Nonparametric Regression on Riemannian manifold under α-Mixing process

Wiem Nefzi, Anne-Françoise Yao, Salah KHARDANI

Nonparametric Regression on Riemannian manifold under α-Mixing process

Nefzi Wiem[1] **Khardani Salah**[1] and **Yao Anne Françoise**[2]

[1] Faculté des Sciences Tunis, Université El-Manar, Laboratoire Modélisation Mathématiques, Fiabilité Statistique et Analyse Stochastique, Tunisie.

[2] Laboratoire de Mathématiques Blaise Pascal, CNRS UMR 6620, Campus des Cézeaux, 63171 Aubière Cedex, France

Abstract

The main focus of our paper is to investigate the behavior of the kernel estimator for the regression function between a real-valued random variable Y and a random variable X, where X takes values in a Riemannian submanifold. The estimator is adapted from the article of Pelletier (2006). Additionally, we study data that adheres to the α-mixing condition, which imposes valuable constraints on the dependence structure of the observations. Specifically, we provide the rate of convergence in mean square error, enabling us to assess the precision and efficiency of the estimator.

Key words: Non-parametric regression estimator, Kernel method, Riemannian manifolds, Mixing condition

1. Introduction

In the field of inferential statistics, the estimation of the regression function has been found to be a valuable tool for fitting data that lies on Riemannian manifolds. Such data find applications across diverse scientific domains, including medical imaging, computer vision, as exemplified by Pennec (2006), and references therein, as well as in medicine, as discussed by Dryden and Mardia (1998). It also has applications in geography, astronomy, biology, and more, where the goal is often to characterize objects and predict their future behavior. These objects can encompass various forms of data, such as directional data (e.g., spheres, as seen in Mardia and Jupp (2000) and examples like animal migration directions, as shown in Jupp and Mardia (1980)), geometric transformations, tensors, and shapes, among others. The analysis of data residing on manifolds requires specialized techniques and methodologies due to their specific properties and geometric characteristics.

In the case of parametric regression estimation, noteworthy studies include those by Downs (2003) who regressed points on the surface of one sphere against points on another, Huang et al. (2010) in Nonlinear Regression Analysis, Hinkle et al. (2012) exploring polynomial regression estimation in Riemannian manifolds, and Hinkle et al. (2014) who developed a framework for polynomial regression on Riemannian manifolds. As always, these regression estimation methods require prior knowledge of the function to be estimated and stands as the most widely adopted regression model, thanks to its simplicity and adaptability to a wide array of phenomena. However, these techniques are limited when dealing with scenarios where the underlying function remains unknown or proves challenging to discern.

As an alternative, nonparametric estimation has gained significant attention in the field of statistics and data analysis in recent years. Specifically, non-parametric methods based on kernel techniques have proven to be effective in addressing this challenge. In this context, Pelletier (2006) has focused on nonparametric estimation

of the regression function on a closed Riemannian manifold using the kernel method when the data set consists of independent observations, drawing inspiration from its original application in density estimation Pelletier (2005). In addition, Nilsson et al. (2007) explores regression techniques on manifolds using kernel dimension reduction. Furthermore, Henry and Rodriguez (2009*a*), Henry and Rodriguez (2009*b*) have made significant contributions to both kernel regression estimation and robust regression on such manifolds. All these references concern the situations that involve independently and identically distributed (*i.i.d.*) observations.

To contribute to this evolving literature, this work focuses on a previously unaddressed scenario: when the data set consists of dependent observations from a stationary Riemannian-valued stochastic process, satisfying a strong mixing condition, with the target variable Y being real-valued. For readers interested in practical applications of $\alpha-$mixing, we recommend the book "Introduction to Strong Mixing Conditions" by Bradley (2007).

Our contributions include explicit expressions for both bias and variance of the regression estimator. Furthermore, we determine the optimal bandwidth, crucial for minimizing the mean squared error (MSE) of our estimator, ensuring accuracy and reliability in predictions.

The rest of this paper is organized as follows: In Section 2, we present the basic assumptions of our framework and the necessary assumptions for the results' proofs. Section 3 presents the asymptotic results of the kernel estimator. Finally, Section 4 provides the detailed proofs.

2. Preliminaries and estimators

2.1 Theoretical framework

We are concerned with a measurable stationary random process $(Z_t = (X_t, Y_t), t \in \mathbb{Z})$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where the Z_t's have the same distribution as (X, Y) admitting an unknown join density $f_{X,Y}$. In this process, X_t takes values on a Riemannian submanifold $\mathcal{M}$ of $\mathbb{R}^d$, and the X_t's are dependent and identically distributed as a random variable X with an unknown density function f. The variable Y_t is a real-valued and integrable variable. The relationship between Y and X is given by the following equation

$$Y = r(X) + \varepsilon$$

where, ε is the random error with $\mathbb{E}(\varepsilon) = 0$, and $r(.)$ denotes the regression function defined as

$$r(x) = \mathbb{E}(Y \mid X = x) = \begin{cases} \frac{\varphi(x)}{f(x)} & \text{if } f(x) > 0 \\ 0 & \text{otherwise} \end{cases}$$

with $\varphi : \mathcal{M} \to \mathbb{R}$ be the function that is defined as $\varphi(x) = \int_{\mathbb{R}} y f_{X,Y}(x, y) dy$.

In the following, we assume that (X_t) satisfies an $\alpha-$mixing condition given by, for any integer $n \geq 1$,

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

where $\mathcal{F}_i^k(X)$ is the σ-field generated by $X_i,\ i \leq j \leq k$.

