Kernel density estimation for a stochastic process with values in a Riemannian manifold

Mohamed Abdillahi Isman, Wiem Nefzi, Papa Mbaye, Salah Khardani, Anne-Françoise Yao

Kernel density estimation for a stochastic process with values in a Riemannian Manifold

Mohamed Abdillahi Isman[1,4], Wiem Nefzi [2], Papa Mbaye[1,3], Salah Khardani[†2] and Anne-Françoise Yao[1,5]

[1] Laboratoire de Mathématiques Blaise Pascal, CNRS UMR 6620, Campus des Cézeaux, 63171 Aubière Cedex, France.
anne.yao@uca.fr, mohamed.abdillahi_isman@doctorant.uca.fr

[2] Faculté des Sciences Tunis, Université El-Manar, Laboratoire de Modélisation mathématique, Statistique et Analyse stochastique M2ΦSAS, Tunisia.
khardani_salah@yahoo.fr, nefziwiem24@gmail.com

[3] Laboratoire Freeland.
papa.mbaye@missioneo.fr

[4] Université de Djibouti, magouleh@gmail.com

[5] Centre de Mathématiques Appliquées, Ecole polytechnique Route de Saclay 91128 Palaiseau Cedex.
anne.yao@uca.fr

[†] Corresponding author : khardani_salah@yahoo.fr

Abstract

This paper is related to the issue of the density estimation of observations with values in a Riemannian submanifold. In this context, Pelletier (2005) proposed a kernel density estimator for independent data. We investigate here the behavior of Pelletier's estimator when the observations are generated from a strictly stationary $\alpha-$mixing process with values in this submanifold. Our study encompasses both pointwise and uniform analyses of the weak and strong consistency of the estimator. Specifically, we give the rate of convergence in terms of mean square error, probability, and almost sure convergence (a.s.). We also give a central-limit theorem and illustrate our proposal through some simulations and a real data application.

Key words: Kernel density estimator, Riemannian manifolds, Mixing condition, Weak and strong consistency, Central limit theorem.

1 Introduction

The problem of unknown density estimation from a kernel approach has been widely treated when the variable of interest lies in an Euclidean space, either for independent or dependent data. However, in many applications, such as biology (for example, protein data as in Mardia et al. (2007) or Mardia et al. (2008)), spatial statistics (for instance, in studies including information related to positions on the Earth), geology, image analysis (see, for example, Pennec (2006) and references therein), medicine, and so on (see also Dryden and Mardia (1998) for additional examples), the Euclidean assumption about the underlying geometry of the observations fails. An alternative is to treat the data as lying on a submanifold, $\mathcal{M}$, which has an unknown structure to be estimated (we refer, for example, Aamari and Levrard (2019) or Berenfeld and Hoffmann (2021) or Khardani and Yao (2022) and references therein). However, $\mathcal{M}$ can be a known shape such as a sphere, a torus, or a cylinder, where the data can be distributed as directional data, as a Von-Mises or a Kent distribution (the reader can see Mardia and Jupp (2000) and references therein for more information on such distributions). Such intrinsically complex (high-dimensional) data structures have led to new challenges in statistical analysis and inference, which require some specific innovative methods and theories.

Actually, Several theoretical works exist on kernel density estimation for specific Riemannian submanifolds such as (finite-dimensional) unit spheres (proposed, for example, by Hall et al. (1987), Kim and Koo (2002), Kerkyacharian et al. (2011), García-Portugués et al. (2020)) or on a torus Di Marzio et al. (2011) or directional data (see Bai et al. (1988) or García-Portugués et al. (2013)), to name but a few.

We focus here on the more general Riemannian manifold framework. In fact, the literature related to the theoretical behavior of the kernel density estimation in this setting is very limited. The key reference is the work of Pelletier (2005). But, we can also quote other references such as Kim and Park (2013), Henry et al. (2013), Berry and Sauer (2017), Cleanthous et al. (2020) and Berenfeld and Hoffmann (2021), Khardani and Yao (2022) and references therein.

All these references concern the case of independent identically distributed (*i.i.d.*) observations. They have shown that the extension of the former Parzen kernel density estimator to the setting of Riemannian manifolds is far from trivial. Furthermore, to our knowledge, the case where the dataset consists of dependent observations has not

been treated yet. This work aims to contribute to addressing this gap. Namely, we study the behavior of Pelletier's kernel density estimator (Pelletier (2005)) applied to observations of a stationary Riemannian-valued stochastic process that satisfies a strong mixing condition (see, for example, Bradley (2007) for more information on practical applications of the $\alpha-$mixing).

The rest of the paper is organized as follows. In Section 2, we present the theoretical framework and outline the necessary conditions required to get the theoretical asymptotic results. Section 3 is devoted to the asymptotic results. Namely, we study the pointwise and uniform weak and strong consistency of the estimator. We also give a Central Limit Theorem. We illustrate the behavior of the estimator in practice through some simulations and a real data application in Section 4. In Section 5, we give a brief discussion and some perspectives on this study. The proofs of the theorems and propositions are given in Section 6.

2 General setting and notations

This work concerns any measurable stationary random process $(X_t, t \in \mathbb{Z})$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values on the Riemannian submanifold $(\mathcal{M}, g)$ of $\mathbb{R}^d$ $(d \geq 2)$. The X_t's are dependent and distributed as a random variable X with an unknown density f on $\mathcal{M}$. We assume that (X_t) satisfies an $\alpha-$mixing condition, defined by

$$\alpha(n) = \sup_k \sup_{A \in \mathcal{F}_1^k(X),\ B \in \mathcal{F}_{k+n}^{\infty}(X)} \{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\}, \quad for\ n \geq 1,$$

where $\mathcal{F}_i^k(X)$ is the $\sigma-$field generated by $\{X_i,\ i \leq j \leq k\}$ and $\alpha(n)$ tends to zero.

Beside the mixing conditions, our estimator runs under some conditions on the submanifold $(\mathcal{M}, g)$.

For each point $x \in \mathcal{M}$, we will set $T_x\mathcal{M}$ the tangent space to $\mathcal{M}$ at x. We assume that $(\mathcal{M}, g)$ is endowed with a measure, μ_g, and is geodesically complete and compact without boundary. Consequently (by the Hopf Rinow Theorem, refer to Spiegel (2016)), $(\mathcal{M}, d_g)$ is a complete metric space, where d_g is the metric induced by g. In the rest of the paper, we will set $d(\cdot, \cdot) = d_g(\cdot, \cdot)$. This allows us to define the exponential map at x, $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ such that for any $v \in T_x\mathcal{M}$, $\exp_x(v) = \gamma_v(1)$ where the function γ_v defined by $\gamma_v(t) = \exp_x(tv)$, $t \in \mathbb{R}$ is the unique geodesic with $\gamma_v(0) = x$ and $\dot{\gamma}_v(0) = v$. Its inverse, denoted by $\exp_x^{-1}$, is a map from the image,

$\mathrm{Im}\,(\exp_x)$ to $T_x\mathcal{M}$. For more information on these notions, the reader can refer to, for example, Gallot et al. (2004) or Chavel (2006).

We will denote by 0_x and μ_x, respectively, the null vector and Lebesgue measure in $T_x\mathcal{M}$. For simplicity, we will write $\int_{T_x\mathcal{M}} J(v)dv = \int_{T_x\mathcal{M}} J(v)d\mu_x(v)$ for any integrable function J defined on $T_x\mathcal{M}$. The inner product (related to g) in $T_x\mathcal{M}$ will be defined by $\langle u, v\rangle = g(u, v)$ for any $u, v \in T_x\mathcal{M}$, and the associated norm by $||.||$. The unit sphere on $T_x\mathcal{M}$ will be referred as $S_x^{d-1} = \{v \in T_x\mathcal{M}, \|v\| = 1\}$, when $B\,(x, h) = \{y, d(x, y) \leq h\}$ and $B\,(h) := B\,(0_x, h) = \{u, ||u|| \leq h\}$ will denote the balls of radius h centered at x and 0_x, respectively.

We assume that the injectivity radius of $\mathcal{M}$ is such that, $inj(\mathcal{M}) > 0$ and we consider only regular balls in $\mathcal{M}$. We recall that a regular (or convex) ball $B\,(x, h)$ is defined such that $h < h^*$, where $h^* = \min\{inj(\mathcal{M}), \frac{\pi}{2\sqrt{\kappa}}\}$ and κ is the supremum of sectional curvatures of $\mathcal{M}$ (see, for example, Gallot et al. (2004) for definition) such that if this upper bound is positive, and $\kappa = 0$ otherwise. Then, $B(x, h) = \exp_x B\,(h)$.

The above properties naturally lead to the following assertion (which will be helpful for the proofs): For any continuous function $\psi : \mathcal{M} \to \mathbb{R}$ with support on $B(x, h)$ where $h < h^*$, we have

$$\mu_g\,(\psi) = \int_{\mathcal{M}} \psi\,(y)\,d\mu(y) = \int_{\exp_x(B(h))} \psi\,(y)\,d\mu(y) = \int_{B(h)} \psi\,(\exp_x(v))\,|g_x(v)|^{1/2}\,dv,$$

where $|g_x(v)|^{1/2}$ denotes the determinant of $g_x(v)$ with g_x defined as $g_x(v) = (g_{ij}(v))$, $g_{ij}(v) = g\,(\partial x_i, \partial x_j)$ (or $g_x = \sum_{i,j} g_{ij}dx^i dx^j$) is the local expression of g in the coordinate system $(x^1, \ldots, x^d)$. We recall that $|g_x(v)|^{1/2} = \frac{d\mu_g}{d\mu_x}\,(\exp_x(v)) = \frac{d\mu_{\exp_x^* g}}{d\mu_x}\,(v)$ is the density of $\mu_{\exp_x^* g}$ with respect to μ_x on $T_x\,(\mathcal{M})$ (for more details, see Chavel (2006), p.18 or Gallot et al. (2004) p. 165). We recall the another expression: $|g_x(\exp^{-1}(y))|^{1/2} = \theta_x(\exp^{-1}(y))$, where $\theta_x(.)$ is called the volume density function. In the literature, $\theta_x(y) = \theta_x(\exp_x^{-1}(y))$ by abuse of notation (see also le Brigant and Puechmorel (2019)).

The transported density of $\exp_x^{-1}(X)$ with respect to the Lebesgue measure in $T_x\,(\mathcal{M})$ is given by

$$f_{T_x}(v) = f(\exp_x(v))\,|g_x(v)|^{1/2} \quad \forall\, v \in T_x\mathcal{M}. \tag{2.1}$$

Note that $f_{T_x}(0_x) = f(x)\left|g_x(0_x)\right|^{1/2} = f(x)$ since $\left|g_x(0_x)\right|^{1/2} = 1$.
We have the following expansion for any $u \in T_x(\mathcal{M})$

$$\left|g_x(u)\right|^{1/2} = 1 - \frac{Ric_x(u,u)}{6} + O\left(\|u\|^3\right),$$

where Ric_x is the Ricci tensor at x. This expansion can be found, for example, in Karcher (1977), p.191 (or Chavel (1993), p.91).
In this work, we aim to study the asymptotic behavior of the kernel estimator of Pelletier (2005):

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_n^d}\frac{1}{\theta_x(X_i)} K\left(\frac{d(x,X_i)}{h_n}\right),$$

based on observations of the process $(X_t, t \in \mathbb{Z})$, X_i, $i = 1, ..., n$.
In what follows, C (or sometimes c) will represent a positive constant that does not affect the outcomes of our results.

Assumptions

We obtain some theoretical results under the following assumptions.
H1: $K : \mathbb{R}^d \to \mathbb{R}_+$ is a bounded and continuous map such that

1. K satisfies a Lipschitz condition.
2. $supp\ K = [0;1]$.
3. $\int K(\|x\|)dx = 1$.
4. $\int xK(\|x\|)dx = 0$ or at least the vector, $\int xK\left(\|x\|\right)dx$ is orthogonal to span$\{\text{grad}f(x)\}$.
5. $\int \|x\|^2 K(\|x\|)dx < \infty$.

H2: $\alpha(n) \leq Cn^{-\nu}$ for some $\nu > 2$.
H3: $h_n \to 0$, $nh_n^d \to \infty$, $nh_n^{d+4} \to \infty$ and $h_n < \frac{h^*}{2}$ as $n \to \infty$.
H4: f is bounded, twice continuously differentiable at any $x \in \mathcal{M}$ and $\|\text{Hess}f\left(x\right)\|_{HS} < \infty$, where $\|\cdot\|_{HS}$ is the Hilbert Schmidt norm (one can also consider the uniform norm of Hess$f\left(x\right)$).
H5: $\forall i, j$, the joint density $f_{i,j}$ of (X_i, X_j) exists is such that

$$\sup_{i,j}\ \sup_{u,v \in \mathcal{M}\times\mathcal{M}} \left|f_{i,j}(u,v) - f(u)f(v)\right| < M, \quad \text{for some } M > 0.$$

Comments on these assumptions

Since this work is related to situations where the data are both with values in a Riemannian manifold and dependent, the above assumptions appear to be a mix of conditions in both cases to derive consistent results for kernel density estimators in this setting. Namely,

1. Hypotheses **H2** and **H5** are classical conditions when dealing with the study of the kernel density estimator for dependent data.

