Robust Nonparametric Regression on Riemannian Manifolds *

Guillermo Henry; Daniela Rodriguez
FCEyN, Universidad de Buenos Aires and CONICET, Argentina.

Abstract

In this paper, we introduce two families of robust kernel–based regression estimators when the regressors are random objects taking values in a Riemannian manifold. The first proposal is a local M-estimator based on kernel methods, adapted to the geometry of the manifold. For the second proposal the weights are based on k-nearest neighbor kernel methods. Strong uniform consistency results as well as the asymptotically normality of both families are established. Finally, a Monte Carlo study is carried out to compare the performance of the robust proposed estimators with the classical ones, in normal and contaminated samples and a cross-validation method is discussed.

AMS Subject Classification 1990: Primary 62F35, Secondary 62H25.
Key words and phrases: Nonparametric Regression, Robust Estimation, Kernel Weights, k-nearest neighbor weights, Riemannian Manifolds.

1 Introduction

Nonparametric inference has gained a lot of attention, in recent years, in order to explore the nature of complex nonlinear phenomena. The idea of nonparametric inference is to leave the data to show the structure lying beyond them, instead of imposing one. Nadaraya (1964) and Watson (1964), introduced kernel–based estimators for the regression function $r(\mathbf{x}) = E(y|\mathbf{x})$, when dealing with independent observations $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$ such that $y_i \in I\!R$, $\mathbf{x}_i \in I\!R^d$. Nearest neighbor with kernel methods for the regression function were introduced by Collomb (1980).

Both of them are a weighted average of the response variables and thus, they are highly sensitive to large fluctuations of the responses. As mentioned by several authors, the treatment of outliers is an important step in highlighting features of a data set since extreme points can affect the scale and the shape of any estimate of the regression function based on

*Corresponding author: Daniela Rodriguez. Email: drodri@dm.uba.ar. This research was partially supported by Grants X-094 from the Universidad de Buenos Aires, PID 5505 from CONICET and PAV 120 and PICT 21407 from ANPCYT, Argentina.

local averaging, leading to possible wrong conclusions. This has motivated the interest in combining the ideas of robustness with those of smoothed regression, to develop procedures which will be resistant to deviations from the central model in nonparametric regression models. As it is well known, robust estimators can be obtained via local $M-$estimates. The first proposal of robust estimates for nonparametric regression was given by Cleveland (1979) who adapted a local polynomial fit by introducing weights to deal with large residuals. See also, Tsybakov (1982) and Härdle (1984), who studied pointwise asymptotic properties of a robust version of the Nadaraya–Watson method. These results were extended to $M-$type scale equivariant kernel estimates by Härdle and Tsybakov (1988) and by Boente and Fraiman (1989) who also considered robust equivariant nonparametric estimates using nearest neighbor weights. A review of several methods leading to robust nonparametric regression estimators can be seen in Härdle (1990).

The proposals mentioned above assume that the predictors $\mathbf{x}$ belong to a subset of $\mathbb{R}^d$ with non empty interior and therein, the Euclidean structure of $\mathbb{R}^d$ is considered. However, in many applications, the predictors $\mathbf{x}$ take values on a Riemannian manifold more than on $\mathbb{R}^d$ and this structure of the explanatory variables needs to be taken into account when considering neighborhoods around a fixed point $\mathbf{x}$. Several authors such as, Mardia and Jupp (2000), Hall *et al.* (1987) and Fischer *et al.* (1993) discussed methods for spherical and circular data analysis while generalizations to different types of manifolds have been studied by Lee and Ruymgaart (1996), Hendriks (1990) and Hendriks *et al.* (1993). An approach based on the Riemannian geodesic distance on the manifold was considered by Pelletier (2005) for the problem of estimating the density of random objects on a compact Riemannian manifold and also by Pelletier (2006) for that of estimating the regression function which is the aim of our paper. More precisely, let (M, g) be a closed Riemannian manifold of dimension d and let $(y, \mathbf{x})$ be a random vector such that $y \in \mathbb{R}$ and $\mathbf{x} \in M$. The classical nonparametric setting assumes that the response variables have finite expectation and focuses on the estimation of the regression function $r(p) = E(y|\mathbf{x} = p)$. Pelletier's (2006) idea was to build an analogue of a kernel on (M, g) by using a positive function of the geodesic distance on M, which is then normalized by the volume density function of (M, g) to take into account for curvature. Under standard assumptions on the kernel and the bandwidth sequence, Pelletier (2006) derives an expression for the asymptotic pointwise bias and variance as well as an expression for the asymptotic integrated mean square error.

As in the Euclidean setting, the estimators introduced by Pelletier (2006) are a weighted average of the response variables y_i with weights depending on the distance between $\mathbf{x}_i$ and p implying that they will suffer from the same lack of robustness that the Nadaraya–Watson estimators with carriers in the Euclidean space $\mathbb{R}^d$. In this paper we consider two families of robust estimators for the regression function when the explanatory variables $\mathbf{x}_i$ take values on a Riemannian manifold (M, g). The first family combine the ideas of robust smoothing in Euclidean spaces with the kernel weights introduced in Pelletier (2005). The second one generalizes to our setting the proposal given by Boente and Fraiman (1989), who considered robust nonparametric estimates using nearest neighbor weights when the predictors $\mathbf{x}$ are on $\mathbb{R}^d$. Therefore, the aim of this paper is to introduce local $M-$estimators adapted to regressors lying on a $d-$dimensional Riemannian manifold and to study their asymptotic properties. It is worth mentioning that robust estimators for directional data were considered among others by He (1992), Ko and Guttorp (1988) and also by Agostinelli (2007) who studied robust methods for circular data analysis.

As in Pelletier (2006), these two families of estimators, kernel or k-nearest neighbor with kernel, will be consistent with the respective estimators on Euclidean spaces, i.e., they reduce to local $M-$estimators based on standard kernel or k-nearest with kernel weights when M is $I\!R^d$. Moreover, they converge at the same rate as the Euclidean kernel estimators. This result generalizes the conclusions obtained by Pelletier (2006) using the pointwise mean square error.

This paper is organized as follows. In Section 2, we introduce two versions of robust local $M-$estimators of the regression function, adapted to the fact that the explanatory variables $\mathbf{x}_i$ take values on a Riemannian manifold. Uniform consistency of the proposed estimators is derived in Section 3 while in Section 4 the asymptotic distributions are obtained. In Section 5, the behavior of the classical and robust approach are compared through a Monte Carlo study under normality and contamination, for moderate sample sizes. Proofs are given in the Appendix.

2 Robust nonparametric estimates based on kernel method

2.1 Preliminaries

Let (M, g) be a $d-$dimensional oriented Riemannian manifold without boundary. Denote by d_g the distance induced by g and by $inj_g M = \inf_{p\in M} \sup\{s \in I\!R > 0 : B_s(p)$ is a normal ball$\}$ the injectivity radius of (M, g).

Throughout this paper, we will assume that (M, g) is complete, i.e., (M, d_g) is a complete metric space, and that $inj_g M$ is strictly positive. Some examples of Riemannian manifolds with positive injectivity radius are $I\!R^d$ with g the canonical metric ($inj_g I\!R^d = \infty$) and the d dimensional sphere S^d with the metric induced by the canonical metric of $I\!R^d$ ($inj_g S^d = \pi$). It is also well known that compact Riemannian manifolds have positive injectivity radius. So the Projective spaces and Grassmannian are examples of manifolds with this property. The Orthogonal group and the Special group of $I\!R^{d\times d}$ are examples of the important family of compact Lie groups, and thus they have positive injectivity radius. Moreover, complete and simply connected Riemannian manifolds with non positive sectional curvature, like the Poincaré ball and the Poincaré halfspace, have also this property. If a complete Riemannian manifold (M, g) have a positive lower bound for its Ricci curvature (i.e. $Ricc_p(v) \geq r > 0$ for all $p \in M$ and $v \in T_pM$), then $inj_g M > 0$. This is because the diameter of M is finite and M result a compact manifold. Finally, an important family of examples is deduced from the following well known fact: If the sectional curvature κ_M of a complete Riemannian manifold (M, g) satisfies that $0 < r_1 \leq \kappa_M \leq r_2$, then $inj_g M > 0$, see Do Carmo (1988). Some standard results on differential geometry can be seen for instance in Berger *et. al* (1971), Besse (1978) and Boothby (1975).

From now on, we will denote by $B_s(p)$ the normal ball in (M, g) centered at p with radius s. Then, $B_s(0_p) = exp_p^{-1}(B_s(p))$ is an open neighborhood of 0_p in T_pM, the tangent space of M at p, and so it has a natural structure of differential manifold. We are going to consider the Riemannian metrics g' and g'' in $B_s(0_p)$, where $g' = exp_p^*(g)$ is the pullback of g by the exponential map and g'' is the canonical metric induced by g_p in $B_s(0_p)$. Let $w \in B_s(0_p)$, for any chart $(\bar{U}, \bar{\psi})$ of $B_s(0_p)$ such that $w \in \bar{U}$, the volumes of the parallelepiped spanned by $\{(\partial/\partial\bar{\psi}_1|_w), \ldots, (\partial/\partial\bar{\psi}_d|_w)\}$ with respect to the metrics g' and g'' are given by $|\det g' ((\partial/\partial\bar{\psi}_i|_w), (\partial/\partial\bar{\psi}_j|_w))|^{1/2}$ and $|\det g'' ((\partial/\partial\bar{\psi}_i|_w), (\partial/\partial\bar{\psi}_j|_w))|^{1/2}$ respectively. The quotient between these two volumes is independent of the selected chart. So, given $q \in B_s(p)$, if $w = exp_p^{-1}(q) \in B_s(0_p)$ we can define the volume density function, $\theta_p(q)$, on (M, g) as

$$\theta_p(q) = \frac{|\det g' ((\partial/\partial\bar{\psi}_i|_w), (\partial/\partial\bar{\psi}_j|_w))|^{1/2}}{|\det g'' ((\partial/\partial\bar{\psi}_i|_w), (\partial/\partial\bar{\psi}_j|_w))|^{1/2}}$$

for any chart $(\bar{U}, \bar{\psi})$ of $B_s(0_p)$ that contains $w = exp_p^{-1}(q)$. For instance, if we consider the exponential chart (U, ψ) of (M, g) induced by an orthonormal basis $\{v_1, \ldots, v_d\}$ of T_pM and U a normal neighborhood of p then

$$\theta_p(q) = \left| \det g_q \left(\frac{\partial}{\partial \psi_i}\Big|_q, \frac{\partial}{\partial \psi_j}\Big|_q \right) \right|^{\frac{1}{2}},$$

where $\partial/\partial\psi_i|_q = D_{\alpha_i(0)} exp_p(\dot{\alpha}_i(0))$ with $\alpha_i(t) = exp_p^{-1}(q) + tv_i$ for $q \in U$. Note that the volume density function $\theta_p(q)$ is not defined for all p and q in M, but only for those points such that $d_g(p, q) < inj_gM$. It is worth noticing that, when M is $\mathbb{R}^d$ with the canonical metric, then $\theta_p(q) = 1$ for all $p, q \in \mathbb{R}^d$.

