Big)\Big) .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\ &= 2.\mathbb{E}\Big[\sin^2\Big(\frac{1}{2}.\Big(\arctan\Big(\frac{\widehat{g}_{2,\widehat{h}_2}(u_x)}{\widehat{g}_{1,\widehat{h}_1}(u_x)}\Big) - \arctan\Big(\frac{g_2(u_x)}{\widehat{g}_{1,\widehat{h}_1}(u_x)}\Big)\Big)\Big) .1_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]. \end{aligned}$$

We conclude by using the second case since $\widehat{g}_{1,\widehat{h}_1}(u_x)$ (resp. $g_1(u_x)$) and $\widehat{g}_{2,\widehat{h}_2}(u_x)$ (resp. $g_2(u_x)$) play a symmetric role.
Note that under Assumption 3.3, the case $g_1(u_x) = g_2(u_x) = 0$ cannot occur. □

5.4.3. Proof of Theorem 3.11

Before tackling the proof of Theorem 3.11, the next lemma shows that the von Mises density satisfies condition (16).

Lemma 5.7. *The von Mises density with location parameter μ and concentration parameter κ satisfies condition (16).*

Proof of Lemma 5.7. We recall the expression of the von Mises density with location parameter $\mu \in [-\pi,\pi)$ and concentration parameter $\kappa > 0$:

$$f_{vM(\mu,\kappa)}(\theta) = c(\kappa).e^{\kappa\cos(\theta-\mu)}, \quad \theta \in [-\pi,\pi),$$

with $c(\kappa)$ the normalizing constant. Let us prove that $f_{vM(\mu,\kappa)}$ satisfies:

$$\exists\, p_* > 0 : \int_{-\pi}^{\pi} f_{vM(\mu,\kappa)}(t)\log\frac{f_{vM(\mu,\kappa)}(t)}{f_{vM(\mu,\kappa)}(t+\theta)}dt \le p_*\theta^2,$$

for all $\theta \in \mathbb{R}$. We have that

$$
\begin{aligned}
&\int_{-\pi}^{\pi} f_{vM(\mu,\kappa)}(t) \log \frac{f_{vM(\mu,\kappa)}(t)}{f_{vM(\mu,\kappa)}(t+\theta)} dt \\
&= c(\kappa).\kappa. \int_{-\pi}^{\pi} e^{\kappa \cos(t-\mu)}.\big(\cos(t-\mu) - \cos(t+\theta-\mu)\big) dt \\
&= 2.c(\kappa).\kappa.\big(\sin\frac{\theta}{2}\big). \int_{-\pi}^{\pi} e^{\kappa \cos(t-\mu)}. \sin\big(t-\mu+\frac{\theta}{2}\big) dt \\
&= 2.c(\kappa).\kappa.\big(\sin\frac{\theta}{2}\big) \int_{-\pi}^{\pi} e^{\kappa \cos(t-\mu)}.\Big(\sin(t-\mu).\cos\frac{\theta}{2} + \sin\frac{\theta}{2}.\cos(t-\mu)\Big) dt \\
&= 2.c(\kappa).\kappa.\Big(\sin\frac{\theta}{2}\Big)^2 \underbrace{\int_{-\pi}^{\pi} e^{\kappa \cos(t-\mu)}. \cos(t-\mu) dt}_{=:C(\kappa)>0} \\
&\leq 2.c(\kappa).\kappa.C(\kappa).\frac{\theta^2}{4},
\end{aligned}
$$

for all $\theta \in \mathbb{R}$. Then, with

$$
p_* = \frac{c(\kappa)\kappa C(\kappa)}{2},
$$

we have for any $\theta \in \mathbb{R}$,

$$
\int_{-\pi}^{\pi} f_{vM(\mu,\kappa)}(t) \log \frac{f_{vM(\mu,\kappa)}(t)}{f_{vM(\mu,\kappa)}(t+\theta)} dt \leq p_* \theta^2. \qquad \square
$$

Proof of Theorem 3.11. To prove the lower bound stated in Theorem 3.11, we follow the lines of Section 2.5 in [27] for the regression at a point. The differences with our problem lie in the circular response and the randomness of the X_i's. We consider $m_0(t) = 0$ and $m_1(t) = Lh_n^\beta K(\frac{t-x}{h_n})$ with $h_n = c_0 n^{-\frac{1}{2\beta+1}}$ and $K : \mathbb{R} \longmapsto \mathbb{S}^1$ satisfying:

$$
K \in \tilde{\Sigma}(\beta, 1/2) \cap C^\infty(\mathbb{R}), \quad K(t) > 0 \iff t \in]-1/2, 1/2[.
$$

Such functions K exist. For instance, for a sufficiently small $a > 0$, one can take

$$
K(t) = a.\exp\left(-\frac{1}{1-4t^2}\right).1_{[-0.5;0.5]}(t).
$$

We have now to check three points which are developed in the sequel.

