Nonparametric regression estimation on closed Riemannian manifolds

Bruno Pelletier
Institut de Mathématiques et de Modélisation de Montpellier
UMR CNRS 5149, Equipe de Probabilités et Statistique
Université Montpellier II
Place Eugène Bataillon
34095 Montpellier Cedex 5, France.

pelletier@math.univ-montp2.fr

Abstract

The nonparametric estimation of the regression function of a real-valued random variable Y on a random object X valued in a closed Riemannian manifold M is considered. A regression estimator which generalizes kernel regression estimators on Euclidean sample spaces is introduced. Under classical assumptions on the kernel and the bandwidth sequence, the asymptotic bias and variance are obtained, and the estimator is shown to converge at the same L^2-rate as kernel regression estimators on Euclidean spaces.

Index Terms **— Nonparametric regression, Kernel regression, Riemannian manifolds, L^2-convergence.**

1 Introduction

Sample spaces which have a more complex structure than the Euclidean space $\mathbb{R}^d$ arise in a variety of contexts and motivate the adaptation of popular nonparametric density or regression estimation techniques on $\mathbb{R}^d$, such as kernel smoothing (see e.g.,[26, 28] for a review). This includes the case of functional statistics, where the regressors are curves or functions taking values in infinite-dimensional metric spaces or semi-normed vector spaces [11, 10, 8], and the case of random objects valued in manifolds of various kind, such as the circle S^1, the sphere S^2, and the Stiefel manifold.

The circle and the sphere arise as sample spaces in axial and directional statistics. Several specific statistical methodologies have been developed

[23, 29, 20], as well as adaptations of existing nonparametric density or regression estimation techniques, including trigonometric Fourier series [9], kernel smoothing [14, 12], and deconvolution [16, 15, 18]. The analysis of orientation data leads to the Stiefel manifold as the sample space. This manifold is defined as the set of orthonormal m-frames in $\mathbb{R}^d$, and includes the sphere and the orthogonal group as special cases. Prentice [25] introduces an extension of spherical regression to this setting, and Lee and Ruymgaart [22] consider nonparametric regression estimation using caps, i.e. intersections of closed balls in the ambient space with the manifold, following an earlier work on density estimation on compact submanifolds of a Euclidean space [19]. Parallely, Chikuse [6] adapts the kernel method for density estimation on the Stiefel manifold. For a larger class of manifolds, Hendriks [17] generalizes the method of Fourier series for density estimation on closed (i.e., compact and without boundary) Riemannian manifolds, using an expansion on the eigenfunctions of the Laplace-Beltrami operator.

In this paper, we consider the nonparametric estimation of a regression function on a closed Riemannian manifold by using a generalization of the kernel method, that we recently introduced in [24] for density estimation. The idea is to build an analogue of a kernel on M by using a positive function of the geodesic distance on M, which is then normalized by the volume density function of M to account for curvature. This kind of estimator has interesting properties: i) its expression is consistent with standard kernel estimators on Euclidean spaces, i.e. when (M, g) is $(\mathbb{R}^d, \delta)$ the estimator expression reduces to the one of a standard kernel estimator on $(\mathbb{R}^d, \delta)$; ii) it converges at the same rate as the Euclidean kernel estimator ; and iii) provided the bandwidth is small enough, the kernel is centered on the observation, i.e. the observation is an intrinsic mean of its associated kernel.

In Section 2 we define the kernel regression estimator of a real-valued random variable Y on a random object X valued in M, and we introduce several assumptions on the kernel and the bandwidth sequence. For materials on differential and Riemannian geometry, we refer the reader to [5, 21, 13, 30]. In Section 3, the asymptotic pointwise bias and variance are obtained under regular assumptions on the bandwidth sequence. This leads immediately to the asymptotic pointwise mean squared error, using the decomposition in terms of squared bias and variance [27]. Finally, we give an integrated version of this result.

2 Definitions and assumptions

Let (M, g) be a closed (i.e. compact and without boundary) Riemannian manifold of dimension d. We assume that (M, g) is complete, i.e., (M, d_g) is a complete metric space, where d_g denotes the Riemannian distance.

Let X be a random object on M, i.e. a measurable map from a probability space $(\Omega, \mathcal{A}, P)$ to $(M, \mathcal{B})$, where $\mathcal{B}$ denotes the Borel σ-field of M. We assume that the image measure of P by X is absolutely continuous with respect to the Riemannian volume measure v_g, and that it admits an a.s. continuous density f on M. Let Y be a real-valued random variable, and let $r : M \ni p \mapsto r(p) = \mathbb{E}[Y|X = p] \in \mathbb{R}$ be the regression function of Y on X. Let $s : M \to \mathbb{R}$ be the function defined by $s(p) = r(p)f(p)$, for all $p \in M$. We consider the nonparametric estimation of the regression function r based on an i.i.d. sample $(X_1, Y_1), ..., (X_n, Y_n)$ of pairs valued in $M \times \mathbb{R}$ and of the same law as (X, Y).