In order to do a Monte Carlo study (Section 5.1) we will need to calculate the volume density on the cylinder. In the following we make a sketch for the construction of it. Let $(\mathcal{C}_1, g)$ be the cylinder of radius 1 endowed with the metric induced by the canonical metric of $\mathbb{R}^3$. Let (V, ψ) be the chart induced by the parametrization (U, φ), where $U = \{(-\pi, \pi) \times \mathbb{R}\}$ and $\varphi(r, s) = (cos(r), sen(r), s)$. Note that the map φ is a local isometry, so the geodesics of U are lifted by φ to geodesics on the cylinder. We mean that if $b \in U$ and a is a direction, then $c(t) = \varphi(at + b)$ is the geodesic that pass at time zero for $\varphi(b)$ with velocity $(-a_1.sen(b_1), a_1 cos(b_1), a_2)$. It well know that the geodesics of the cylinder are parametrization (with $\|\dot{c}\| = constant$) of circles, straight lines and helices. Let p and q in V and (r_1, s_1) and (r_2, s_2) such that $\varphi(r_1, s_1) = p$ and $\varphi(r_2, s_2) = q$. Then we have that $d_g(p, q) = d_2((r_1, s_1), (r_2, s_2))$ if $d_2((r_1, s_1), (r_2, s_2)) < \pi$, where d_2 is the euclidean distance of $\mathbb{R}^2$. Consider for p the chart of $B_\pi(0_p)$ $(\bar{V}, \bar{\psi})$ given by $\bar{\psi}^{-1}(r, s) = (cos(r_1 + r), sen(r_1 + r), s_1 + s)$. Using this chart an easy computation verify that $\theta_p \equiv 1$.

Finally, we show another example of the volume density θ. Let S^2 be the two dimensional sphere of radius 1 and $p \in S^2$. Let v and w such that $\|v\| = \|w\| = 1$ and $\{v, w, p\}$ is an orthonormal basis of $\mathbb{R}^3$. Consider the exponential chart (U, ψ) induced by the parametrization $\psi^{-1} : B_\pi(0) \longrightarrow S^2 \setminus \{-p\}$ given by $\psi^{-1}(s, t) = cos(r)p + sen(r)(sv + tw)r^{-1}$ if $(s, t) \neq (0, 0)$ and $\psi^{-1}(0, 0) = p$ where $r = \sqrt{s^2 + t^2}$. Note that $r = d_g(p, q)$ if

$q = exp_p(sv + tw)$. Then we have that

$$A_{(s,t)} = \frac{\partial}{\partial \psi_1}|_{\psi^{-1}(s,t)} = \frac{-sen(r)s}{r} p + \frac{cos(r)rs^2 + t^2 sen(r)}{r^3} v + \frac{[cos(r)r - sen(r)]st}{r^3} w$$

$$B_{(s,t)} = \frac{\partial}{\partial \psi_2}|_{\psi^{-1}(s,t)} = \frac{-sen(r)t}{r} p + \frac{[cos(r)r - sen(r)]st}{r^3} v + \frac{cos(r)rt^2 + s^2 sen(r)}{r^3} w$$

Since $g(A_{(s,t)}, B_{(s,t)}) = [r^2 - sen^2(r)]st\ r^{-4}$, $g(A_{(s,t)}, A_{(s,t)}) = (s^2 r^2 + sen^2(r)t^2)\ r^{-4}$ and $g(B_{(s,t)}, B_{(s,t)}) = (t^2 r^2 + sen^2(r)s^2)\ r^{-4}$, we obtain that the volume density on the sphere is

$$\theta_p(q) = \frac{|sen(d_g(p,q))|}{d_g(p,q)} \quad \text{for } q \neq p, -p.$$

See also, Besse (1978) for a discussion on the volume density function.

2.2 The robust regression estimators

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let us consider $\mathcal{B}$ the Borel $\sigma-$field of M. Let $\mathbf{x}$ be a random object defined on M, i.e., a measurable map $\mathbf{x} : \Omega \to M$ and y be a random variable on $I\!R$. Denote by $\Psi : I\!R \to I\!R$ a strictly increasing, bounded and continuous function and by $F(y|\mathbf{x} = p)$, the conditional distribution function of y given $\mathbf{x} = p$.

For each $p \in M$ denote by $r_\Psi(p)$ the solution with respect to a of $\eta(p, a) = 0$ where

$$\eta(p, a) = E\left(\Psi\left(\frac{y - a}{\sigma(p)}\right)\Big| \mathbf{x} = p\right) ,$$

with $\sigma(p)$ is a robust measure of the conditional scale, for instance,

$$\sigma(p) = \text{MAD}_c(p) = \text{median}(|y - m(p)|), \tag{1}$$

and $m(p) = \text{median}(y|\mathbf{x} = p)$ the median of $F(y|\mathbf{x} = p)$. Theorem 2.1 of Boente and Fraiman (1989) states that if the score function Ψ is a strictly increasing bounded continuous score function, the robust location conditional functional r_Ψ exists, is unique and measurable. Moreover, if the conditional distribution function $F(y|\mathbf{x} = p)$ is symmetric around $r(p)$ and Ψ is odd, then $r_\Psi(p) \equiv r(p)$. This setting includes the nonparametric regression model

$$y = r(\mathbf{x}) + \sigma(\mathbf{x})\varepsilon \tag{2}$$

where the error ε has symmetric distribution $F_0(\cdot)$ and is independent of $\mathbf{x}$. As mentioned in Boente and Fraiman (1989), $r_\Psi(p)$ is a natural extension of the conditional expectation $E(y|\mathbf{x})$ to a setting where no moment conditions are required and thus, in a robust setting it represents the functional to be estimated. When considering $\Psi(t) = \text{sgn}(t)$, the target is then the conditional median, while for general score functions Ψ, the target is the robust location conditional functional $r_\Psi(p)$ related to Ψ. It is worth noticing that the symmetry assumption required to the error's distribution is needed if we want to guarantee that all robust location conditional functionals are Fisher–consistent, i.e., $r_\Psi(p) = r(p)$, and so, to ensure that the robust estimators introduced will be consistent to the regression function $r(p)$. For a discussion regarding the choice of the score function leading to the conditional location functionals, see He *et al.* (2002).

Let $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$ be i.i.d. random vectors valued in $I\!R \times M$ with the same distribution as $(y, \mathbf{x})$. As mentioned above, our aim is to estimate robustly the nonparametric regression function r. As in the Euclidean case, a natural approach is to define the robust estimators of r by considering estimators of the conditional distribution function $F(y|\mathbf{x} = p)$ adapted to fact that the carriers belong to a Riemannian manifold. Following, Pelletier's (2006) proposal and the proposal given by Boente and Fraiman (1989), we will consider two families of estimators of $F(y|\mathbf{x} = p)$.

Estimators based on kernel weight. They are defined as

$$F_n(y|\mathbf{x} = p) = \sum_{i=1}^n w_{ni}(p) I_{(-\infty, y]}(y_i), \tag{3}$$

with

$$w_{ni}(p) = \frac{\dfrac{1}{\theta_{\mathbf{x}_i}(p)} K\left(\dfrac{d_g(p, \mathbf{x}_i)}{h_n}\right)}{\displaystyle\sum_{j=1}^n \dfrac{1}{\theta_{\mathbf{x}_j}(p)} K\left(\dfrac{d_g(p, \mathbf{x}_j)}{h_n}\right)},$$

where the bandwidth h_n is a sequence of real positive numbers such that $\lim_{n\to\infty} h_n = 0$ and $h_n < inj_g M$ for all n. This last requirement on the bandwidth guarantees that (3) is defined for all $p \in M$ (see, Pelletier (2006)).

Estimators based on k-nearest neighbor kernel weights. These estimators are defined through

$$\widetilde{F}_n(y|\mathbf{x} = p) = \sum_{i=1}^n \widetilde{w}_{ni}(p) I_{(-\infty, y]}(y_i),$$

with

$$\widetilde{w}_{ni}(p) = \frac{\dfrac{1}{\theta_{\mathbf{x}_i}(p)} K\left(\dfrac{d_g(p, \mathbf{x}_i)}{H_n(p)}\right)}{\displaystyle\sum_{j=1}^n \dfrac{1}{\theta_{\mathbf{x}_j}(p)} K\left(\dfrac{d_g(p, \mathbf{x}_j)}{H_n(p)}\right)},$$

where $H_n(p)$ is the distance d_g between p and the k-nearest neighbor of p among $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and $k = k_n$ is a sequence of non–random integers such that $\lim_{n\to\infty} k_n = \infty$.

In both cases, $K : I\!R \to I\!R$ is a non-negative function such that $\int K(u)du < \infty$ and $\theta_p(q)$ denotes the volume density function on (M, g).

The robust nonparametric estimators of the regression function can be defined as follows. Denote by $\sigma_n(p)$ and $\widetilde{\sigma}_n(p)$ the preliminary local robust scale estimates corresponding to $F_n(y|\mathbf{x}=p)$ and $\widetilde{F}_n(y|\mathbf{x}=p)$ respectively. For instance, $\sigma_n(p)$ and $\widetilde{\sigma}_n(p)$ can be taken as the local median of the absolute deviations from the local median, i.e., the MAD, defined in (1), with respect to the distributions $F_n(y|\mathbf{x}=p)$ and $\widetilde{F}_n(y|\mathbf{x}=p)$. Then, the robust regression estimators are the solutions $r_n(p)$ and $\widetilde{r}_n(p)$ of

$$\sum_{i=1}^{n} w_{in}(p)\Psi\left(\frac{y_i - r_n(p)}{\sigma_n(p)}\right) = 0. \tag{4}$$

and

$$\sum_{i=1}^{n} \widetilde{w}_{in}(p)\Psi\left(\frac{y_i - \widetilde{r}_n(p)}{\widetilde{\sigma}_n(p)}\right) = 0. \tag{5}$$

respectively.

