



**HAL**  
open science

# Adaptive warped kernel estimation for nonparametric regression with circular responses

Tien-Dat Nguyen, Thanh Mai Pham Ngoc, Vincent Rivoirard

► **To cite this version:**

Tien-Dat Nguyen, Thanh Mai Pham Ngoc, Vincent Rivoirard. Adaptive warped kernel estimation for nonparametric regression with circular responses. 2022. hal-03633917v1

**HAL Id: hal-03633917**

**<https://hal.science/hal-03633917v1>**

Preprint submitted on 7 Apr 2022 (v1), last revised 27 Dec 2023 (v2)

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

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

# Adaptive warped kernel estimation for nonparametric regression with circular responses

Tien-Dat Nguyen\*, Thanh Mai Pham Ngoc†, Vincent Rivoirard‡

April 1, 2022

## 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.

**Keywords**— circular data, nonparametric regression, warping method, kernel rule, adaptive minimax estimation, Goldenshluger-Lepski procedure.

## 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 [15], Jammalamadaka and SenGupta [11], Ley and Verdebout [14] and recent advances are collected in Pewsey and García-Portugués [18]. 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. Assume that we have an independent identically distributed (i.i.d in the sequel) sample  $\{(X_i, \Theta_i)\}_{i=1}^n$  distributed as  $(X, \Theta)$ , where  $\Theta$  is a circular random variable, and  $X$  is a random variable with density  $f_X$  supported on  $\mathbb{R}$  such that the cumulative distribution function of  $X$  is invertible on  $\mathbb{R}$ . It means 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  $\Theta_i$ .

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 [10], Johnson and Wehly [12] or Fisher and

---

\*Laboratoire de Mathématiques, UMR 8628, Université Paris Sud, 91405 Orsay Cedex France; E-mail: tien-dat.nguyen@math.u-psud.fr

†Laboratoire de Mathématiques, UMR 8628, Université Paris Sud, 91405 Orsay Cedex France; E-mail: thanh.pham.ngoc@math.u-psud.fr

‡Ceremade, CNRS, UMR 7534, Université Paris-Dauphine, PSL Research University, 75016 Paris, France; E-mail: Vincent.Rivoirard@dauphine.fr

Lee [7]. 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. [21]. Presnell et al. [21] 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 [18]. In order to get a more flexible approach, nonparametric paradigm has been considered, first in the pioneering work by Di Marzio et al. [16] and more recently in Meilán-Vila et al. [17] 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 condition 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 [16] or [17] consist in estimating the arctangent function of the ratio of the trigonometric moments of  $\Theta$  (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. [16] 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. [17], 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.* Assuming that the cumulative distribution function (c.d.f.) of the design  $X$  is known and invertible, warping methods used in this paper consists 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 [9]. 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 3.5 and 3.10 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  $\theta = \text{atan2}(w_1, w_2)$  defined as follows:

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

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

with  $\arctan$  taking values in  $[-\pi/2, \pi/2]$ . In particular  $\arctan(+\infty) = \pi/2$  and  $\arctan(-\infty) = -\pi/2$ .

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  $\text{atan2}(0, 0)$

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:

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

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

## 2 The estimation procedure

We denote  $F_X : y \in \mathbb{R} \mapsto \mathbb{P}(X \leq y)$  the cumulative distribution function (c.d.f.) of  $X$  and in the sequel, we assume that  $F_X$  is known and is invertible, so that  $F_X(\mathbb{R}) = (0, 1)$ .

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, we use the angular distance

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

Notice that this angular distance corresponds to the usual Euclidean norm in  $\mathbb{R}^2$ . Indeed, the angles  $\theta_1$  and  $\theta_2$  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

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

Hence, using the angular distance, we look for a function  $m$  minimizing the following risk:

$$L(m(X)) := \mathbb{E}[1 - \cos(\Theta - m(X)) | X].$$

Minimizing  $\mathbb{E}[1 - \cos(\Theta - m(X)) | X]$  with respect to  $m(X) \in [-\pi, \pi)$  is equivalent to maximizing  $\mathbb{E}[\cos(\Theta - m(X)) | X]$  with respect to  $m(X) \in [-\pi, \pi)$ . Now for  $x \in \mathbb{R}$ , let

$$m_1(x) := \mathbb{E}(\sin(\Theta) | X = x) \quad \text{and} \quad m_2(x) := \mathbb{E}(\cos(\Theta) | X = x). \quad (1)$$

Moreover, write

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

where  $\gamma \in [-\pi, \pi)$  is defined by

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

Thus, we have

$$\operatorname{argmin}_{m(X) \in [-\pi, \pi)} L(m(X)) = \operatorname{argmax}_{m(X) \in [-\pi, \pi)} \cos(m(X) - \gamma) = \gamma = \operatorname{atan2}(m_1(X), m_2(X)). \quad (2)$$

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  $\Theta$  given  $X$  and we tackle the problem by estimating the function

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

with  $m_1$  and  $m_2$  defined in (1). For this purpose, we use warped estimators.

## 2.2 Warping strategy

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

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

with  $K : \mathbb{R} \rightarrow \mathbb{R}$  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.

Warping methods consist in introducing the auxiliary function  $g := m \circ F_X^{\leq -1}$ , with  $F_X^{\leq -1}$  the inverse of  $F_X$ . The strategy then boils 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 Kerkycharian and Picard [13], Pham Ngoc [19], 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. 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  $g_2$ .

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 2.1.** Let  $K : \mathbb{R} \rightarrow \mathbb{R}$  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  $\int_{\mathbb{R}} K(y) dy = 1$ .

Now, let  $g_1, g_2 : (0, 1) \mapsto \mathbb{R}$  be defined by

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

so that  $m_1 = g_1 \circ F_X$  and  $m_2 = g_2 \circ F_X$ . Furthermore, setting for  $u \in (0, 1)$ ,

$$g(u) := \text{atan2}(g_1(u), g_2(u)),$$

we observe that  $m = g \circ F_X$ . Now, for  $u \in (0, 1)$ , we estimate  $g_1(u)$  and  $g_2(u)$  by

$$\widehat{g}_{1,h_1}(u) := \frac{1}{n} \sum_{i=1}^n \sin(\Theta_i) \cdot K_{h_1}(u - F_X(X_i)) \quad \text{and} \quad \widehat{g}_{2,h_2}(u) := \frac{1}{n} \sum_{i=1}^n \cos(\Theta_i) \cdot K_{h_2}(u - F_X(X_i)) \quad (3)$$

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

Now, we estimate  $g$  by

$$\widehat{g}_h(u) := \text{atan2}(\widehat{g}_{1,h_1}(u), \widehat{g}_{2,h_2}(u)), \quad u \in (0, 1), \quad (4)$$

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

$$\widehat{m}_{1,h_1}(x) := \widehat{g}_{1,h_1}(F_X(x)) = \frac{1}{n} \sum_{i=1}^n \sin(\Theta_i) \cdot K_{h_1}(F_X(x) - F_X(X_i)), \quad (5)$$

and

$$\widehat{m}_{2,h_2}(x) := \widehat{g}_{2,h_2}(F_X(x)) = \frac{1}{n} \sum_{i=1}^n \cos(\Theta_i) \cdot K_{h_2}(F_X(x) - F_X(X_i)). \quad (6)$$

We then obtain an estimator of  $m(x)$  at  $x \in \mathbb{R}$  by setting

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

### 2.3 Bandwidth selection

To study the pointwise quadratic-risk of the estimator  $\widehat{m}_h(x)$  whose expression is

$$\mathbb{E}\left[\left(\widehat{m}_h(x) - m(x)\right)^2\right] = \mathbb{E}\left[\left(\widehat{g}_h(F_X(x)) - g(F_X(x))\right)^2\right],$$

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

$$\mathcal{H}_n := \left\{ h = k^{-1} : k \in \mathbb{N}^*, h \leq h_{\max}, n.h > \max\left(\frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}^2}{\|K\|_{\infty}^2}; 1\right) \cdot \log(n) \right\}. \quad (7)$$

We have  $\text{Card}(\mathcal{H}_n) \approx n/\log n$ . In the following, we apply the method proposed by Goldenshluger and Lepski in [9] 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

$$A_j(h_j, v) := \sup_{h'_j \in \mathcal{H}_n} \left\{ |\widehat{g}_{j, h_j, h'_j}(v) - \widehat{g}_{j, h'_j}(v)| - \sqrt{\widetilde{V}_j(n, h'_j)} \right\}_+, \quad (8)$$

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

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

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

$$\widehat{h}_j = \underset{h_j \in \mathcal{H}_n}{\text{argmin}} \left\{ A_j(h_j, v) + \sqrt{\widetilde{V}_j(n, h_j)} \right\}. \quad (9)$$

The criterion (9) is inspired from [9], in order to mimic the optimal "bias-variance" trade-off in the decomposition

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

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

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

(see (22)), so it is natural to estimate it by an estimator of the form  $|(K_{h_j} * \widehat{g}_{j, h'_j})(v) - \widehat{g}_{j, h'_j}(v)|$ . Thus, the estimator of the bias term is  $A_j(h_j, v)$ , defined in (8), 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:

$$\widehat{g}_{\widehat{h}}(v) := \text{atan2}(\widehat{g}_{1, \widehat{h}_1}(v), \widehat{g}_{2, \widehat{h}_2}(v)), \quad (10)$$

where we denote  $\widehat{h} := (\widehat{h}_1, \widehat{h}_2)$ , and

$$\widehat{g}_{1, \widehat{h}_1}(v) := \frac{1}{n} \sum_{i=1}^n \sin(\Theta_i) \cdot K_{\widehat{h}_1}(v - F_X(X_i)), \quad \text{and} \quad \widehat{g}_{2, \widehat{h}_2}(v) := \frac{1}{n} \sum_{i=1}^n \cos(\Theta_i) \cdot K_{\widehat{h}_2}(v - F_X(X_i)). \quad (11)$$

As a consequence, at  $x \in \mathbb{R}$ , the adaptive estimators for  $m_1$  and  $m_2$  are

$$\widehat{m}_{1, \widehat{h}_1}(x) := (\widehat{g}_{1, \widehat{h}_1} \circ F_X)(x) = \frac{1}{n} \sum_{i=1}^n \sin(\Theta_i) \cdot K_{\widehat{h}_1}(F_X(x) - F_X(X_i)), \quad (12)$$

and

$$\widehat{m}_{2, \widehat{h}_2}(x) := (\widehat{g}_{2, \widehat{h}_2} \circ F_X)(x) = \frac{1}{n} \sum_{i=1}^n \cos(\Theta_i) \cdot K_{\widehat{h}_2}(F_X(x) - F_X(X_i)). \quad (13)$$

We finally define the adaptive estimator for  $m(x)$  at  $x \in \mathbb{R}$  by

$$\widehat{m}_{\widehat{h}}(x) := \text{atan2}(\widehat{m}_{1, \widehat{h}_1}(x), \widehat{m}_{2, \widehat{h}_2}(x)). \quad (14)$$

### 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  $\widehat{g}_{\widehat{h}} := \text{atan2}(\widehat{g}_{1,\widehat{h}_1}, \widehat{g}_{2,\widehat{h}_2})$  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 (7). Let  $j \in \{1, 2\}$  and  $q \geq 1$  and assume that  $\min\{c_{0,1}; c_{0,2}\} \geq 16(2+q)^2 \cdot (1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})^2$ . Then, for  $n$  sufficiently large, with a probability larger than  $1 - 4.n^{-q}$ ,*

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

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.