2.2 Geometrical setting

We consider a Riemannian submanifold $(\mathcal{M}, g)$ with a tangent space $T_x\mathcal{M}$ at any point $x \in \mathcal{M}$. The submanifold $(\mathcal{M}, g)$ is endowed with a measure μ_g and satisfies the properties of being geodesically complete, compact, and without a boundary. According to the Hopf Rinow Theorem, the metric space $(\mathcal{M}, d_g)$ is complete , where d_g is the metric induce by g. To analyze distances on $\mathcal{M}$ effectively, we define the exponential map $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ at each point x. For more information, the readers are invited to refer to Khardani and Yao (2022). To ease the reading, we will use d to refer to d_g.

Next, we introduce the notation 0_x to represent the null vector in the tangent space $T_x\mathcal{M}$. The inner product in $T_x\mathcal{M}$ is defined as $\langle u, v\rangle = g(u, v)$ for any $u, v \in T_x\mathcal{M}$, and the associated norm is denoted as $\|.\|$. Additionally,

we use the notation $B(x,h)$ to represent the ball of radius h centered at x, and $B(0_x,h)$ represents the ball of radius h centered at 0_x.

We assume that the injectivity radius of $\mathcal{M}$ is such that $inj(\mathcal{M}) > 0$, and we only consider regular balls in $\mathcal{M}$. A regular (or convex) ball, denoted as $B(x,h)$, satisfies $h < h^*$, where $h^* = \min\{inj(\mathcal{M}), \frac{\pi}{2\sqrt{\kappa}}\}$. Here, κ represents the supremum of sectional curvatures of $\mathcal{M}$ (refer to Gallot et al. (2004) for the definition). If this upper bound is positive, we set $\kappa = 0$ otherwise. It is worth noting that $B(x,h) = \exp_x(B(0_x,h))$. To ease the reading, we will use $B(h)$ to refer to $B(0_x,h)$.

Based on the properties mentioned above, we can deduce that for any continuous function $\psi : \mathcal{M} \to \mathbb{R}$ with support on $B(x,h)$, the following equation holds

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

where $|g_x(v)|^{1/2}$ represents the determinant of $g_x(v)$, with $g_x(v)$ is more detailed in Chavel (2006), p.18 and Gallot et al. (2004), p.165). Alternatively, we can denote $\left|g_x(\exp^{-1}(y))\right|^{1/2} = \theta_x(\exp^{-1}(y))$, where $\theta_x(.)$ is called the volume density function. In the literature, $\theta_x(y)$ is also denoted as $\theta_x(\exp_x^{-1}(y))$ by abuse of notation (see also le Brigant and Puechmorel (2019)).

The density of f_{T_x} with respect to the Lebesgue measure in $T_x(\mathcal{M})$ is given by :

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

where 1_{V_x} is the indicator function of V_p. Note that $f_{T_x}(0_x) = f(x)\, |g_x(0)|^{1/2} = f(x)$ since $|g_x(0)|^{1/2} = 1$.

The study of the bias terms of the estimators below, requires some regularity condition on f, φ (assumption **H5** below) and $|g_x(\cdot)|^{1/2}$, for which we have the following expansion which can be found for example in Karcher (1977), p. 191 (or Chavel (1993), p.91): for any $u \in T_x(\mathcal{M})$

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

where Ric_x is the Ricci tensor at x.

In the following, we will denote by grad and Hess the gradient and the Hessian operators respectively.

2.3 The kernel estimator of interest

In this subsection, we aim to study the behavior of kernel estimator of Pelletier (2006). Let (X_i, Y_i) be n observations drawn from the process $(Z_t = (X_t, Y_t), t \in \mathbb{Z})$ which is defined above. Therefore, the kernel regression estimator is defined as follows, for each $x \in \mathcal{M}$

$$r_n(x) = \begin{cases} \frac{\varphi_n(x)}{f_n(x)} & \text{if } f_n(x) > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $f_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_n^d} \frac{1}{\theta_x(X_i)} K\left(\frac{d(x,X_i)}{h_n}\right)$ and $\varphi_n(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i \frac{1}{h_n^d} \frac{1}{\theta_x(X_i)} K\left(\frac{d(x,X_i)}{h_n}\right)$ are respectively the kernel estimators of f and φ. In addition, K and h_n are respectively the kernel function and the bandwidth that satisfy the following assumptions.

2.4 Assumptions

The results presented below are based on the following assumptions:

H1: $K : \mathbb{R}^d \longrightarrow \mathbb{R}_+$ is a bounded and continuous map such that :

1. $supp\ K = [0;1]$,

2. $\int K(\|x\|)dx = 1$,
3. $\int xK(\|x\|)dx = 0$ or at least the vector, $\int xK(\|x\|)\,dx$ is orthogonal to span$\{\text{grad}f(x)\}$
4. $\int \|x\|^2 K(\|x\|)dx < \infty$.

H2: The mixing coefficient $\alpha(n)$ tends to zero at a polynomial rate, specifically $\alpha(n) = O(n^{-v})$ for some $v > 2$.

H3: The bandwidth $h_n > 0$ and satisfies the following assumptions

$nh_n^d \longrightarrow \infty$ and $h_n < \frac{h^*}{2}$, $n \to \infty$.

H4:

1. The function $\sigma_2(\cdot) = \mathbb{E}\left(Y^2 \middle| X = \cdot\right)$ is both continuous and bounded away from zero at x,
2. Y is bounded,
3. r is bounded and two-times continuously differentiable at x.