1. Let us prove that $m_1 \in \tilde{\Sigma}(\beta, L)$ (the function m_0 obviously belongs to $\tilde{\Sigma}(\beta, L)$). For $l = \lfloor \beta \rfloor$ we have

$$
m_1^{(l)}(t) = Lh_n^{\beta-l} K^{(l)}\left(\frac{t-x}{h_n}\right),
$$

then, with $u = \frac{t-x}{h_n}$ and $u' = \frac{t'-x}{h_n}$,

$$
d_c(m_1^{(l)}(t), m_1^{(l)}(t')) = 1 - \cos(m_1^{(l)}(t) - m_1^{(l)}(t'))
$$

$$
\begin{aligned}
&\le 2\sin^2((m_1^{(l)}(t) - m_1^{(l)}(t'))/2) \\
&\le \frac{1}{2}|m_1^{(l)}(t) - m_1^{(l)}(t')|^2 \\
&= \frac{L^2}{2} h_n^{2(\beta-l)} |K^{(l)}(u) - K^{(l)}(u')|^2 \\
&\le \frac{L^2}{8} h_n^{2(\beta-l)} |u-u'|^{2(\beta-l)} = \frac{L^2}{8}|t-t'|^{2(\beta-l)}.
\end{aligned}
$$

Then, $m_1 \in \tilde{\Sigma}(\beta, L)$.

2. Let us show that $d_c(m_0(x), m_1(x)) \ge 4s_n^2$.
   We have that $m_1(x) = Lh_n^\beta K(0) = Lc_0^\beta K(0) n^{-\frac{\beta}{2\beta+1}}$, hence for n sufficiently large, $m_1(x) \in [0, \frac{\pi}{2}]$. Hence using that $\sin(t) \ge \frac{2}{\pi}t$ for $t \in [0, \frac{\pi}{2}]$, we get

$$
\begin{aligned}
d_c(m_0(x), m_1(x)) &= 1 - \cos(m_1(x)) = 2\sin^2(m_1(x)/2) \\
&\ge 2\left(\frac{2}{\pi}\right)^2 \frac{m_1^2(x)}{4} \\
&= \frac{2}{\pi^2} L^2 c_0^{2\beta} K^2(0) n^{-\frac{2\beta}{2\beta+1}}
\end{aligned}
$$

   then the condition is fulfilled with $s_n = \frac{1}{\sqrt{2\pi}} Lc_0^\beta K(0) n^{-\frac{\beta}{2\beta+1}} =: A\psi_n$.

3. Using the classical reduction to a two test hypotheses problem for the pointwise regression problem, we get for any estimator T_n:

$$
\begin{aligned}
&\sup_{m \in \tilde{\Sigma}(\beta,L)} \mathbb{E}_m[\psi_n^{-2} d_c(T_n, m(x))] \qquad (38) \\
&\ge A^2 \max_{m \in \{m_0, m_1\}} \mathbb{P}_m(d_c(T_n, m(x)) \ge A^2 {\psi_n}^2) \\
&\ge \frac{A^2}{2} \mathbb{E}_{X_1,\dots,X_n}\Big[\mathbb{P}_{m_0}(d_c(T_n, m_0(x)) \ge A^2\psi_n^2 | X_1, \dots, X_n) \\
&\quad + \mathbb{P}_{m_1}(d_c(T_n, m_1(x)) \ge A^2\psi_n^2 | X_1, \dots, X_n)\Big] \\
&\ge \frac{A^2}{2} \mathbb{E}_{X_1,\dots,X_n}\Big[\inf_\psi \Big\{\mathbb{P}_{m_0}(\psi \ne 0 | X_1, \dots, X_n) \\
&\quad + \mathbb{P}_{m_1}(\psi \ne 1 | X_1, \dots, X_n)\Big\}\Big], \qquad (39)
\end{aligned}
$$

   where $\inf_\psi$ denotes the infimum over all tests ψ taking values in $\{0,1\}$. We have used that $\sqrt{d_c}$ is a true distance on $\mathbb{S}^1$, so that it satisfies the triangular inequality.
   Now let us fix the X_i's. The minimum average probability $\overline{p}_{e,1}$ is defined as (see page 116 in [27]):

$$
\overline{p}_{e,1} := \frac{1}{2}\inf_\psi \Big\{\mathbb{P}_{m_0}(\psi \ne 0 | X_1, \dots, X_n) + \mathbb{P}_{m_1}(\psi \ne 1 | X_1, \dots, X_n)\Big\}.
$$