Let $K : \mathbb{R}_+ \to \mathbb{R}$ be a positive and continuous map such that:

(K1) $\int_{\mathbb{R}^d} K(\|u\|)d\lambda(u) = 1$,

(K2) $\int_{\mathbb{R}^d} uK(\|u\|)d\lambda(u) = 0$,

(K3) $\int_{\mathbb{R}^d} \|u\|^2 K(\|u\|)d\lambda(u) < \infty$,

(K4) $suppK = [0; 1]$,

where λ denotes the Lebesgue measure on $\mathbb{R}^d$.

For p and q two points of M, let $\theta_p(q)$ be the volume density function on M roughly defined by [1, p. 154]:

$$\theta_p : q \mapsto \theta_p(q) = \frac{\mu_{exp_p^* g}}{\mu_{g_p}}(exp_p^{-1}(q)),$$

i.e., the quotient of the canonical measure of the Riemannian metric $exp_p^* g$ on $T_p(M)$ (pullback of g by the map exp_p) by the Lebesgue measure of the Euclidean structure g_p on $T_p(M)$. This definition makes sense for q in a neighborhood of p, yet the volume density function may be defined globally by recursing to Jacobi fields [30, p. 219]. In terms of geodesic normal coordinates at p, $\theta_p(q)$ equals the square root of the determinant of the metric g expressed in these coordinates at $exp_p^{-1}(q)$, and for p and q in a normal neighborhood U of M, we have $\theta_p(q) = \theta_q(p)$ [30, p. 221].

We define the estimators f_n and s_n of f and s respectively by:

$$f_n(p) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n^d}\frac{1}{\theta_{X_i}(p)}K\left(\frac{d_g(p,X_i)}{h_n}\right), \tag{1}$$

$$s_n(p) = \frac{1}{n}\sum_{i=1}^{n}Y_i\frac{1}{h_n^d}\frac{1}{\theta_{X_i}(p)}K\left(\frac{d_g(p,X_i)}{h_n}\right), \tag{2}$$

for all $p \in M$, and where h_n is the bandwidth. Finally, we define the regression estimator r_n of r by:

$$r_n(p) = \frac{s_n(p)}{f_n(p)} \tag{3}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}Y_i\frac{1}{h_n^d}\frac{1}{\theta_{X_i}(p)}K\left(\frac{d_g(p,X_i)}{h_n}\right)}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n^d}\frac{1}{\theta_{X_i}(p)}K\left(\frac{d_g(p,X_i)}{h_n}\right)}, \tag{4}$$

if $f_n(p) \neq 0$, and by $r_n(p) = 0$ otherwise.

To allow for simple computations in normal charts, we impose the following condition on the bandwidth:

(A0) $h_n < inj_g(M)$,

where $inj_g(M)$ is the injectivity radius of M. Note that since M is compact, the injectivity radius of M is strictly positive, by the theorem of Whitehead. Under this condition, there exists, for each $p \in M$, a normal coordinate neighborhood at p containing the closed ball $B_M(p, h_n)$ in M.

As mentioned in the Introduction, these estimators have interesting properties. First, when (M, g) is the Euclidean space $(\mathbb{R}^d, \delta)$, we have $\theta_p(q) = 1, \forall p, q \in M$. Hence in that case, f_n, s_n, and r_n reduce to standard kernel estimators with isotropic kernels. Second, the kernels introduced in Eq. (1) and Eq. (2) are centered on the observation, in the sense that X_i is an intrinsic mean of the probability measure associated with the density $\frac{1}{h_n^d}\frac{1}{\theta_{X_i}(p)}K\left(\frac{d_g(p,X_i)}{h_n}\right)$. An intrinsic mean of a probability measure μ on a Riemannian manifold M is a minimizer of the energy functional [21, 2, 3]

$$F(p) = \int_M d_g^2(q,p)d\mu(q).$$

This statement holds whenever the bandwidth is small enough. More specifically, we have the following result [24].

Proposition 2.1 *Let q be an arbitrary point of M. Let $\mathcal{K} : \cup_{p\in M} G_p^2 M \to \mathbb{R}$ be the sectional curvature of M, where $G_p^2 M$ is the set of 2-dimensional subspaces of $T_p(M)$. Let $\delta = \sup \mathcal{K}$ be the supremum of the sectional curvatures in M. Let μ be a probability measure on M, absolutely continuous w.r.t. the Riemannian volume measure, and with a density $f_q(p)$ defined by*

$$f_q(p) = \frac{1}{h^d} \frac{1}{\theta_q(p)} K \left(\frac{d_g(p, q)}{h} \right),$$

where K satisfies (K1)-(K4), where $h < \min\{\frac{inj_g(M)}{2}, \frac{\pi}{4\sqrt{\delta}}\}$, and where we set $\frac{\pi}{4\sqrt{\delta}} = +\infty$ when $\delta \leq 0$. Then q is the unique intrinsic mean of μ.