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

**Assumption 3.2.** Assume that

$$|m_1(x)| = |g_1(F_X(x))| \geq \delta_1 > 0, \quad \text{and} \quad |m_2(x)| = |g_2(F_X(x))| \geq \delta_2 > 0.$$

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

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

We also consider the following assumption on the kernel  $K$ :

**Assumption 3.4.** The kernel  $K$  is of order  $\mathcal{L} \in \mathbb{R}_+$ , i.e.

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

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

**Theorem 3.5.** *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 3.4 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  $\min\{c_{0,1}; c_{0,2}\} \geq 16(2+q)^2 \cdot (1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})^2$ . Then, under Assumption 3.2, for  $n$  sufficiently large,*

$$\mathbb{E} \left[ |\widehat{m}_{\widehat{h}}(x) - m(x)|^2 \right] \leq \left( \frac{1}{\delta_1^2} + \frac{1}{\delta_2^2} \right) \cdot (C_1^2 + C_2^2) \cdot \max\{\psi_n^2(\beta_1), \psi_n^2(\beta_2)\},$$

where

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

$C_1$  is a constant depending on  $\beta_1, L_1, c_{0,1}, K$  and  $C_2$  is a constant depending on  $\beta_2, L_2, c_{0,2}, K$ .

A proof of Theorem 3.5 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]). In this case, the obtained rate is much faster than the rate  $(\log n/n)^{\beta/(2\beta+2)}$  obtained for bivariate regression function estimation in the isotropic case. This justifies the introduction of specific methodologies for circular regression.

**Remark 3.6.** Eventually, to obtain the fully computable estimator, we replace the c.d.f.  $F_X$  by its natural estimate  $\widehat{F}_n(y) := n^{-1} \sum_{i=1}^n 1_{X_i \leq y}$ . Following arguments of [4], this replacement should not change the final rate of convergence of our nonparametric estimator.



## 3.2 Minimax lower bounds

To establish minimax lower bounds, we need to consider a statistical model which relates the  $\Theta_i$ 's and  $X_i$ 's, namely we assume:

$$\Theta_i = m(X_i) + \zeta_i \pmod{2\pi}, \quad i = 1, \dots, n \quad (15)$$

where  $m$  is a circular regression function and the  $\zeta_i$ 's are centered i.i.d. random angles, independent of the  $X_i$ 's. We will assume that Model (15) satisfies the following assumption.

**Assumption 3.7.** *The design points  $X_i$ 's are i.i.d. random variables with density  $\mu(\cdot)$  on  $[0, 1]$  such that there exists  $\mu_0 < \infty$  and  $\mu(x) \leq \mu_0 \forall x \in [0, 1]$  and the errors  $\zeta_i$  have common density  $p_\zeta(\cdot)$  on  $\mathbb{S}^1$  with respect to the Lebesgue measure on  $\mathbb{S}^1$ , verifying*

$$\exists p_* > 0, \exists y_0 > 0 : \int p_\zeta(u) \log \frac{p_\zeta(u)}{p_\zeta(u+y)} du \leq p_* y^2, \quad \forall |y| \leq y_0. \quad (16)$$

In the sequel, the function  $m$  belongs to the class  $\tilde{\Sigma}(\beta, L)$  defined as the set of functions  $f : [0, 1] \mapsto \mathbb{S}^1$  such that the derivative  $f^{(l)}$ ,  $l = \lfloor \beta \rfloor$  exists and verifies

$$|f^{(l)}(x) - f^{(l)}(y)| \leq L|x - y|^{\beta-l}, \quad \forall x, y \in [0, 1].$$

**Remark 3.8.** *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 3.9.** *The classical von Mises density satisfies condition (16). See Equation (20) where we recall the definition of the von Mises density.*

We obtain the following lower bound:

**Theorem 3.10.** *Let  $\beta > 0$  and  $L > 0$ . Under Assumptions 3.7, we have*

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

where  $c$  depends only on  $\beta, L, p_*$  and  $\mu_0$  and the infimum is taken over all possible estimates based on observations  $(\Theta_i, X_i)_{i=1, \dots, n}$ .

According to Remark 3.8, Theorem 3.10 entails that the lower bound for the minimax risk for functions  $m : \mathbb{R} \mapsto \mathbb{S}^1$  such that  $m \in \mathcal{H}(\beta, L)$  is  $n^{-\frac{2\beta}{2\beta+1}}$ . Now let us connect this result to the upper bound obtained in Theorem 3.5. As the function  $\text{atan2}$  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  $\beta$  of  $m$  will be the minimum of the smoothness of  $g_1$  and the smoothness of  $g_2$ . This shows that the rate achieved in Theorem 3.5 by our estimator is optimal up to a log factor. We conjecture that the logarithmic term is unavoidable in minimax rates even if it remains an open problem.

## 4 Numerical simulations

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

$$\text{M1. } \Theta = \text{atan2}(2X - 1, X^2 + 2) + \zeta \pmod{2\pi}, \quad (17)$$

$$\text{M2. } \Theta = \text{atan2}(-2X + 1, X^2 - 1) + \zeta \pmod{2\pi}, \quad (18)$$

$$\text{M3. } \Theta = \arccos(X^5 - 1) + 3 \cdot \arcsin(X^3 - X + 1) + \zeta \pmod{2\pi}, \quad (19)$$

where the circular error,  $\zeta$ , is distributed according to a von Mises distribution  $vM(0, 10)$  and is independent from  $X$ . Recall that the von Mises distribution  $vM(\tilde{\mu}, \tilde{\kappa})$  has density

$$\theta \in [-\pi, \pi) \mapsto \frac{1}{2\pi \cdot I_0(\tilde{\kappa})} \cdot \exp(\tilde{\kappa} \cdot \cos(\theta - \tilde{\mu})), \quad (20)$$



with  $I_0(\tilde{\kappa})$  the modified Bessel function of the first kind and of order  $0^1$ , a location parameter  $\tilde{\mu} \in [-\pi, \pi)$  and concentration  $\tilde{\kappa} > 0$ . See e.g [14, Section 2.2.2]. 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  $(\Theta_i, X_i)_{i=1, \dots, n}$  with the same distribution as  $(\Theta, X)$ . To implement the Goldenshluger-Lepski methodology, we shall consider either the Gaussian kernel defined by  $y \mapsto K(y) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}}$ , or the Epanechnikov kernel  $K$  defined by  $y \mapsto K(y) = \frac{3}{4} \cdot (1 - y^2) \cdot 1_{|y| \leq 1}$ . 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, we make simulations for the general case of unknown design distribution, i.e. the final estimators are computed using  $y \in \mathbb{R} \mapsto \hat{F}_n(y) := \frac{1}{n} \sum_{j=1}^n 1_{X_j \leq y}$  instead of  $F_X$ .

#### 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 quadratic pointwise risk

$$\mathcal{R} := \left| \hat{m}_{\hat{h}}(x) - m(x) \right|^2 \quad (21)$$

with  $\hat{m}_{\hat{h}}(x) = \text{atan2}(\hat{g}_{1, \hat{h}_1}(\hat{F}_n(x)), \hat{g}_{2, \hat{h}_2}(\hat{F}_n(x)))$ . 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.

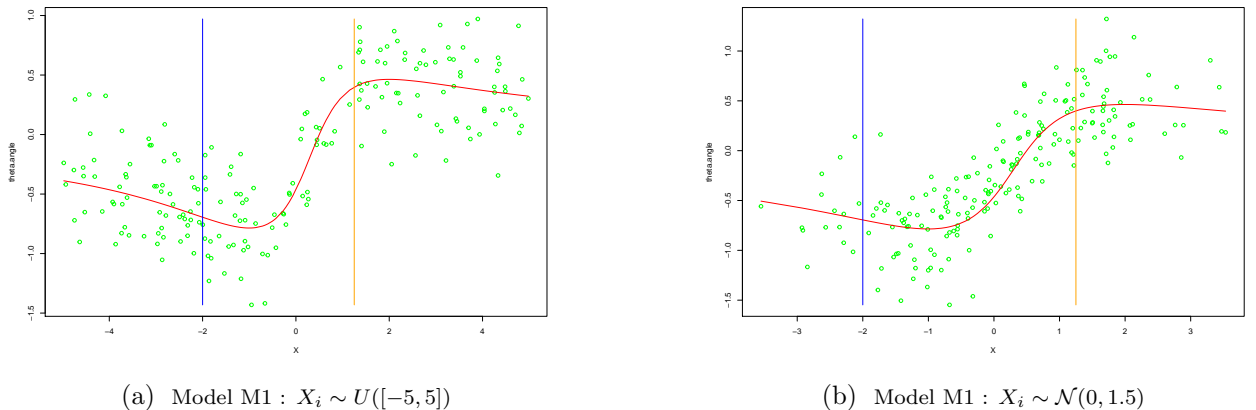


Figure 1: Illustration of model M1 with  $n = 200$  for two different density functions of the design. Simulated data  $(\Theta_i)_{i=1}^n$  are displayed in green points. The red curve represents the regression function  $m$ , while the blue vertical line displays the point  $x = -2$  and the orange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

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

To select  $\hat{h}_1$  and  $\hat{h}_2$ , 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

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

<sup>1</sup>The modified Bessel function of the first kind and of order  $\alpha \geq 0$ ,  $I_\alpha(z)$  for  $z > 0$ , admits the integral representation

$$I_\alpha(z) = \frac{1}{\pi} \int_0^\pi \exp(z \cdot \cos(\theta)) \cdot \cos(\alpha\theta) d\theta - \frac{\sin(\alpha\pi)}{\pi} \int_0^\infty \exp(-z \cdot \cosh(y) - \alpha y) dy.$$

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

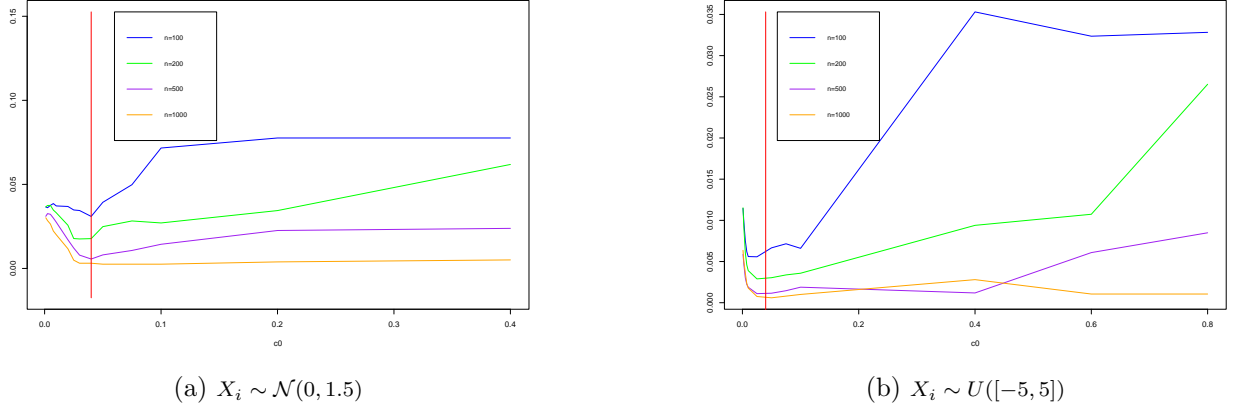


Figure 2: Plot of the Monte Carlo estimation of the function  $c_0 \in G_{c_0} \mapsto \mathcal{R}(c_0)$ , based on 50 runs, for  $x = -2$ , the Gaussian and the uniform designs and by using the Epanechnikov kernel for  $n \in \{100; 200; 500; 1000\}$ . The red vertical line displays the point  $c_0 = 0.04$ .