H5: The functions f and φ are bounded and twice continuously differentiable at any $x \in \mathcal{M}$. Additionally, $\|\text{Hess}f(x)\|_{HS} < \infty$ as well as $\|\text{Hess}\varphi(x)\|_{HS} < \infty$.

H6: $\forall 1, j$, the joint density $f_{1,j}$ of (X_1, X_j) exists is such that

$$\sup_j \sup_{u,v\in\mathcal{M}\times\mathcal{M}} |f_{1,j}(u,v) - f(u)f(v)| < M. \quad \text{for all } j \geq 2$$

for some $M > 0$.

2.5 comments

- Assumptions **H1**, **H4** and **H5** are considered classical to get consistency results in the context of kernel estimation. In particular, assumption **H5** allows the Taylor expansion: Let s satisfies **H5** then

$$s(\exp_x(u)) = s(x) + < \text{grad}s(x), u > + \frac{1}{2}\text{Hess}s(x)(u,u) + o(\|u\|^2)$$

  for all $u \in T_x\mathcal{M}$. Thus, for all $h > 0$ and $\|v\| \leq 1$, we have

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

- The assumptions **H2** and **H6** are also classical in scenarios involving dependent data.
- Concerning **H3**, it is an adaptation of classical assumptions on the bandwidth in Riemannian manifolds setting. Namely, we need the constraint $h_n < \frac{h^*}{2}$ to ensure that x is locally a central point when using the kernel estimator (see for example Karcher (1977) and Le (2001), or Pelletier (2005) for more details).

3. Auxiliary results

Lemma 3.1. *Under the assumptions **H1** and **H5**, we have*

$$\mathbb{E}f_n(x) - f(x) = \frac{h_n^2}{2}\left[\int_{B(1)} K(\|v\|)Hessf(x)(v,v)dv + o(1)\right] \tag{2}$$

and

$$\mathbb{E}\varphi_n(x) - \varphi(x) = \frac{h_n^2}{2}\left[\int_{B(1)} K(\|v\|)Hess\varphi(x)(v,v)dv + o(1)\right] \tag{3}$$

Lemma 3.2. *Under the assumptions **H1, H2, H3, H5** and **H6**, the variance of $f_n(x)$ is given by*

$$V\left(f_n(x)\right) = \frac{1}{nh_n^d}\left[f(x)\int_{B(1)} K^2(\|v\|)dv + o(1)\right] \tag{4}$$

and then

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

*In addition, under the assumptions **H1** to **H6**, we obtain the variance of $\varphi_n(x)$*

$$V(\varphi_n(x)) = \frac{1}{nh_n^d}\left[f.\sigma_2(x)\int_{B(1)} K^2(\|v\|)dv + o(1)\right] \tag{5}$$

and then

$$nh_n^d V(\varphi_n(x)) \underset{n\to\infty}{\longrightarrow} f.\sigma_2(x)\int_{B(1)} K^2(\|v\|)dv.$$

where $\sigma_2(x) = \mathbb{E}\left(Y^2\middle|X = x\right)$.

4. Asymptotic results

In this section, we provide explicit expressions of the bias and variance associated with r_n. Additionally, we present the optimal bandwidth that minimizes the mean squared error of the regression estimator.

Proposition 4.1. *Under the assumptions of Lemma 3.1, we can derive the following expression for the bias of $r_n(x)$*

$$\mathbb{E}r_n(x) - r(x) = \frac{h_n^2}{2f(x)}\left(\int_{B(1)} K(\|v\|)(Hess\varphi(x)(v,v) - r(x)Hessf(x)(v,v))dv\right) + o\left(h_n^2\right). \tag{6}$$

Furthermore, considering Lemma 3.1 and Lemma 3.2, we can establish that, for all $f(x) > 0$, the variance of $r_n(x)$

$$V(r_n(x)) = \frac{1}{nh_n^d}\left(\frac{\sigma_2(x) - r^2(x)}{f(x)}\int_{B(1)} K^2(\|v\|)dv + o\left(1\right)\right). \tag{7}$$

As Corollary of this proposition we obtain an optimal rate in MSE meaning for $r_n(x)$.

4.1 Rate of convergence in Mean Squared Error

Corollary 4.1. *(Optimal bandwidth) To minimize the mean squared error (MSE) of r_n, under the assumption **H1**, the optimal bandwidth h_n must be chosen as $h_{n,opt} = C_0\, n^{\frac{-1}{d+4}}$, where C is a constant. The MSE can then be expressed as*

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

where $C = C_0^4 C_1 + C_0^{-d} C_2$ with $C_0 = \left(\frac{dC_2}{4C_1}\right)^{\frac{1}{4+d}}$. Here, constants C_1 and C_2 are respectively given by

$C_1 = \frac{1}{4(f(x))^2}\left(\int_{B(1)} K(\|v\|)(Hess\varphi(x)(v,v) - r(x)Hessf(x)(v,v))dv\right)$ *and* $C_2 = \frac{\sigma_2(x)-(r(x))^2}{f(x)}\int_{B(1)} K^2(\|v\|)dv$.