Example 2.2 Let the manifold (M, g) be $(\mathbb{R}^d, \delta)$, where δ denotes the usual Euclidean metric, and consider the canonical identification of the tangent space $T_p M$ at some point p of $(\mathbb{R}^d, \delta)$, with $\mathbb{R}^d$. Note that any two tangent spaces at different points are also canonically identified. This defines trivialy a normal chart, the domain of which is the entire manifold. In this chart, the components of the metric tensor form the identity matrix, and for all $p, q \in \mathbb{R}^d$, we have $\theta_p(q) = 1$. Then in a system $(x^1, ..., x^d)$ of normal coordinates, and denoting by $\|x\|$ the length of x considered as a vector of $T_0\mathbb{R}^d$, we obtain that:

$$r_n(x) = \frac{\frac{1}{n} \sum_{i=1}^n Y_i \frac{1}{h_n^d} K \left(\frac{\|x - X_i\|}{h_n} \right)}{\frac{1}{n} \sum_{i=1}^n \frac{1}{h_n^d} K \left(\frac{\|x - X_i\|}{h_n} \right)},$$

which is the expression of a standard kernel regression estimator with isotropic kernels. The bandwidth may be arbitrarily large since $inj_\delta \left(\mathbb{R}^d\right) = \infty$. The concept of intrinsic mean used in Proposition 2.1 reduces to the one of the standard mean of a random vector, which naturally is unique, and for all $h > 0$ since $(\mathbb{R}^d, \delta)$ has constant sectional curvature equal to 0.

Example 2.3 Consider the surface in $\mathbb{R}^3$ defined on the unit disk D^1 by

$$(u, v) \mapsto \left(\frac{u}{\rho} \sin \rho, \frac{v}{\rho} \sin \rho, \cos \rho \right),$$

where $\rho = \sqrt{u^2 + v^2}$, which is the upper-hemisphere of S^2 considered as a submanifold of $\mathbb{R}^3$. We now illustrate the construction of a kernel centered at the point p of the surface of coordinates $(0, 0, 1)$ in $\mathbb{R}^3$. Note that (u, v) is a system of normal coordinates at p : $\exp_p((u, v))$ is the point on the geodesic issued from p in the direction of (u, v), and located at the distance ρ from p.

The coordinates vector fields are given by:

$$\begin{aligned}\partial_u &= \left(\frac{u^2}{\rho^2}\cos\rho + \frac{v^2}{\rho^3}\sin\rho, \frac{uv}{\rho^2}\cos\rho - \frac{uv}{\rho^3}\sin\rho, -\frac{u}{\rho}\sin\rho\right),\\ \partial_v &= \left(\frac{uv}{\rho^2}\cos\rho - \frac{uv}{\rho^3}\sin\rho, \frac{v^2}{\rho^2}\cos\rho + \frac{u^2}{\rho^3}\sin\rho, -\frac{v}{\rho}\sin\rho\right),\end{aligned}$$

when $\rho \neq 0$ and by $\partial_u = (1,0,0)$ and $\partial_v = (0,1,0)$ otherwise. In this chart, the components of the metric tensor are given by: $g_{p,11} = \frac{u^2}{\rho^2} + \frac{v^2}{\rho^4}\sin^2\rho$, $g_{p,22} = \frac{v^2}{\rho^2} + \frac{u^2}{\rho^4}\sin^2\rho$, and $g_{p,12} = \frac{uv}{\rho^2} - \frac{uv}{\rho^4}\sin^2\rho$, when $\rho \neq 0$, and by $g_{p,ij} = \delta_{ij}$ when $\rho = 0$. Then in this chart, $|g_p|(u,v) = \frac{\sin^2\rho}{\rho^2}$ when $\rho \neq 0$, and $|g_p|(0,0) = 1$. Note that it is a special case that $|g_p|$ only depends on the distance from p. Then, the expression of a kernel centered at p is $f_p(u,v) = \frac{1}{h^2}\frac{\rho}{|\sin\rho|}K\left(\frac{\rho}{h}\right)$.

3 Convergence properties

In this section, we study the L^2-convergence of r_n. First we derive asymptotic expansions of the pointwise bias and variance of f_n and s_n in Theorem 3.1 and Theorem 3.2. Then we formulate results on the asymptotic pointwise bias and variance of r_n in Theorem 3.3 and Theorem 3.4, leading to the result in Theorem 3.5 on the asymptotic pointwise mean squared error, the proofs of whose mainly follow the lines devised by Collomb [7] (see also [26]). Finally, we give an integrated version of this result. For the pointwise convergence results at a point $p \in M$, we will need the following assumptions:

(A1) $\lim h_n = 0$ and $\lim nh_n^d = \infty$;

(A2) Y is bounded;

(A3) f is two-times continuously differentiable at p and $f(p) > 0$;

(A4) r is two-times continuously differentiable at p;

(A5) $\varphi(p) \stackrel{\text{def}}{=} \mathbb{E}\{Y^2|X=p\}$ is continuous at p.