2. Assumption **H3** is the Riemannian manifolds counterpart of the classical assumptions on the bandwidth for dependent or independent data. Except that here, we require the constraint $h_n < \frac{h^*}{2}$. Actually, because of dealing with Riemannian manifolds, we mainly need the condition $h_n < h^*$ (as recall in Section 2). However, the constraint $h_n < \frac{h^*}{2}$ ensures that x is locally a central point when using the kernel estimator in practice (see, for example, Karcher (1977) and Le (2001), or Pelletier (2005) for more details).

3. Assumptions **H1** and **H4** are classical regularity conditions on K and f, which are helpful for controlling the bias terms (3.1) obtained using the Taylor expansion, for all $h > 0$ and $\|v\| \leq 1$, we have

$$f(\exp_x(hv)) = f(x) + h < \text{grad} f(x), v > + \frac{1}{2} h^2 \text{Hess} f(x)(v, v) + o(h^2), \quad (2.2)$$

where grad and Hess denote the gradient and the Hessian operators, respectively.

3 Asymptotic results

In this section, we give some theoretical results on the asymptotic behavior of $\widehat{f}_n(x)$. Namely, we study both the weak and strong consistency of $\widehat{f}_n(x)$ in terms of Mean Squared Error (MSE), probability and almost sure convergence.

3.1 Weak and Strong Consistency

The bias term

The asymptotic behavior of the bias term has been studied in Pelletier (2005). However, we give here an explicit expression

$$b(x) := \mathbb{E}\widehat{f}_n(x) - f(x) = \frac{h_n^2}{2} \int_{B(1)} K(\|v\|) \text{Hess} f\,(x)\,(v, v) dv + o(h_n^2). \quad (3.1)$$

See also Henry and Rodriguez (2009) or Kim and Park (2013) for some other expressions. Unlike the bias, all the results below are specific theoretical contributions of this paper.

Asymptotic behavior of the variance of the estimator

Proposition 3.1 *Under the assumptions **H1**, **H2** and **H5**, the variance of $\widehat{f}_n(x)$ is given by*

$$V\left(\widehat{f}_n(x)\right) := \mathbb{E}\left[\left(\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x)\right)^2\right] = \frac{1}{nh_n^d}\left[f(x)\int_{B(1)} K^2(\|v\|)dv + o(1)\right] \quad (3.2)$$

and then

$$nh_n^d V\left(\widehat{f}_n(x)\right) \underset{n\to\infty}{\longrightarrow} f(x)\int_{B(1)} K^2(\|v\|)dv.$$

The results (3.1) and (3.2) lead to the following results.

Rate of Convergence in Mean Squared Error

For each $x \in \mathcal{M}$, we set $MSE(x) := \mathbb{E}\left((\widehat{f}_n(x) - f(x))^2\right) = b^2(x) + V\left(\widehat{f}_n(x)\right)$.

Theorem 3.1 *Under the assumptions **H1** to **H5**,*

$$MSE(x) \leq C_{x,f} \times \left(h_n^4 + \frac{1}{nh_n^d}\right),$$

*and since $\mathcal{M}$ is compact, and f and K are bounded (see **H1** and **H4**), we have*

$$\sup_{x\in\mathcal{M}} MSE(x) \leq C \times \left(h_n^4 + \frac{1}{nh_n^d}\right).$$

An optimal rate in MSE meaning for $\widehat{f}_n(x)$ is deduced from Theorem 3.1 in the next corollary.

Corollary 3.2 *The bandwidth which minimizes the function $x \mapsto MSE(x)$, under the assumption **H1** is given by*

$$h_{n,opt} = C\ n^{\frac{-1}{d+4}},$$

and the corresponding MSE is given by

$$MSE(x) = C_{x,f}\ n^{\frac{-4}{4+d}} + o\left(n^{\frac{-4}{d+4}}\right),$$

where $C_{x,f} = C^4 C_1 + C^{-d} C_2$ *with* $C = \left(\frac{dC_2}{4C_1}\right)^{\frac{1}{4+d}}$, $C_1 = \left(\int_{B(1)} Hessf(x)(v,v)K(\|v\|)dv\right)^2$ *and* $C_2 = f(x)\int_{B(1)} K^2(\|v\|)dv$.

Convergence in probability

The Markov inequality leads to the following results.

Theorem 3.3 *Under the assumptions* ***H1*** *to* ***H5****, for a given* $x \in \mathcal{M}$

$$|\widehat{f}_n(x) - f(x)| \longrightarrow 0 \ \textit{in probability,}$$

and since $\mathcal{M}$ *is compact by assumption, we have*

$$\sup_{x\in\mathcal{M}} |\widehat{f}_n(x) - f(x)| \longrightarrow 0 \ \textit{in probability.}$$

We now present some pointwise and uniform rates of convergence in probability, with some additional conditions.

Theorem 3.4 *Under the assumptions* ***H1*** *to* ***H5****, if* $\alpha(n) \leq Cn^{-\nu}$ *with* $\nu > 2$ *(as stated in* ***H2****) and if* $\frac{nh_n^{d\frac{\nu+1}{\nu-1}}}{\log n} \to \infty$*, then for a given* $x \in \mathcal{M}$

$$\widehat{f}_n(x) - f(x) = O_p\left(h_n^2 + \sqrt{\frac{\log n}{nh_n^d}}\right). \tag{3.3}$$

Moreover, since $\mathcal{M}$ *is compact, if* $\nu > d+1$ *and* $n^{-1}(\log n)h_n^{-d\frac{\nu+d+3}{\nu-d-1}} \to 0$*, we have*

$$\sup_{x\in\mathcal{M}} |\widehat{f}_n(x) - f(x)| = O_p\left(h_n^2 + \sqrt{\frac{\log n}{nh_n^d}}\right). \tag{3.4}$$

Almost sure convergence

The following theorem is the Riemannian Manifold counterpart of Theorem 2.1 and Theorem 2.2 of Boente and Fraiman (1988). Further insights can also be found in Bosq (1998) and the references therein.

Theorem 3.5 *Under the assumptions* ***H1*** *to* ***H5****, if* $\frac{nh_n^d}{\log n} \to \infty$ *and* $\alpha(n) \leq Cn^{-\nu}$ *with* $\nu > 1$, $n^{-1}h_n^{-d\frac{\nu+1}{\nu-1}}(\log n)^{\frac{2\nu}{\nu-1}} g(n)^{\frac{1}{\nu-1}} \to 0$ *where* $g(n) = \log n \, (\log\log n)^{1+\varepsilon}$ *for some small* $\varepsilon > 0$*, then for all given* $x \in \mathcal{M}$*, we have*

$$|\widehat{f}_n(x) - f(x)| \longrightarrow 0 \ \ a.s. \tag{3.5}$$

Additionally, since $\mathcal{M}$ *is compact, we have*

$$\sup_{x\in\mathcal{M}} |\widehat{f}_n(x) - f(x)| \longrightarrow 0 \ \ a.s. \tag{3.6}$$

With some additional conditions, we get the following rates of convergence, almost surely.

Theorem 3.6 *Under the assumptions* ***H1*** *to* ***H5****, if* $\alpha(n) \leq Cn^{-\nu}$ *with* $\nu > 3$ *and* $nh_n^{\frac{d(\nu+1)}{\nu-3}} (\log n)^{-\frac{\nu-1}{\nu-3}} g(n)^{\frac{-2}{\nu-3}} \to \infty$ *with* $h_n < \frac{h^*}{2}$ *and* $g(n) = \log n \, (\log\log n)^{1+\varepsilon}$ *for some small* $\varepsilon > 0$*, then for each given* $x \in \mathcal{M}$*, we have*

$$\widehat{f}_n(x) - f(x) = O_{a.s.}\left(h_n^2 + \sqrt{\frac{\log n}{nh_n^d}}\right), \tag{3.7}$$

and because $\mathcal{M}$ *is compact, we have*

$$\sup_{x\in\mathcal{M}} |\widehat{f}_n(x) - f(x)| = O_{a.s.}\left(h_n^2 + \sqrt{\frac{\log n}{nh_n^d}}\right). \tag{3.8}$$

3.2 Asymptotic normality

We now state a Central Limit Theorem (CLT) for each given $x \in \mathcal{M}$.

Theorem 3.7 *Under the assumptions* ***H1*** *to* ***H5****, we have*

$$\sqrt{nh_n^d}\left(\widehat{f}_n(x) - f(x)\right) \longrightarrow \mathcal{N}\left(0, \sigma^2(x)\right),$$

where $\sigma^2(x) = f(x)\int_{B(1)} K^2\left(\|x\|\right) dx$.

Corollary 3.8 *Based on $\widehat{f}_n$, we easily get a plug-in estimator $\widehat{\sigma}_n^2$ of $\sigma^2(x)$. Under the assumptions of Theorem 3.7, we have a confidence interval (CI) for $f(x)$ at a given level $1-\alpha$*

$$\left[\widehat{f}_n(x) - \frac{z_{1-\frac{\alpha}{2}}\widehat{\sigma}_n(x)}{\sqrt{nh_n^d}}, \widehat{f}_n(x) + \frac{z_{1-\frac{\alpha}{2}}\widehat{\sigma}_n(x)}{\sqrt{nh_n^d}}\right],$$

where $z_{1-\frac{\alpha}{2}}$ denotes the $1-\frac{\alpha}{2}$ quantile of $\mathcal{N}(0,1)$.

4 Numerical experiments and real data illustration

Our aim here is to illustrate the behavior of our estimator through simulations and a real data application. This topic can be explored in further studies. Actually, as any kernel estimator, we face the well-known curse of dimensionality issue, which can be more acute in the Riemannian manifold setting due to supplementary conditions on the bandwidth ($h_n < h^*$ or $h_n < \frac{h^*}{2}$). Our strategy is then to give here some information that highlight the impact of these restrictions on the convergence of the estimator. We precise this idea through the following remarks and discussions.

4.1 Preliminaries

Unlike the classical (Euclidean-based) counterpart of the kernel estimator, the additional constraints imposed on h_n can be detrimental in practice due to the well-known curse of dimensionality. In fact, even if h_n^d should tend to 0 as usually with kernel estimation, the main hypothesis governing $\widehat{f}_n$ is that $nh_n^d \to \infty$. In the Euclidean setting, to avoid the dimensionality issue, one commonly uses a trade-off by choosing h_n to be sometimes very far from zero if n is not large enough to ensure $nh_n^d \to \infty$. However, in our context, the condition $h_n < \frac{h^*}{2}$ or at least $h_n < h^*$ makes this approach impractical. It is important to note that the hypothesis $h_n < h^*$ is a crucial geometric condition for any ball $B(x, h_n)$ (see Section 2 for more details). Nonetheless, the condition $h_n < \frac{h^*}{2}$ ensures that every x is a central point of the distribution, such that for all $y \in \mathcal{M}$:

$$f_{K,h_n}(y) = \frac{1}{h_n^d \theta_x(y)} K\left(\frac{d(x,y)}{h_n}\right).$$

Thus, this condition appears in some way as a practical statistical condition to ensure that locally, x is sufficiently surrounded by some x_i's (we refer to the Proposition 2.2 of Pelletier (2005) for more details).

Clearly, these constraints do not allow for practical use of the classical trick for small sample sizes, which consists in taking h_n^d far enough from 0 to balance the low values of n and ensure that $nh_n^d \to \infty$. Thus, the only lever here is to take n large enough or take at least a h_n that prevents nh_n^d from approaching zero, even if it means dealing with $\frac{h^*}{2} < h_n < h^*$ (not that it is the same idea that leads to taking h_n is sometimes very big in practice in the Euclidean case).
This motivates our simulation study, where we investigate the impact of the theoretical constraints h_n on the performance of $\widehat{f}_n$ assessed through specific performance metrics.

Let Me denote a metric. We set

$$h_{opt,\text{global}} = \arg\min_{h>0} Me\left(\widehat{f}_n\right),$$

and

$$h_{opt} = \arg\min_{0<h<\frac{h^*}{2}} Me\left(\widehat{f}_n\right) \text{ or } h_{opt} = \arg\min_{0<h<h^*} Me\left(\widehat{f}_n\right).$$

As previously mentioned, the condition $h_n < \frac{h^*}{2}$ is (theoretically) less crucial than $h_n < h^*$ and $nh_n^d \to \infty$. Therefore, we will prefer $h_n < h^*$ if it is necessary to ensure $nh_n^d \to \infty$ (at least to avoid the stability drawbacks of the estimator when nh_n^d is closed to zero), thus minimizing the curse of dimensionality issues.

Now, since we do not necessarily have $h_{opt} = h_{opt,\text{global}}$, depending on the situation, we will have the following cases.

Case 1 $h_{opt,\text{global}} > h^*$. This case is the worst one, which leads to a loss of performance (compared with the classical bandwidth selection).