Theorem 4.1. *Under the assumptions **H1** to **H6**, since $\mathcal{M}$ is compact, we have*

$$\sup_{x\in\mathcal{M}} \mathbb{E}\left(r_n(x) - r(x)\right)^2 \leq C \times \left(h_n^4 + \frac{1}{nh_n^d}\right)$$

4.2 Convergence in probability

Theorem 4.2. *Under the assumptions **H1**, to **H6**, for a given $x \in \mathcal{M}$*

$$|r_n(x) - r(x)| \longrightarrow 0 \ \textbf{\textit{in probability}}, \tag{8}$$

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

$$\sup_{x\in\mathcal{M}}|r_n(x) - r(x)| \longrightarrow 0 \ \textbf{\textit{in probability}}. \tag{9}$$

Remark 4.1. *Note that Theorem 4.2 is a consequence of Corollary 4.1 and Theorem 4.1 through Markov's inequality.*

5. Proofs

5.1 Proof of Lemma 3.1

Proof of (2)

$$\begin{aligned}
\mathbb{E}f_n(x) - f(x) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n^d}\mathbb{E}\left(\frac{1}{\theta_x(X_i)}K\left(\frac{d(x,X_i)}{h_n}\right)\right) - f(x) \\
&= \frac{1}{h_n^d}\int_{\mathcal{M}}\frac{1}{\theta_x(y)}K\left(\frac{d(x,y)}{h_n}\right)f(y)d\mu_g(y) - f(x) \\
&= \frac{1}{h_n^d}\int_{\mathcal{M}}\frac{1}{\theta_x(y)}K\left(\frac{d(x,y)}{h_n}\right)[f(y) - f(x)]\,d\mu_g(y)
\end{aligned}$$

Taking the integral over $B(x, h_n)$, we obtain

$$\begin{aligned}
\mathbb{E}f_n(x) - f(x) &= \frac{1}{h_n^d}\int_{B(x,h_n)}\frac{1}{\theta_x(y)}K\left(\frac{d(x,y)}{h_n}\right)(f(y) - f(x))\,d\mu_g(x) \\
&= \frac{1}{h_n^d}\int_{B(h_n)}K\left(\frac{\|v\|}{h_n}\right)(f(\exp_x(v)) - f(x))\,dv \\
&= \int_{B(1)}K(\|v\|)(f(\exp_x(h_n v)) - f(x))\,dv
\end{aligned}$$

Applying equation (1), under assumptions **H1** and **H5**, we obtain

$$\mathbb{E}f_n(x) - f(x) = \frac{h_n^2}{2}\int_{B(1)}K(\|v\|)\mathrm{Hess}f(x)(v,v)dv + o(h_n^2)$$

Proof of (3): For all $x \in \mathcal{M}$, we have

$$\begin{aligned}
\mathbb{E}\varphi_n(x) - \varphi(x) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\frac{1}{h_n^d}\frac{1}{\theta_x(X_i)}K\left(\frac{d(x,X_i)}{h_n}\right)\right) - \varphi(x) \\
&= \frac{1}{h_n^d}\mathbb{E}\left(Y_1\frac{1}{\theta_x(X_1)}K\left(\frac{d(x,X_1)}{h_n}\right)\right) - \varphi(x) \\
&= \frac{1}{h_n^d}\mathbb{E}\left(\frac{1}{\theta_x(X_1)}K\left(\frac{d(x,X_1)}{h_n}\right)\mathbb{E}(Y_1|X_1 = y)\right) - \varphi(x) \\
&= \frac{1}{h_n^d}\int_{\mathcal{M}}r(y)\frac{1}{\theta_x(y)}K\left(\frac{d(x,y)}{h_n}\right)f(y)d\mu_g(y) - \varphi(x) \\
&= \frac{1}{h_n^d}\int_{\mathcal{M}}\frac{1}{\theta_x(y)}K\left(\frac{d(x,y)}{h_n}\right)(\varphi(y) - \varphi(x))d\mu_g(y).
\end{aligned}$$

Taking the integral over $B(x, h_n)$, we obtain

$$\begin{aligned}\mathbb{E}\varphi_n(x) - \varphi(x) &= \frac{1}{h_n^d}\int_{B(x,h_n)} \frac{1}{\theta_x(y)} K\left(\frac{d(x,y)}{h_n}\right)(\varphi(y) - \varphi(x))\, d\mu_g(x) \\ &= \frac{1}{h_n^d}\int_{B(h_n)} K\left(\frac{\|v\|}{h_n}\right)(\varphi(\exp_x(v)) - \varphi(x))\, dv \\ &= \int_{B(1)} K(\|v\|)\,(\varphi(\exp_x(h_n v)) - \varphi(x))\, dv\end{aligned}$$

Using (1) and under assumptions **H1** and **H5**, we get

$$\mathbb{E}\varphi_n(x) - \varphi(x) = \frac{h_n^2}{2}\int_{B(1)} K(\|v\|)\mathrm{Hess}\varphi(x)(v,v)dv + o(h_n^2)$$

5.2 Proof of Lemma 3.2

Proof of (4) For all $x \in \mathcal{M}$, we have

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

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 $\mathbb{E}(Z_i(x)) = 0,\ \forall i = 1, ..., n$. Therefore,

$$V(f_n(x)) = \frac{1}{n^2h_n^{2d}}\sum_{i=1}^n V(Z_i) + \frac{2}{n^2h_n^{2d}}\sum_{j\neq i} Cov(Z_i, Z_j) = A + B.$$

Expression of A:

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

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

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

Thus,

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

Hence,

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

Regarding the second term, we have

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

Using (1), under assumptions **H1** and **H5**, we get

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

The expression of B is as follows

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

where $E_1 = \{(i,j) \mid 1 \leq |j-i| \leq \beta_n\}$ and $E_2 = \{(i,j) \mid \beta_n + 1 \leq |j-i| \leq n-1\}$, with $\beta_n = o(n)$.