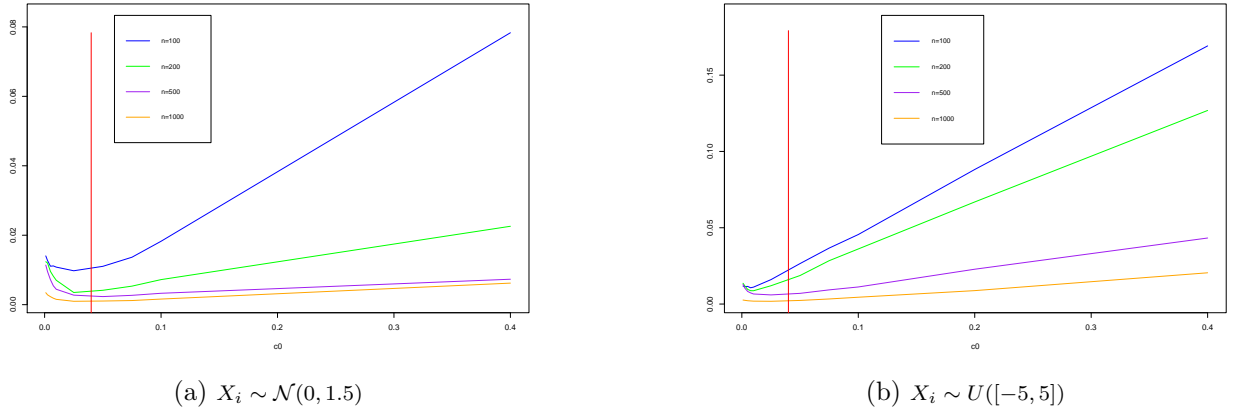


Figure 3: Plot of the Monte Carlo estimation of the function  $c_0 \in G_{c_0} \mapsto \mathcal{R}(c_0)$ , based on 50 runs, for  $x = 1.25$ , the Gaussian and the uniform designs and by using the Epanechnikov kernel for  $n \in \{100; 200; 500; 1000\}$ . The red vertical line displays the point  $c_0 = 0.04$ .

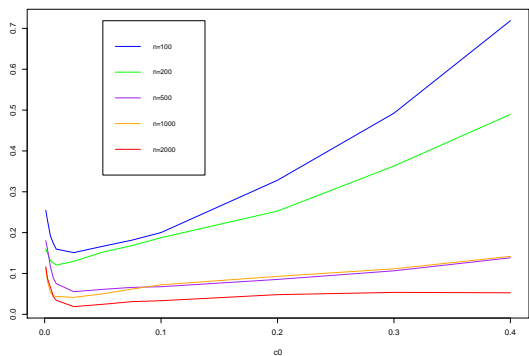
associated numerical illustrations are provided in Figure 4 for  $X \sim U([-5, 5])$  and for  $X \sim \mathcal{N}(0, 1.5)$ . This brief numerical study shows that the choice  $c_{0,1} = c_{0,2} = 0.04$  is convenient for each numerical scheme.

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

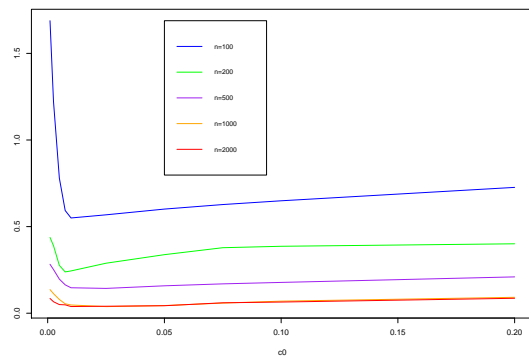
We do no longer assume that  $c_{0,1} = c_{0,2}$ . For  $n = 200$ , we compute the risk  $\mathcal{R}$  defined in (21) as a function of  $(c_{0,1}, c_{0,2})$  on the following discretization grid

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

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.

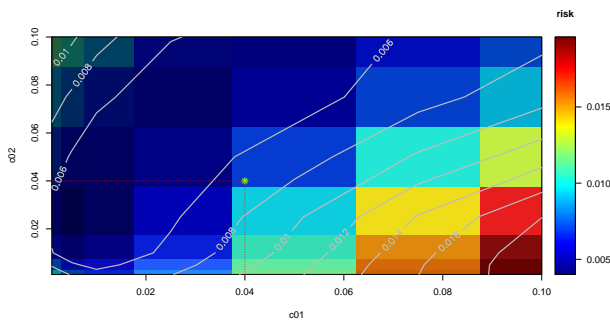


(a) Model M1 :  $X_i \sim U([-5, 5])$ .

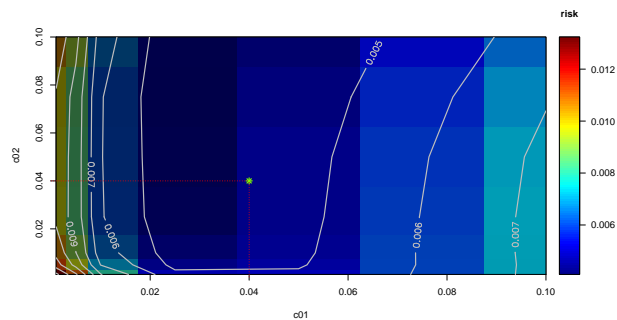


(b) Model M1 :  $X_i \sim \mathcal{N}(0, 1.5)$ .

Figure 4: Plot of the Monte Carlo estimation of the function  $c_0 \in G_{c_0} \mapsto \mathcal{R}(c_0)$ , based on 50 runs, for  $x = -2$ , the Gaussian and the uniform designs and by using the Gaussian kernel for  $n \in \{100; 200; 500; 1000; 2000\}$ .



(a) estimation at  $x = -2$

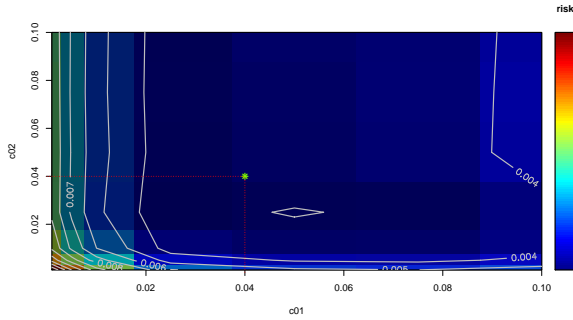


(b) estimation at  $x = 1.25$

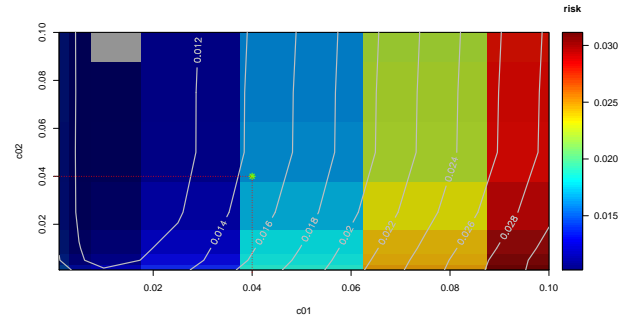
Figure 5: 2D-representation of the Monte Carlo estimation of the function  $(c_{0,1}, c_{0,1}) \in G_{c_0} \times G_{c_0} \mapsto \mathcal{R}(c_{0,1}, c_{0,2})$ , based on 50 runs, for Model M1 with  $X_i \sim \mathcal{N}(0, 1.5)$ ,  $n = 200$  at  $x = -2$  and  $x = 1.25$  by using the Epanechnikov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

## 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  $x = 0.95$ . Moreover, to make a comparison with our adaptive estimator, as proposed in [16], we also compute the Nadaraya-Watson (NW) estimator  $\hat{m}_h^{NW}$  and the version of the local linear (LL) estimator proposed by [16, Section 4.2] denoted by  $\hat{m}_h^{LL}$ . Cross-Validation is used to select the bandwidth parameter for  $\hat{m}_h^{NW}$  and  $\hat{m}_h^{LL}$ . Boxplots in Figures 7, 8, 9 and 10 show that the performances of our estimator are quite satisfying.

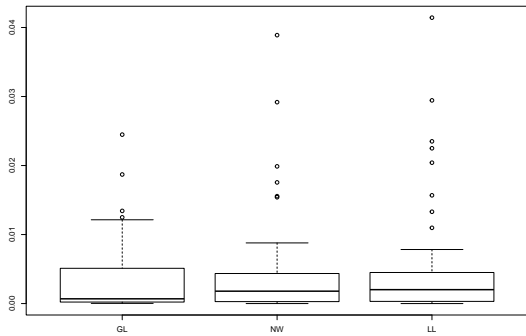


(a) estimation at  $x = -2$

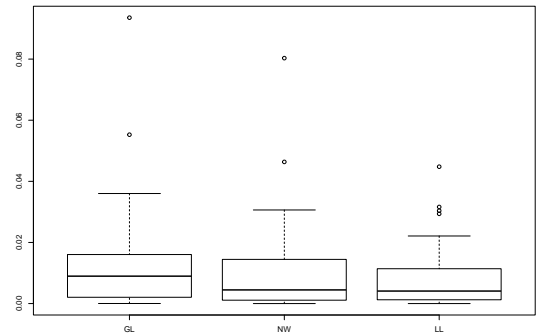


(b) estimation at  $x = 1.25$

Figure 6: 2D-representation of the Monte Carlo estimation of the function  $(c_{0,1}, c_{0,1}) \in G_{c_0} \times G_{c_0} \mapsto \mathcal{R}(c_{0,1}, c_{0,2})$ , based on 50 runs, for Model M1 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  $c_{0,1} = c_{0,2} = 0.04$ .



(a) Model M1: estimation at  $x = -2$



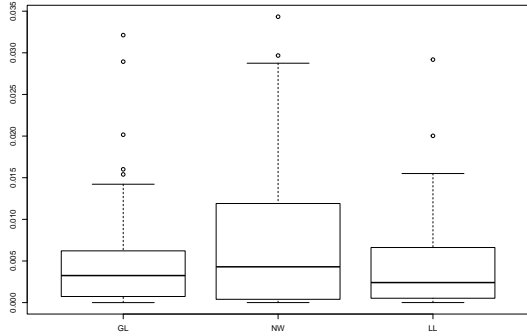
(b) Model M1: estimation at  $x = 1.25$

Figure 7: Boxplots of the estimated risk with 50 runs for the GL, NW and LL methodologies for  $n = 200$  and  $X \sim U([-5, 5])$  by using the Epanechnikov kernel.

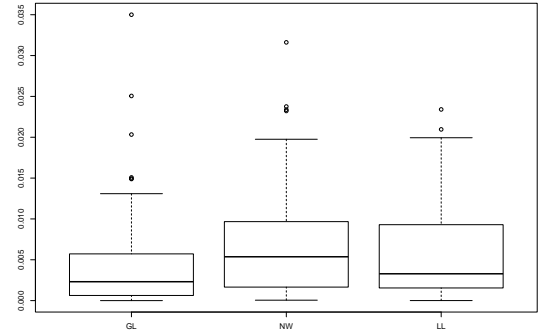
We finally repeat the previous numerical experiments but with the use of the Gaussian kernel: Figures 11 and 12 are the analogs of Figures 7 and 8. Figure 13a shows the numerical simulation for model M2 with  $X \sim \mathcal{N}(0, 1.5)$  and we estimate  $m(x)$  at  $x = 1.05$ . Figure 13b shows the 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.