The estimator defined by Pelletier (2006) corresponds to the choice $\Psi(t) = t$ with the estimators of the conditional distribution based on kernel weights defined in (3). On the other hand, local kernel medians corresponds to $\Psi(t) = sg(t)$. It is worth noticing that, when M equals $I\!R^d$ with the canonical metric, r_n and $\widetilde{r}_n$ reduce to the local $M-$estimators with standard kernel weights or k-nearest with kernel introduced in Boente and Fraiman (1989). This fact was pointed out by Pelletier (2006) for the linear kernel estimator, i.e., when $\Psi(t) = t$ and when the weights corresponds to the kernel weights $w_{in}(p)$.

Being a conditional location estimator, the estimators $r_n(p)$ or $\widetilde{r}_n(p)$ can be computed iteratively using reweighting, as described for the location setting in Chapter 2 of Maronna et al. (2006). In what follows, we briefly describe the procedure to compute $r_n(p)$. Let $r_n^{(0)}(p)$ be an initial robust regression estimator, for instance, the local median and $\sigma_n(p)$ a robust dispersion estimator such as the conditional MAD. Given $r_n^{(\ell)}(p)$, the weights involved in the reweighting algorithm are computed as $w_{in}(p)\; W\left(\left(y_i - r_n^{(\ell)}(p)\right)/\sigma_n(p)\right)$, with $W(t) = \Psi(t)/t$ for $t \neq 0$ and $W(0) = \Psi'(0)$. Thus, at step $(\ell+1)$, we define

$$r_n^{(\ell+1)}(p) = \frac{\displaystyle\sum_{i=1}^{n} w_{in}(p)\; W\left(\frac{y_i - r_n^{(\ell)}(p)}{\sigma_n(p)}\right) y_i}{\displaystyle\sum_{i=1}^{n} w_{in}(p)\; W\left(\frac{y_i - r_n^{(\ell)}(p)}{\sigma_n(p)}\right)}$$

and we iterate until the convergence criterion is achieved. It is worth noticing that a similar algorithm can be consider for $\widetilde{r}_n(p)$ replacing $w_{in}(p)$ by $\widetilde{w}_{in}(p)$.

An important issue in any smoothing procedure is the choice of the smoothing parameter. Under a nonparametric regression model with carriers in an Euclidean space, i.e., when M is $I\!R^d$ with the canonical metric, two commonly used approaches are L^2 cross–validation and plug–in methods. However, these procedures may not be robust and their sensitivity to anomalous data was discussed by several authors, including Wang and Scott (1994), Boente *et al.* (1997), Cantoni and Ronchetti (2001) and Leung (2005). Wang and Scott (1994) note that, in the presence of outliers, the least squares cross–validation function is nearly constant on its whole domain and thus, essentially worthless for the purpose of choosing a bandwidth. To solve this problem, they proposed an L^1 cross–validation method. On the other hand, for an homoscedastic nonparametric regression model with unidimensional carriers, Leung (2005) considered a robust cross–validation procedure by bounding large residuals using a bounded score function, such as Huber's function. The robustness issue remains valid for the estimators considered in this paper. With a small bandwidth, a small number of outliers with similar values of $\mathbf{x}_i$ could easily drive the estimate of r to dangerous levels. Therefore, one may consider to adapt to this setting the robust cross-validation approach described in Leung (2005). However, the asymptotic properties of data–driven estimators require further careful investigation and are beyond the scope of this paper.

3 Consistency

Let U be an open set of M, we denote by $C^k(U)$ the set of k times continuously differentiable functions from U to $\mathbb{R}$. As in Pelletier (2006), we assume that the image measure of P by $\mathbf{x}$ is absolutely continuous with respect to the Riemannian volume measure ν_g and we denote by f its density on M with respect to ν_g. To derive strong consistency results of the estimates $r_n(p)$ and $\widetilde{r}_n(p)$ defined in (4) and (5) respectively, we will consider the following set of assumptions.

*H*1. $\Psi : \mathbb{R} \to \mathbb{R}$ is an odd, strictly increasing, bounded and continuously differentiable function, such that $t\Psi'(t) \leq \Psi(t)$ for $t > 0$.

*H*2. $F(y|\mathbf{x} = p)$ is symmetric around $r(p)$ and a continuous function of y for each p.

*H*3. Let M_0 be a compact set on M such that:

i) f is a bounded function such that $\inf_{p\in M_0} f(p) = A > 0$.

ii) $\inf\limits_{\substack{p\in M_0\\ q\in M_0}} \theta_p(q) = B > 0$.

*H*4. The following equicontinuity condition holds

$$\forall\varepsilon > 0, \quad \exists\delta > 0 : |y - y'| < \delta \Rightarrow \sup_{p\in M_0} |F(y|\mathbf{x} = p) - F(y'|\mathbf{x} = p)| < \varepsilon \,.$$

*H*5. For any open set U_0 of M such that $M_0 \subset U_0$,

i) f is of class C^2 on U_0.

ii) $F(y|\mathbf{x} = p)$ is uniformly Lipschitz in U_0, that is, there exists a constant $C > 0$ such that $|F(y|\mathbf{x} = p) - F(y|\mathbf{x} = q)| \leq C\, d_g(p, q)$ for all $p, q \in U_0$ and $y \in \mathbb{R}$.

*H*6. $K : \mathbb{R} \to \mathbb{R}$ is a bounded nonnegative Lipschitz function of order one, with compact support $[0, 1]$ satisfying $\int_{\mathbb{R}^d} \mathbf{u}K(\|\mathbf{u}\|)d\mathbf{u} = \mathbf{0}$ and $0 < \int_{\mathbb{R}^d} \|\mathbf{u}\|^2 K(\|\mathbf{u}\|)d\mathbf{u} < \infty$.

$H7$. The sequence h_n is such that $h_n \to 0$, $nh_n^d \to \infty$, $nh_n^d/\log n \to \infty$ as $n \to \infty$.

$H8$. a) The estimator $\sigma_n(p)$ of $\sigma(p)$ satisfies $\sigma_n(p) \xrightarrow{a.s.} \sigma(p)$ as $n \to \infty$ for all $p \in M_0$.

b) The estimator $\widetilde{\sigma}_n(p)$ of $\sigma(p)$ satisfies $\widetilde{\sigma}_n(p) \xrightarrow{a.s.} \sigma(p)$ as $n \to \infty$ for all $p \in M_0$.

$H9$. The kernel $K(u)$ verifies $K(uz) \geq K(z)$ for all $u \in (0,1)$.

$H10$. The sequence k_n verifies that $k_n \to \infty$, $k_n/n \to 0$ and $k_n/\log n \to \infty$ as $n \to \infty$

Remark 3.1. Assumption $H1$ is a standard condition in a robustness framework. Boundness of the score function Ψ allows to derive the weak continuity of the robust conditional functional r_Ψ as shown in Theorem 2.2 of Boente and Fraiman (1989). Therefore, by applying this functional to weak consistent estimators of the conditional distribution, we obtain pointwise consistent and asymptotically strongly robust estimators of the regression function r. Differentiability of the score function is needed in order to obtain uniform consistency results over compact sets. As mentioned above, assumption $H2$ and the oddness of the score function guarantee Fisher–consistency. If the regression model (2) holds, this assumption can be replaced by $E\left(\Psi\left(\epsilon/\sigma\right)\right) = 0$, for any $\sigma > 0$. The fact that $\theta_p(p) = 1$ for all $p \in M$ guarantees that $H3$ holds for a small compact neighborhood of p. $H4$ and $H5$ are needed in order to derive strong uniform consistency results (see, for instance, Boente and Fraiman (1991)). Assumption $H6$ is a standard assumption when dealing kernel estimators. It is easy to see that Assumption $H8$ is satisfied, when we consider $\sigma_n(p)$ or $\widetilde{\sigma}_n(p)$ as the local MAD, i.e, the MAD, defined in (1), with respect to the distribution $F_n(y|\mathbf{x} = p)$ or $\widetilde{F}_n(y|\mathbf{x} = p)$.

Theorem 3.1. *Under $H3$ to $H8$ a), we have that* $\sup_{p \in M_0} |F_n(y|\mathbf{x} = p) - F(y|\mathbf{x} = p)| \xrightarrow{a.s.} 0$.

Remark 3.2. An inspection of the proof of Theorem 3.1 allows to conclude that the kernel density estimation on Riemannian manifolds defined in Pelletier (2005), is strong uniformly consistent on any compact set. Effectively, Pelletier (2005) defined a kernel density estimator of the probability density on Riemannian manifolds as

$$f_n(p) = \frac{1}{nh_n^d} \sum_{j=1}^{n} \frac{1}{\theta_{\mathbf{x}_j}(p)} K\left(\frac{d_g(p, \mathbf{x}_j)}{h_n}\right) .$$

In the proof of Theorem 3.1, it is shown that $\sup_{p \in M_0} |f_n(p) - E(f_n(p))| \xrightarrow{a.s.} 0$. Arguing as in Theorem 3.2 in Pelletier (2005) and using the compactness of M_0 we also have that,

$$\sup_{p \in M_0} |E(f_n(p) - f(p))| \leq Ch^2 \int \|\mathbf{u}\|^2 K(\|\mathbf{u}\|) d\mathbf{u}$$

and so, we can conclude that $\sup_{p \in M_0} |f_n(p) - f(p)| \xrightarrow{a.s.} 0$.

Theorem 3.2. *Under $H3$ to $H6$ and $H8$ b) to $H10$ we have that*

$$\sup_{p \in M_0} |\widetilde{F}_n(y|\mathbf{x} = p) - F(y|\mathbf{x} = p)| \xrightarrow{a.s.} 0.$$

Theorem 3.3.

a) *Assume that $H1$ to $H8$ a) holds. Then,* $\sup_{p\in M_0} |r_n(p) - r(p)| \xrightarrow{a.s.} 0$ *as* $n \to \infty$.

b) *Assume that $H1$ to $H6$ and $H8$ b) to $H10$ holds. Then,* $\sup_{p\in M_0} |\widetilde{r}_n(p) - r(p)| \xrightarrow{a.s.} 0$ *as* $n \to \infty$.