To compute $J_{1,n}$ and $J_{2,n}$, we need the following control

$$\begin{aligned} \left|Cov(Z_i(x), Z_j(x))\right| &= \left|\mathbb{E}(Z_i(x) Z_j(x))\right| \\ &= \left| \mathbb{E}\left(\frac{1}{\theta_x(X_i)\theta_x(X_j)} K\left(\frac{d(x,X_i)}{h_n}\right) K\left(\frac{d(x,X_j)}{h_n}\right)\right) - \mathbb{E}\left(\frac{1}{\theta_x(X_i)} K\left(\frac{d(x,X_i)}{h_n}\right)\right)^2 \right| \\ &= \left| \int \int \frac{1}{\theta_x(y)\theta_x(z)} K\left(\frac{d(x,y)}{h_n}\right) K\left(\frac{d(x,z)}{h_n}\right) f(y,z)\, d\mu_g(y) d\mu_g(z) \right) \\ &- \left[\int \frac{1}{\theta_x(y)} K\left(\frac{d(x,y)}{h_n}\right) f(y) d\mu_g(y)) \right]^2 \Bigg| \\ &= \left| \int \int \frac{1}{\theta_x(y)\theta_x(z)} K\left(\frac{d(x,y)}{h_n}\right) K\left(\frac{d(x,z)}{h_n}\right) [f(y,z) - f(y)f(z)]\, d\mu_g(y) d\mu_g(z) \right|. \end{aligned}$$

Finally,

$\left|Cov(Z_i(x), Z_j(x))\right| = h_n^{2d} \left| \int_{B(1)} \int_{B(1)} K(\|u\|) K(\|v\|) \left(f\left(\exp_x(h_n u), \exp_x(h_n v)\right) - f\left(\exp_x(h_n u)\right) f\left(\exp_x(h_n v)\right)\right) dv du \right|$.

Thus, under assumptions **H1** and **H6**, we obtain:

$$|Cov(Z_i(x), Z_j(x))| \leq M\, h_n^{2d} = O\left(h_n^{2d}\right).$$

So that

$$J_{1,n} \leq \frac{2M}{n}\beta_n = O\left(n^{-1}\beta_n\right).$$

For E_2, we can use the modified Davydov inequality for mixing processes (see Rio (1999), p.10, formula 1.12a). This gives, for all $i \neq j$

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

Hence,

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

By choosing $\beta_n = h_n^{\frac{-2d}{\nu}}$ under assumption **H2**, we obtain

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

Finally, we obtain

$$V\left(f_n(x)\right) = \frac{1}{nh_n^d}\left(f(x)\int_{B(1)} K^2\left(\|v\|\right)dv + o(1)\right)$$

Proof of (5) For all $x \in \mathcal{M}$, we have

$$\varphi_n(x) = \varphi_n(x) - \mathbb{E}\left(\varphi_n(x)\right) + \mathbb{E}(\varphi_n(x)).$$

Therefore, we can express the variance of $\varphi_n(x)$ as follows

$$V(\varphi_n(x)) = V\left(\varphi_n(x) - \mathbb{E}\varphi_n(x)\right) = V\left(\frac{1}{nh_n^d}\sum_{i=1}^{n} L_i(x)\right)$$

where $L_i(x) = \frac{Y_i}{\theta_x(X_i)}K\left(\frac{d(x,X_i)}{h_n}\right) - \mathbb{E}\left(\frac{Y_i}{\theta_x(X_i)}K\left(\frac{d(x,X_i)}{h_n}\right)\right)$, for all $i = 1, ..., n$. It is clear that the L_i's are dependent and identically distributed with expectation $\mathbb{E}(L_i(x)) = 0$. Then,

$$V(\varphi_n(x)) = \frac{1}{n^2h_n^{2d}}\sum_{i=1}^{n} V\left(L_i(x)\right) + \frac{2}{n^2h_n^{2d}}\sum_{j\neq i} Cov\left(L_i(x), L_j(x)\right) = D + E$$

Starting with the expression of D, we have

$$\begin{aligned} D &= \frac{1}{nh_n^{2d}}\mathbb{E}\left(L_1^2(x)\right) \\ &= \frac{1}{nh_n^{2d}}\mathbb{E}\left(\frac{Y_1^2}{\theta_x^2(X_1)}K^2\left(\frac{d(x,X_1)}{h_n}\right)\right) - \frac{1}{nh_n^{2d}}\mathbb{E}^2\left(\frac{Y_1}{\theta_x(X_1)}K\left(\frac{d(x,X_1)}{h_n}\right)\right) \\ &= \mathcal{I}_{1,n} - \mathcal{I}_{2,n} \end{aligned}$$

where

$$\begin{aligned} \mathcal{I}_{1,n} &= \frac{1}{nh_n^{2d}}\mathbb{E}\left(\frac{Y_1^2}{\theta_x^2(X_1)}K^2\left(\frac{d(x,X_1)}{h_n}\right)\right) \\ &= \frac{1}{nh_n^{2d}}\mathbb{E}\left(\frac{1}{\theta_x^2(X_1)}K^2\left(\frac{d(x,X_1)}{h_n}\right)\mathbb{E}\left[Y_1^2|X_1 = y\right]\right) \\ &= \frac{1}{nh_n^{2d}}\int_{\mathcal{M}} \frac{\sigma_2(y)}{\theta_x^2(y)}K^2\left(\frac{d(x,y)}{h_n}\right)f(y)d\mu_g(y) \\ &= \frac{1}{nh_n^{2d}}\int_{B_{\mathcal{M}}(x,h_n)} \frac{f.\sigma_2(y)}{\theta_x^2(y)}K^2\left(\frac{d(x,y)}{h_n}\right)d\mu_g(y) \end{aligned}$$