We have for the Kullback Leibler divergence (still with the X_i's fixed)

$$\mathbf{K}(\mathbb{P}_{m_0}, \mathbb{P}_{m_1}) = \int \log\left(\frac{d\mathbb{P}_{m_0}}{d\mathbb{P}_{m_1}}\right) d\mathbb{P}_{m_0} = \sum_{i=1}^{n} \int \log \frac{p_\zeta(y)}{p_\zeta(y - m_1(X_i))} p_\zeta(y) dy. \tag{40}$$

There exists n_0 such that $\forall n > n_0$, $Lh_n^\beta K_{\max} \leq y_0$ where $K_{\max} = \max_t |K(t)|$. Using (40) and (16), we have:

$$\begin{aligned}\mathbf{K}(\mathbb{P}_{m_0}, \mathbb{P}_{m_1}) &\leq p_* \sum_{i=1}^{n} m_1^2(X_i) \\ &\leq p_* L^2 h_n^{2\beta} K_{\max}^2 \sum_{i=1}^{n} 1_{\left\{\left|\frac{X_i - x}{h_n}\right| \leq \frac{1}{2}\right\}}.\end{aligned}$$

Now taking the expectation and using that the density of the X_i's is bounded by μ_0, we get:

$$\begin{aligned}\mathbb{E}_{X_1,\ldots,X_n} \mathbf{K}(\mathbb{P}_{m_0}, \mathbb{P}_{m_1}) &\leq p_* L^2 K_{\max}^2 h_n^{2\beta} n \mathbb{P}\left(\left|\frac{X_1 - x}{h_n}\right| \leq \frac{1}{2}\right) \\ &\leq p_* L^2 K_{\max}^2 \mu_0 h_n^{2\beta+1} n.\end{aligned}$$

For $\alpha < 2\log(2)$, since $h_n = c_0 n^{-\frac{1}{2\beta+1}}$, setting

$$c_0 = \left(\frac{\alpha}{p_* \mu_0 L^2 K_{\max}^2}\right)^{\frac{1}{2\beta+1}},$$

we get that

$$\mathbb{E}_{X_1,\ldots,X_n} \mathbf{K}(\mathbb{P}_{m_0}, \mathbb{P}_{m_1}) \leq \alpha.$$

As in Lemma 2.10 of [27], we introduce the function $\mathcal{H}(t) = -t\log(t) - (1-t)\log(1-t)$ for $t \in (0,1)$ and $\mathcal{H}(0) = \mathcal{H}(1) = 0$. Inequality (2.70) of [27] with $M = 1$ gives

$$\mathbb{E}_{X_1,\ldots,X_n}[\mathcal{H}(\overline{p}_{e,1})] \geq \log(2) - \frac{1}{2}\mathbb{E}_{X_1,\ldots,X_n} \mathbf{K}(\mathbb{P}_{m_0}, \mathbb{P}_{m_1}) \geq \log(2) - \frac{\alpha}{2},$$

since $\mathbb{E}_{X_1,\ldots,X_n} \mathbf{K}(\mathbb{P}_{m_0}, \mathbb{P}_{m_1}) \leq \alpha$. Since $\mathcal{H}$ is concave, $\mathcal{H}(\mathbb{E}_{X_1,\ldots,X_n}[\overline{p}_{e,1}]) \geq \mathbb{E}_{X_1,\ldots,X_n}[\mathcal{H}(\overline{p}_{e,1})]$ and

$$\mathbb{E}_{X_1,\ldots,X_n}[\overline{p}_{e,1}] \geq \mathcal{H}^{-1}\big(\log 2 - \frac{\alpha}{2}\big) > 0,$$

with, for any $t > 0$, $\mathcal{H}^{-1}(t) = \min\{p \in (0, \frac{1}{2}] : \mathcal{H}(p) \geq t\}$. Hence we deduce using (38)

$$\sup_{m \in \tilde{\Sigma}(\beta, L)} \mathbb{E}_m\big[\psi_n^{-2} |d_c(T_n, m(x))\big] \geq A^2 \mathcal{H}^{-1}\big(\log 2 - \frac{\alpha}{2}\big),$$

where the right hand side is a positive constant. This concludes the proof of Theorem 3.11. $\square$

6. Conclusion

Considering nonparametric regression for circular data, we derive minimax convergence rates and prove near optimal properties of our kernel estimate combined with a warping strategy on anisotropic Hölder classes of functions for pointwise estimation. The bandwidth parameter is selected by using a data-driven Goldenshluger-Lepski type procedure. After tuning hyperparameters of our estimate, we show that it remains very competitive with respect to existing methods.