The proof of Theorem 3.3. follows easily using Theorem 3.1, Theorem 3.2 and similar arguments to those considered in Theorem 3.3 in Boente and Fraiman (1991). We will give a sketch of the proof in the Appendix.

4 Asymptotic normality

Denote by $V \subset M$ an open neighborhood of p and by $\kappa(q)$ and $\lambda(q)$ the functions

$$\kappa(q) = E\left(\Psi'\left(\frac{y - r(p)}{\sigma(p)}\right)\Big|\mathbf{x} = q\right) \quad \text{and} \quad \lambda(q) = E\left(\Psi^2\left(\frac{y - r(p)}{\sigma(p)}\right)\Big|\mathbf{x} = q\right) .$$

We will denote by $\mathcal{V}_s$ the ball in $I\!R^d$ centered at the origin with radius s and $\mu(\mathcal{V}_1)$ is the Lebesgue measure of the unit ball in $I\!R^d$.

In order to derive the asymptotic distribution of the estimator $r_n(p)$ defined in (4), we will consider the following set of assumptions.

*A*1. The function Ψ is twice continuously differentiable with bounded derivatives. Its second derivative Ψ'' verifies that, for some positive constants c, D and ϵ, $|\Psi''(t)| \le c|t|^{-(2+\epsilon)}$ for $|t| > D$.

*A*2. $f(p) > 0$ and f is of class C^2 on V.

*A*3. The sequence h_n is such that $h_n \to 0$, $nh_n^d \to \infty$ and there exists $0 \le \beta < \infty$ $n^{1/(d+4)} h_n \to \beta$ as $n \to \infty$.

*A*4. There exists a continuous symmetric distribution function F_0 such that $F(y|\mathbf{x} = p) = F_0\left((y - r(p))/\sigma(p)\right)$ with r and σ such that

   i) r is of class C^2 on V.

   ii) $|\sigma(q) - \sigma(p)| < C\,[d_g(p,q)]^{\frac{1}{2}}$ for $q \in V$ and some constant $C > 0$.

*A*5. $\int \Psi'(z)\, dF_0(z) \neq 0$.

*A*6. $\kappa(q) \in C^2(V)$ and $\lambda(q)$ is continuous at p.

*A*7. The sequence k_n is such that $k_n \to \infty$, $k_n/n \to 0$ and there exists $0 \le \beta < 0$ such that $k_n^{1/d} n^{1/(d+4)-1/d} \to \beta(f(p)\mu(\mathcal{V}_1))^{1/d}$.

A8. The kernel K is twice continuously differentiable and verifies:

i) $0 < \int |K'(t)t|dt < \infty$, $\int (K'(t)t)^2 dt < \infty$ and $|t|^2 K'(t) \to 0$ as $|t| \to \infty$.

ii) $|t|^4 K'(t) \to 0$ as $|t| \to \infty$.

Theorem 4.1. *Assume A1 to A6 and H6, and that* $r_n(p) \xrightarrow{p} r(p)$, $\sigma_n(p) \xrightarrow{p} \sigma(p)$. *Then, we have that*

$$\sqrt{nh_n^d}(r_n(p) - r(p)) \xrightarrow{\mathcal{D}} N(b(p), V(p))$$

with

$$b(p) = \frac{\beta^{(d+4)/2}}{2f(p)} \frac{\int_{\mathcal{V}_1} K(\|\mathbf{u}\|) u_1^2 \, d\mathbf{u}}{\int_{\mathcal{V}_1} K(\|\mathbf{u}\|) d\mathbf{u}} L(p)$$

and

$$V(p) = \frac{\sigma^2(p)}{f(p)} \frac{\int \Psi^2(z) dF_0(z)}{[\int \Psi'(z) dF_0(z)]^2} \frac{\int_{\mathcal{V}_1} K^2(\|\mathbf{u}\|) d\mathbf{u}}{[\int_{\mathcal{V}_1} K(\|\mathbf{u}\|) d\mathbf{u}]^2}$$

where $\mathbf{u} = (u_1, \ldots, u_d)$ *and* $L(p) = \sum_{i=1}^{d} \frac{\partial^2 \phi \circ \psi^{-1}}{\partial u_i u_i}\Big|_{u=0}$ *with* $\phi(q) = f(q)(r(q) - r(p))$ *and* $(B_h(p), \psi)$ *some exponential chart induced by an orthonormal basis of* T_pM.

From the proof of Theorem 4.1, it follows easily that $L(p)$ is well defined.

In order to derive the asymptotic distribution of $\widetilde{r}_n(p)$, we will first study the asymptotic behavior of h_n^d / H_n^d where $h_n^d = k_n/(nf(p)\mu(\mathcal{V}_1))$. Note that if we consider $\widetilde{f}_n(p) = k_n/(nH_n^d \mu(\mathcal{V}_1))$ a careful inspection of the proof of Theorem 3.2 allows to conclude that $\widetilde{f}_n(p)$ is a consistent estimator of $f(p)$. Theorem 4.2 state that this estimator is also asymptotically normally distributed as it is the case when M is $\mathbb{R}^d$.

Theorem 4.2. *Assume A2, A7 and H6, and that let* $h_n^d = k_n/(nf(p)\mu(\mathcal{V}_1))$. *Then, we have that*

$$\sqrt{k_n}\left(\frac{h_n^d}{H_n^d} - 1\right) \xrightarrow{\mathcal{D}} N(b_1(p), 1)$$

with

$$b_1(p) = (\beta^{d+4} f(p)\mu(\mathcal{V}_1))^{\frac{1}{2}} \left\{ \frac{\tau}{6d+12} + \frac{\int_{\mathcal{V}_1} u_1^2 \, d\mathbf{u} \; L_1(p)}{f(p)\mu(\mathcal{V}_1)} \right\}$$

where $\mathbf{u} = (u_1, \ldots, u_d)$, τ *is the Ricci curvature,*

$$L_1(p) = \sum_{i=1}^{d} \left(\frac{\partial^2 f \circ \psi^{-1}}{\partial u_i u_i}\Big|_{u=0} + \frac{\partial f \circ \psi^{-1}}{\partial u_i}\Big|_{u=0} \frac{\partial \theta_p \circ \psi^{-1}}{\partial u_i}\Big|_{u=0} \right)$$

and $(B_h(p), \psi)$ *is some exponential chart induced by an orthonormal basis of* T_pM.

The following result is a consequence of Theorem 4.1 and Theorem 4.2 and follows using analogous arguments to those considered in Theorem 2 in Boente and Fraiman (1990).

Theorem 4.3. *Assume A1, A2, H10 and A4 to A8 and that* $\widetilde{r}_n(p) \xrightarrow{p} r(p)$, $\widetilde{\sigma}_n(p) \xrightarrow{p} \sigma(p)$. *Then, we have that*

$$\sqrt{k_n}(\widetilde{r}_n(p) - r(p)) \xrightarrow{\mathcal{D}} N(b_2(p), V_2(p))$$

with

$$b_2(p) = \frac{(\beta^{d+4}\mu(\mathcal{V}_1))^{1/2}}{2(f(p))^{1/2}} \frac{\int_{\mathcal{V}_1} K(\|\mathbf{u}\|)u_1^2\, d\mathbf{u}}{\int_{\mathcal{V}_1} K(\|\mathbf{u}\|)d\mathbf{u}}\; L(p)$$

and

$$V_2(p) = \sigma^2(p)\mu(\mathcal{V}_1)\frac{\int \Psi^2(z)dF_0(z)}{[\int \Psi'(z)dF_0(z)]^2}\frac{\int_{\mathcal{V}_1} K^2(\|\mathbf{u}\|)d\mathbf{u}}{[\int_{\mathcal{V}_1} K(\|\mathbf{u}\|)d\mathbf{u}]^2}$$

where $\mathbf{u} = (u_1, \ldots, u_d)$ *and* $L(p) = \sum_{i=1}^{d} \frac{\partial^2 \phi \circ \psi^{-1}}{\partial u_i u_i}\Big|_{u=0}$ *with* $\phi(q) = f(q)(r(q) - r(p))$ *and* $(B_h(p), \psi)$ *some exponential chart induced by an orthonormal basis of* T_pM.

5 Monte Carlo Study

This section contains the results of a preliminary simulation study designed to evaluate the performance of the robust procedure defined in Section 2.2. We only consider in this study the estimators based on kernel weights. The aim of this study is to compare the behavior of the classical and robust estimators under normal samples and under contamination.

We performed 1000 replications generating independent samples of size $n = 200$. In all cases, the carriers $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent with support in the cylinder with radius 1 and height between $(-2, 2)$ denoted $\mathcal{C}_{1,(-2,2)}$. The predictors were generated as $\mathbf{x}_i = (\cos(x_{1,i}), \sin(x_{1,i}), x_{2,i})$ with $x_{1,i}$ and $x_{2,i}$ independent such that $x_{1,i} \sim U(-\pi, \pi)$ and $x_{2,i} \sim U(-2, 2)$. The response regression function was taken as $r(\mathbf{x}) = 4 - x_2^2 + \sin(x_1)$.

The non-contaminated case, indicated by C_0, corresponds to normally distributed errors ε_i with mean 0 and standard deviation 1. Besides, the so–called contaminations C_1 and C_2, which correspond to select a distribution in a neighborhood of the central normal distribution, are defined as $\varepsilon \sim 0.9N(0,1) + 0.1\mathcal{C}(0,1)$ and $\varepsilon \sim 0.9N(0,1) + 0.1\mathcal{C}(5,0.5)$ respectively, where $\mathcal{C}(0,\sigma)$ indicates the distribution Cauchy centered in 0 with scale σ. The contamination C_1 corresponds to inflating the errors and thus, will affect the variance of the regression estimates while the goal of C_2 is to introduce a bias in the estimation.

As described in Section 2.2, we consider the local median, as initial estimate in the iterative procedure to compute r_n. In all cases, for smoothing, the kernel was taken as the quadratic kernel $K(t) = (15/16)(1 - t^2)^2 I(|x| < 1)$. We have considered two choices for the smoothing parameter $h = 0.5$ and $h = 1$. Besides, the robust estimators were computed using as score function $\Psi(u) = (1 - (u/c)^2)^2 I(|u| < c)$, the bisquare function, with tuning constant $c = 4.685$. Even if this choice of Ψ does not correspond to an increasing function, it leads to consistent and asymptotically normally estimators if we choose the solution which minimizes $\sum_{i=1}^n w_{in}(p)\rho\left((y_i - a)/\sigma_n(p)\right)$ as in the location case.