Using the fact that for any $v \in T_x$, $\theta_x(\exp_x(v)) = |g_x(v)|^{\frac{1}{2}}$, we can rewrite $\mathcal{I}_{1,n}$ as

$$\begin{aligned}\mathcal{I}_{1,n} &= \frac{1}{nh_n^{2d}} \int_{B(h_n)} \frac{f.\sigma_2(\exp_x(v))}{|g_x(v)|^{\frac{1}{2}}} K^2\left(\frac{\|v\|}{h_n}\right) dv \\ &= \frac{1}{nh_n^{d}} \int_{B(1)} \frac{f.\sigma_2(\exp_x(h_n v))}{|g_x(v)|^{\frac{1}{2}}} K^2\left(\|v\|\right) dv.\end{aligned}$$

Note that $|g_x(v)|^{\frac{1}{2}} = 1 + O(\|v\|^2)$ by assumption, hence

$$\int_{B(1)} \frac{f.\sigma_2}{\theta_x}(\exp_x(h_n v)) K^2\left(\|v\|\right) dv \to \int_{B(1)} f.\sigma_2(x) K^2\left(\|v\|\right) dv,$$

which gives us

$$nh_n^d \mathcal{I}_{1,n} \longrightarrow f.\sigma_2(x) \int_{B(1)} K^2\left(\|v\|\right) dv.$$

Regarding the second term, we have

$$\begin{aligned}\mathcal{I}_{2,n} &= \frac{1}{nh_n^{2d}} \mathbb{E}^2\left(\frac{Y_1}{\theta_x(X_1)} K\left(\frac{d(x,X_1)}{h_n}\right)\right) \\ &= \frac{1}{nh_n^{2d}} \left(\int_{B(x,h_n)} \frac{1}{\theta_x(y)} K\left(\frac{d(x,y)}{h_n}\right) \varphi(y) d\mu_g(y)\right)^2 \\ &= \frac{1}{n} \left(\int_{B(1)} K\left(\|v\|\right) \varphi(\exp_x(h_n v)) dv\right)^2.\end{aligned}$$

Using (1) and assuming **H1** and **H5** , we get

$$\mathcal{I}_{2,n} = \frac{\varphi^2(x)}{n} + o(1) = O\left(n^{-1}\right).$$

Next, we can decompose the expression of E as follows

$$E = \frac{2}{n^2 h_n^{2d}} \sum_{(i,j)\in E_1}^{n} Cov(L_i(x), L_j(x)) + \frac{2}{n^2 h_n^{2d}} \sum_{(i,j)\in E_2}^{n} Cov(L_i(x), L_j(x)) = \mathcal{J}_{1,n} + \mathcal{J}_{2,n}$$

where $E_1 = \{(i,j) \mid 1 \le |j-i| \le \beta_n\}$ and $E_2 = \{(i,j) \mid \beta_n + 1 \le |j-i| \le n-1\}$.

To compute $\mathcal{J}_{1,n}$ and $\mathcal{J}_{2,n}$, we need the following control

$$\begin{aligned}\left|Cov(L_i(x), L_j(x))\right| &= |\mathbb{E}(L_i(x)L_j(x))| \\ &= |\mathbb{E}\left(\frac{Y_i\, Y_j}{\theta_x(X_i)\theta_x(X_j)} K\left(\frac{d(x,X_i)}{h_n}\right) K\left(\frac{d(x,X_j)}{h_n}\right)\right) \\ &- \mathbb{E}\left(\frac{Y_i}{\theta_x(X_i)} K\left(\frac{d(x,X_i)}{h_n}\right)\right) \mathbb{E}\left(\frac{Y_j}{\theta_x(X_j)} K\left(\frac{d(x,X_j)}{h_n}\right)\right)| \\ &= \left| \int\int \frac{r(y)r(z)}{\theta_x(y)\theta_x(z)} K\left(\frac{d(x,y)}{h_n}\right) K\left(\frac{d(x,z)}{h_n}\right) (f(y,z) - f(y)f(z)) d\mu_g(y) d\mu_g(z) \right|\end{aligned}$$

Finally, $\left|Cov(L_i(x), L_j(x))\right| =$
$h_n^{2d} \left| \int_{B(1)} \int_{B(1)} r(\exp_x(h_n u)) r(\exp_x(h_n v)) K(\|u\|) K(\|v\|) \left(f\left(\exp_x(h_n u), \exp_x(h_n v)\right) - f\left(\exp_x(h_n u)\right) f\left(\exp_x(h_n v)\right)\right) dv du \right|$.