## 5 Proofs

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

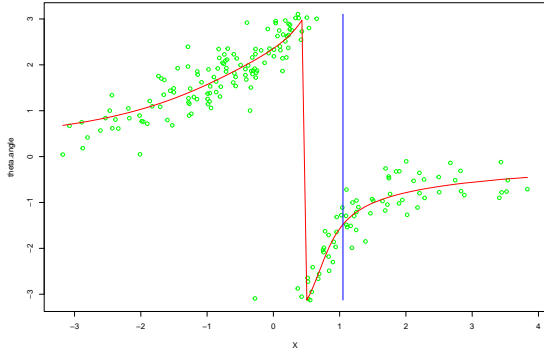


(a) Model M1: estimation at  $x = -2$

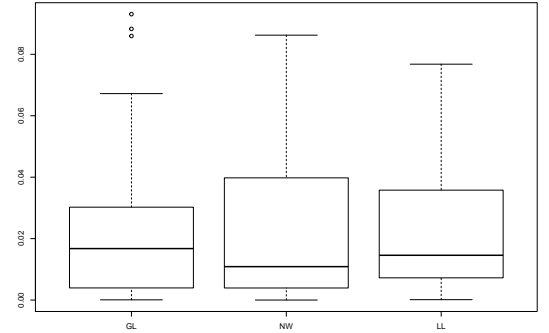


(b) Model M1: estimation at  $x = 1.25$

Figure 8: Boxplots of the estimated risk with 50 runs for the GL, NW and LL methodologies for  $n = 200$  and  $X \sim \mathcal{N}(0, 1.5)$  by using the Epanechnikov kernel



(a) Model M2:  $X_i \sim \mathcal{N}(0, 1.5)$



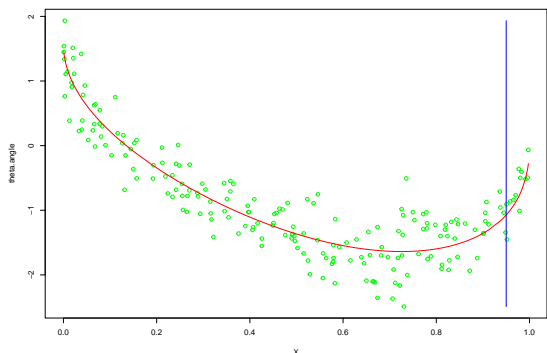
(b) Boxplots for GL, NW and LL methodologies

Figure 9: (a): 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 by using the Epanechnikov kernel.

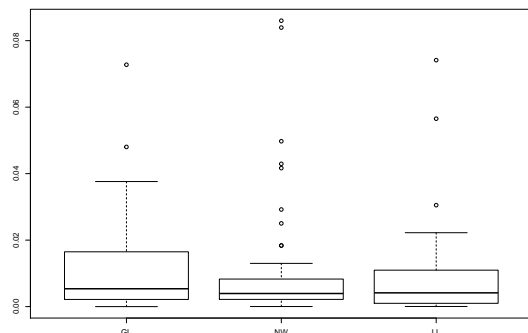
## 5.1 Preliminary results

In this section, we study several preliminary results for  $\hat{g}_{1,h_1}$  and  $\hat{g}_{2,h_2}$  defined in (3). First of all, via the warping method, we observe

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

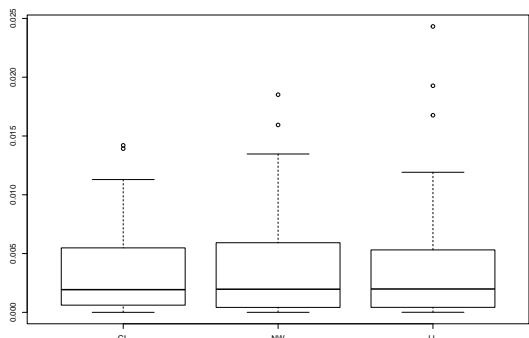


(a) Model M3:  $X_i \sim U([0, 1])$

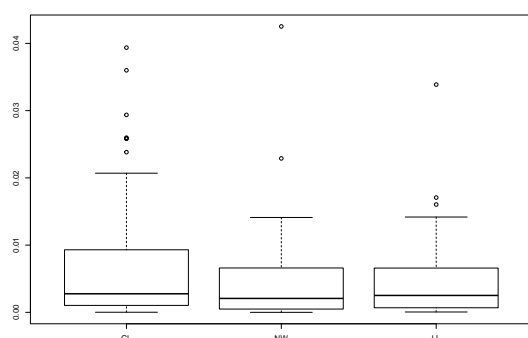


(b) Boxplots for GL, NW and LL methodologies

Figure 10: (a): 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 by using the Epanechnikov kernel.



(a) Model M1: estimation at  $x = -2$



(b) Model M1: estimation at  $x = 1.25$

Figure 11: Boxplots of the estimated risk with 50 runs for the GL, NW and LL methodologies for  $n = 200$  and  $X \sim U([-5, 5])$  by using the Gaussian kernel.

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

$$\int_{F_X(\mathbb{R})} g_1(w) \cdot K_{h_1}(u_x - w) dw = \int_{u_x - A \cdot h_{\max}}^{u_x + A \cdot h_{\max}} g_1(w) \cdot K_{h_1}(u_x - w) dw = (K_{h_1} * g_1)(u_x). \quad (22)$$

Thus, we obtain

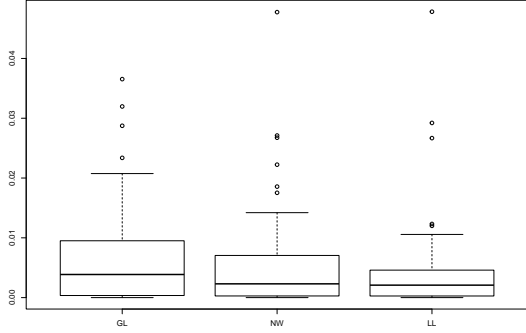
$$\mathbb{E}(\hat{g}_{1,h_1}(u_x)) = (K_{h_1} * g_1)(u_x)$$

and similarly,

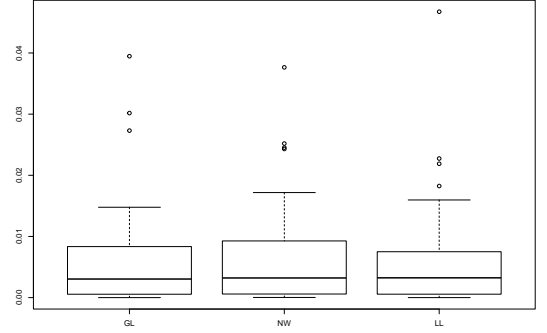
$$\mathbb{E}(\hat{g}_{2,h_2}(u_x)) = (K_{h_2} * g_2)(u_x). \quad (23)$$

We obtain upper bounds for the bias and variance terms.

**Lemma 5.1.** *Suppose that  $g_1$  belongs to  $\mathcal{H}(\beta_1, L_1)$  and  $g_2$  belongs to  $\mathcal{H}(\beta_2, L_2)$ , with  $L_1, L_2, \beta_1, \beta_2 \in \mathbb{R}_+^*$ . Assume that the kernel  $K$  satisfies Assumption 3.4 with an index  $\mathcal{L} \in \mathbb{R}_+$  such that  $\mathcal{L} \geq \max(\beta_1, \beta_2)$ . Then,*

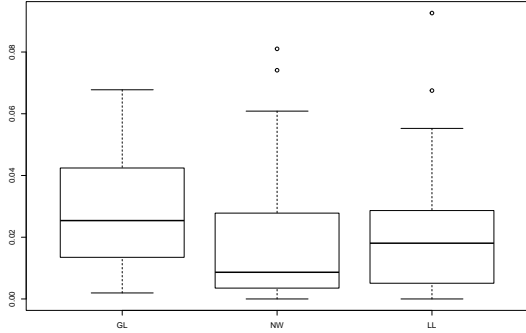


(a) Model M1: estimation at  $x = -2$

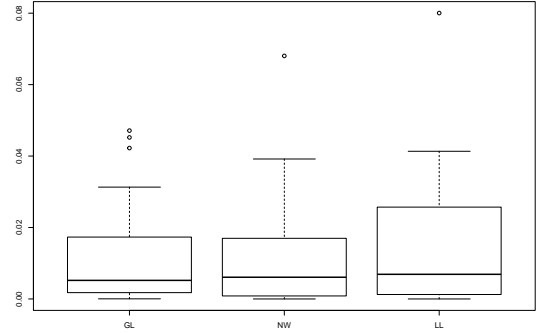


(b) Model M1: estimation at  $x = 1.25$

Figure 12: Boxplots of the estimated risk with 50 runs for the GL, NW and LL methodologies 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  $x = 0.95$

Figure 13: Boxplots of the estimated risk with 50 runs for the GL, NW and LL methodologies for  $n = 200$  by using the Gaussian kernel.

for any  $(h_1, h_2) \in \mathcal{H}_n^2$ ,

$$\begin{aligned} \left| \mathbb{E}(\hat{g}_{1,h_1}(u_x)) - g_1(u_x) \right| &\leq C_{K,\mathcal{L}} \cdot L_1 \cdot h_1^{\beta_1}, \quad \text{and} \quad \text{Var}(\hat{g}_{1,h_1}(u_x)) \leq \frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}^2}{n \cdot h_1}, \\ \left| \mathbb{E}(\hat{g}_{2,h_2}(u_x)) - g_2(u_x) \right| &\leq C_{K,\mathcal{L}} \cdot L_2 \cdot h_2^{\beta_2}, \quad \text{and} \quad \text{Var}(\hat{g}_{2,h_2}(u_x)) \leq \frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}^2}{n \cdot h_2}, \end{aligned}$$

with  $C_{K,\mathcal{L}}$  the constant defined in Assumption 3.4.

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  $\hat{g}_{1,h_1}$  and  $\hat{g}_{2,h_2}$ .

**Definition 5.2.** For  $n \in \mathbb{N}^*$ ,  $p \geq 1$ , and  $h_1, h_2 > 0$ , we define for an arbitrary  $v \in F_X(\mathbb{R})$  the following two events

$$\Omega_{1,n}(v, h_1) := \left\{ \left| \hat{g}_{1,h_1}(v) - \mathbb{E}[\hat{g}_{1,h_1}(v)] \right| \leq c_1(p) \cdot \sqrt{\tilde{V}_1(n, h_1)} \right\}$$

and

$$\Omega_{2,n}(v, h_2) := \left\{ \left| \hat{g}_{2,h_2}(v) - \mathbb{E}[\hat{g}_{2,h_2}(v)] \right| \leq c_1(p) \cdot \sqrt{\tilde{V}_2(n, h_2)} \right\},$$



with  $c_1(p)$  satisfying  $c_1(p) \cdot \sqrt{\min\{c_{0,1}; c_{0,2}\}} \geq 4p$ .  
 Furthermore, for  $L_1, L_2, \beta_1, \beta_2 > 0$ , we also introduce

$$E_{1,n}(v, h_1) := \{|\widehat{g}_{1,h_1}(v) - g_1(v)| \leq \Phi_1(n, h_1)\} \text{ and } E_{2,n}(v, h_2) := \{|\widehat{g}_{2,h_2}(v) - g_2(v)| \leq \Phi_2(n, h_2)\}$$

where

$$\Phi_1(n, h_1) := c_1(p) \cdot \sqrt{\widetilde{V}_1(n, h_1)} + C_{K,\mathcal{L}} \cdot L_1 \cdot h_1^{\beta_1}$$

and

$$\Phi_2(n, h_2) := c_1(p) \cdot \sqrt{\widetilde{V}_2(n, h_2)} + C_{K,\mathcal{L}} \cdot L_2 \cdot h_2^{\beta_2}.$$

Then, the following proposition gives a concentration inequality for  $\widehat{g}_{1,h_1}(u_x)$  and  $\widehat{g}_{2,h_2}(u_x)$ .