5.1 Simulation results

The performance of an estimate r_n of r is measured using an approximation to the MISE where

$$\text{MISE}(r_n) = E\int (r(q) - r_n(q))^2 dq.$$

The value of the MISE was approximated by Monte Carlo as $\sum_{i=1}^{1000} M(r_n^i)/1000$ where r_n^i corresponds to the estimators of r computed at the ith replication and

$$M(r_n^i) = \frac{1}{800}\sum_{l=1}^{20}\sum_{j=1}^{40}(r(\mathbf{z}_{lj}) - r_n^i(\mathbf{z}_{lj}))^2,$$

with $\mathbf{z}_{lj}$ a grid of 800 equispaced points in $\mathcal{C}_{1,(-2,2)}$. We have selected $\mathbf{z}_{lj} = (\cos\theta_l, \sin\theta_l, a_j)$ with $\theta_l \in (-\pi, \pi)$ for $1 \le l \le 20$ and $a_j \in (-2, 2)$ for $1 \le j \le 40$.

Table 1 gives the values of the approximation of the MISE, denoted MISE for the sake of simplicity, for the classical and robust nonparametric estimators when considering normal samples and under C_1 and C_2, for the selected bandwidths.

	$h = 0.5$		$h = 1$	
	MISE$(r_{n,\mathrm{C}})$	MISE$(r_{n,\mathrm{R}})$	MISE$(r_{n,\mathrm{C}})$	MISE$(r_{n,\mathrm{R}})$
C_0	0.4183	0.4240	0.2015	0.2038
C_1	1832.452	2.7432	555.7038	0.2287
C_2	459.7590	2.1784	139.7073	0.5069

Table 1: Estimated mean integrated Square Error (MISE) for the classical estimator ($r_{n,\mathrm{C}}$) and the robust estimator ($r_{n,\mathrm{R}}$).

The simulation study confirms the inadequate behavior of the classical estimators under the considered contaminations and in particular, how it leads to an increased mean integrated square error when anomalous data are present. On the other hand, for normal errors the robust estimates of the regression function r behaves almost as its linear relative, showing only a small lack of efficiency. It is worth noticing that the MISE for normal samples corresponding to $h = 0.5$ is more than twice that related to $h = 1$, showing that in this case, $h = 1$ provides a more adequate smoothing parameter. Moreover, when the smallest bandwidth is selected, the MISE of the robust estimators are increased six and five times under C_1 and C_2, respectively. This can be explained by the fact than in some of these small neighborhoods the proportion of outliers in the neighborhood exceeds the breakdown point of the $M-$estimator. However, even in this situation, MISE$(r_{n,\mathrm{R}})$ is much lower than that of the linear estimators. In fact when $h = 0.5$, the MISE of the classical estimator is 668 times larger than that of the robust procedure under C_1 and more than 200 times under C_2. On the other hand, when $h = 1$, the MISE of the robust estimators under C_1 is almost the same of that obtained under C_0, while, due to the bias introduced by the asymmetric contamination C_2, MISE$(r_{n,\mathrm{R}})$ is increased 2.5 times under C_2. The linear estimator is more sensitive to C_1 than to C_2 due to the large scale of the contaminating distribution. In both cases, the MISE($r_{n,\mathrm{C}}$) shows that the results obtained with the linear estimators are not reliable giving mean square errors more than 2000 and 300 times larger than those corresponding to the robust procedure, under C_1 and C_2, respectively.

This extreme behavior of the linear kernel estimator shows its inadequacy when one suspects that the sample can contain outliers.

5.2 Selection of the smoothing parameter

As mention in Section 2.2, an important issue in any smoothing procedure is the choice of the smoothing parameter. In order to complete the simulation study considered in the previous section, we included a classic method and a robust method for the choice of the bandwidth as in Leung (2005).

The classical least square cross-validation method constructs an asymptotically optimal data-driven bandwidth and thus, adaptive data-driven estimators, by minimizing

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\{y_i - r_{n,c_{-i}}(\mathbf{x}_i)\}^2$$

where $r_{n,c_{-i}}(h)$ denote the classic estimator computed with bandwidth h using all the data except $(y_i, \mathbf{x}_i)$. In this way the classical bandwidth can be chosen as $\widehat{h}_{n,c} = \operatorname{argmin} CV(h)$. On the other hand, as in Leung (2005), a robust cross-validation rules is defined as

$$RCV(h) = \frac{1}{n}\sum_{i=1}^{n}\Psi^2(y_i - r_{n,\mathrm{R}_{-i}}(\mathbf{x}_i))$$

where Ψ is a bounded score function as the Huber's function and $r_{n,\mathrm{R}_{-i}}(h)$ denote the robust estimator computed with bandwidth h using all the data except $(y_i, \mathbf{x}_i)$. Then, the robust cross validation selector $\widehat{h}_{n,\mathrm{R}}$ minimizes $RCV(h)$.

In order to study the sensitivity of the resulting estimator to the bandwidth, we have generated the sample according to C_0, C_1 and C_2, as in Section 5.1. The smoothing parameters were selected on a grid of 40 equally space values h between 0.01 and 2. In the robust cross-validation method we consider the Huber's score function with tunning constant $c = 1.345$. We performance only 200 replications taking into account that the robust cross-validation procedure is very expensive computationally. The performance of the estimators are measured as in the previous subsection, using an approximation to the MISE calculated in the same grid of points .

Table 2 gives the values of the approximation of the MISE for the classical and robust nonparametric estimators when considering normal samples, the contaminations C_1 and C_2 and the cross-validation selectors.

	MISE($r_{n,\mathrm{C}}$)	MISE($r_{n,\mathrm{R}}$)
C_0	0.21450	0.2372
C_1	17.0951	0.2550
C_2	4.9210	0.5396

Table 2: Estimated mean integrated Square Error (MISE) for the classical estimator ($r_{n,\mathrm{C}}$) and the robust estimator ($r_{n,\mathrm{R}}$) when using cross-validation.

The study of the sensitivity of the data selector shows that for the robust estimators, the results are quite similar to those obtained when using the cross-validation criterium and when using a fixed bandwidth equal 1. When we considered a bandwidth selector, the classical estimator has a better performance than when using a fixed bandwidth. However, under contamination the behavior of the classical estimator is still being inadequate.

6 Concluding Remarks

We have introduced two robust procedures to estimate the regression function when the regressors are random objects taking values in a Riemannian manifolds. Both procedures are strongly consistent and asymptotically normally distributed. Under the considered contaminations, they showed their advantage over the classical estimators defined by Pelletier (2006).

It is worth noticing that the added difficulties when regressors take values in Riemannian manifolds are those imposed by the geometry of the manifold. Given a Riemannian manifold, the volume density function and the geodesic distance can be difficult to calculate in all points of the manifold. In the cylinder's case considered in the simulation study, or in the sphere's case considered by Pelletier (2006), which can be the examples appearing more often in applications, their particular geometries facilitate the calculation.

Acknowledgments

We would like to thank Professor Graciela Boente for helpful discussions and for useful suggestions. We wish to thank two anonymous referees for valuable comments which led to an improved version of the original paper.

Appendix

From now on, we will denote by $d\nu_g$ the usual volume element induced by the metric g and the orientation of M.

Proof of Theorem 3.1. Let us begin by fixing some notation. Given $y \in I\!R$, denote $Z_j = I_{(-\infty,y]}(y_j)$ and for $i = 0, 1$, let $S_{in}(p) = \sum_{j=1}^{n} V_{ij} / \left(n h_n^d \right)$ with

$$V_{ij} = Z_j^i \frac{1}{\theta_{\mathbf{x}_j}(p)} K\left(\frac{d_g(p, \mathbf{x}_j)}{h_n} \right) - S_i(p) \qquad \text{and} \qquad S_i(p) = E\left(Z_1^i \frac{1}{\theta_{\mathbf{x}_1}(p)} K\left(\frac{d_g(p, \mathbf{x}_1)}{h_n} \right) \right) .$$

Note that

$$\sup_{p \in M_0} |F_n(y|\mathbf{x} = p) - F(y|\mathbf{x} = p)| \leq \left[\sup_{p \in M_0} |S_{1n}(p)| + \sup_{p \in M_0} |S_1(p) - S_0(p) F(y|\mathbf{x} = p)| \right.$$

$$+ \quad \sup_{p\in M_0} |S_{0n}(p)F(y|\mathbf{x}=p)| \Bigg] \Big/ \inf_{p\in M_0} (S_{0n}(p)+S_0(p)) \ .$$

Since $\inf_{p\in M_0}(S_{0n}(p)+S_0(p)) \geq \inf_{p\in M_0} S_0(p) - \sup_{p\in M_0}|S_{0n}(p)|$, to conclude the proof it is enough to show that

$$\sup_{p\in M_0} |S_{in}(p)| \xrightarrow{a.s.} 0 \quad \text{for} \quad i=0,1 \ , \tag{6}$$

$$\sup_{p\in M_0} |S_1(p) - S_0(p)F(y|\mathbf{x}=p)| \to 0 \ , \tag{7}$$

$$\inf_{p\in M_0} S_0(p) \geq A_1 > 0 \tag{8}$$

for some $A_1 > 0$.

We begin by proving (8). Using that

$$S_0(p) \ = \ \frac{1}{h_n^d}\int_M \frac{1}{\theta_q(p)} K\left(\frac{d_g(p,q)}{h_n}\right) f(q)d\nu_g(q) = \frac{1}{h_n^d}\int_{B_h(p)} \frac{1}{\theta_q(p)} K\left(\frac{d_g(p,q)}{h_n}\right) f(q)d\nu_g(q)$$

and $H3$ i) we have that

$$S_0(p) \ \geq \ A\frac{1}{h_n^d}\int_{B_h(p)} \frac{1}{\theta_q(p)} K\left(\frac{d_g(p,q)}{h_n}\right) d\nu_g(q) = A\int_{\mathcal{V}_1} K(\|\mathbf{u}\|)d\mathbf{u} = A_1 \ ,$$

which shows (8).