Thus, under the assumptions **H1(i)**, **H4** and **H6**

$$\left|Cov(L_i(x), L_j(x))\right| = O\left(h_n^{2d}\right).$$

For E_1, we have

$$\mathcal{J}_{1,n} = \frac{2}{n^2 h_n^{2d}} \sum_{i=1}^{n} \sum_{1 \le |j-i| \le \beta_n} Cov(L_i(x), L_j(x)) \le 2n^{-1}\beta_n = O\left(n^{-1}\beta_n\right).$$

For E_2, we use the modified Davydov's inequality for mixing processes (see Rio (1999), p.10, formula 1.12a). For all $i \neq j$, this leads to

$$|Cov(L_i(x), L_j(x))| \le c\alpha(|i-j|).$$

Therefore

$$\mathcal{J}_{2,n} \le \frac{2}{n^2 h_n^{2d}} \sum_{i=1} \sum_{\beta_n+1 \le j-i \le n-1} c\alpha(|i-j|) = \frac{2c}{h_n^{2d}} \int_{\beta_n+1}^{n-1} s^{-\nu} ds \le 2c\, h_n^{-2d} \beta_{n+1}^{1-\nu} \le \frac{2c}{1-\nu} h_n^{-2d} \beta_n^{1-\nu} = O\left(h_n^{-2d}\beta_n^{1-\nu}\right).$$

By choosing $\beta_n = h_n^{\frac{-2d}{\nu}}$ under the assumption **H2**, we have

$$\mathcal{J}_{1,n} + \mathcal{J}_{2,n} = o\left(\frac{1}{nh_n^{\frac{2d}{\nu}}}\right) = o\left(\frac{1}{nh_n^d}\right).$$

Finally we obtain

$$V(\varphi_n(x)) = \frac{1}{nh_n^d}\left[f.\sigma_2(x)\int_{B(1)} K^2(\|x\|)dx + o(1)\right].$$

5.3 Proof of the proposition 4.1

Proof of (6)

In order to compute the bias of the regression estimator, we use the following decomposition

$$r_n(x) - r(x) = \frac{\varphi_n(x)}{f_n(x)} - \frac{\varphi(x)}{f(x)} = B_n(x)\frac{f(x)}{f_n(x)}$$

where $B_n(x)$ is given as

$$B_n(x) = \frac{1}{f(x)}\left(\varphi_n(x) - \varphi(x)\right) - \frac{r(x)}{f(x)}\left(f_n(x) - f(x)\right). \tag{10}$$

Since f_n converges asymptotically to f in a normal distribution, to compute $\mathbb{E}r_n(x) - r(x)$, it is sufficient to compute

$$\mathbb{E}B_n(x) = \frac{1}{f(x)}\mathbb{E}\left(\varphi_n(x) - \varphi(x)\right) - \frac{r(x)}{f(x)}\mathbb{E}\left(f_n(x) - f(x)\right).$$

By using equations (2) and (3), we can straightforward obtain

$$\mathbb{E}r_n(x) - r(x) = \mathbb{E}B_n(x) = \frac{h_n^2}{2f(x)}\left(\int_{B(1)} K(\|v\|)(\text{Hess}\varphi(x)(v,v) - r(x)\text{Hess}f(x)(v,v))dv\right) + o\left(h_n^2\right).$$

Proof of (7)

Using the same decomposition (10), we obtain

$$V\left(r_n(x)\right) = V(B_n(x)) = \frac{1}{f^2(x)}V(\varphi_n(x)) + \frac{r^2(x)}{f^2(x)}V\left(f_n(x)\right) - \frac{2r(x)}{f^2(x)}Cov\left(\varphi_n(x), f_n(x)\right)$$

where we can compute $Cov\left(f_n(x), \varphi_n(x)\right)$ to obtain the desired result. We define

$$\begin{aligned} Cov\left(f_n(x), \varphi_n(x)\right) &= \frac{1}{n^2 h_n^{2d}} \sum_{i=1}^{n} Cov\left(\frac{1}{\theta_x(X_i)} K\left(\frac{d(x, X_i)}{h_n}\right), Y_i \frac{1}{\theta_x(X_i)} K\left(\frac{d(x, X_i)}{h_n}\right)\right) \\ &+ \frac{1}{n^2 h_n^{2d}} \sum_{i \neq j} Cov\left(\frac{1}{\theta_x(X_i)} K\left(\frac{d(x, X_i)}{h_n}\right), Y_j \frac{1}{\theta_x(X_j)} K\left(\frac{d(x, X_j)}{h_n}\right)\right) \\ &= \mathcal{K}_{1,n} + \mathcal{K}_{2,n} \end{aligned}$$

Starting with $\mathcal{K}_{1,n}$

$$\begin{aligned} \mathcal{K}_{1,n} &= \frac{1}{n h_n^{2d}} \int_{\mathcal{M}} \mathbb{E}(Y_1 | X_1 = y) \frac{1}{\theta_x^2(y)} K^2\left(\frac{d(x, y)}{h_n}\right) f(y) d\mu_g(y) \\ &= \frac{1}{n h_n^{2d}} \int_{B_{\mathcal{M}}(x, h_n)} r(y) \frac{1}{\theta_x^2(y)} K^2\left(\frac{d(x, y)}{h_n}\right) f(y) d\mu_g(y) \\ &= \frac{1}{n h_n^{d}} \int_{B(1)} \frac{\varphi}{\theta_x}(\exp_x(h_n v)) K^2(\|v\|) dv, \end{aligned}$$

By performing the same calculations as in the proof of $\mathcal{I}_{1,n}$, we obtain

$$n h_n^d \mathcal{K}_{1,n} \longrightarrow \varphi(x) \int_{B(1)} K^2(\|v\|) dv \quad \text{as } n \to \infty.$$

For $\mathcal{K}_{2,n}$, we follow the same steps as in the proof of the expression E. Thus, we obtain

$$\mathcal{K}_{2,n} = o\left(\frac{1}{n h_n^d}\right).$$

Therefore, we have

$$Cov\left(f_n(x), \varphi_n(x)\right) = \frac{1}{n h_n^d} \varphi(x) \int_{B(1)} K^2(\|x\|) dx + o\left(\frac{1}{n h_n^d}\right).$$

Consequently, we obtain

$$V(r_n(x)) = \frac{1}{n h_n^d} \left(\frac{\sigma_2(x) - r^2(x)}{f(x)} \int_{B(1)} K^2(\|x\|) dx + o(1)\right).$$

where $o(1)$ is independent to x.