**Proposition 5.3.** *For  $p \geq 1$  and bandwidths  $h_1, h_2 \in \mathcal{H}_n$  (defined in (7)), we have:*

$$\mathbb{P}\left(\left(\Omega_{1,n}(u_x, h_1)\right)^c\right) \leq 2 \cdot n^{-p}, \quad \text{and} \quad \mathbb{P}\left(\left(\Omega_{2,n}(u_x, h_2)\right)^c\right) \leq 2 \cdot n^{-p}.$$

Consequently, suppose further that  $g_1$  belongs to  $\mathcal{H}(\beta_1, L_1)$ , and  $g_2$  belongs to  $\mathcal{H}(\beta_2, L_2)$ , with  $L_1, L_2, \beta_1, \beta_2 \in \mathbb{R}_+^*$  and the kernel  $K$  satisfies Assumption 3.4 with an index  $\mathcal{L} \in \mathbb{R}_+$  such that  $\mathcal{L} \geq \max(\beta_1, \beta_2)$ , then we get:

$$\mathbb{P}\left(\left(E_{1,n}(u_x, h_1)\right)^c\right) \leq 2 \cdot n^{-p} \quad \text{and} \quad \mathbb{P}\left(\left(E_{2,n}(u_x, h_2)\right)^c\right) \leq 2 \cdot n^{-p}.$$

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  $\widehat{g}_{1,h}(u_x)$  at  $u_x = F_X(x)$ , using (22), we can write

$$\begin{aligned} \mathbb{E}(\widehat{g}_{1,h_1}(u_x)) - g_1(u_x) &= \frac{1}{h_1} \cdot \int_{u_x - A \cdot h_{\max}}^{u_x + A \cdot h_{\max}} K\left(\frac{u_x - z}{h_1}\right) \cdot g_1(z) dz - g_1(u_x) \\ &= \int_{-A}^A K(w) \cdot (g_1(u_x - h_1 \cdot w) - g_1(u_x)) dw. \end{aligned} \quad (24)$$

Since  $g_1$  belongs to  $\mathcal{H}(\beta_1, L_1)$ , using a Taylor expansion for  $g_1$ , we get for  $w \in [-A, A]$ ,

$$g_1(u_x - h_1 \cdot w) = g_1(u_x) + g_1'(u_x) \cdot (-h_1) \cdot w + \dots + \frac{(-h_1 \cdot w)^{\lfloor \beta_1 \rfloor}}{(\lfloor \beta_1 \rfloor)!} \cdot g_1^{(\lfloor \beta_1 \rfloor)}(u_x - \tau \cdot h_1 \cdot w), \quad 0 \leq \tau \leq 1.$$

Then, under Assumption 3.4 with an index  $\mathcal{L} \in \mathbb{R}_+$  satisfying  $\mathcal{L} \geq \max(\beta_1, \beta_2)$ , from (24) one gets

$$\begin{aligned} &\int_{-A}^A K(w) \cdot (g_1(u_x - h_1 \cdot w) - g_1(u_x)) dw \\ &= \int_{-A}^A K(w) \cdot \frac{(-h_1 \cdot w)^{\lfloor \beta_1 \rfloor}}{(\lfloor \beta_1 \rfloor)!} \cdot g_1^{(\lfloor \beta_1 \rfloor)}(u_x - \tau \cdot h_1 \cdot w) dw \\ &= \int_{-A}^A K(w) \cdot \frac{(-h_1 \cdot w)^{\lfloor \beta_1 \rfloor}}{(\lfloor \beta_1 \rfloor)!} \cdot (g_1^{(\lfloor \beta_1 \rfloor)}(u_x - \tau \cdot h_1 \cdot w) - g_1^{(\lfloor \beta_1 \rfloor)}(u_x)) dw. \end{aligned}$$

This implies that with  $0 \leq \tau \leq 1$ , with  $C_{K,\mathcal{L}}$  the constant defined in Assumption 3.4,

$$\begin{aligned} \left| \mathbb{E}(\widehat{g}_{1,h_1}(u_x)) - g_1(u_x) \right| &\leq \int_{-A}^A |K(w)| \cdot \frac{|h_1 \cdot w|^{\lfloor \beta_1 \rfloor}}{(\lfloor \beta_1 \rfloor)!} \cdot \left| g_1^{(\lfloor \beta_1 \rfloor)}(u_x - \tau \cdot h_1 \cdot w) - g_1^{(\lfloor \beta_1 \rfloor)}(u_x) \right| dw \\ &\leq \int_{-A}^A |K(w)| \cdot \frac{|h_1 \cdot w|^{\lfloor \beta_1 \rfloor}}{(\lfloor \beta_1 \rfloor)!} \cdot L_1 \cdot |\tau \cdot h_1 \cdot w|^{\beta_1 - \lfloor \beta_1 \rfloor} dw \quad (\text{since } g_1 \in \mathcal{H}(\beta_1, L_1)) \\ &\leq L_1 \cdot h_1^{\beta_1} \cdot \int_{-A}^A |K(w)| \cdot |w|^{\beta_1} dw \\ &\leq L_1 \cdot h_1^{\beta_1} \cdot \int_{-A}^A |K(w)| \cdot (1 + |w|)^{\beta_1} dw \leq L_1 \cdot h_1^{\beta_1} \cdot C_{K,\mathcal{L}}. \end{aligned}$$

Similarly, we obtain for  $\widehat{g}_{2,h_2}$ :

$$\left| \mathbb{E}(\widehat{g}_{2,h_2}(u_x)) - g_2(u_x) \right| \leq L_2 \cdot h_2^{\beta_2} \cdot C_{K,\mathcal{L}}.$$

For the variance of  $\widehat{g}_{1,h_1}(u_x)$ , one gets:

$$\begin{aligned} \text{Var}(\widehat{g}_{1,h_1}(u_x)) &= \mathbb{E} \left[ \left( \widehat{g}_{1,h_1}(u_x) - \mathbb{E}(\widehat{g}_{1,h_1}(u_x)) \right)^2 \right] = \text{Var} \left( \frac{1}{n} \cdot \sum_{k=1}^n \sin(\Theta_k) \cdot K_{h_1}(u_x - F_X(X_k)) \right) \\ &\leq \frac{1}{n} \cdot \mathbb{E} \left[ (\sin(\Theta))^2 \cdot [K_{h_1}(u_x - F_X(X))]^2 \right] \\ &\leq \frac{1}{n} \cdot \mathbb{E} \left( [K_{h_1}(u_x - F_X(X))]^2 \right) \leq \frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}^2}{n \cdot h_1}, \end{aligned}$$

and similarly, for  $\widehat{g}_{2,h_2}(u_x)$ . This concludes the proof of Lemma 5.1.  $\square$

### 5.3 Proof of Proposition 5.3

We shall use the following version of Bernstein inequality (see [6, Lemma 2]).

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

$$\mathbb{P}(|S_n| \geq n \cdot \eta) \leq 2 \cdot \max \left( \exp \left( -\frac{n \cdot \eta^2}{4 \cdot V} \right), \exp \left( -\frac{n \cdot \eta}{4 \cdot b} \right) \right),$$

with  $\text{Var}(T_1) \leq V$  and  $|T_1| \leq b$ , 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 [6, Proposition 6]. First of all, we define random variables  $Z_k(u_x) := \cos(\Theta_k) \cdot K_{h_2}(u_x - F_X(X_k))$ , for  $1 \leq k \leq n$ , and hence,  $\widehat{g}_{2,h_2}(u_x) = \frac{1}{n} \sum_{k=1}^n Z_k(u_x)$ . Notice that  $\mathbb{E}(Z_k(u_x)) = (K_{h_2} * g_2)(u_x)$  (see (23)). Since  $\|K\|_\infty < +\infty$ , we then have for any  $v \in F_X(\mathbb{R})$ :

$$|Z_k(v)| = |\cos(\Theta_k) \cdot K_{h_2}(v - F_X(X_k))| \leq \frac{\|K\|_\infty}{h_2} =: b(h_2), \quad (25)$$

and  $\text{Var}(Z_k(v)) = n \cdot \text{Var}(\widehat{g}_{2,h_2}(v)) \leq n \cdot \frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}^2}{n \cdot h_2} =: n \cdot V_0(n, h_2)$ . Now, applying Lemma 5.4 to the  $Z_k(v)$ 's, we obtain for  $\eta(h_2) > 0$ ,

$$\begin{aligned} \mathbb{P} \left( \left| \widehat{g}_{2,h_2}(v) - \mathbb{E}(\widehat{g}_{2,h_2}(v)) \right| \geq \eta(h_2) \right) &= \mathbb{P} \left( \left| \sum_{k=1}^n Z_k(v) - \mathbb{E}(Z_k(v)) \right| \geq n \cdot \eta(h_2) \right) \\ &\leq 2 \cdot \max \left\{ \exp \left( -\frac{n \cdot (\eta(h_2))^2}{4n \cdot V_0(n, h_2)} \right); \exp \left( -\frac{n \cdot \eta(h_2)}{4b(h_2)} \right) \right\}. \end{aligned}$$

For  $p \geq 1$ , choose  $\eta(h_2) = c_1(p) \cdot \sqrt{\widetilde{V}_2(n, h_2)}$ , with

$$\widetilde{V}_2(n, h_2) = c_{0,2} \cdot \log(n) \cdot V_0(n, h_2). \quad (26)$$

Then,

$$\begin{aligned} \mathbb{P} \left( \left| \widehat{g}_{2,h_2}(v) - \mathbb{E}(\widehat{g}_{2,h_2}(v)) \right| \geq c_1(p) \cdot \sqrt{\widetilde{V}_2(n, h_2)} \right) \\ \leq 2 \cdot \max \left\{ \exp \left( -\frac{n \cdot c_1(p)^2 \cdot \widetilde{V}_2(n, h_2)}{4n \cdot V_0(n, h_2)} \right); \exp \left( -\frac{n \cdot c_1(p) \cdot \sqrt{\widetilde{V}_2(n, h_2)}}{4b(h_2)} \right) \right\}. \end{aligned} \quad (27)$$

We then choose  $c_1(p)$ .

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

$$\frac{n \cdot c_1(p)^2 \cdot \tilde{V}_2(n, h_2)}{4n \cdot V_0(n, h_2)} = \frac{n \cdot c_1(p)^2 \cdot c_{0,2} \cdot \log(n) \cdot V_0(n, h_2)}{4n \cdot V_0(n, h_2)} \geq p \cdot \log(n), \quad (28)$$

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

+ Secondly, we can write

$$\frac{n \cdot c_1(p) \cdot \sqrt{\tilde{V}_2(n, h_2)}}{4b(h_2)} = \frac{c_1(p) \cdot \sqrt{c_{0,2}}}{4} \cdot \sqrt{\log(n)} \cdot \frac{n \cdot \sqrt{V_0(n, h_2)}}{b(h_2)}$$

and for  $h_2 \in \mathcal{H}_n$ ,

$$n \cdot \frac{\sqrt{V_0(n, h_2)}}{b(h_2)} = n \cdot \frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}}{\sqrt{n \cdot h_2}} \cdot \frac{h_2}{\|K\|_\infty} = \sqrt{n \cdot h_2} \cdot \frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}}{\|K\|_\infty} > \sqrt{\log(n)},$$

then, we have

$$\frac{n \cdot c_1(p) \cdot \sqrt{\tilde{V}_2(n, h_2)}}{4b(h_2)} = \frac{c_1(p) \cdot \sqrt{c_{0,2}}}{4} \cdot \sqrt{\log(n)} \cdot \frac{n \cdot \sqrt{V_0(n, h_2)}}{b(h_2)} \geq \frac{c_1(p) \cdot \sqrt{c_{0,2}}}{4} \cdot \log(n) \geq p \cdot \log(n), \quad (29)$$

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

Now, combining (28) and (29), we get from (27) for any  $p \geq 1$ :

$$\mathbb{P} \left( \left( \Omega_{2,n}(u_x, h_2) \right)^c \right) := \mathbb{P} \left( \left| \hat{g}_{2,h_2}(u_x) - \mathbb{E}(\hat{g}_{2,h_2}(u_x)) \right| > c_1(p) \cdot \sqrt{\tilde{V}_2(n, h_2)} \right) \leq 2 \cdot n^{-p}.$$

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

$$\left| \hat{g}_{2,h_2}(u_x) - \mathbb{E}(\hat{g}_{2,h_2}(u_x)) \right| \leq c_1(p) \cdot \sqrt{\tilde{V}_2(n, h_2)}.$$

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

$$\begin{aligned} \left| \hat{g}_{2,h_2}(u_x) - g_2(u_x) \right| &\leq \left| \hat{g}_{2,h_2}(u_x) - \mathbb{E}(\hat{g}_{2,h_2}(u_x)) \right| + \left| \mathbb{E}(\hat{g}_{2,h_2}(u_x)) - g_2(u_x) \right| \\ &\leq c_1(p) \cdot \sqrt{\tilde{V}_2(n, h_2)} + L_2 \cdot h_2^{\beta_2} \cdot C_{K,\mathcal{L}}. \end{aligned}$$