It is easy to see that (7) follows using $H5$ and a Taylor's expansion of order two, as in Pelletier (Theorem 3.2, 2006). On the other hand, Bernstein's inequality implies that, for $n \geq n_0$ and for some positive constant α, we have

$$P\left(|S_{in}(p)| > \varepsilon\right) \leq 2e^{-nh^d\alpha}. \tag{9}$$

Let us consider a finite collection of balls, $\{B_i = B_{h^\gamma}(p_i)\}_{1\leq i\leq \ell}$, centered at points $p_i \in M$ with radius h^γ with $\gamma > d+2$, such that $M_0 \subset \cup_{i=1}^{\ell} B_{h^\gamma}(p_i)$. Then, $\ell = O(h^{-\gamma})$ and

$$\sup_{p\in M_0} |S_{in}(p)| \leq \max_{1\leq j\leq l} \sup_{p\in B_j} |S_{in}(p) - S_{in}(p_j)| + \max_{1\leq j\leq l} |S_{in}(p_j)|. \tag{10}$$

Using that K is a Lipschitz function with Lipschitz constant $\|K\|_L$, straightforward calculation lead to $|S_{in}(p) - S_{in}(p_j)| < 2\|K\|_L \ h_n^{\gamma-(d+1)}$ for any $p \in B_j$, which entails that for n large enough, let us say, for $n > n_1$, we have

$$\max_{1\leq j\leq l} \sup_{p\in B_j} |S_{in}(p) - S_{in}(p_j)| < \varepsilon. \tag{11}$$

Through (10) and (11), for $n > n_0$ we have that

$$P\left(\sup_{p\in M_0} |S_{in}(p)| > 2\varepsilon\right) \leq P\left(\max_{1\leq j\leq l} |S_{in}(p_j)| > \varepsilon\right)$$

and by (9) imply that

$$P\left(\max_{1\le j\le l}|S_{in}(p_j)|>\varepsilon\right)\le 2\ell\, e^{-nh^d\alpha}.$$

Then, if $\delta_n = nh_n^d/\log n$ we get that $e^{-nh^d\alpha} = n^{-\alpha\delta_n}$. By $H7$ we have that for $n \ge n_1$, $nh_n^d > 1$, therefore, $\ell n^{-\alpha\delta_n} \le Cn^{\gamma/d-\alpha\delta_n}$ for $n \ge n_2$. Since $\delta_n \to \infty$, we have that for $n \ge n_3$, $\gamma/d - \delta_n\alpha < -2$. Hence, for $n \ge \max\{n_0, n_1, n_2, n_3\}$ and some constant C, we get

$$P\left(\sup_{p\in M_0}|S_{in}(p)|>2\varepsilon\right)\le Cn^{-2}$$

which shows that $\sum_{n=1}^{\infty} P\left(\sup_{p\in M_0}|S_{in}(p)|>2\varepsilon\right) < \infty$ for $i = 0, 1$ concluding the proof. □

Proof of Theorem 3.2. As in Theorem 3.1, given $y \in I\!R$, let $Z_i = I_{(-\infty,y]}(y_i)$ and denote by

$$\widehat{r}_n(p,\delta_n)=\frac{\sum_{i=1}^n Z_i\frac{1}{\theta_{\mathbf{x}_i}(p)}K\left(\frac{d_g(p,\mathbf{x}_i)}{\delta_n}\right)}{\sum_{i=1}^n \frac{1}{\theta_{\mathbf{x}_i}(p)}K\left(\frac{d_g(p,\mathbf{x}_i)}{\delta_n}\right)}$$

and

$$\widehat{f}_n(p,\delta_n)=\frac{1}{n\delta_n^d}\sum_{i=1}^n\frac{1}{\theta_{\mathbf{x}_i}(p)}K\left(\frac{d_g(p,\mathbf{x}_i)}{\delta_n}\right)\ .$$

Note that if $\delta_n = \delta_n(p)$ verifies $\delta_{1n} \le \delta_n(p) \le \delta_{2n}$ for all $p \in M_0$ where δ_{1n} and δ_{2n} satisfy $H7$; Theorem 3.1. and Remark 3.2 entail that

$$\sup_{p\in M_0}|\widehat{r}_n(p,\delta_n)-F(y|\mathbf{x}=p)|\xrightarrow{a.s.}0$$

and

$$\sup_{p\in M_0}|\widehat{f}_n(p,\delta_n)-f(p)|\xrightarrow{a.s.}0.$$

Now, the proof follows as in the Theorem 3.2 in Boente and Fraiman (1991). Effectively, for $0 < \beta < 1$ we define

$$\widehat{r}^{-}(p,\beta)=\frac{\sum_{i=1}^n Z_i\frac{1}{\theta_{\mathbf{x}_i}(p)}K\left(\frac{d_g(p,\mathbf{x}_i)}{\beta^{\frac{d}{2}}h_n}\right)}{\sum_{i=1}^n \frac{1}{\theta_{\mathbf{x}_i}(p)}K\left(\frac{d_g(p,\mathbf{x}_i)}{\beta^{-\frac{d}{2}}h_n}\right)}$$

and

$$\widehat{r}^{+}(p,\beta)=\frac{\sum_{i=1}^n Z_i\frac{1}{\theta_{\mathbf{x}_i}(p)}K\left(\frac{d_g(p,\mathbf{x}_i)}{\beta^{-\frac{d}{2}}h_n}\right)}{\sum_{i=1}^n \frac{1}{\theta_{\mathbf{x}_i}(p)}K\left(\frac{d_g(p,\mathbf{x}_i)}{\beta^{\frac{d}{2}}h_n}\right)}$$

where $h_n^d = k_n/(n\, f(p)\mu(\mathcal{V}_1))$ and $\mu(\mathcal{V}_r)$ denote the Lebesgue measure of the ball in $I\!R^d$ with radius r centered at the origin. By Assumption $H3$ i) and $H10$ we have that there exist sequences δ_{1n} and δ_{2n} verifying Assumption $H8$ such that $\delta_{1n} \leq h_n \leq \delta_{2n}$. Then,

$$\sup_{p \in M_0} \left| \sum_{i=1}^n \frac{1}{\theta_{\mathbf{x}_i}(p)} K\left(\frac{d_g(p,\mathbf{x}_i)}{\beta^{\frac{d}{2}} h_n}\right) \Big/ \sum_{i=1}^n \frac{1}{\theta_{\mathbf{x}_i}(p)} K\left(\frac{d_g(p,\mathbf{x}_i)}{\beta^{-\frac{d}{2}} h_n}\right) - \beta \right| \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

which implies

$$\sup_{p \in M_0} |\widehat{r}^-(p,\beta) - \beta F(y|\mathbf{x}=p)| \xrightarrow{a.s.} 0 \quad \text{and} \quad \sup_{p \in M_0} |\widehat{r}^+(p,\beta) - \beta^{-1} F(y|\mathbf{x}=p)| \xrightarrow{a.s.} 0. \quad (12)$$

For all $0 < \beta < 1$ and $\varepsilon > 0$, let us consider the sets

$$\begin{aligned}
A_n(\varepsilon) &= \{ \sup_{p \in M_0} |\widetilde{F}(y|\mathbf{x}=p) - F(y|\mathbf{x}=p)| < \varepsilon \}, \\
S_n^-(\beta,\varepsilon) &= \{ \sup_{p \in M_0} |\widehat{r}^-(p,\beta) - F(y|\mathbf{x}=p)| < \varepsilon \}, \\
S_n^+(\beta,\varepsilon) &= \{ \sup_{p \in M_0} |\widehat{r}^+(p,\beta) - F(y|\mathbf{x}=p)| < \varepsilon \}, \\
S_n(\beta) &= \{\widehat{r}^-(p,\beta) \leq \widetilde{F}(y|\mathbf{x}=p) \leq \widehat{r}^+(p,\beta) \text{ for all } p \in M_0\},
\end{aligned}$$

then $S_n(\beta) \cap S_n^-(\beta,\varepsilon) \cap S_n^+(\beta,\varepsilon) \subset A_n(\varepsilon)$. Therefore, given $0 \leq \varepsilon < \frac{3}{2}$ let $\beta_\varepsilon = 1 - \frac{\varepsilon}{3}$ and

$$\begin{aligned}
G_n^-(\varepsilon) &= \{ \sup_{p \in M_0} |\widehat{r}^-(p,\beta_\varepsilon) - \beta_\varepsilon F(y|\mathbf{x}=p)| < \varepsilon/3 \}, \\
G_n^+(\varepsilon) &= \{ \sup_{p \in M_0} |\widehat{r}^+(p,\beta_\varepsilon) - \beta_\varepsilon^{-1} F(y|\mathbf{x}=p)| < \varepsilon/3 \}, \\
G_n(\varepsilon) &= \{\beta_\varepsilon^{\frac{d}{2}} h_n \leq H_n(p) \leq \beta_\varepsilon^{-\frac{d}{2}} h_n \text{ for all } p \in M_0\}.
\end{aligned}$$

Then, we have that

$$G_n(\varepsilon) \subset S_n(\beta_\varepsilon), \quad G_n^-(\varepsilon) \subset S_n^-(\beta_\varepsilon,\varepsilon) \quad \text{and} \quad G_n^+(\varepsilon) \subset S_n^+(\beta_\varepsilon,\varepsilon)$$

therefore $G_n(\varepsilon) \cap G_n^-(\varepsilon) \cap G_n^+(\varepsilon) \subset A_n(\varepsilon)$.