Case 2 $h_{opt,\text{global}} < \frac{h^*}{2}$. This is the ideal situation, where $h_{opt} = h_{opt,\text{global}}$.

Case 3 $h_{opt,\text{global}} \in [\frac{h^*}{2}, h^*]$. This case is a trade-off between **Case 1** and **Case 2** where only the less crucial condition ($h_n < \frac{h^*}{2}$) is not satisfied.

Comments

In Case 1, since we cannot have $h_{opt} = h_{opt,\text{global}}$, we are interested in understanding the behavior of $\widehat{f}_n$ when h_{opt} is set to either $\frac{h^*}{2}$ or h^*. These settings are considered for the stability of the estimator mentioned previously. For this case, we will choose $h_{opt} = h^* - \varepsilon > 0$, where $\varepsilon > 0$ is a very small positive number, to avoid boundary issues.
In Case 2, we will naturally have $h_{opt} = h_{opt,\text{global}}$.

In Case 3, the issue will be to choose between $h_{opt} = h_{opt,\text{global}}$ and $h_{opt} = \frac{h^*}{2}$. The gap between the values of the metric at both points will guide our choice.

Metrics of performance

We deal with the following metrics of performance: both are used only for the simulations, while the last one is used in the real data application part.

- The Mean Squared Error, $MSE(\widehat{f}_n) := \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{f}_n(X_i) - f(X_i)\right)^2$, is a classical metric.
- The Median Squared Error, $MedSE(\widehat{f}_n) := \text{median}_{i=1,\dots,n}\left(\widehat{f}_n(X_i) - f(X_i)\right)^2$. This metric serves as a first step to understanding the distribution of the error, as illustrated in Henry et al. (2013).
- The Mean Integrate Squared Error, $MISE(\widehat{f}_n) = \mathbb{E}\|\widehat{f}_n - f\|_2^2$, is particularly relevant since the function f is unknown in this setting. Furthermore, the $MISE$ can be expressed as

  $$MISE = E(ISE),$$

  where ISE (Integrate Squared Error) is such that

  $$ISE = \int (\widehat{f}_n(x) - f(x))^2 d\mu(x).$$

Besides these metrics, the results of Theorem 3.7 and Corollary 3.8 allow us to give a confidence interval, here of level 95 % of the estimation of $\widehat{f}_{n,opt}$ (with respect to h_{opt}) $\left[\widehat{f}_{n,\text{inf95\%}}, \widehat{f}_{n,\text{sup95\%}}\right]$.

In all the application part, we will deal with two Riemannian manifolds: the unit sphere, $\mathbb{S}^2$, and a cylinder in $\mathbb{R}^3$ with radius one (and a given height). Each is endowed with its respective metric: for the unit sphere, the metric is given by $d(x,y) = \arccos\left(x^T y\right)$ and for the cylinder, the metric is given by $d(x_1,x_2) = \sqrt{(z_1-z_2)^2 + (\theta_1-\theta_2)^2}$ where $x_i = (z_i, \theta_i)$ for $i = 1, 2$ in cylindrical coordinates. u^T will denote the transpose of a vector u in $\mathbb{R}^p$, $p \in \{2, 3\}$.

We recall that $h^* = \pi$ if $\mathcal{M}$ is a cylinder, and $h^* = \pi/2$ if $\mathcal{M}$ is a unit sphere. Additionally, for any v in the tangent space,

$$g_x(v)|^{1/2} = \begin{cases} 1 & \text{for} \quad \text{any cylinder} \\ \frac{\sin(||v||)}{||v||} & \text{for} \quad \text{the unit sphere} \end{cases}$$

All the applications have been run using the Epanechnikov's kernel: $K(x) = \frac{3}{4}(1-x^2)1_{\{|x|\leq 1\}}$, which satisfies Assumption **H1**. This assumption states that K should be a kernel in $T_x\mathcal{M}$, unlike directional kernels, which are densities in $\mathcal{M}$.

4.2 Simulations study

In this part, as $\alpha-$mixing processes, we consider first-order Autoregressive process (AR(1)), $(X_t, t \in \mathbb{Z})$ with values either on the unit sphere or a cylinder. Namely, we are interested in the estimation of the marginal density of the X_t's based on $n = 500$ observations. We notice that we have run our procedure with various values of n and get some results where, as expected, $\widehat{f}_n$ performs better for $n > 500$ and underperforms for smaller values of n. In this sense, the behavior of $\widehat{f}_n$ versus n is similar to the Euclidean counterpart. Then, we focus on the behavior of $\widehat{f}_n$ versus h_n while considering the constraints discussed in Section 4.1.

Two models on $\mathbb{S}^2$

We are interested here in two AR(1) models in $\mathbb{S}^2$ denoted by Model 1 and Model 2. In both cases, we will deal with a process (X_t) such that for each t,

$$X_t = (\sin(\theta_t)\cos(\phi_t),\, \sin(\theta_t)\sin(\phi_t),\, \cos(\theta_t))^T, \tag{4.1}$$

with

$$\theta_t = \rho\theta_{t-1} + \varepsilon_t, \qquad \theta_0 = \varepsilon_0, \tag{4.2}$$

and

$$\phi_t = \arctan(v_y/v_x),$$

$V = (v_x, v_y, v_z)$ is a given random variable, and (ε_t) is a white noise distributed as $\arccos(v_z)$, whose distribution will be specified below, depending on the model.

Model 1: AR(1) based on a Von Mises-Fisher distribution on $\mathbb{S}^2$

We consider here a process of type (4.1) with (θ_t) verifying the equation (4.2) with $\rho = 0.1$ and V distributed as a Von Mises-Fisher distribution $VMF(\mu, \kappa)$, with density defined by

$$f_{\mu,\kappa}(X) = \left(\frac{\kappa}{2}\right)^{1/2} I_{1/2}(\kappa) \exp\{\kappa X^T \mu\},$$

such that the parameter of concentration is $\kappa = 2$, the mean parameter is $\mu^T = (0, 1, 0)$, and $I_{1/2}(\kappa)$ is the modified Bessel function defined by $I_{1/2}(\kappa) = \left(\frac{\kappa\pi}{2}\right) \sinh(\kappa)$ for $\kappa \in \mathbb{R}$. For background on such distributions, we refer, for example, to Mardia and Sutton (1978), Mardia and Jupp (2000), García-Portugués et al. (2020), Jammalamadaka and Sengupta (2001). We have simulated some observations distributed according to Model 1 using the r_vMF function from the *rotasym* package in the R software, as proposed by García-Portugués et al. (2020) (Figues 4.1, 4.2).

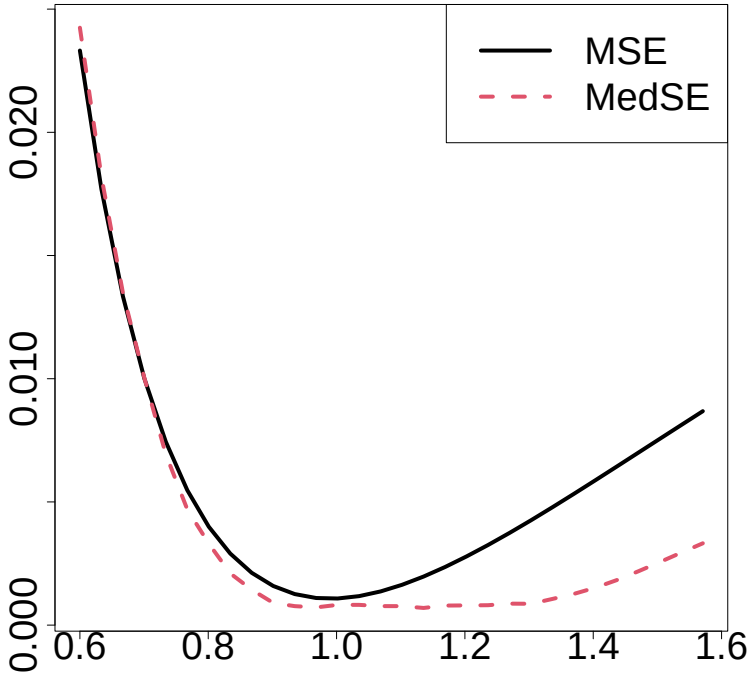


Figure 4.1: MSE and MedSE versus h_n for Model 1. This figure shows that $h_{\text{opt, global}} \in \left(\frac{h^*}{2}, h^*\right)$ for both MSE and $MedSE$, with $h_{\text{opt, global}} \simeq 1$ and $h_{\text{opt, global}} \simeq 1.13$ for MSE and MedSE, respectively.

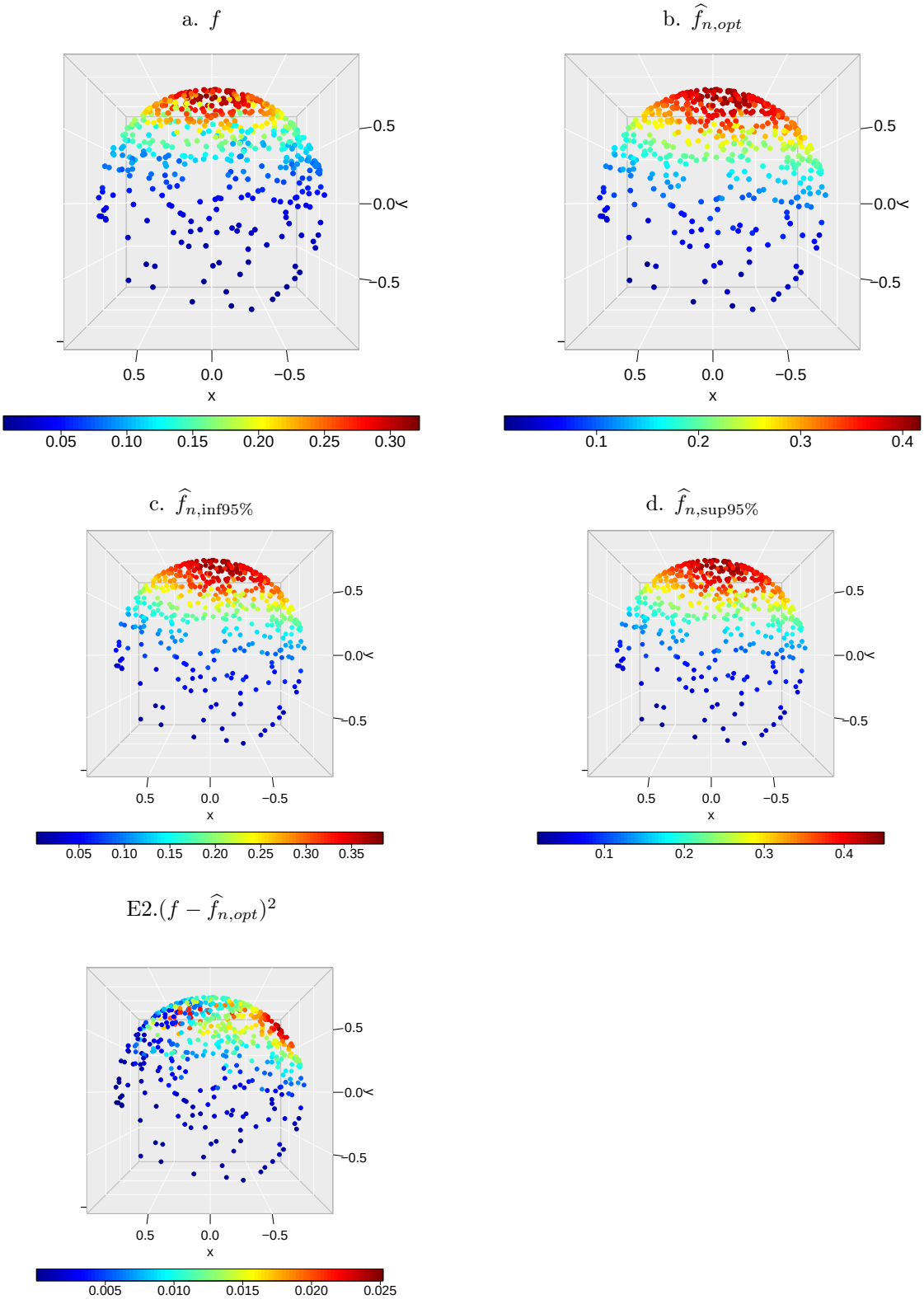


Figure 4.2: Results on the behavior of $\widehat{f}_{n,\text{opt}}$ in Model 1 where $h_{opt} = \frac{\pi}{4}$ and $h^* = \pi/2$.

Model 2: AR(1) based on a Uniform distribution on $\mathbb{S}^2$

In this case, we have simulated the data considering (4.2) with $\rho = 0.1$, θ_0 distributed uniformly in $[0, \pi]$ ($U(0, \pi)$), and (ϕ_t) i.i.d observations from the uniform distribution on $[0, 2\pi]$ ($U(0, 2\pi)$). Here, $h^* = \pi/2$. The results are given in Figure 4.4. Since $f = \frac{1}{4\pi}$ is a constant function, the 3D representation is not really interesting here. Instead, we represent the distribution of the values of $\widehat{f}_{n,opt}$ and we compare it with the true density, $f = \frac{1}{4\pi}$. Note that in Figure 4.3, unlike the MSE, the $MedSE$ presents a local minimum smaller than $\frac{h^*}{2}$.