Recall that  $\Phi_2(n, h_2) = c_1(p) \cdot \sqrt{\tilde{V}_2(n, h_2)} + L_2 \cdot h_2^{\beta_2} \cdot C_{K,\mathcal{L}}$ , therefore, we finally obtain for  $p \geq 1$  that

$$\mathbb{P} \left( \left( E_{2,n}(u_x, h_2) \right)^c \right) = \mathbb{P} \left( \left| \hat{g}_{2,h_2}(u_x) - g_2(u_x) \right| > \Phi_2(n, h_2) \right) \leq 2 \cdot n^{-p}. \quad (30)$$

By similar arguments we obtain the desired result for  $\hat{g}_{1,h_1}$ . 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.** *Under the Assumptions of Proposition 3.1, for all  $h_1, h'_1, h_2, h'_2 \in \mathcal{H}_n$ , for all  $p \geq 1$ , for  $u_x = F_X(x)$ ,*

$$\mathbb{P} \left( \left| \hat{g}_{1,h_1,h'_1}(u_x) - \mathbb{E}[\hat{g}_{1,h_1,h'_1}(u_x)] \right| > c_1(p) \cdot \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \sqrt{\tilde{V}_1(n, h'_1)} \right) \leq 2 \cdot n^{-p},$$

and

$$\mathbb{P} \left( \left| \hat{g}_{2,h_2,h'_2}(u_x) - \mathbb{E}[\hat{g}_{2,h_2,h'_2}(u_x)] \right| > c_1(p) \cdot \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \sqrt{\tilde{V}_2(n, h'_2)} \right) \leq 2 \cdot n^{-p},$$

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

*Proof of Corollary 5.5.* We define random variables  $\tilde{Z}_k(u_x) := \cos(\Theta_k) \cdot (K_{h'_2} * K_{h_2})(u_x - F_X(X_k))$ , for  $1 \leq k \leq n$ , and hence,  $\hat{g}_{2,h_2,h'_2}(u_x) = \frac{1}{n} \sum_{k=1}^n \tilde{Z}_k(u_x)$ . Since  $\|K\|_\infty < +\infty$ , we then have:

$$\left| \tilde{Z}_k(u_x) \right| = \left| \cos(\Theta_k) \cdot (K_{h'_2} * K_{h_2})(u_x - F_X(X_k)) \right| \leq \frac{\|K\|_\infty \cdot \|K\|_{\mathbb{L}^1(\mathbb{R})}}{h_2} =: \tilde{b}(h_2),$$

and  $\text{Var}(\tilde{Z}_k(u_x)) = n \cdot \text{Var}(\hat{g}_{2,h_2,h'_2}(u_x)) \leq n \cdot \frac{\|K\|_{\mathbb{L}^1(\mathbb{R})}^2 \cdot \|K\|_{\mathbb{L}^2(\mathbb{R})}^2}{n \cdot h_2} =: n \cdot \tilde{V}_0(n, h_2)$ .

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

$$\left| \hat{g}_{2,h_2,h'_2}(u_x) - \mathbb{E}[\hat{g}_{2,h_2,h'_2}(u_x)] \right| \leq \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot c_1(p) \cdot \sqrt{\tilde{V}_2(n, h_2)}.$$

Similarly, we get the desired concentration for  $\left| \hat{g}_{1,h_1,h'_1}(u_x) - \mathbb{E}[\hat{g}_{1,h_1,h'_1}(u_x)] \right|$ . This concludes the proof of Corollary 5.5.  $\square$

Now, we can start to prove Proposition 3.1.

*Proof of Proposition 3.1.* We follow the strategy proposed in [6, Theorem 2]. The target is to find an upper bound for  $\left| \hat{g}_{2,\hat{h}_2}(u_x) - g_2(u_x) \right|$ . Let  $h_2 \in \mathcal{H}_n$  be fixed. We consider the following decomposition:

$$\left| \hat{g}_{2,\hat{h}_2}(u_x) - g_2(u_x) \right| \leq \underbrace{\left| \hat{g}_{2,\hat{h}_2}(u_x) - \hat{g}_{2,h_2,\hat{h}_2}(u_x) \right|}_{=: I_{g_2,1}} + \underbrace{\left| \hat{g}_{2,h_2,\hat{h}_2}(u_x) - \hat{g}_{2,h_2}(u_x) \right|}_{=: I_{g_2,2}} + \left| \hat{g}_{2,h_2}(u_x) - g_2(u_x) \right|.$$

By the definition of  $A_2(h_2, u_x)$ , we have

$$\begin{aligned} I_{g_2,1} &= \left| \hat{g}_{2,\hat{h}_2}(u_x) - \hat{g}_{2,h_2,\hat{h}_2}(u_x) \right| = \left| \hat{g}_{2,\hat{h}_2}(u_x) - \hat{g}_{2,h_2,\hat{h}_2}(u_x) \right| - \sqrt{\tilde{V}_2(n, \hat{h}_2)} + \sqrt{\tilde{V}_2(n, \hat{h}_2)} \\ &\leq \sup_{h'_2 \in \mathcal{H}_n} \left\{ \left| \hat{g}_{2,h'_2}(u_x) - \hat{g}_{2,h_2,h'_2}(u_x) \right| - \sqrt{\tilde{V}_2(n, h'_2)} \right\}_+ + \sqrt{\tilde{V}_2(n, \hat{h}_2)} \\ &= A_2(h_2, u_x) + \sqrt{\tilde{V}_2(n, \hat{h}_2)}. \end{aligned}$$

And similarly, by the definition of  $A_2(\hat{h}_2, u_x)$ ,

$$\begin{aligned} I_{g_2,2} &= \left| \hat{g}_{2,h_2,\hat{h}_2}(u_x) - \hat{g}_{2,h_2}(u_x) \right| \leq \sup_{h'_2 \in \mathcal{H}_n} \left\{ \left| \hat{g}_{2,h'_2,\hat{h}_2}(u_x) - \hat{g}_{2,h'_2}(u_x) \right| - \sqrt{\tilde{V}_2(n, h'_2)} \right\}_+ + \sqrt{\tilde{V}_2(n, h_2)} \\ &= A_2(\hat{h}_2, u_x) + \sqrt{\tilde{V}_2(n, h_2)}. \end{aligned}$$

Therefore, by using the definition of  $\hat{h}_2$ , we get

$$I_{g_2,1} + I_{g_2,2} \leq A_2(h_2, u_x) + \sqrt{\tilde{V}_2(n, \hat{h}_2)} + A_2(\hat{h}_2, u_x) + \sqrt{\tilde{V}_2(n, h_2)} \leq 2 \cdot \left[ A_2(h_2, u_x) + \sqrt{\tilde{V}_2(n, h_2)} \right].$$

Hence, we obtain

$$\left| \hat{g}_{2,\hat{h}_2}(u_x) - g_2(u_x) \right| \leq 2 \cdot A_2(h_2, u_x) + 2 \cdot \sqrt{\tilde{V}_2(n, h_2)} + \left| \hat{g}_{2,h_2}(u_x) - g_2(u_x) \right|. \quad (31)$$

Now, to study  $A_2(h_2, u_x)$ , we can write:

$$\begin{aligned} \hat{g}_{2,h'_2}(u_x) - \hat{g}_{2,h_2,h'_2}(u_x) &= \hat{g}_{2,h'_2}(u_x) - \mathbb{E}[\hat{g}_{2,h'_2}(u_x)] - (\hat{g}_{2,h_2,h'_2}(u_x) - \mathbb{E}[\hat{g}_{2,h_2,h'_2}(u_x)]) \\ &\quad + \mathbb{E}[\hat{g}_{2,h'_2}(u_x)] - \mathbb{E}[\hat{g}_{2,h_2,h'_2}(u_x)], \end{aligned}$$

and, we have  $\mathbb{E}[\widehat{g}_{2,h'_2}(u_x)] = (K_{h'_2} * g_2)(u_x)$  as well as  $\mathbb{E}[\widehat{g}_{2,h_2,h'_2}(u_x)] = \mathbb{E}[K_{h'_2} * \widehat{g}_{2,h_2}(u_x)] = (K_{h'_2} * K_{h_2} * g_2)(u_x)$ .

Thus,

$$\begin{aligned} |\widehat{g}_{2,h'_2}(u_x) - \widehat{g}_{2,h_2,h'_2}(u_x)| - \sqrt{\widetilde{V}_2(n, h'_2)} &\leq |\widehat{g}_{2,h'_2}(u_x) - \mathbb{E}[\widehat{g}_{2,h'_2}(u_x)]| - \frac{\sqrt{\widetilde{V}_2(n, h'_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \\ &\quad + |\widehat{g}_{2,h_2,h'_2}(u_x) - \mathbb{E}[\widehat{g}_{2,h_2,h'_2}(u_x)]| - \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \frac{\sqrt{\widetilde{V}_2(n, h'_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \\ &\quad + |\mathbb{E}[\widehat{g}_{2,h'_2}(u_x)] - \mathbb{E}[\widehat{g}_{2,h_2,h'_2}(u_x)]|. \end{aligned}$$

However, for any  $h'_2 \in \mathcal{H}_n$ ,

$$|\mathbb{E}[\widehat{g}_{2,h'_2}(u_x)] - \mathbb{E}[\widehat{g}_{2,h_2,h'_2}(u_x)]| = |K_{h'_2} * (g_2 - K_{h_2} * g_2)(u_x)| \leq \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \|g_2 - K_{h_2} * g_2\|_\infty.$$

Hence, incorporating this bound in the definition of  $A_2(h_2, u_x)$ , we obtain

$$\begin{aligned} A_2(h_2, u_x) &= \sup_{h'_2 \in \mathcal{H}_n} \left\{ |\widehat{g}_{2,h'_2}(u_x) - \widehat{g}_{2,h_2,h'_2}(u_x)| - \sqrt{\widetilde{V}_2(n, h'_2)} \right\}_+ \\ &\leq \sup_{h'_2 \in \mathcal{H}_n} \left\{ |\widehat{g}_{2,h'_2}(u_x) - \mathbb{E}[\widehat{g}_{2,h'_2}(u_x)]| - \frac{\sqrt{\widetilde{V}_2(n, h'_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \right\}_+ \end{aligned} \quad (32)$$

$$\begin{aligned} &+ \sup_{h'_2 \in \mathcal{H}_n} \left\{ |\widehat{g}_{2,h_2,h'_2}(u_x) - \mathbb{E}[\widehat{g}_{2,h_2,h'_2}(u_x)]| - \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \frac{\sqrt{\widetilde{V}_2(n, h'_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \right\}_+ \\ &\quad + \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \|g_2 - K_{h_2} * g_2\|_\infty. \end{aligned} \quad (33)$$

From Corollary 5.5, for  $h_2, h'_2 \in \mathcal{H}_n$ ,

$$\mathbb{P} \left( |\widehat{g}_{2,h_2,h'_2}(u_x) - \mathbb{E}[\widehat{g}_{2,h_2,h'_2}(u_x)]| > c_1(p) \cdot \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \sqrt{\widetilde{V}_2(n, h'_2)} \right) \leq 2 \cdot n^{-p}.$$

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

$$\mathbb{P} \left( \sup_{h'_2 \in \mathcal{H}_n} \left\{ |\widehat{g}_{2,h'_2}(u_x) - \mathbb{E}[\widehat{g}_{2,h'_2}(u_x)]| - \frac{\sqrt{\widetilde{V}_2(n, h'_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \right\}_+ > 0 \right) \leq 2 \cdot \sum_{h_2 \in \mathcal{H}_n} n^{-p} \leq 2 \cdot n^{1-p},$$

as  $\text{Card}(\mathcal{H}_n) \leq n$ . In the same way for all  $h_2 \in \mathcal{H}_n$ ,

$$\mathbb{P} \left( \sup_{h'_2 \in \mathcal{H}_n} \left\{ |\widehat{g}_{2,h_2,h'_2}(u_x) - \mathbb{E}[\widehat{g}_{2,h_2,h'_2}(u_x)]| - \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \frac{\sqrt{\widetilde{V}_2(n, h'_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \right\}_+ > 0 \right) \leq 2 \cdot n^{1-p}.$$