On the other hand, using that $\lim_{r \to 0} V(B_r(p))/r^d \mu(\mathcal{V}_1) = 1$, where $V(B_r(p))$ denotes the volume of the geodesic ball centered at p with radius r (see Gray and Vanhecke (1979)) and similar arguments those considered in Devroye and Wagner (1977), we get that $I_{G_n^c(\varepsilon)} \xrightarrow{a.s.} 0$. This fact and (12) imply that $I_{A_n^c(\varepsilon)} \xrightarrow{a.s.} 0$ which concludes the proof $\square$

Proof of Theorem 3.3. a) This proof can be divided in three steps. The first one is using Theorem 3.1, $H5$ ii) and the equicontinuity condition given in $H4$ to get, as in Theorem 3.1 in Boente and Fraiman (1991) that

$$\sup_{y \in I\!R} \sup_{p \in M_0} |F_n(y|\mathbf{x}=p) - F(y|\mathbf{x}=p)| \xrightarrow{a.s.} 0. \quad (13)$$

The second step is derived from $H3$, $H4$ and $H5$. We can easily prove that there exist positive constants a, b and n_0 such that $a < \sigma_n(p) < b \; \forall p \in M_0$, for all $n \geq n_0$. These results can be obtained from Lemma 3.1 in Boente and Fraiman (1991). Finally, the last step follows from the following bound

$$\begin{aligned}\sup_{p\in M_0}\left|\int \Psi\left(\frac{y-(r_n(p)+t)}{\sigma_n}\right)dF(y|\mathbf{x}=p) - \int \Psi\left(\frac{y-(r_n(p)+t)}{\sigma_n}\right)dF_n(y|\mathbf{x}=p)\right| \\ \leq C(\Psi)\sup_{p\in M_0}\sup_{y\in I\!R}|F(y|\mathbf{x}=p)-F_n(y|\mathbf{x}=p)|, \quad (14)\end{aligned}$$

where $C(\Psi)$ is the total variation of Ψ. Denote by

$$\lambda(p,t,\sigma) = \int \Psi\left(\frac{y-t}{\sigma}\right)dF(y|\mathbf{x}=p)$$

and

$$\lambda_n(p,t,\sigma) = \int \Psi\left(\frac{y-t}{\sigma}\right)dF_n(y|\mathbf{x}=p).$$

Then by the first step, the left hand side of (14) converges to 0 a.s. as $n \to \infty$. By the second step there exist a and b such that $a < \sigma_n(p) < b \; \forall p \in M_0$ for all $n > n_0$.
Given $\varepsilon > 0$, $H1$, $H2$ and the continuity of $\lambda(p, r(p)+\varepsilon, \sigma)$ imply that

$$\lambda_1 = \sup_{a<\sigma<b}\sup_{p\in M_0}\lambda(p,r(p)+\varepsilon,\sigma) < 0 < \inf_{a<\sigma<b}\inf_{p\in M_0}\lambda(p,r(p)-\varepsilon,\sigma) = \lambda_2$$

and for n large enough

$$\lambda_n(p,r(p)+\varepsilon,\sigma_n(p)) < \lambda_1/2 < 0 < \lambda_2/2 < \lambda_n(p,r(p)-\varepsilon,\sigma_n(p))$$

for all $p \in M_0$. So, we can conclude that $\sup_{p\in M_0}|r_n(p)-r(p)| < \varepsilon$.

b) The proof of this part is similar that a) but using Theorem 3.2.□

Proof of Theorem 4.1. Denote $W_{in}(p) = (\theta_{\mathbf{x}_i}(p))^{-1}K\left(d_g(\mathbf{x}_i,p)/h_n\right)$ and

$$\begin{aligned} S_n(t,\sigma) &= \frac{1}{nh_n^d}\sum_{i=1}^n W_{in}(p)\Psi'\left(\frac{y_i-t}{\sigma}\right) \\ T_n(t,\sigma) &= \frac{1}{nh_n^d}\sum_{i=1}^n W_{in}(p)\Psi\left(\frac{y_i-t}{\sigma}\right) \\ R_n(\sigma) &= \frac{1}{nh_n^d}\sum_{i=1}^n W_{in}(p)\Psi\left(\frac{y_i-r(\mathbf{x}_i)}{\sigma}\right). \end{aligned}$$

Using a Taylor's expansion of order one, we obtain that

$$\sqrt{nh_n^d}(r_n(p)-r(p)) = \frac{\sigma_n(p)}{S_n(\xi_n(p),\sigma_n(p))}\sqrt{nh_n^d}\,T_n(r(p),\sigma_n(p))$$

where $\xi_n(p)$ is an intermediate point. Thus, it is enough to show that

$$\sqrt{nh_n^d}\, R_n(\sigma(p)) \xrightarrow{\mathcal{D}} N\left(0, f(p)\int \Psi^2(z)dF_0(z)\int_{\mathcal{V}_1} K^2(\|\mathbf{u}\|)d\mathbf{u}\right), \quad (15)$$

$$S_n(\xi_n(p),\sigma_n(p)) \xrightarrow{p} f(p)\int \Psi'(z)dF_0(z)\int K(\|\mathbf{u}\|)d\mathbf{u}\ , \quad (16)$$

$$\sqrt{nh_n^d}\, (T_n(r(p),\sigma_n(p)) - R_n(\sigma_n(p)) \xrightarrow{p} \frac{\beta^{(d+4)/2}\int K(\|u\|)u_1^2 d\mathbf{u}}{2\sigma(p)}\ L(p)\int \Psi'(u)dF_0(u), \quad (17)$$

$$\sqrt{nh_n^d}\, (R_n(\sigma_n(p)) - R_n(\sigma(p)) \xrightarrow{p} 0\ . \quad (18)$$

The proof of (15) follows easily using the Lindeberg Central Limit Theorem. On the other hand, (16) follows using similar arguments to those considered in Pelletier (2005), *A*6 and a Taylor's expansion of order two around q of the function $f(q)\kappa(q)$. The proof of (18) follows as in Boente and Fraiman (1990).

It remains to prove (17). Using a Taylor's expansion of order two of $T_n(r(p),\sigma_n(p))$ around $(y_i - r(\mathbf{x}_i))/\sigma_n(\mathbf{x}_i)$ and a Taylor's expansion of order one around $(y_i - r(\mathbf{x}_i))/\sigma(\mathbf{x}_i)$ we have that $\sqrt{nh_n^d}(T_n(r(p),\sigma_n(p)) - R_n(\sigma_n(p)) = T_{1n} + T_{2n} + T_{3n} + T_{4n}$ with

$$\begin{aligned}
T_{1n} &= \frac{1}{\sqrt{nh_n^d}}\sum_{i=1}^n W_{in}(p)\Psi'\left(\frac{y_i - r(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}\right)\frac{r(\mathbf{x}_i) - r(p)}{\sigma_n(p)} \\
T_{2n} &= \frac{1}{2{\sigma_n}^2(p)\sqrt{nh_n^d}}\sum_{i=1}^n W_{in}(p)\Psi''(\delta_i)(r(\mathbf{x}_i) - r(p))(\sigma(\mathbf{x}_i) - \sigma(p))\varepsilon_i \\
T_{3n} &= \frac{1}{2{\sigma_n}^2(p)\sqrt{nh_n^d}}\sum_{i=1}^n W_{in}(p)\Psi''(\delta_i)(r(\mathbf{x}_i) - r(p))(\sigma(p) - \sigma_n(p))\varepsilon_i \\
T_{4n} &= \frac{1}{2{\sigma_n}^2(p)\sqrt{nh_n^d}}\sum_{i=1}^n W_{in}(p)\Psi''(\xi_i)(r(\mathbf{x}_i) - r(p))^2
\end{aligned}$$

where $\varepsilon_i = (y_i - r(\mathbf{x}_i))/\sigma(\mathbf{x}_i)$, δ_i and ξ_i stand for intermediate points. Using *A*1, *A*2, *A*4, the fact that $\int K(\|\mathbf{u}\|)u_i u_j\ d\mathbf{u} = 0$ if $i \neq j$ and $\int K(\|\mathbf{u}\|)u_i^2\ d\mathbf{u} = \int K(\|\mathbf{u}\|)u_j^2\ d\mathbf{u}$; it is easy to see that,

$$E(\sigma_n(p)T_{1n}) = \frac{\sqrt{nh_n^{d+4}}}{2}\int \Psi'(u)dF_0(u)\left\{\sum_{i,j=1}^d \frac{\partial^2 \phi\circ\psi^{-1}}{\partial u_i \partial u_j}|_{u=0}\int K(\|\mathbf{u}\|)u_i u_j\ d\mathbf{u} + o(h_n^3)\right\}$$

and $\mathrm{Var}(\sigma_n(p)T_{1n}) \to 0$. Then, $T_{1n} \xrightarrow{p} \beta^{(d+4)/2}\int K(\|u\|)u_1^2 d\mathbf{u}\, L(p)\ \int \Psi'(u)dF_0(u)/(2\sigma(p))$. Finally, *A*1, *A*2, *A*4 and the consistency of the scale estimator entail that $T_{in} \xrightarrow{p} 0$ for i=2,3,4, concluding the proof. □

Proof of Theorem 4.2. Denote $b_n = h_n^d/(1 + zk_n^{-1/2})$, then $P(\sqrt{k_n}(h_n^d/H_n^d - 1) \leq z) = P(H_n^d \geq b_n)$.

Defined Bernoulli random variables Z_i such that $Z_i = 1$ when $d_g(p, \mathbf{x}_i) \leq b_n^{1/d}$ and $Z_i = 0$ elsewhere.Then, we have that $P(H_n^d \geq b_n) = P(\sum_{i=1}^n Z_i \leq k_n)$. Let $q_n = P(d_g(p, \mathbf{x}_i) \leq b_n^{1/d})$ the expected value of Z_i, therefore

$$P\left(\sum_{i=1}^n Z_i \leq k_n\right) = P\left(\frac{1}{\sqrt{nq_n}} \sum_{i=1}^n (Z_i - E(Z_i)) \leq \frac{1}{\sqrt{nq_n}}(k_n - nq_n)\right).$$

It's easy to see that $q_n \to 0$ and $nq_n \to \infty$ as $n \to \infty$, then using the Lindeberg Central Limit Theorem we easily obtain that $(nq_n)^{-1/2} \sum_{i=1}^n (Z_i - E(Z_i))$ is asymptotically normal with mean zero and variance one. Hence, it is enough to show that $(nq_n)^{-1/2}(k_n - nq_n) \xrightarrow{p} z + b_1(p)$.