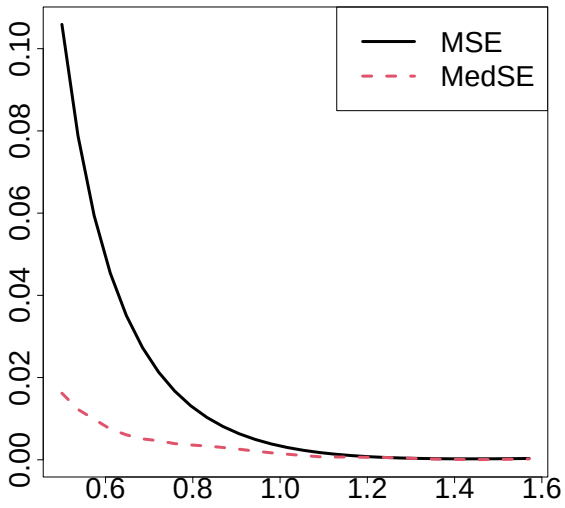
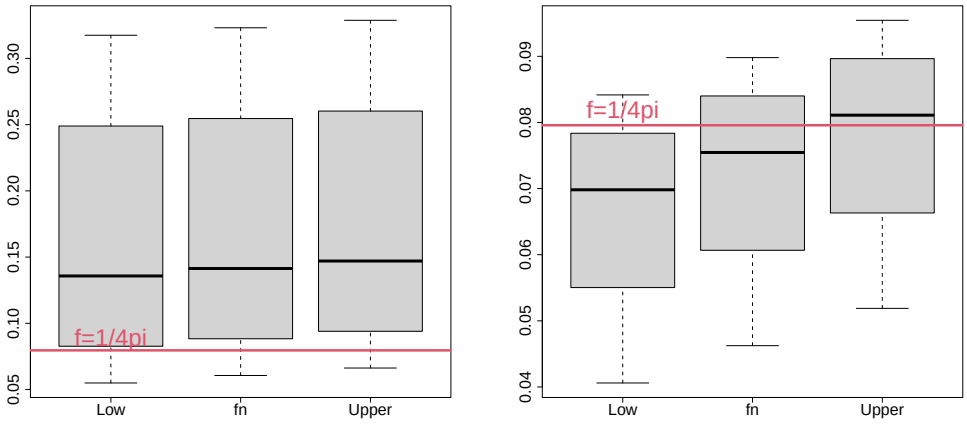


Figure 4.3: Values of MSE and $MedSE$ for Model 2. This figure shows that $h_{opt,\text{global}} \simeq 1.42 \in]\frac{h^*}{2}, h^*[$.



a. $\widehat{f}_n$, $h_{opt} = \frac{\pi}{4}$ b. $\widehat{f}_n$ for $h_{opt} \simeq 1.42$

Figure 4.4: Results on the behavior of $\widehat{f}_n$ for Model 2: a. represents results concerning the case where we choose $h_{opt} = \frac{h^*}{2} = \frac{\pi}{4}$. b. represents results concerning the case where we choose $h_{opt} = h_{opt,\text{global}}$.

Model 3: AR(1) based on a cylindrical Von-Mises distribution.

We deal here with a random process (X_t) such that

$$X_t = (\cos(\theta_t),\, \sin(\theta_t), z_t)\,,$$

where (θ_t) generated as in (4.2), with $\rho = 0.1$, $\theta_0 = \varepsilon_0$, $(\varepsilon_t)$and (z_t) are white noises respectively distributed as $\arccos(v_y/v_x)$ respectively v_z, where $V = (v_x, v_y, v_z)$ follows a Von Mises distribution with density given by

$$f_{VM}(\theta, z) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa cos(\theta - \mu_0)} \frac{1}{\sqrt{2\pi\sigma_c}} e^{-\left[\frac{(z-\mu_c)^2}{2\sigma_c^2}\right]} 1_{\{\theta \in [0, 2\pi[, z \in \mathbb{R}\}},$$

where $\mu_0 = 0$, $\kappa = 2$, $\sigma_c = \sigma^2(1 - ||R||^2)$, and $\mu_c = \mu + \sqrt{\kappa\sigma^2} R^T m_{\theta,\mu_0}$ with $m^T_{\theta,\mu_0} = (\cos(\theta) - \cos\mu_0, \sin(\theta) - \sin\mu_0)$, $R = (0.5, 0.5)$ (which satisfies the constraint $||R|| \leq 1$ as suggested in Mardia and Sutton (1978)), $\sigma = \sqrt{3}$ and $\mu = 1$. I_0 is the modified Bessel function of the first kind and order 0 defined by $I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} e^{\kappa cos(\theta)} d\theta$. We refer, for example, to Mardia and Sutton (1978) for more details on such distribution. We have simulated observations distributed according to Model 3 using the *rvonmises* function from the *Rfast* package in the R software (Figures 4.5, 4.6).

We have run our procedure on some observations simulated from the current model.

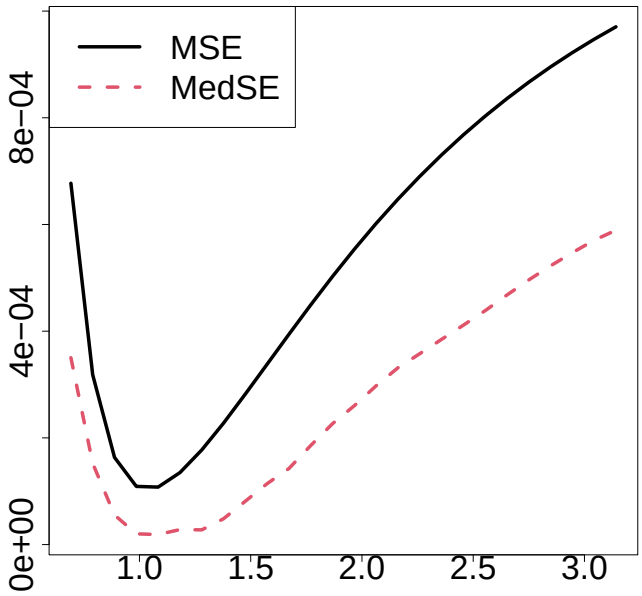


Figure 4.5: Values of MSE and $MedSE$ for Model 3. This figure shows that $h_{opt} = h_{opt,\text{global}} \simeq 1.08$.

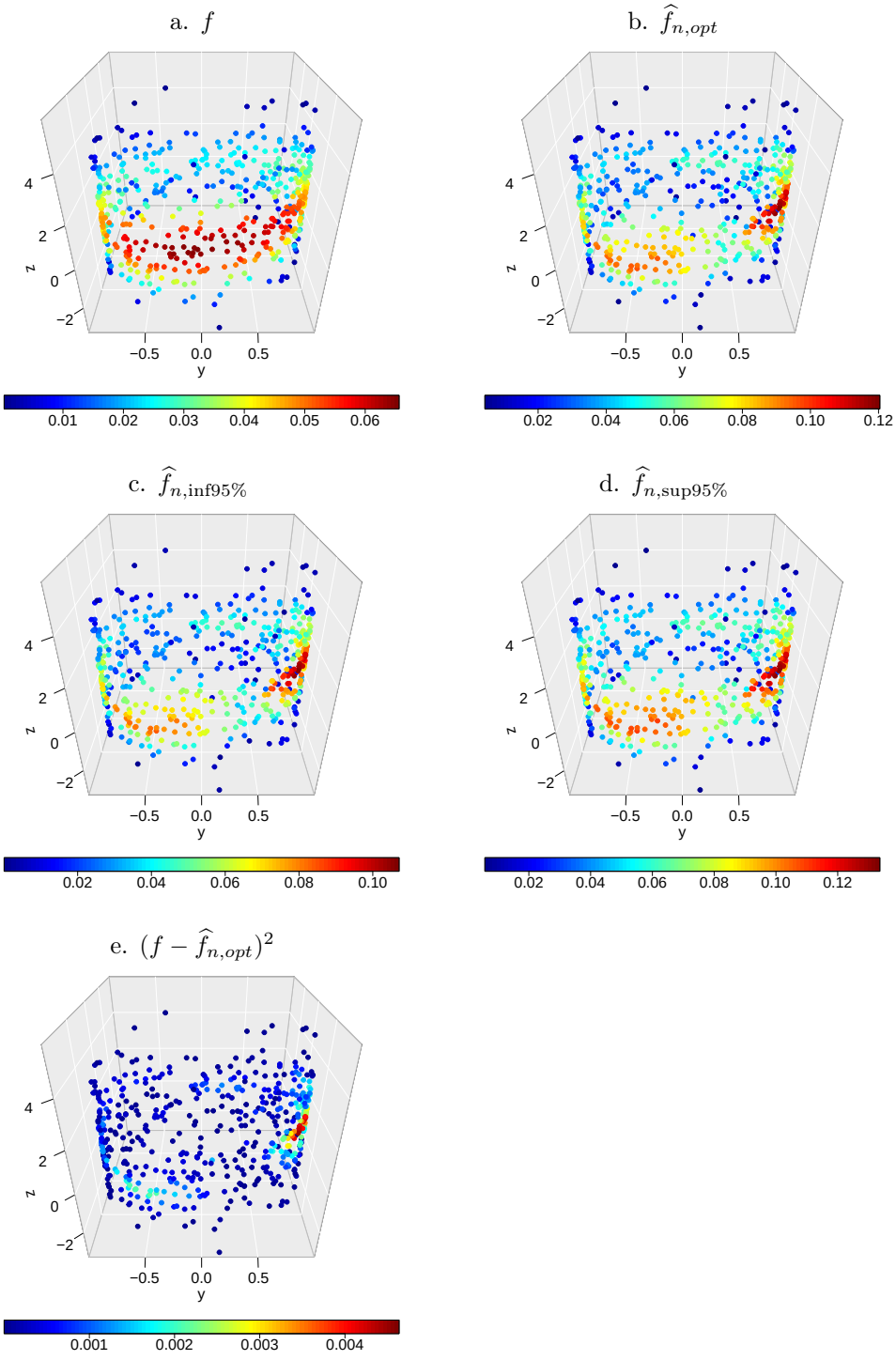


Figure 4.6: Results for $\widehat{f}_{n,\text{opt}}$ with Model 3, where $h_{\text{opt}} = 1.08$.

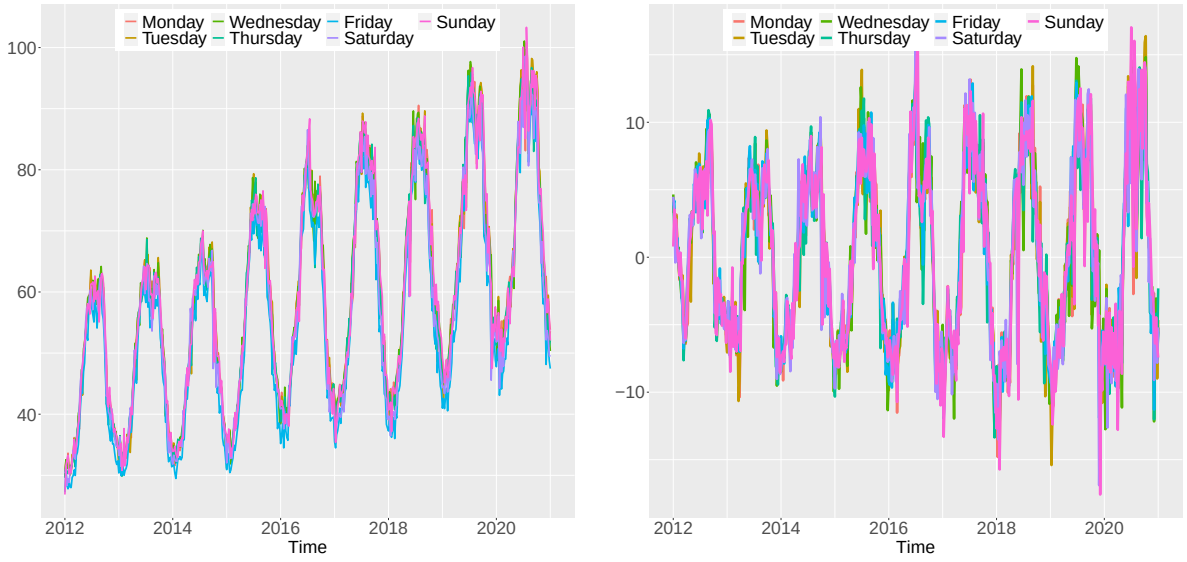
Based on the results shown in Figure 4.6, we conclude that the estimator performs well.