Consequently, the following set

$$\begin{aligned} \widetilde{A}_2 &:= \left\{ \sup_{h'_2 \in \mathcal{H}_n} \left\{ |\widehat{g}_{2,h'_2}(u_x) - \mathbb{E}[\widehat{g}_{2,h'_2}(u_x)]| - \frac{\sqrt{\widetilde{V}_2(n, h'_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \right\}_+ = 0 \right\} \\ &\cap \left\{ \forall h_2 \in \mathcal{H}_n, \sup_{h'_2 \in \mathcal{H}_n} \left\{ |\widehat{g}_{2,h_2,h'_2}(u_x) - \mathbb{E}[\widehat{g}_{2,h_2,h'_2}(u_x)]| - \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \frac{\sqrt{\widetilde{V}_2(n, h'_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \right\}_+ = 0 \right\} \end{aligned}$$

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

Combining inequalities (31) and (32), we have on  $\tilde{A}_2$ :

$$\begin{aligned} |\hat{g}_{2,\hat{h}_2}(u_x) - g_2(u_x)| &\leq 2.A_2(h_2, u_x) + 2.\sqrt{\tilde{V}_2(n, h_2)} + |\hat{g}_{2,h_2}(u_x) - g_2(u_x)| \\ &\leq 2.\|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \|g_2 - K_{h_2} * g_2\|_\infty + 2.\sqrt{\tilde{V}_2(n, h_2)} + |\hat{g}_{2,h_2}(u_x) - g_2(u_x)|, \end{aligned}$$

but still on  $\tilde{A}_2$ , one gets  $|\hat{g}_{2,h_2}(u_x) - \mathbb{E}[\hat{g}_{2,h_2}(u_x)]| - \frac{\sqrt{\tilde{V}_2(n, h_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \leq 0$ , so

$$\begin{aligned} |\hat{g}_{2,h_2}(u_x) - g_2(u_x)| &\leq |\mathbb{E}[\hat{g}_{2,h_2}(u_x)] - g_2(u_x)| + |\hat{g}_{2,h_2}(u_x) - \mathbb{E}[\hat{g}_{2,h_2}(u_x)]| - \frac{\sqrt{\tilde{V}_2(n, h_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \\ &\quad + \frac{\sqrt{\tilde{V}_2(n, h_2)}}{(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})})} \\ &\leq \|g_2 - K_{h_2} * g_2\|_\infty + \sqrt{\tilde{V}_2(n, h_2)}. \end{aligned}$$

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

$$|\hat{g}_{2,\hat{h}_2}(u_x) - g_2(u_x)| \leq (1 + 2\|K\|_{\mathbb{L}^1(\mathbb{R})}) \cdot \|g_2 - K_{h_2} * g_2\|_\infty + 3.\sqrt{\tilde{V}_2(n, h_2)}.$$

By similar arguments, we obtain the desired result for  $\hat{g}_{1,\hat{h}_1}(u_x)$ . This concludes the proof of Proposition 3.1.  $\square$

#### 5.4.2 Proof of Theorem 3.5

First, we establish a concentration result for  $\hat{g}_{1,\hat{h}_1}(u_x)$  and  $\hat{g}_{2,\hat{h}_2}(u_x)$  as follows:

**Corollary 5.6.** *For  $\beta_1, \beta_2, L_1, L_2 > 0$ , suppose that  $g_1$  belongs to  $\mathcal{H}(\beta_1, L_1)$  and  $g_2$  belongs to  $\mathcal{H}(\beta_2, L_2)$ . Then, under the assumptions of Proposition 3.1, for  $q \geq 1$ ,*

$$\begin{aligned} \mathbb{P}\left((\tilde{E}_{1,n}(u_x, \hat{h}_1))^c\right) &:= \mathbb{P}\left(|\hat{g}_{1,\hat{h}_1}(u_x) - g_1(u_x)| > C_1.\psi_n(\beta_1)\right) \leq 4.n^{-q}, \\ \mathbb{P}\left((\tilde{E}_{2,n}(u_x, \hat{h}_2))^c\right) &:= \mathbb{P}\left(|\hat{g}_{2,\hat{h}_2}(u_x) - g_2(u_x)| > C_2.\psi_n(\beta_2)\right) \leq 4.n^{-q}, \end{aligned}$$

where  $C_1$  is a constant (depending on  $\beta_1, L_1, c_{0,1}, K$ ) and  $C_2$  is a constant (depending on  $\beta_2, L_2, c_{0,2}, K$ ). In the previous expressions,  $\psi_n(\beta_1) = (\log(n)/n)^{\frac{\beta_1}{2\beta_1+1}}$  and  $\psi_n(\beta_2) = (\log(n)/n)^{\frac{\beta_2}{2\beta_2+1}}$ .

*Proof of Corollary 5.6.* Since  $g_1$  belongs to  $\mathcal{H}(\beta_1, L_1)$ , from Lemma 5.1, we have

$$\|g_1 - K_{h_1} * g_1\|_\infty \leq L_1.h_1^{\beta_1}.C_{K,\mathcal{L}}.$$

From Proposition 3.1, this implies that with a probability greater than  $1 - 4.n^{-q}$ , one gets for any  $h_1 \in \mathcal{H}_n$ :

$$|\hat{g}_{1,\hat{h}_1}(u_x) - g_1(u_x)| \leq (1 + 2\|K\|_{\mathbb{L}^1(\mathbb{R})}) \cdot L_1.h_1^{\beta_1}.C_{K,\mathcal{L}} + 3.\sqrt{\tilde{V}_1(n, h_1)}. \quad (34)$$

In (34), we take  $h_1$  so that  $h_1^{-1}$  is an integer and  $h_1$  of order  $\left(\frac{\log(n)}{n}\right)^{\frac{1}{2\beta_1+1}}$ . Since  $1/(2\beta_1 + 1) < 1$ ,  $h_1 \in \mathcal{H}_n$ , for  $n$  large enough. As a result, we obtain with probability greater than  $1 - 4.n^{-q}$ , that

$$|\hat{g}_{1,\hat{h}_1}(u_x) - g_1(u_x)| \leq C_1.\psi_n(\beta_1),$$

with a constant  $C_1$  (depending on  $\beta_1, L_1, c_{0,1}, K$ ) and  $\psi_n(\beta_1) = (\log(n)/n)^{\frac{\beta_1}{2\beta_1+1}}$ .

By similar arguments, we obtain the desired result for  $\hat{g}_{2,\hat{h}_2}(u_x)$ . This concludes the proof of Corollary 5.6.  $\square$

Now, we start to prove Theorem 3.5.

*Proof of Theorem 3.5.* We have  $\mathbb{E}\left[|\widehat{m}_{\widehat{h}}(x) - m(x)|^2\right] = \mathbb{E}\left[|\widehat{g}_{\widehat{h}}(u_x) - g(u_x)|^2\right]$ . First, on the event  $\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)$ , for  $n$  sufficiently large satisfying  $C_2 \cdot \psi_n(\beta_2) \leq \delta_2/2$  and  $C_1 \cdot \psi_n(\beta_1) \leq \delta_1/2$ , we have  $|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)| \leq C_2 \cdot \psi_n(\beta_2) < \frac{|g_2(u_x)|}{2}$  and  $|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)| \leq C_1 \cdot \psi_n(\beta_1) < \frac{|g_1(u_x)|}{2}$ . Thus, we get  $\widehat{g}_{2,\widehat{h}_2}(u_x) \cdot g_2(u_x) > 0$  and  $\widehat{g}_{1,\widehat{h}_1}(u_x) \cdot g_1(u_x) > 0$ . Then, we have

$$\begin{aligned}
& \mathbb{E}\left[|\widehat{g}_{\widehat{h}}(u_x) - g(u_x)|^2 \cdot \mathbb{1}_{\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)}\right] \\
&= \mathbb{E}\left[\left|\operatorname{atan2}(\widehat{g}_{1,\widehat{h}_1}(u_x), \widehat{g}_{2,\widehat{h}_2}(u_x)) - \operatorname{atan2}(g_1(u_x), g_2(u_x))\right|^2 \cdot \mathbb{1}_{\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)}\right] \\
&= \mathbb{E}\left[\left|\arctan\left(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right) - \arctan\left(\frac{g_1(u_x)}{g_2(u_x)}\right)\right|^2 \cdot \mathbb{1}_{\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)}\right] \\
&\leq 2 \cdot \mathbb{E}\left[\left|\arctan\left(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right) - \arctan\left(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right)\right|^2 \cdot \mathbb{1}_{\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)}\right] \tag{35} \\
&\quad + 2 \cdot \mathbb{E}\left[\left|\arctan\left(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right) - \arctan\left(\frac{g_1(u_x)}{g_2(u_x)}\right)\right|^2 \cdot \mathbb{1}_{\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)}\right].
\end{aligned}$$

For  $n$  sufficiently large,  $|\widehat{g}_{2,\widehat{h}_2}(u_x)| \geq |g_2(u_x)| - |g_2(u_x) - \widehat{g}_{2,\widehat{h}_2}(u_x)| > \delta_2 - C_2 \cdot \psi_n(\beta_2) \geq \delta_2/2$  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 (35), since on  $\widetilde{E}_{1,n}(u_x, \widehat{h}_1)$  one has  $|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)| \leq C_1 \cdot \psi_n(\beta_1)$

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

Moreover, for the second term in (35), since  $\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)} \cdot \frac{g_1(u_x)}{g_2(u_x)} > 0$  on  $\widetilde{E}_{2,n}(u_x, \widehat{h}_2)$ , we have

$$\begin{aligned}
& \mathbb{E}\left[\left|\arctan\left(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\right) - \arctan\left(\frac{g_1(u_x)}{g_2(u_x)}\right)\right|^2 \cdot \mathbb{1}_{\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)}\right] \\
&= \mathbb{E}\left[\left|\arctan\left(\frac{\widehat{g}_{2,\widehat{h}_2}(u_x)}{g_1(u_x)}\right) - \arctan\left(\frac{g_2(u_x)}{g_1(u_x)}\right)\right|^2 \cdot \mathbb{1}_{\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)}\right] \\
&\leq \frac{1}{|g_1(u_x)|^2} \cdot \mathbb{E}\left[|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)|^2 \cdot \mathbb{1}_{\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)}\right] \\
&\leq \frac{1}{\delta_1^2} \cdot \mathbb{E}\left[C_2^2 \cdot \psi_n(\beta_2)^2 \cdot \mathbb{1}_{\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)}\right] \leq \frac{1}{\delta_1^2} \cdot C_2^2 \cdot \psi_n(\beta_2)^2,
\end{aligned}$$

by using Corollary 5.6.