Denote by $\mu_n = n \int_{B_{b_n^{1/d}}(p)} (f(q) - f(p)) d\nu_g(q)$. Note that $\mu_n = n\ q_n - w_n$ with $w_n = n\ f(p) V(B_{b_n^{1/d}}(p))$ and $V(B_r(p))$ the volume of geodesic ball centered at p with radius r and thus,

$$\frac{1}{\sqrt{nq_n}}(k_n - nq_n) = w_n^{-1/2}(k_n - w_n)\left(\frac{w_n}{w_n + \mu_n}\right)^{1/2} + \frac{\mu_n}{w_n^{1/2}}\left(\frac{w_n}{w_n + \mu_n}\right)^{1/2}.$$

We will prove that

i) $\dfrac{\mu_n}{w_n} \to 0$

ii) $w_n^{-1/2}(k_n - w_n) \to z + \beta^{\frac{d+4}{2}} \frac{\tau}{6d+12} (f(p)\mu(\mathcal{V}_1))^{1/2}$

iii) $\dfrac{\mu_n}{w_n^{1/2}} \to \dfrac{\beta^{\frac{d+4}{2}}}{(f(p)\mu(\mathcal{V}_1))^{1/2}} \displaystyle\int_{\mathcal{V}_1} u_1^2\ d\mathbf{u}\ L_1(p)$

i) Let ψ some exponential chart induced by an orthonormal basis of T_pM. Then, we note that

$$\frac{1}{\mu(\mathcal{V}_{b_n^{1/d}})} \int_{B_{b_n^{1/d}}(p)} f(q) d\nu_g(q) = \frac{1}{\mu(\mathcal{V}_{b_n^{1/d}})} \int_{\mathcal{V}_{b_n^{1/d}}} f \circ \psi^{-1}(\mathbf{u}) \theta_p \circ \psi^{-1}(\mathbf{u}) d\mathbf{u}$$

Therefore, the Lebesgue's Differentiation Theorem and the fact that $\dfrac{V(B_{b_n^{1/d}}(p))}{\mu(\mathcal{V}_{b_n^{1/d}})} \to 1$ imply i).

ii) From Gray and Vanhecke (1979), we have that

$$V(B_r(p)) = r^d \mu(\mathcal{V}_1)(1 - \frac{\tau}{6d+12} r^2 + O(r^4))$$

with τ the Ricci's curvature. Hence, we obtain that

$$\begin{aligned} w_n^{-1/2}(k_n - w_n) &= \frac{w_n^{-1/2}\ k_n\ z\ k_n^{-1/2}}{1 + z k_n^{-1/2}} + \frac{w_n^{-1/2}\ \tau b_n^{2/d}\ k_n}{(6d+12)(1 + z k_n^{-1/2})} + w_n^{-1/2}\ k_n\ O(b_n^{4/d}) \\ &= A_n + B_n + C_n. \end{aligned}$$

Note that $A_n = \frac{z}{1+zk_n^{-1/2}} \left(\frac{k_n}{nf(p)b_n\mu(\mathcal{V}_1)}\right)^{1/2} \left(\frac{b_n\mu(\mathcal{V}_1)}{V(B_{b_n^{1/d}}(p))}\right)^{1/2}$. It's easy to see that $A_n \to z$. On the other hand, as

$$w_n^{-1/2}\, b_n^{2/d}\, k_n = \frac{k_n n^{-1/2} b_n^{2/d-1/2}}{(f(p)\mu(\mathcal{V}_1))^{1/2}} \left(\frac{b_n\mu(\mathcal{V}_1)}{V(B_{b_n^{1/d}}(p))}\right)^{1/2},$$

$A7$ implies that $B_n \to \tau\, \beta^{(d+4)/2}/(6d+12)\, (f(p)\mu(\mathcal{V}_1))^{1/2}$. A similar argument shows that $C_n \to 0$. In order to prove iii), we use a second Taylor expansion that leads to,

$$\begin{aligned}
\int_{B_{b_n^{1/d}}(p)} (f(q)-f(p))d\nu_g(q) &= \sum_{i=1}^{d} \frac{\partial f\circ\psi^{-1}}{\partial u_i}|_{u=0} b_n^{1+1/d} \int_{\mathcal{V}_1} u_i\, \theta_p\circ\psi^{-1}(b_n^{1/d}\mathbf{u})\, d\mathbf{u} \\
&+ \sum_{i,j=1}^{d} \frac{\partial^2 f\circ\psi^{-1}}{\partial u_i \partial u_j}|_{u=0} b_n^{1+2/d} \int_{\mathcal{V}_1} u_i u_j\, \theta_p\circ\psi^{-1}(b_n^{1/d}\mathbf{u})\, d\mathbf{u} \\
&+ O(b_n^{1+3/d}).
\end{aligned}$$

Using again a Taylor expansion on $\theta_p\circ\psi^{-1}(\cdot)$ at 0 we have that

$$\begin{aligned}
\int_{B_{b_n^{1/d}}(p)} (f(q)-f(p))d\nu_g(q) &= b_n^{1+2/d} \int_{\mathcal{V}_1} u_1^2\, d\mathbf{u} \sum_{i=1}^{d} \frac{\partial f\circ\psi^{-1}}{\partial u_i}|_{u=0} \frac{\partial \theta_p\circ\psi^{-1}}{\partial u_i}|_{u=0} \\
&+ b_n^{1+2/d} \int_{\mathcal{V}_1} u_1^2\, d\mathbf{u} \sum_{i,j=1}^{d} \frac{\partial^2 f\circ\psi^{-1}}{\partial u_i \partial u_j}|_{u=0} + O(b_n^{1+3/d}) \\
&= b_n^{1+2/d} \int_{\mathcal{V}_1} u_1^2\, d\mathbf{u}\, L_1(p) + O(b_n^{1+3/d})
\end{aligned}$$

and $A7$ allow to conclude iii).□

References

[1] Agostinelli, C. (2007), 'Robust estimation for circular data', *Computational Statistics and Data Analysis* ,51, 5867-5875.

[2] Berger, M., Gauduchon, P., and Mazet, E. (1971), *Le spectre d' une variété Riemannienne*. Springer-Verlag.

[3] Besse, A. (1978), *Manifolds all of whose geodesics are closed*. Springer-Verlag.

[4] Boente, G., and Fraiman, R. (1989), 'Robust nonparametric regression estimation', *Journal of Multivariate Analysis*, 29, 180-198.

[5] Boente, G., and Fraiman, R. (1990), 'Asymptotic distribution of robust estimators for nonparametric models from mixing processes', *Annals of Statistics*, 18, 891-906.

[6] Boente, G., and Fraiman, R. (1991), 'Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes', *International Statistical Review*, **59**, 355-372.

[7] Boente, G., Fraiman, R., and Meloche, J. (1997), 'Robust plug-in bandwidth estimators in nonparametric regression', *Journal of Statistical Planning and Inference*, 57, 109-142.

[8] Boothby, W. M. (1975), *An introduction to differentiable manifolds and Riemannian geometry*, New York: Academic Press.

[9] Cantoni, E., and Ronchetti, E. (2001), 'Resistant selection of the smoothing parameter for smoothing splines', *Statistics and Computing*, 11, 141-146.

[10] Cleveland, W. (1979), 'Robust locally weighted regression and smoothing scatter plots', *Jornal of the American Statistical Association*, 74, 829-836.

[11] Collomb, G. (1980), 'Estimation de la regression par la méthode des k points les plus proches avec noyau: Quelques propiétés de convergence ponctuelle', *Lecture Notes in Mathematics*, 821 , 159-175.

[12] Devroye, L., and Wagner, T.J. (1977), 'The strong uniform consistency of nearest neighbor density estimates', *Annals of Statistics*,3, 536-540.

[13] Do Carmo, M. (1988), *Geometria Riemaniana*. Proyecto Euclides, IMPA. 2da edición.

[14] Fischer, N.I., Lewis, T., and Embleton, B.J.J. (1993), *Statistical Analysis of Spherical Data*, Cambridge: University Press.

[15] Gray, A., and Vanhecke, L. (1979), 'Riemannian geometry as determined by the volumes of small geodesic balls', *Acta Math.*, 142, 157-198.

[16] Ko, D., and Guttorp, P. (1988), 'Robustness of estimators for directional data', *Annals of Statistics*, 16, 609-618.

[17] Hall, P., Watson, G.S., and Cabrera, J. (1987), 'Kernel density estimation with spherical data', *Biometrika*, 74, 751762.

[18] Härdle, W. (1984), 'Robust regression function estimation', *Journal of Multivariate Analysis*, 14, 169-180.

[19] Härdle, W. (1990), *Applied nonparametric regression*, Cambridge: University Press.

[20] Härdle, W., and Tsybakov, A.B. (1988), 'Robust nonparametric regression with simultaneous scale curve estimation', *Annals of Statistics*, 16, 120-135.

[21] He, X. (1992), 'Robust statistics of directional data: a survey', *Nonparametric Statistics and Related Topics*, eds A. M. Saleh, 87-95, New York: North Holland.

[22] He, X., Zhu, Z., and Fung, W. (2002), 'Estimation in a semiparametric model for longitudinal data with unspecified dependence structure', *Biometrika*, 89, 579-590.

[23] Hendriks, H. (1990), Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions', *Annals of Statistics*, 18, 832849.

[24] Hendriks, H., Janssen, J.H.M., and Ruymgaart, F.H. (1993), 'Strong uniform convergence of density estimators on compact Euclidean manifolds', *Statistics and Probability Letters*, 16, 305311.

[25] Lee, J.M., and Ruymgaart, F.H. (1996), 'Nonparametric curve estimation on Stiefel manifolds', *Journal of Nonparametric Statistics*, 6, 5768.

[26] Leung, D. (2005), 'Cross-validation in nonparametric regression with outliers', *Annals of Statistics*, 33, 2291-2310.

[27] Mardia, K., and Jupp, P. (2000), *Directional Data*, New York: Wiley.

[28] Maronna, R., Martin, D. and Yohai, V. (2006), *Robust Statistics: Theory and Methods*, New York: Wiley.

[29] Nadaraya, E. A. (1964), 'On estimating regression', *Theory Probality and Applications*, 9, 141-142.

[30] Pelletier, B. (2005), 'Kernel density estimation on Riemannian maniflods', *Statistics and Probability Letters*, 73, 297-304.

[31] Pelletier, B. (2006), 'Nonparametric regression estimation on closed Riemannian manifolds', *Journal of Nonparametric Statistics*, 18, 57-67.

[32] Tsybakov, A.B. (1982), 'Robust estimates of a function' *Peredaci Informacii*, 18, 39-52. English translation in 'Problems Inform Transmission', 18, 190-201.

[33] Wang, F., and Scott, D. (1994), 'The L_1 method for robust nonparametric regression', *Journal of the American Statistical Association*, 89, 65-76.

[34] Watson, G.S. (1964), 'Smooth regression analysis', *Sankhya Ser. A*, 26, 359-392.

Guillermo Henry
Departamento de Matemática, FCEyN, Universidad de Buenos Aires
Ciudad Universitaria, Pabellón I, Buenos Aires, C1428EHA, Argentina
e-mail address: ghenry@dm.uba.ar

Daniela Rodriguez
Instituto de Cálculo, FCEyN, Universidad de Buenos Aires
Ciudad Universitaria, Pabellón II, Buenos Aires, C1428EHA, Argentina
e-mail address: drodrig@dm.uba.ar