4.3 Real data application: the electricity demand at Djibouti

Our aim in this subsection is to provide some tools which help to understand the behavior of the process of weekly Electricity demand in Djibouti-ville, the main city of Djibouti. The dataset, provided by the company *Electricité De Djibouti*(EDD), consists of daily mean electricity demands recorded from January 1, 2012, to December 31, 2020. It pertains to low-voltage (LV) electricity network customers with subscribed powers from 1kva to 9kva. We are particularly interested in the distribution of the electricity demand for each day of the week: Monday,..., Sunday.

We denote by $(X_{t,j})$ the process representing the level of electricity demand, where $j \in \{\text{Monday}, ..., \text{Sunday}\}$. The observations are presented in Figure 4.7.

In a first step study, we have stationarized each time series, the $(X_{t,j})$'s by removing the deterministic trend and seasonal components. We assume then that the remaining part of $(X_{t,j})$'s are strictly stationary processes. From now on, we identify the $(X_{t,j})$'s with the stationarized version. This previous study has shown that the $(X_{t,j})$'s are periodic with a period of 6-months. Actually, our study concerns the Phase-Amplitude representation of $(X_{t,j})$'s (from their Fourier transform).



Original time series Stationarized version

Figure 4.7: The dataset.

Amplitude, Phase representation

The motivation for such a study arises from its widespread applications such as geology and oceanography, where the study of time series (or signals) commonly involves their representation using Fourier transforms. These transformations lead to complex signals that can be represented either by their real and imaginary parts or by their phase and amplitude representations. In this context, our objective is to illustrate our estimator from this point of view.

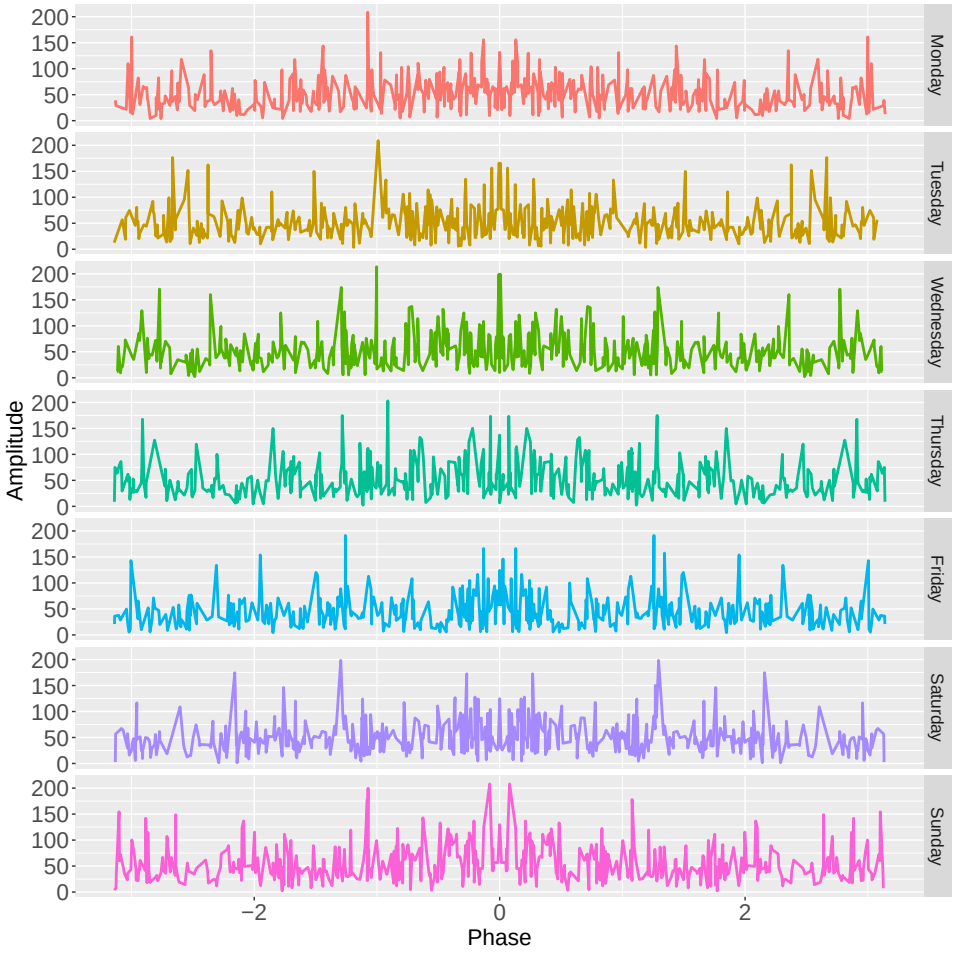


Figure 4.8: Amplitude versus Phase representation for each day.

That is, we are interested in the estimation of the marginal distributions of the $X_{t,j}$'s

(for any j) based on their phase and amplitude representations. Then, we identify each $(X_{t,j})$ with its phase and amplitude series, $X_{t,j} := (\theta_{t,j}, z_{t,j})$ with $(z_{t,j})$ and $(\theta_{t,j})$ being the amplitude(modulus) and the phase(argument) processes, respectively. The observations exclude phase values of 0 and 2π, as illustrated in Figure 4.8. Thus, each $(X_{t,j})$ has an unknown distribution $f^{(j)}$, supported on a cylinder, where $z_{t,j}$ takes values along the "amplitude line" and $\theta_{t,j}$ takes values on $\mathbb{S}^1$. Similar behavior of the daily electricity demand can be modeled using analogous distributions $f^{(j)}$.

As in the simulation part, our strategy consists of first estimating the optimal bandwidth, h_{opt}, and then computing the corresponding $\widehat{f}_{n,opt}$. To determine h_{opt}, we proceed with a leave-one-out cross-validation procedure. It is well known (see Chiu (1991)) that optimizing the $MISE$ is equivalent to optimizing the ISE. The $\widetilde{ISE}$, defined as

$$\widetilde{ISE} = \int \widehat{f}_n^2(x)d\mu(x) - 2E(\hat{f}_n(X)).$$

is a key criterion in this optimization process. We have approximated $\int \widehat{f}_n^2(x)d\mu(x)$ using a Monte-Carlo approach based on observations of the uniform distribution on the cylinder of interest.

The optimization of the $\widetilde{ISE}$ for each j has led to the optimal bandwidths, which are presented in Table 1. The results on the estimation of the $f^{(j)}$'s, $j \in \{\text{Monday}, ..., \text{Sunday}\}$ are presented in Figure 4.10 and Figure 4.11.
It is worth noting that other representations of $\widehat{f}_{n,opt}^{(j)}$ could be presented here, but we have chosen those smoothed versions obtained through local linear smoothing. For clarity, we have avoided highly underrepresented values of $(\theta_{t,j}, z_{t,j})$ (amplitude> 100).

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1.33	1.55	1.12	1.23	1.55	1.55	0.91

Table 1: The optimal bandwidth, h_{opt}, for each day.

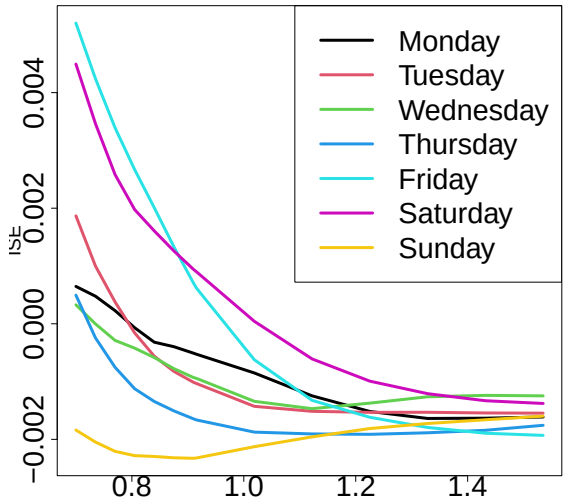


Figure 4.9: Optimization of the $\widetilde{ISE}$'s for each $j \in \{\text{Monday}, \ldots, \text{Sunday}\}$.

Here, we focus on the more restrictive condition $h_{opt} < \frac{h^*}{2}$. The results on Figure 4.9 show that the shape of the $\widetilde{ISE}$ depends on the day of the week, and we have $h_{opt} = h_{opt,\text{global}}$ for Monday, Wednesday, Thursday and Sunday. However, for Tuesday, Friday and Saturday, $h_{opt,\text{global}} \in]\frac{h^*}{2}, \infty[$. In this cases, we computed the corresponding $\widehat{f}_{n,opt}$ with $h_{opt} = 1.55$ to avoid boundary issues. Obviously, the density depends on the day of the week, and further detailed interpretations can be made from a signal analysis perspective.

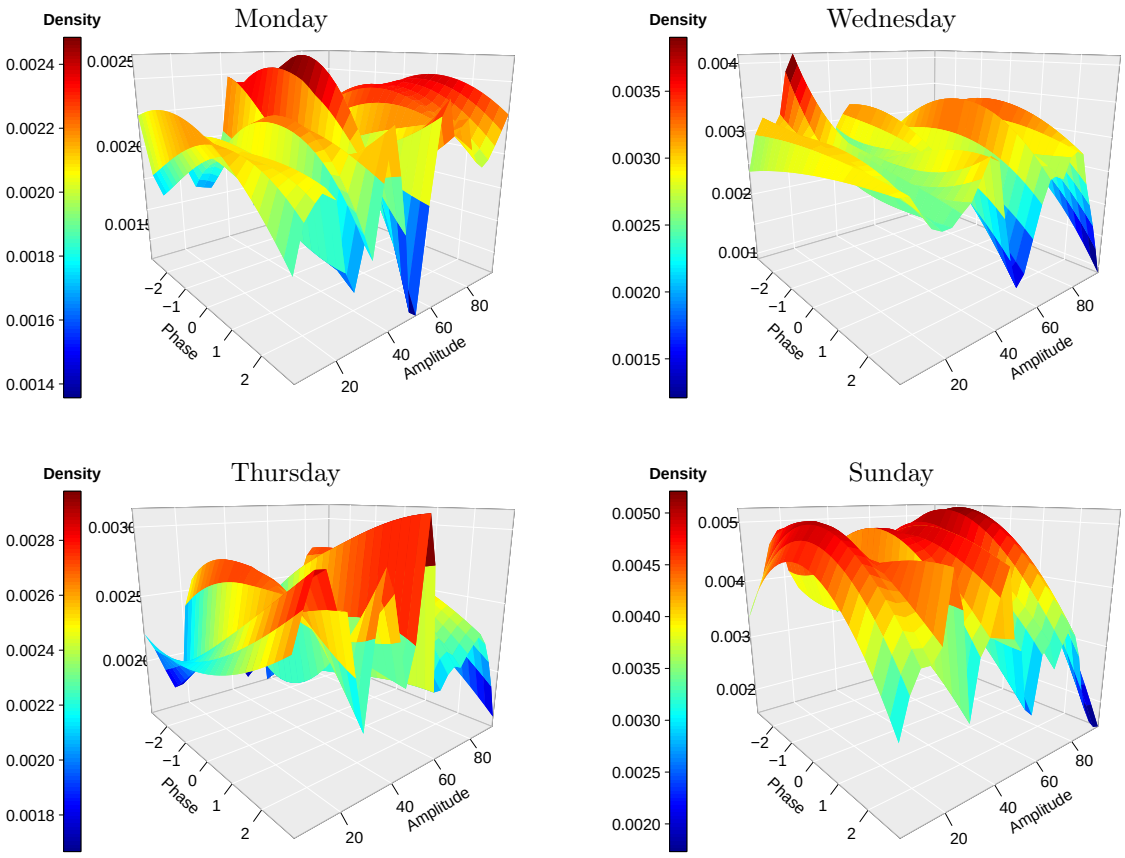


Figure 4.10: $\widehat{f}^{(j)}_{n,opt}$'s for $j \in \{\text{Monday}, \text{Wednesday}, \text{Thursday}, \text{Sunday}\}$.