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  $C_2 \cdot \psi_n(\beta_2) \leq \delta_2/2$  and  $C_1 \cdot \psi_n(\beta_1) \leq \delta_1/2$ , we obtain

$$\mathbb{E}\left[|\widehat{g}_{\widehat{h}}(u_x) - g(u_x)|^2 \cdot \mathbb{1}_{\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)}\right] \leq 8 \cdot \left(\frac{1}{\delta_2^2} + \frac{1}{\delta_1^2}\right) \cdot (C_1^2 + C_2^2) \cdot \max\left(\psi_n(\beta_1)^2, \psi_n(\beta_2)^2\right).$$



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

$$\begin{aligned} & \mathbb{E} \left[ \left| \hat{g}_{\hat{h}}(u_x) - g(u_x) \right|^2 \cdot \mathbb{1}_{(\tilde{E}_{2,n}(u_x, \hat{h}_2))^c \cup (\tilde{E}_{1,n}(u_x, \hat{h}_1))^c} \right] \\ &= \mathbb{E} \left[ \left| \operatorname{atan2}(\hat{g}_{1, \hat{h}_1}(u_x), \hat{g}_{2, \hat{h}_2}(u_x)) - \operatorname{atan2}(g_1(u_x), g_2(u_x)) \right|^2 \cdot \mathbb{1}_{(\tilde{E}_{2,n}(u_x, \hat{h}_2))^c \cup (\tilde{E}_{1,n}(u_x, \hat{h}_1))^c} \right] \\ &\leq 4\pi^2 \cdot \mathbb{P} \left( (\tilde{E}_{2,n}(u_x, \hat{h}_2))^c \right) + 4\pi^2 \cdot \mathbb{P} \left( (\tilde{E}_{1,n}(u_x, \hat{h}_1))^c \right) \leq 2.4\pi^2 \cdot 4 \cdot n^{-q}, \end{aligned}$$

by Corollary 5.6. For  $q \geq 1$ , we get that  $n^{-q}$  is negligible in comparison with  $C_1^2 \cdot \psi_n(\beta_1)^2 = C_1^2 \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2\beta_1}{2\beta_1+1}}$  and  $C_2^2 \cdot \psi_n(\beta_2)^2 = C_2^2 \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2\beta_2}{2\beta_2+1}}$ . This concludes the proof of Theorem 3.5.  $\square$

### 5.4.3 Proof of Theorem 3.10

Before tackling the proof of Theorem 3.10, the next lemma shows that the von Mises density with zero mean direction satisfies condition (16).

**Lemma 5.7.** *The von Mises density with mean direction  $\mu$  and concentration parameter  $\kappa$  satisfies condition (16).*

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

$$f_{vM}(x) = c(\kappa) e^{\kappa \cos(x-\mu)} \mathbb{1}_{x \in [-\pi, \pi]}, \quad x \in \mathbb{R}$$

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

$$\exists p_* > 0 : \int_{-\pi}^{\pi} f_{vM}(x) \log \frac{f_{vM}(x)}{f_{vM}(x+y)} dx \leq p_* y^2,$$

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

$$\begin{aligned} \int_{-\pi}^{\pi} f_{vM}(x) \log \frac{f_{vM}(x)}{f_{vM}(x+y)} dx &= c(\kappa) \kappa \int_{-\pi}^{\pi} e^{\kappa \cos(x-\mu)} (\cos(x-\mu) - \cos(x+y-\mu)) dx \\ &= 2c(\kappa) \kappa \sin \frac{y}{2} \int_{-\pi}^{\pi} e^{\kappa \cos(x-\mu)} \sin(x-\mu + \frac{y}{2}) dx \\ &= 2c(\kappa) \kappa \sin \frac{y}{2} \int_{-\pi}^{\pi} e^{\kappa \cos(x-\mu)} \left( \sin(x-\mu) \cos \frac{y}{2} + \sin \frac{y}{2} \cos(x-\mu) \right) dx \\ &= 2c(\kappa) \kappa \left( \sin \frac{y}{2} \right)^2 \underbrace{\int_{-\pi}^{\pi} e^{\kappa \cos(x-\mu)} \cos(x-\mu) dx}_{=C(\kappa) > 0} \\ &\leq 2c(\kappa) \kappa C(\kappa) \frac{y^2}{4} \end{aligned}$$

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

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

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

$$\int_{-\pi}^{\pi} f_{vM}(x) \log \frac{f_{vM}(x)}{f_{vM}(x+y)} dx \leq p_* y^2. \quad \square$$

*Proof of Theorem 3.10.* To prove the lower bound stated in Theorem 3.10, we follow the lines of Section 2.5 in [22] 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  $m_0(x) = 0$  and  $m_1(x) = Lh_n^\beta K\left(\frac{x-x_0}{h_n}\right)$  with  $h_n = c_0 n^{-\frac{1}{2\beta+1}}$  and  $K : \mathbb{R} \mapsto \mathbb{S}^1$  satisfying:

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

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

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

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

1. Let us prove that  $m_j \in \tilde{\Sigma}(\beta, L)$ . For  $l = \lfloor \beta \rfloor$  we have

$$m_1^{(l)}(x) = Lh_n^{\beta-l} K^{(l)}\left(\frac{x-x_0}{h_n}\right)$$

then, with  $u = \frac{x-x_0}{h_n}$  and  $u' = \frac{x'-x_0}{h_n}$ ,

$$\begin{aligned} |m_1^{(l)}(x) - m_1^{(l)}(x')| &= Lh_n^{\beta-l} |K^{(l)}(u) - K^{(l)}(u')| \\ &\leq Lh_n^{\beta-l} |u - u'|^{\beta-l}/2 = L|x - x'|^{\beta-l}/2. \end{aligned}$$

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

2. Let us show that  $|m_0(x_0) - m_1(x_0)| \geq 2s_n$ . We have

$$|m_0(x_0) - m_1(x_0)| = |m_1(x_0)| = Lh_n^\beta K(0) = Lc_0^\beta K(0)n^{-\frac{\beta}{2\beta+1}},$$

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

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

$$\begin{aligned} \sup_{m \in \tilde{\Sigma}(\beta, L)} \mathbb{E}_m[\psi_n^{-2} |T_n - m(x_0)|^2] &\geq A^2 \max_{m \in \{m_0, m_1\}} \mathbb{P}_m(|T_n - m(x_0)| \geq A\psi_n) \\ &\geq \frac{A^2}{2} \mathbb{E}_{X_1, \dots, X_n} \left[ \mathbb{P}_{m_0}(|T_n - m_0(x_0)| \geq A\psi_n | X_1, \dots, X_n) \right. \\ &\quad \left. + \mathbb{P}_{m_1}(|T_n - m_1(x_0)| \geq A\psi_n | X_1, \dots, X_n) \right] \\ &\geq \frac{A^2}{2} \mathbb{E}_{X_1, \dots, X_n} \left[ \inf_{\psi} \left\{ \mathbb{P}_{m_0}(\psi \neq 0 | X_1, \dots, X_n) \right. \right. \\ &\quad \left. \left. + \mathbb{P}_{m_1}(\psi \neq 1 | X_1, \dots, X_n) \right\} \right], \end{aligned} \quad (36)$$

where  $\inf_{\psi}$  denotes the infimum over all tests  $\psi$  taking values in  $\{0, 1\}$ .

Now let us fix the  $X_i$ 's. The minimum average probability  $\bar{p}_{e,1}$  is defined as (see page 116 in [22]):

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

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

$$\mathbf{K}(\mathbb{P}_{m_0}, \mathbb{P}_{m_1}) = \int \log \left( \frac{d\mathbb{P}_{m_0}}{d\mathbb{P}_{m_1}} \right) d\mathbb{P}_{m_0} = \sum_{i=1}^n \int \log \frac{p_{\zeta}(y)}{p_{\zeta}(y - m_1(X_i))} p_{\zeta}(y) dy. \quad (37)$$

There exists  $n_0$  such that  $\forall n > n_0$ , we have  $nh_n \geq 1$  and  $Lh_n^\beta K_{\max} \leq y_0$  where  $K_{\max} = \max_x K(x)$ . Using (37) and (16), we have:

$$\begin{aligned}\mathbf{K}(\mathbb{P}_{m_0}, \mathbb{P}_{m_1}) &\leq p_* \sum_{i=1}^n m_1^2(X_i) \\ &\leq p_* L^2 h_n^{2\beta} K_{\max}^2 \sum_{i=1}^n 1_{\left\{ \left| \frac{X_i - x_0}{h_n} \right| \leq \frac{1}{2} \right\}}.\end{aligned}$$

Now taking the expectation and using that the density of the  $X_i$ 's is bounded by  $\mu_0$ , we get:

$$\mathbb{E}_{X_1, \dots, X_n} \mathbf{K}(\mathbb{P}_{m_0}, \mathbb{P}_{m_1}) \leq p_* L^2 K_{\max}^2 h_n^{2\beta} n \mathbb{P} \left( \left| \frac{X_1 - x_0}{h_n} \right| \leq \frac{1}{2} \right) \leq p_* L^2 K_{\max}^2 \mu_0 h_n^{2\beta+1} n.$$

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

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

As in Lemma 2.10 of [22], we introduce the function  $\mathcal{H}(x) = -x \log(x) - (1-x) \log(1-x)$  for  $x \in (0, 1)$  and  $\mathcal{H}(0) = \mathcal{H}(1) = 0$ . Inequality (2.70) of [22] with  $M = 1$  gives

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

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

$$\mathbb{E}_{X_1, \dots, X_n} \bar{p}_{e,1} \geq \mathcal{H}^{-1} \left( \log 2 - \frac{\alpha}{2} \right) > 0,$$

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 (36)

$$\sup_{m \in \tilde{\Sigma}(\beta, L)} \mathbb{E}_m [\psi_n^{-2} |T_n - m(x_0)|^2] \geq A^2 \mathcal{H}^{-1} \left( \log 2 - \frac{\alpha}{2} \right),$$

where the right hand side is a positive constant. This concludes the proof of Theorem 3.10.  $\square$

## 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 [8]. 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 [8], 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.

## References

- [1] M. Alonso-Pena and R. Crujeiras. Nonparametric multimodal regression for circular data. *preprint*, 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.
- [3] G. Chagny. Penalization versus Goldenshluger-Lepski strategies in warped bases regression. *ESAIM Probab. Stat.*, 17: 328-358, 2013.
- [4] G. Chagny. Adaptive warped kernel estimators. *Scandinavian Journal of Statistics*, 42(2): 336-360, 2015.
- [5] G. Chagny, T. Laloë and R. Servien. Multivariate adaptive warped kernel estimation. *Electron. J. Statist.*, 13(1): 1759-1789, 2019.
- [6] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.*, 49(2): 569-609, 2013.
- [7] N.I. Fisher and A.J. Lee. Regression models for angular responses. *Biometrics*, 48: 665-677, 1992.
- [8] M. Di Marzio, A. Panzera and C.C. Taylor. Nonparametric Regression for Spherical Data. *Journal of the American Statistical Association*, 109:748-763, 2014.
- [9] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3): 1608-1632, 2011.
- [10] A.L. Gould. A regression technique for angular variates. *Biometrics*, 25: 683-700, 1969.
- [11] S.R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*. World Scientific, Singapore, 2001.
- [12] R.A. Johnson and T.E. Wehly. Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73: 602-606, 1978.
- [13] G. Kerkycharian and D. Picard. Regression in random design and warped wavelets. *Bernoulli* 10(6):1053-1105, 2004.
- [14] C. Ley and T. Verdebout. *Modern Directional Statistics*. (1st ed.). Chapman and Hall/CRC, 2017.
- [15] K.V. Mardia and P.E. Jupp. *Directional Statistics*. New York, NY: John Wiley, 2000.
- [16] M. Di Marzio, A. Panzera and C.C. Taylor. Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, 40: 238-255, 2013.
- [17] A. Meilán-Vila, M. Francisco-Fernández, R.M. Crujeiras and A. Panzera. Nonparametric multiple regression estimation for circular response. *TEST*, 30: 650-672, (2021).
- [18] A. Pewsey and E. García-Portugués. Recent advances in directional statistics. *TEST*, 30: 1-58, 2021.
- [19] T.M. Pham Ngoc. Regression in random design and Bayesian warped wavelets estimators. *Electron. J. Stat*, 3: 1084-1112, 2009.
- [20] T.M. Pham Ngoc. Adaptive optimal kernel density estimation for directional data. *J. Multivariate Anal.*, 173: 248-267, 2019.
- [21] B. Presnell, S.P. Morrison and R.C. Littell. Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, 93: 1068-1077, 1998.
- [22] A.B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009.