5.4 Proof of corollary 4.1

To establish this result, we use the conventional method of decomposing the mean squared error,

$$MSE(x) = \mathbb{E}\left(r_n(x) - r(x)\right)^2 = b^2(x) + V\left(r_n(x)\right) \tag{11}$$

where $b(x) := \mathbb{E}(r_n(x)) - r(x)$ is the bias of the kernel estimator.

Base on the results derived from (6) and (7), we get

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

where $C_1 = \frac{1}{4(f(x))^2} \left(\int_{B(1)} K(\|v\|) (\text{Hess}\varphi(x)(v, v) - r(x) \text{Hess} f(x)(v, v)) dv\right)$ and $C_2 = \frac{\sigma_2(x) - (r(x))^2}{f(x)} \int_{B(1)} K^2(\|v\|) dv$.

Then, we can achieve the goal of minimizing the MSE by focusing on minimizing the term $C_1 h_n^4 + C_2 \frac{1}{n h_n^d}$. Hence, we get

$$h_{n,opt} = C_0 \, n^{\frac{-1}{d+4}}$$

where $C_0 = \left(\frac{d C_2}{4 C_1}\right)^{\frac{1}{4+d}}$.

5.5 Proof of Theorem 4.1

The MSE decomposition in (11) leads to

$$\sup_{x\in\mathcal{M}}\mathbb{E}\left(r_n(x)-r(x)\right)^2 \leq \sup_{x\in\mathcal{M}} b^2(x) + \sup_{x\in\mathcal{M}} V\left(r_n(x)\right)$$

Now, by Assumption **H4** and for the same reasons provided by Pelletier (2005)(in p.302), we have, for all $v \in B(1)$, $|f(\exp_x(h_n v)) - f(x) + h_n \langle \mathrm{grad} f(x), v\rangle| \leq Ch_n^2\|v\|^2$. Therefore, by **H1**, we have $|\mathbb{E}f_n(x) - f(x)| \leq Ch_n^2$. Using the same reasons and by adding the fact that Y is bounded in **H4**, we get $|\mathbb{E}\varphi_n(x) - \varphi(x)| \leq Ch_n^2$, where $C > 0$ is independent of x in both cases. Consequently, due to the compactness of $\mathcal{M}$, the following bound holds

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

for some $C > 0$ and $f(x) > 0$.

Now, using the fact that $o(1)$ is independent of x and f, σ_2, r and K are bounded $f(x) > 0$ in (7), we get straightforward

$$\sup_{x\in\mathcal{M}}|V(r_n(x))| \leq C\frac{1}{nh_n^d}$$

which achieved the proof.

References

Bradley, R. C. (2007), *Introduction to strong mixing conditions*, Kendrick press.

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

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

Downs, T. (2003), 'Spherical regression', *Biometrika* **90**(3), 655–668.

Dryden, I. and Mardia, K. (1998), 'Statistical shape analysis.,(wiley: New york, ny.)'.

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

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

Henry, G. and Rodriguez, D. (2009*b*), 'Robust nonparametric regression on riemannian manifolds', *Journal of Nonparametric Statistics* **21**(5), 611–628.

Hinkle, J., Fletcher, P. T. and Joshi, S. (2014), 'Intrinsic polynomials for regression on riemannian manifolds', *Journal of Mathematical Imaging and Vision* **50**, 32–52.

Hinkle, J., Muralidharan, P., Fletcher, P. T. and Joshi, S. (2012), Polynomial regression on riemannian manifolds, *in* 'Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III 12', Springer, pp. 1–14.

Huang, H.-H., Hsiao, C. K., Huang, S.-Y., Peterson, P., Baker, E. and McGaw, B. (2010), 'Nonlinear regression analysis', *International encyclopedia of education* **2010**, 339–346.

Jupp, P. E. and Mardia, K. V. (1980), 'A general correlation coefficient for directional data and related regression problems', *Biometrika* **67**(1), 163–173.

Karcher, H. (1977), 'Riemannian center of mass and mollifier smoothing', *Communications on pure and applied mathematics* **30**(5), 509–541.

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

le Brigant, A. and Puechmorel, S. (2019), 'Approximation of densities on riemannian manifolds', *Entropy* **21**(1).
URL: *https://www.mdpi.com/1099-4300/21/1/43*

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

Mardia, K. and Jupp, P. (2000), 'Directional data'.

Nilsson, J., Sha, F. and Jordan, M. I. (2007), Regression on manifolds using kernel dimension reduction, *in* 'Proceedings of the 24th international conference on Machine learning', pp. 697–704.

Pelletier, B. (2005), 'Kernel density estimation on riemannian manifolds', *Statistics & probability letters* **73**(3), 297–304.

Pelletier, B. (2006), 'Non-parametric regression estimation on closed riemannian manifolds', *Journal of Nonparametric Statistics* **18**(1), 57–67.

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

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