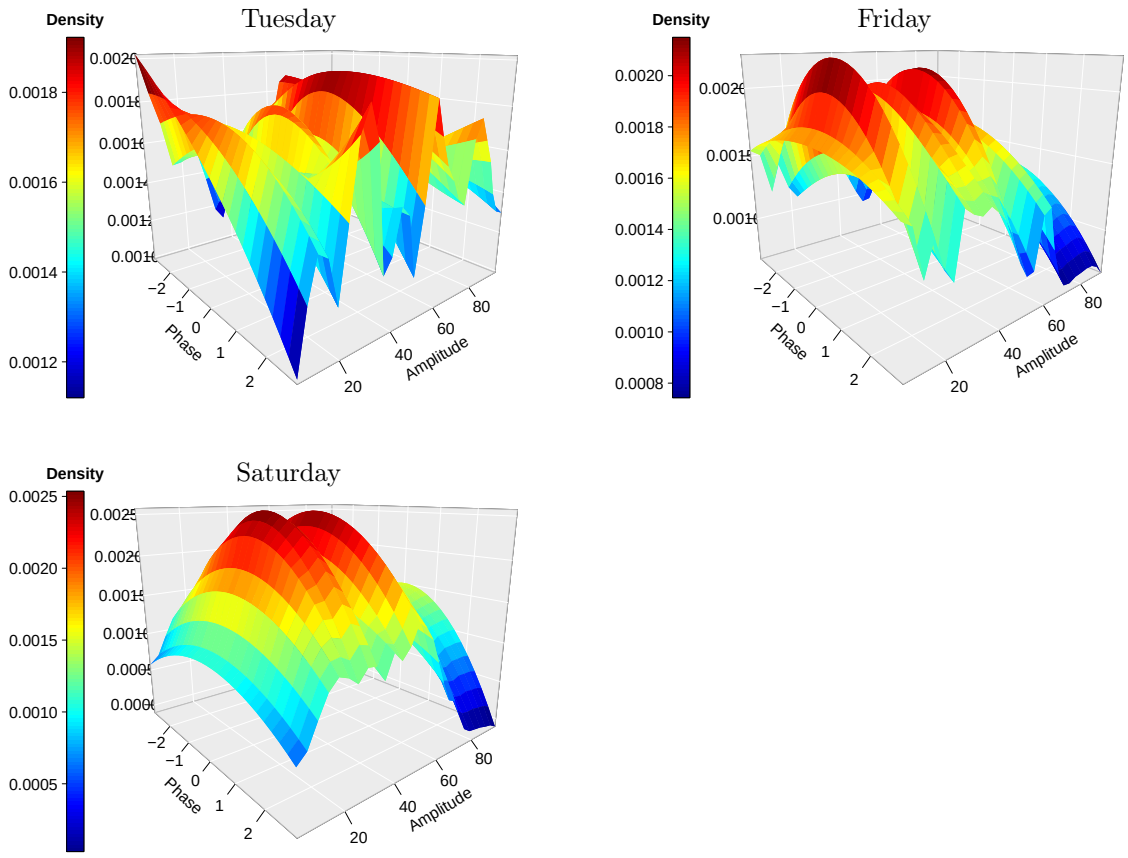


Figure 4.11: $\widehat{f}^{(j)}_{n,opt}$'s for $j \in \{\text{Tuesday}, \text{Friday}, \text{Saturday}\}$.

5 Conclusion

In this work, we study the behavior of the former kernel estimator of Pelletier (2005) when used to estimate the marginal distribution of an $\alpha-$mixing process. We give both weak and strong consistency rates and check the estimator's asymptotic normality.

In the application, we faced some classical issues related to kernel models in the Riemannian manifold setting. Namely, the curse of dimensionality issue which is more acute in this setting. Actually, it is known that if X_i live in a submanifold, they lie on a subspace whose dimension is smaller than that of the ambient space.

Several strategies are proposed in the literature to estimate the latter. Typically, one can proceed with methods such as upstream global or local dimensional reduction or more methods that involve estimating the dimension of the manifold itself. As a first step towards this concept, further developments can be considered by extending the estimator, as suggested in Henry et al. (2013), Wu and Wu (2021), Aamari and Levrard (2019). For additional context on relevant issues, the reader is referred to the introduction of Berenfeld and Hoffmann (2021), which provides numerous key references that support the discussion.

Acknowledgments

This work was supported by the University of El-Manar, Tunis (Tunisia), under Grant PHC-Utique 20G-1503, the Partnership between the University of Djibouti and the University of Clermont Auvergne (France) and Campus France. The authors thank the company "Electricité De Djibouti" (EDD) which has provided the dataset through the Partnership between the University of Djibouti and the University of Clermont Auvergne (France).

References

Aamari, E. and Levrard, C. (2019). Non-asymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177 – 204. 1, 5

Bai, Z., Rao, C., and Zhao, L. (1988). Kernel estimators of density function of directional data. *Journal of Multivariate Analysis*, 27(1):24–39. 1

Berenfeld, C. and Hoffmann, M. (2021). Density estimation on an unknown submanifold. *Electronic Journal of Statistics*, 15(1):2179 – 2223. 1, 5

Berry, T. and Sauer, T. (2017). Density estimation on manifolds with boundary. *Computational Statistics and Data Analysis*, 107:1–17. 1

Boente G and Fraiman R. (1988) Consistency of a nonparametric estimate of a density function for dependent variables. *Journal of Multivariate Analysis*;25(1):90–99. 3.1

Bosq, D. (1998). *Nonparametric statistics for stochastic processes. Estimation and Prediction.* Lecture Notes in Stat., Vol. 110, 2nd. ed., Springer, New York. 3.1, 6

Bradley, R. C. (2007). *Introduction to strong mixing conditions.* Heber City, Utah: Kendrick Press. 1

Chavel, I. (1993). Riemannian geometry-a modern introduction cambridge univ. 2

Chavel, I. (2006). *Riemannian Geometry, A Modern Introduction. Second Edition.* Cambridge University Press. 2

Chiu, S.-T. (1991). Bandwidth selection for kernel density estimation. *The Annals of Statistics*, pages 1883–1905. 4.3

Cleanthous, G., Georgiadis, A. G., Kerkyacharian, G., Petrushev, P., and Picard, D. (2020). Kernel and wavelet density estimators on manifolds and more general metric spaces. *Bernoulli*, 26(3):1832 – 1862. 1

Dabo-Niang, S. and Yao, A.-F. (2007). Kernel regression estimation for continuous spatial processes. *Mathematical Methods of Statistics*, 16(4):298–317. 6, 6

Di Marzio, M., Panzera, A., and Taylor, C. C. (2011). Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*, 141(6):2156–2173. 1

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis.* Wiley New York. 1

Gallot, S., Hulin, D., and Lafontaine, J. (2004). Riemmanian geometry. 2

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013). Kernel density estimation for directional–linear data. *Journal of Multivariate Analysis*, 121:152–175. 1

García-Portugués, E., Paindaveine, D., and Verdebout, T. (2020). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *Journal of the American Statistical Association*, 115(532):1873–1887. 1, 4.2

Hall, P., Watson, G., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762. 1

Henry, G., Muñoz, A., and Rodriguez, D. (2013). Locally adaptive density estimation on riemannian manifolds. *SORT-Statistics and Operations Research Transactions*, pages 111–130. 1, 4.1, 5

Henry, G. and Rodriguez, D. (2009). Kernel density estimation on riemannian manifolds: Asymptotic results. *Journal of Mathematical Imaging and Vision*, 34:235–239. 3.1

Jammalamadaka, S. R. and Sengupta, A. (2001). *Topics in circular statistics*, volume 5. world scientific. 4.2

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541. 2, 2

Kerkyacharian, G., Ngoc, T. M. P., and Picard, D. (2011). Localized spherical deconvolution. *The Annals of Statistics*, 39(2):1042 – 1068. 1

Khardani, S. and Yao, A. F. (2022). Nonparametric recursive regression estimation on riemannian manifolds. *Statistics & Probability Letters*, 182:109274. 1

Kim, P. T. and Koo, J.-Y. (2002). Optimal spherical deconvolution. *Journal of Multivariate Analysis*, 80(1):21–42. 1

Kim, Y. T. and Park, H. S. (2013). Geometric structures arising from kernel density estimation on riemannian manifolds. *Journal of Multivariate Analysis*, 114:112–126. 1, 3.1

Le, H. (2001). Locating fréchet means with application to shape spaces. *Advances in Applied Probability*, 33(2):324–338. 2

le Brigant, A. and Puechmorel, S. (2019). Approximation of densities on Riemannian manifolds. *Entropy*, 21(1). 2

Mardia, K. V., Hughes, G., Taylor, C. C., and Singh, H. (2008). A multivariate von mises distribution with applications to bioinformatics. *Canadian Journal of Statistics*, 36(1):99–109. 1

Mardia, K. V. and Jupp, P. (2000). *Directional data*. New York: Wiley. 1, 4.2

Mardia, K. V. and Sutton, T. W. (1978). A model for cylindrical variables with applications. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(2):229–233. 4.2, 4.2

Mardia, K. V., Taylor, C. C., and Subramaniam, G. K. (2007). Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics*, 63(2):505–512. 1

Pelletier, B. (2005). Kernel density estimation on Riemannian manifolds. *Statistics & Probability Letters*, 73(3):297–304. (document), 1, 2, 2, 3.1, 4.1, 5, 6, 6, 6, 6

Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154. 1

Rio, E. (1999). *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31. Springer Science & Business Media. 6

Wu, H.-T. and Wu, N. (2021). Strong uniform consistency with rates for kernel density estimators with general kernels on manifolds. 5

Spiegel, D. (2016). The hopf-rinow theorem. *Notes available online*. 2

O'neill, Barrett. (1983). Semi-Riemannian geometry with applications to relativity. *Academic press*, 25(1):127–154.

6 Proofs.

For all $x \in \mathcal{M}$, we have

$$V\left(\widehat{f}_n(x)\right) = V\left(\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x)\right) = V\left(\frac{1}{nh_n^d}\sum_{i=1}^{n} Z_i(x)\right),$$

where

$$Z_i(x) = \frac{1}{\theta_x(X_i)}K\left(\frac{d(x,X_i)}{h_n}\right) - \mathbb{E}\left(\frac{1}{\theta_x(X_i)}K\left(\frac{d(x,X_i)}{h_n}\right)\right).$$

It is clear that the $(Z_i)_i$ are dependent and identically distributed with expectation $\mathbb{E}(Z_i) = 0,\ \forall i = 1, ..., n$. Therefore,

$$\begin{aligned} V\left(\widehat{f}_n(x)\right) &= \frac{1}{n^2h_n^{2d}}\sum_{i=1}^{n} V(Z_i(x)) + \frac{1}{n^2h_n^{2d}}\sum_{j\neq i} Cov(Z_i(x), Z_j(x)) \\ &:= A + B. \end{aligned}$$

The expression of A was studied in Pelletier (2005), but we aim here to give an explicit expression.

$$\begin{aligned} A &= \frac{1}{nh_n^{2d}} V(Z_1(x)) \\ &= \frac{1}{nh_n^{2d}} \int_{\mathcal{M}} \frac{1}{\theta_x^2(y)} K^2\left(\frac{d(x,y)}{h_n}\right) f(y) d\mu_g(y) - \frac{1}{nh_n^{2d}} \left(\int_{\mathcal{M}} \frac{1}{\theta_x(y)} K\left(\frac{d(x,y)}{h_n}\right) f(y) d\mu_g(y)\right)^2 \\ &:= I_{1,n}(x) - I_{2,n}(x). \end{aligned}$$

Note that for any $v \in T_x$, $\theta_x(\exp_x(v)) = |g_x(v)|^{\frac{1}{2}} = 1 + O(||v||^2)$ by assumption, then we have the first term on the right is such that

$$\begin{aligned} I_{1,n}(x) &= \frac{1}{nh_n^{2d}} \int_{B(x,h_n)} \frac{1}{\theta_x^2(y)} K^2\left(\frac{d(x,y)}{h_n}\right) f(y) d\mu_g(y) \\ &= \frac{1}{nh_n^{2d}} \int_{B(h_n)} \frac{1}{\theta_x(\exp_x(v))} K^2\left(\frac{\|v\|}{h_n}\right) f\left(\exp_x(v)\right) dv \\ &= \frac{1}{nh_n^{d}} \int_{B(1)} \frac{1}{\theta_x(\exp_x(h_n v))} K^2\left(\|v\|\right) f\left(\exp_x(h_n v)\right) dv. \end{aligned}$$

Therefore

$$\int_{B(1)} \frac{f}{\theta_x}(\exp_x(h_n v)) K^2\left(\|v\|\right) dv \to f(x) \int_{B(1)} K^2\left(\|v\|\right) dv.$$

Hence,

$$nh_n^d I_{1,n}(x) \to f(x) \int_{B(1)} K^2\left(\|v\|\right) dv.$$

Similarly, it is easy to see that

$$\begin{aligned} I_{2,n}(x) &= \frac{1}{nh_n^{2d}} \left(\int \frac{1}{\theta_x(y)} K\left(\frac{d(x,y)}{h_n}\right) f(y) d\mu_g(y)\right)^2 \\ &= \frac{1}{n} \left(\int_{B(1)} K\left(\|v\|\right) f(\exp_x(h_n v)) dv\right)^2. \end{aligned}$$

Using the Taylor expansion, under the assumption **H1** we get

$$I_{2,n}(x) = \frac{f^2(x)}{n} + o(1) = O(n^{-1}).$$

The expression of B is such that

$$\begin{aligned} B &= \frac{1}{n^2 h_n^{2d}} \sum_{j \neq i} \mathbb{E}(Z_i(x) Z_j(x)) \\ &= \frac{1}{n^2 h_n^{2d}} \sum_{(i,j) \in E_1} \mathbb{E}(Z_i(x) Z_j(x)) + \frac{1}{n^2 h_n^{2d}} \sum_{(i,j) \in E_2} \mathbb{E}(Z_i(x) Z_j(x)) \\ &:= J_{1,n}(x) + J_{2,n}(x), \end{aligned}$$

where $E_1 = \{(i,j) \mid 0 < |j-i| \leq \beta_n\}$ and $E_2 = \{(i,j) \mid \beta_n + 1 \leq |j-i| \leq n-1\}$, such that $\beta_n = o(n)$. We have

$$\begin{aligned} \left|Cov(Z_i(x), Z_j(x))\right| &= \left|\mathbb{E}(Z_i(x) Z_j(x))\right| \\ &\leq \int\int \frac{1}{\theta_x(y)\theta_x(z)} K\left(\frac{d(x,y)}{h_n}\right) K\left(\frac{d(x,z)}{h_n}\right) \\ &\times \left|f_{i,j}(y,z) - f(y)f(z)\right| d\mu_g(y) d\mu_g(z) \\ &\leq M h_n^{2d} \int_{B(1)} \int_{B(1)} K(\|u\|) K(\|v\|) dv du \quad (By\ \mathbf{H5}) \\ &\leq M\ h_n^{2d} \quad (By\ \mathbf{H1}). \end{aligned}$$

Therefore, $J_{1,n}(x) \leq \frac{M}{n}\beta_n = O(n^{-1}\beta_n) = o(1)$. For E_2, we use the modified Davydov inequality for mixing processes (see Rio (1999), p.10, formula 1.12a). This leads, for all $i \neq j$, to

$$|Cov(Z_i(x), Z_j(x))| \leq c\alpha(|i-j|).$$

Then, $J_{2,n}(x) \leq \frac{c}{n^2 h_n^{2d}} \sum_{i=1}^{n} \sum_{\beta_n + 1 < |i-j| \leq n-1} \alpha(|i-j|) < \frac{c}{n h_n^{2d}} \int_{\beta_n+1}^{n-1} s^{-\nu} ds = O(n^{-1}\beta_n^{1-\nu} h_n^{-2d})$.