As a natural extension, it could be very challenging to investigate our regression problem with a response on the sphere $\mathbb{S}^2$ or more generally on the unit hypersphere $\mathbb{S}^{d-1}$. The case of predictors $X \in \mathbb{S}^{d-1}$ and a response $\Theta \in \mathbb{S}^{d-1}$ has been tackled in [10]. The spherical context is of course more complicated than the circular one and the arctangent function approach used here is not easily generalizable in the spherical setting. In [10], Di Marzio et al. proposed a local constant estimator by smoothing on each component of the response. Once again no rates of convergence were obtained. Hence, in a future work, a first task would be to obtain convergence rates and then investigate adaptation issue.

Acknowledgments

The authors would like to warmly thank the Associate Editor and the anonymous referees for very valuable comments and suggestions. Parts of this work were conducted during T.D.N.'s PhD study at Laboratoire de Mathématiques d'Orsay, UMR 8628, Université Paris-Saclay, 91405 Orsay, France.

References

[1] M. Alonso-Pena and R. Crujeiras. Nonparametric multimodal regression for circular data. Submitted, arXiv:2012.09915, 2020.

[2] N. Bochkina and T. Sapatinas. Minimax rates of convergence and optimality of Bayes factor wavelet regression estimators under pointwise risks. *Statistica Sinica*, 19: 1389–1406, 2009. MR2589188

[3] G. Chagny. Penalization versus Goldenshluger-Lepski strategies in warped bases regression. *ESAIM Probab. Stat.*, 17: 328–358, 2013. MR3066383

[4] G. Chagny. Adaptive warped kernel estimators. *Scandinavian Journal of Statistics*, 42(2): 336–360, 2015. MR3345108

[5] G. Chagny, T. Laloë and R. Servien. Multivariate adaptive warped kernel estimation. *Electron. J. Statist.*, 13(1): 1759–1789, 2019. MR3959872

[6] B. Charlier. Necessary and sufficient condition for the existence of a Fréchet mean on the circle. *ESAIM Probab. Stat.*, 17: 635–649, 2013. MR3126155

[7] C. Chesneau and T. Willer. Estimation of a cumulative distribution function under interval censoring "case 1" via warped wavelets. *Comm. Statist. Theory Methods*, 44(17): 3680–3702, 2015. MR3397157

[8] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.*, 49(2): 569–609, 2013. MR3088382

[9] N.I. Fisher and A.J. Lee. Regression models for angular responses. *Biometrics*, 48: 665–677, 1992. MR1187598

[10] M. Di Marzio, A. Panzera and C.C. Taylor. Nonparametric Regression for Spherical Data. *Journal of the American Statistical Association*, 109: 748–763, 2014. MR3223747

[11] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist*, 39(3): 1608–1632, 2011. MR2850214

[12] A.L. Gould. A regression technique for angular variates. *Biometrics*, 25: 683–700, 1969.

[13] S.R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*. World Scientific, Singapore, 2001. MR1836122

[14] R.A. Johnson and T.E. Wehlry. Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73: 602–606, 1978. MR0514163

[15] G. Kerkyacharian and D. Picard. Regression in random design and warped wavelets. *Bernoulli*, 10(6): 1053–1105, 2004. MR2108043

[16] O.V. Lepski One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, 35: 459–470, 1990. MR1091202

[17] O.V. Lepski and V. G. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, 25(6): 2612–2546, 1997. MR1604408

[18] C. Ley and T. Verdebout. *Modern Directional Statistics* (1st ed.). Chapman and Hall/CRC, 2017. MR3752655

[19] K.V. Mardia and P.E. Jupp. *Directional Statistics*. John Wiley, New York, NY, 2000. MR1828667

[20] M. Di Marzio, A. Panzera and C.C. Taylor. Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, 40: 238–255, 2013. MR3066413

[21] P. Massart. The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *Ann. Probab.*, 18(3): 1269–1283, 1990. MR1062069

[22] A. Meilán-Vila, M. Francisco-Fernández, R.M. Crujeiras and A. Panzera. Nonparametric multiple regression estimation for circular response. *TEST*, 30: 650–672, (2021). MR4297272

[23] A. Pewsey and E. García-Portugués. Recent advances in directional statistics. *TEST*, 30: 1–58, 2021. MR4242171

[24] T.M. Pham Ngoc. Regression in random design and Bayesian warped wavelets estimators. *Electron. J. Stat*, 3: 1084–1112, 2009. MR2566182

[25] T.M. Pham Ngoc. Adaptive optimal kernel density estimation for directional data. *J. Multivariate Anal.*, 173: 248–267, 2019. MR3923431

[26] B. Presnell, S.P. Morrison and R.C. Littel. Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, 93: 1068–1077, 1998. MR1649201

[27] A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009. MR2724359