Choosing $\beta_n = h_n^{\frac{-2d}{\nu}}$ allows us to get, under Assumption **H2**,

$$J_{1,n}(x) + J_{2,n}(x) = o\left(\frac{1}{n h_n^{\frac{2d}{\nu}}}\right) = \frac{1}{n h_n^d} o(1),$$

where $o(1)$ is independent of x. Finally, we obtain

$$V\left(\widehat{f}_n(x)\right) = \frac{1}{n h_n^d} \left(f(x) \int_{B(1)} K^2\left(\|v\|\right) dv + o(1)\right). \tag{6.1}$$

Proof of Theorem 3.1

Clearly

$$\sup_{x\in\mathcal{M}} MSE(x) \leq \sup_{x\in\mathcal{M}} b^2(x) + \sup_{x\in\mathcal{M}} V\left(\widehat{f}_n(x)\right).$$

By Assumption **H4** and for the same reasons as provided by Pelletier (2005)(on p.302), for all $v \in B(1)$, we have

$$|f(\exp_x(h_n v)) - f(x) + h_n \langle \text{grad} f(x), v\rangle| \leq C h_n^2 \|v\|^2.$$

Therefore, under **H1**, we deduce that $|b(x)| \leq C h_n^2$, where $C > 0$ is independent of x. Additionally, due to the compactness of $\mathcal{M}$, the following bound holds

$$\sup_{x\in\mathcal{M}} |b(x)| \leq C h_n^2.$$

Now, using the fact that $o(1)$ is independent of x and f and K are bounded in (3.2), we get straightforwardly

$$\sup_{x\in\mathcal{M}} |V(\widehat{f}_n(x))| \leq C \frac{1}{n h_n^d},$$

which achieved the proof.

Proof of Corollary 3.2

Based on results (3.1) and (3.2), for all $x \in \mathcal{M}$, we have

$$MSE(x) = C_1 h_n^4 + C_2 \frac{1}{n h_n^d} + o\left(h_n^4 + \frac{1}{n h_n^d}\right),$$

where $C_1 = \left(\int_{B(1)} Hess(f(x))(v,v) K(\|v\|) dv\right)^2$ and $C_2 = f(x) \int_{B(1)} K^2(\|v\|) dv$. Then, the goal of minimizing the MSE can be achieved by focusing on minimizing the term $C_1 h_n^4 + C_2 \frac{1}{n h_n^d}$. Hence, we get $h_{n,opt} = C\ n^{\frac{-1}{d+4}}$, where $C = \left(\frac{dC_2}{4C_1}\right)^{\frac{1}{4+d}}$.

The following proofs are based on the following lemma, which is a Large Deviation result for the process $\left(\widehat{f}_n(x) - \mathbb{E}\left(\widehat{f}_n(x)\right), n \in \mathbb{Z}\right)$, as a consequence of Lemma 3.2 of Dabo-Niang and Yao (2007).

Lemma 6.1 *Let $n \geq 1$, without loss of generality, assume that there exists an integer p and a real q such that $n = 2pq$. If **H1**, **H4** and **H5** hold, then for each $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\widehat{f}_n(x) - \mathbb{E}\left(\widehat{f}_n(x)\right)\right| > \varepsilon\right) \leq 4\exp\left(-\frac{\varepsilon}{4A_0}qh_n^d\right) + \frac{8b_2\alpha(p)}{\varepsilon h_n^d}$$

for some positive constant A_0 and b_2.

In the following, let $C_g = \sup_{x\mathcal{M}} C_g(y)$, where $C_g(y) = \sup_{y\in B(x,h_n)} \theta_x^{-1}(y)$ (we refer to page 303 of Pelletier (2005) for the existence of $0 < C_g < \infty$ based on the geometrical assumptions).

proof of Lemma 1.1

Let $\zeta_i(x) = \frac{Z_i(x)}{h_n^d}$. Then, Assumption **H1** ensures that setting $b_2 = \sup_{v\in B(h_n)} K(||v||)\times C_g$, we have for all $x, y \in \mathcal{M}$,

$$\left|\frac{K\left(\frac{d(x,y)}{h_n}\right)}{\theta_x(y)}\right| \leq b_2$$

and consequently, $\zeta_i(x)$ is with 0 mean $|\zeta_i(x)| \leq \frac{b_2}{h_n^d}$. By applying Lemma 3.2 in Dabo-Niang and Yao (2007) with $N = 1$, we get

$$\mathbb{P}\left(\left|\widehat{f}_n(x) - \mathbb{E}\left(\widehat{f}_n(x)\right)\right| > \varepsilon\right) \leq 4\exp\left(-\frac{\varepsilon^2}{4v(q)^2}q\right) + \frac{8b_2\alpha(p)}{\varepsilon h_n^d},$$

where $v(q)^2 = \frac{2}{p^2}\sigma^2(q) + \frac{b_2}{h_n^d}\varepsilon$ and $\sigma^2(q) = \mathbb{E}\left[\left(\sum_{j=1}^{p+1}\zeta_j(x)\right)^2\right] = V\left(\widehat{f}_{p+1}(x)\right)$. From Proposition 3.1, we know that $\sigma^2(q) \leq \frac{a}{ph_n^d}$, where $a = a(K, ||f||_\infty, d, \nu)$. Then,

$$v(q)^2 \leq \frac{b_2}{h_n^d}\varepsilon + \frac{2a}{ph_n^d}.$$

Setting $p = [\varepsilon^{-1}]$, we get $v(q)^2 \leq \frac{A_0\varepsilon}{h_n^d}$, where A_0 is a positive constant.

Proof of Theorem 3.4

Proof of the result (3.3)

Our aim is to show that

$$\left|\widehat{f}_n(x) - \mathbb{E}\left(\widehat{f}_n(x)\right)\right| = O_p\left(\Psi_n\right),$$

where $\Psi_n = \sqrt{\frac{\log n}{nh_n^d}}$.
We apply Lemma 5.1 with $\varepsilon = \eta\Psi_n$, where $\eta > 0$ is a constant. Consequently, we have $\frac{\alpha(p)}{\varepsilon h_n^d} = \frac{\alpha(p)}{\eta h_n^d \Psi_n}$. By setting $p = [\Psi_n^{-1}]$, we have $q \geq \frac{\Psi_n n}{2}$. Since $\Psi_n = \sqrt{\frac{\log n}{nh_n^d}}$, we have $\frac{\varepsilon}{A_0} q h_n^d \geq \frac{\eta \Psi_n^2 n h_n^d}{2A_0} = \frac{\eta}{2A_0}\log n$. Now, let $c = \frac{\eta}{8A_0}$, we obtain

$$\mathbb{P}\left(\left|\widehat{f}_n(x) - \mathbb{E}\left(\widehat{f}_n(x)\right)\right| > \varepsilon\right) \leq C\left(n^{-c} + v_n\right), \tag{6.2}$$

with $v_n = \frac{\alpha(p)}{\Psi_n h_n^d}$. By Assumption **H2**, we have $v_n = \frac{\alpha(p)}{\Psi_n h_n^d} \leq C p^{-\nu} \Psi_n^{-1} h_n^{-d} \sim C\Psi_n^{\nu-1} h_n^{-d} = C\left(n^{-1}\log n\right)^{\frac{\nu-1}{2}} h_n^{-d\frac{\nu+1}{2}}$. Consequently,

$$v_n \leq C\left(n^{-1} h_n^{-\frac{d(\nu+1)}{\nu-1}} \log n\right)^{\frac{\nu-1}{2}}, \tag{6.3}$$

and

$$\mathbb{P}\left(\left|\widehat{f}_n(x) - \mathbb{E}\left(\widehat{f}_n(x)\right)\right| > \varepsilon\right) \leq C\left(n^{-c} + \left(n^{-1} h_n^{-\frac{d(\nu+1)}{\nu-1}} \log n\right)^{\frac{\nu-1}{2}}\right), \tag{6.4}$$

which achieved the proof since $\nu > 1$ and $n^{-1} h_n^{-\frac{d(\nu+1)}{\nu-1}} \log n \to 0$ by assumption.

Proof of the result (3.4)

Note that we have $\sup_{x\in\mathcal{M}} |\widehat{f}_n(x) - f(x)| \leq \sup_{x\in\mathcal{M}} |\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x)| + O_p\left(h_n^2\right)$. Therefore, our aim is to show that

$$\sup_{x\in\mathcal{M}} \left|\widehat{f}_n(x) - \mathbb{E}\left(\widehat{f}_n(x)\right)\right| = O_p\left(\Psi_n\right),$$

where $\Psi_n = \sqrt{\frac{\log n}{nh_n^d}}$. Since $\mathcal{M}$ is compact, it can be covered with μ_n balls which have x_k as a center and a_n as a radius that satisfies: $a_n \leq (h^*)^{-\frac{d}{2}} h_n^{(d+1)} \Psi_n$, and

$\mu_n \leq C(h_n^{(d+1)}\Psi_n)^{-d}$ for some constant $C > 0$. We then have

$$\begin{aligned}\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x) &= \left(\widehat{f}_n(x) - \widehat{f}_n(x_k)\right) + \left(\mathbb{E}\widehat{f}_n(x_k) - \mathbb{E}\widehat{f}_n(x)\right) + \left(\widehat{f}_n(x_k) - \mathbb{E}\widehat{f}_n(x_k)\right) \\ &= S_{1n} + S_{2n} + S_{3n}.\end{aligned}$$

Then,

$$\sup_{x\in\mathcal{M}} |\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x)| \leq \max_{1\leq k\leq \mu_n} \sup_{x\in B(x_k,a_n)} |S_{1n} + S_{2n}| + \max_{1\leq k\leq \mu_n} |S_{3n}|. \tag{6.5}$$

For all $x \in B(x_k, a_n)$, we have

$$\begin{aligned}S_{1n} &= \frac{1}{nh_n^d}\sum_{i=1}^{n}\left(\frac{1}{\theta_x(X_\mathbf{i})}K\left(\frac{d_g(x,X_i)}{h_n}\right) - \frac{1}{\theta_{x_k}(X_i)}K\left(\frac{d_g(x_k,X_i)}{h_n}\right)\right) \\ &\leq \frac{1}{h_n^d}\sup_{y\in B(x,h_n)\cap B(x_k,h_n)}\left|\frac{1}{\theta_x(y)}\right|\left|K\left(\frac{d_g(x,y)}{h_n}\right) - K\left(\frac{d_g(x_k,y)}{h_n}\right)\right| \\ &+ \frac{1}{h_n^d}\sup_{y\in B(x,h_n)\cap B(x_k,h_n)}\left|\frac{1}{\theta_x(y)} - \frac{1}{\theta_{x_k}(y)}\right| K\left(\frac{d_g(x_k,y)}{h_n}\right).\end{aligned}$$

Concerning the first right term, we use the fact that K is Lipschitz (by **H1**). For the second right term, we use the fact that $\theta_x(y) = \theta_y(x)$ for all $x, y \in \mathcal{M}$ (we refer to Pelletier (2005)) and the Taylor expansion in the paper which leads to

$$|\theta_x(y) - 1| \leq C||Ric_x||_{hs}||\exp_x^{-1}(y)||^2$$

where Ric_x is the Ricci tensor at x which is equivalent to

$$|\theta_x(y) - 1| \leq C||Ric_x||_{hs}d(x,y)^2$$

as soon as y belongs to regular ball centered at x. Then, we obtain

$$
\begin{aligned}
|S_{1n}| &\leq \frac{cC_g}{h_n^{d+1}} \sup_{y\in B(x,h_n)\cap B(x_k,h_n)} |d_g(x,y)-d_g(x_k,y)| \\
&+ \frac{C_g}{h_n^d} \sup_{y\in B(x,h_n)\cap B(x_k,h_n)} \frac{1}{\theta_{x_k}(y)} K\left(\frac{d_g(x_k,y)}{h_n}\right) \times |\theta_y(x_k)-\theta_y(x)| \\
&\leq \frac{cC_g}{h_n^{d+1}} d_g(x,x_k) + \frac{C_g^2\|K\|_\infty}{h_n^d} \sup_{y\in B(x,h_n)\cap B(x_k,h_n)} C||Ric_y||_{HS}|| \\
&\times \left(\exp_y^{-1}(x)||^2 + ||\exp_y^{-1}(x_k)||^2\right) \\
&\leq \frac{cC_g}{h_n^{d+1}} d_g(x,x_k) + \sup_{y\in\mathcal{M}} ||Ric_y||_{HS} d(x_k,x)^2 C\frac{C_g^2}{h_n^d}\|K\|_\infty \\
&\leq \left(cC_g + a_n h_n C C_g^2 \sup_{y\in\mathcal{M}} ||Ric_y||_{hs}\|K\|_\infty\right)\frac{Ca_n}{h_n^{d+1}}
\end{aligned}
$$

with c and C some positives constants. Consequently,

$$
\max_{1\leq k\leq \mu_n} \sup_{x\in B(x_k,a_n)} |S_{1n}| = O\left(\Psi_n\right) \text{ in probability,} \tag{6.6}
$$

and the same result holds for S_{2n} since $S_{2n} = \mathbb{E}(S_{1n})$. Now, using (6.4), we obtain

$$
\mathbb{P}\left(\max_{1\leq k\leq \mu_n} |S_{3n}| > \varepsilon\right) \leq C\mu_n \left(n^{-c} + \left(n^{-1}h_n^{-\frac{d(\nu+1)}{\nu-1}}\log n\right)^{\frac{\nu-1}{2}}\right). \tag{6.7}
$$

We have $\mu_n \leq Cn^{\frac{d}{2}} h_n^{-d(\frac{d}{2}+1)} (\log n)^{\frac{-d}{2}}$. Since $nh_n^d \to \infty$ by assumption **H3**, for n large enough, there exists $C>0$ such that $nh_n^d > C$. It follows that $h_n^{-d(\frac{d}{2}+1)} \leq Cn^{\frac{d}{2}+1}$. Consequently,

$$
\mu_n n^{-c} \leq Cn^{d+1-c}(\log n)^{\frac{-d}{2}}, \tag{6.8}
$$

which goes to 0 since $c > d+1$. Next, using (6.3), we get

$$
\mu_n v_n \leq C\left(n^{-1}(\log n) h_n^{-d\frac{\nu+d+3}{\nu-(d+1)}}\right)^{\frac{\nu-(d+1)}{2}}. \tag{6.9}
$$

The proof is then achieved since $\nu > d+1$ and $n^{-1}(\log n)h_n^{-d\frac{\nu+d+3}{\nu-d-1}} \to 0$.

Proof of Theorem 3.5

Proof of the result (3.5)

By Lemma 6.1, we know that for all $\varepsilon > 0$

$$\sum_{n\geq 1} \mathbb{P}\left(\left|\widehat{f}_n(x) - \mathbb{E}\left(\widehat{f}_n(x)\right)\right| > \varepsilon\right) \leq 4\sum_{n\geq 1} A_n + \frac{8b_2}{\varepsilon}\sum_{n\geq 1} B_n, \tag{6.10}$$

where $A_n = \exp\left(-\frac{\varepsilon}{4A_0} q h_n^d\right)$ and $B_n = \frac{\alpha(p)}{h_n^d}$. Let $p = \left[\frac{nh_n^d}{(\log n)^2}\right]$. Then, $\frac{\varepsilon}{4A_0} q h_n^d = \frac{\varepsilon}{4A_0}\frac{n}{2p} h_n^d \geq \frac{\varepsilon}{8A_0}(\log n)^2 > c \log n$, with $c > 2$ for n large enough. Therefore, we conclude that $\sum\limits_{n\geq 1} A_n < \infty$.
Under Assumption **H2**, we have

$$\frac{\alpha(p)}{h_n^d} \leq C\frac{p^{-\nu}}{h_n^d} \sim Cn^{-\nu} h_n^{-d(\nu+1)} (\log n)^{\nu} = C\left(n^{-1} h_n^{-d\frac{\nu+1}{\nu}} (\log n)\right)^{\nu},$$

which goes to 0 by assumption with $\nu > 2$. For some $\varepsilon > 0$, we consider $g(n) = \log n \, (\log\log n)^{1+\varepsilon}$. Then, we deduce that

$$\begin{aligned} B_n n g(n) &\leq C\left(n^{-1} h_n^{-d\frac{\nu+1}{\nu}} (\log n)\right)^{\nu} n g(n) \\ &\leq C\left(n^{\frac{-\nu+1}{\nu}} h_n^{-\frac{d}{2}\frac{\nu+1}{\nu}} \log n \; g(n)^{\frac{1}{\nu}}\right)^{\nu} \\ &\leq C\left(n^{-1} h_n^{-d\frac{\nu+1}{\nu-1}} (\log n)^{\frac{\nu}{\nu-1}} g(n)^{\frac{1}{\nu-1}}\right)^{\nu-1}. \end{aligned}$$

Since $\nu > 1$ and $n^{-1} h_n^{-d\frac{\nu+1}{\nu-1}} (\log n)^{\frac{\nu}{\nu-1}} g(n)^{\frac{1}{\nu-1}} \to 0$, we deduce that $B_n n g(n) \to 0$. Consequently, there exists an integer n_0 such that for all $n > n_0$, we have $B_n n g(n) \leq C$. Then, $\sum_{n>1} B_n \leq \sum_{n\leq n_0} B_n + C \times \sum_{n>n_0} \frac{1}{ng(n)} < \infty$.

Proof of the result (3.6)

Here, we have to prove that

$$\sum_{n\geq 1} \mathbb{P}\left(\sup_{x\in\mathcal{M}} |\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x)| > \varepsilon\right) < \infty.$$

Based on (6.5), it suffices to show that $\sum_{n\geq 1}\mathbb{P}\left(\max_{1\leq k\leq \mu_n}\left|\widehat{f}_n(x_k)-\mathbb{E}\widehat{f}_n(x_k)\right|>\varepsilon\right)<\infty$. By combining (6.7) and (6.10), we have for all $\varepsilon>0$, $\sum_{n\geq 1}\mathbb{P}\left(\max_{1\leq k\leq \mu_n}\left|\widehat{f}_n(x_k)-\mathbb{E}\widehat{f}_n(x_k)\right|>\varepsilon\right)<\infty$. Then, by Borel-Cantelli Lemma, $\max_{1\leq k\leq \mu_n}\left|\widehat{f}_n(x_k)-\mathbb{E}\widehat{f}_n(x_k)\right|\to 0$ a.s. Consequently, the proof is achieved.

Proof of Theorem 3.6

Proof of the result (3.7)

Recall that $g(n)=\log n\,(\log\log n)^{1+\varepsilon}$ for some $\varepsilon>0$, we deduce from (6.3) that

$$\begin{aligned} v_n n g(n) &\leq C\left(nh_n^{\frac{d(\nu+1)}{\nu-1}}(\log n)^{-1}\right)^{\frac{-\nu+1}{2}} ng(n)\\ &\leq C\left(n^{\frac{\nu-3}{\nu-1}}h_n^{\frac{d(\nu+1)}{\nu-1}}(\log n)^{-1}g(n)^{\frac{-2}{\nu-1}}\right)^{\frac{-\nu+1}{2}}\\ &\leq C\left(nh_n^{\frac{d(\nu+1)}{\nu-3}}(\log n)^{-\frac{\nu-1}{\nu-3}}g(n)^{\frac{-2}{\nu-3}}\right)^{\frac{-\nu+1}{2}}. \end{aligned}$$

Since $\nu>3$ and $nh_n^{\frac{d(\nu+1)}{\nu-3}}(\log n)^{-\frac{\nu-1}{\nu-3}}g(n)^{\frac{-2}{\nu-3}}\to\infty$ by assumption, we have $v_n n g(n)\to 0$. Consequently, there exists an integer n_0 such that for $n>n_0$, $v_n n g(n)\leq C$. Then, $\sum_{n>1}v_n\leq\sum_{n\leq n_0}v_n+C\times\sum_{n>n_0}\frac{1}{ng(n)}<\infty$. We achieve the proof by combining this result with Inequality (6.2) and the Borel-Cantelli Lemma.

Proof of the result (3.8)

Our aim is to show that

$$\sum_{n\geq 1}\mathbb{P}(\sup_{x\in\mathcal{M}}|\widehat{f}_n(x)-\mathbb{E}\widehat{f}_n(x)|>\varepsilon)<\infty,$$

with $\varepsilon=\eta\Psi_n$. Since $\mathcal{M}$ is compact, we can use the following inequality

$$\sup_{x\in\mathcal{M}}|\widehat{f}_n(x)-\mathbb{E}\widehat{f}_n(x)|\leq C\sqrt{\frac{\log n}{nh_n^d}}+\max_{1\leq k\leq\mu_n}|S_{3n}|.$$

This leads us to prove

$$\sum_{n\geq 1}\mathbb{P}(\max_{1\leq k\leq \mu_n}|S_{3n}|>\varepsilon)<\infty.$$

Recall that $g(n)=\log n\,(\log\log n)^{1+\varepsilon}$ for some $\varepsilon>0$. From (6.9), we deduce

$$\mu_n v_n n g(n)\leq C\left(nh_n^{\frac{d(\nu+(d+3)}{\nu-(d+3)}}(\log n)^{-\frac{\nu-1}{\nu-3}}g(n)^{\frac{-2}{\nu-(d+3)}}\right)^{-\frac{\nu-(d+3)}{2}}.$$

Since $\nu>d+3$ and $nh_n^{\frac{d(\nu+(d+3)}{\nu-(d+3)}}(\log n)^{-\frac{\nu-1}{\nu-3}}g(n)^{\frac{-2}{\nu-(d+3)}}\to\infty$ by assumption, we have $\mu_n v_n n g(n)\to 0$. Consequently, there exists an integer n_0 such that for $n>n_0$, $\mu_n v_n n g(n)\leq C$. Then, $\sum_{n>1}\mu_n v_n\leq\sum_{n\leq n_0}\mu_n v_n + C\times\sum_{n>n_0}\frac{1}{ng(n)}<\infty$. By applying the Borel-Cantelli Lemma, the desired result yields.

Proof of Theorem 3.7

For all $x\in\mathcal{M}$, we start by decomposing our expression into two terms, $K_{1,n}$ and $K_{2,n}$, as follows

$$\sqrt{nh_n^d}\left(\widehat{f}_n(x)-f(x)\right)=\sqrt{nh_n^d}\left(\widehat{f}_n(x)-\mathbb{E}\widehat{f}_n(x)\right)+\sqrt{nh_n^d}\left(\mathbb{E}\widehat{f}_n(x)-f(x)\right)=K_{1,n}+K_{2,n}.$$

Clearly, $K_{2,n}$ is negligible under the assumption **H3**, whereas $K_{1,n}$ is asymptotically normal.

To establish the asymptotic normality of $K_{1,n}$, dealing with strong mixing random variables under Assumption **H2**, we define

$$S_n(x):=\widehat{f}_n(x)-\mathbb{E}\widehat{f}_n(x)=\frac{1}{nh_n^d}\sum_{i=1}^{n}Z_i(x),$$

where

$$Z_i(x)=\frac{1}{\theta_x(X_i)}K\left(\frac{d(x,X_i)}{h_n}\right)-\mathbb{E}\left(\frac{1}{\theta_x(X_i)}K\left(\frac{d(x,X_i)}{h_n}\right)\right).$$

Since $\mathcal{M}$ is compact, we have $|Z_i(x)|\leq C$ by assumption **H1**. Then, under Assumption **H4**, we can deduce that

$$\mathbb{E}|Z_i(x)|^{\gamma}<\infty,$$

for some $\gamma>2$. Additionally, Assumption **H2** implies the existence of a constant C such that

$$\alpha(k)<Ck^{-\nu},$$

where $\nu > 2$ and $\nu > \frac{\gamma}{\gamma-2}$. We also establish that $\sigma^2(x) = f(x) \int_{B(1)} K^2(\|v\|)dv > 0$ (derived from Proposition 3.1). Utilizing Theorem 1.7 on page 36 of Bosq (1998), we can conclude the following result

$$\sqrt{nh_n^d} S_n(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(x)),$$

which achieved the proof.