In all of the following, we shall use the Einstein summation convention. The closed ball in (M,g) of center p and of radius h will be denoted by $B_M(p,h)$, and the closed ball in $(\mathbb{R}^d,\delta)$ centered at the origin and of radius h will be denoted by $B(h)$.

Theorem 3.1 (Pointwise bias and variance of f_n.) *Let f_n be the density estimator of f defined by Eq. (1). Assume (K1)-(K4), (A0), (A1), and (A3) hold. Then the asymptotic pointwise bias and variance of f_n are given by:*

$$\mathbb{E}f_n(p) - f(p) = \frac{1}{2}\left((\nabla^2 f(p))_{ij}\int_{B(1)} K(\|u\|)u^iu^j du\right)h_n^2 + o(h_n^2), \quad (5)$$

$$Var\left(f_n(p)\right) = \frac{1}{nh_n^d} f(p)\int_{B(1)} K^2\left(\|u\|\right)du + o\left(\frac{1}{nh_n^d}\right). \quad (6)$$

Proof

Proof of Eq. (5):

$$\begin{aligned}\mathbb{E}f_n(p) &= \mathbb{E}\frac{1}{h_n^d}\frac{1}{\theta_X(p)}K\left(\frac{d_g(p,X)}{h_n}\right)\\ &= \int_M \frac{1}{h_n^d}\frac{1}{\theta_q(p)}K\left(\frac{d_g(p,q)}{h_n}\right)f(q)dv_g(q).\end{aligned}$$

The integral may be taken over $B_M(p,h_n)$, and using the relation $\theta_q(p) = \theta_p(q)$ on $B_M(p,h_n)$, we have:

$$\begin{aligned}\mathbb{E}f_n(p) - f(p) &= \int_{B_M(p,h_n)} \frac{1}{h_n^d}\frac{1}{\theta_p(q)}K\left(\frac{d_g(p,q)}{h_n}\right)f(q)dv_g(q) - f(p)\\ &= \int_{B_M(p,h_n)} \frac{1}{h_n^d}\frac{1}{\theta_p(q)}K\left(\frac{d_g(p,q)}{h_n}\right)\left(f(q)-f(p)\right)dv_g(q).\end{aligned}$$

Now we let $x = \exp_p^{-1}(q)$ and we take a covariant Taylor expansion of f around p at the order two, and with the remainder denoted by $R(p,x)$:

$$\begin{aligned}\mathbb{E}f_n(p) - f(p) &= \int_{B_M(p,h_n)} \frac{1}{h_n^d}\frac{1}{\theta_p(q)}K\left(\frac{d_g(p,q)}{h_n}\right)\\ &\quad \left((\nabla f(p))_i x^i + \frac{1}{2}(\nabla^2 f(p))_{ij}x^ix^j + R(p,x)\right)dv_g(q)\\ &= \int_{B_M(p,h_n)} \frac{1}{h_n^d}\frac{1}{\theta_p(q)}K\left(\frac{d_g(p,q)}{h_n}\right)\left(\frac{1}{2}(\nabla^2 f(p))_{ij}x^ix^j + R(p,x)\right)dv_g(q)\\ &= \frac{1}{2}\left((\nabla^2 f(p))_{ij}\int_{B(1)} K(\|u\|)u^iu^j du\right)h_n^2 + o(h_n^2).\end{aligned}$$

Proof of Eq. (6):
The variance of $f_n(p)$ may be expressed as:

$$\begin{aligned}
Var\left(f_n(p)\right) &= \frac{n}{(nh_n^d)^2} Var\left(\frac{1}{\theta_X(p)} K\left(\frac{d_g(p,X)}{h_n}\right)\right) \\
&= \frac{1}{nh_n^d}\mathbb{E}\left\{\frac{1}{h_n^d}\frac{1}{\theta_X^2(p)}K^2\left(\frac{d_g(p,X)}{h_n}\right)\right\} - \frac{1}{n}\mathbb{E}\left\{\frac{1}{h_n^d}\frac{1}{\theta_X(p)}K\left(\frac{d_g(p,X)}{h_n}\right)\right\}^2 \\
&= \frac{1}{nh_n^d}\mathbb{E}\left\{\frac{1}{h_n^d}\frac{1}{\theta_X^2(p)}K^2\left(\frac{d_g(p,X)}{h_n}\right)\right\} - \frac{1}{n}\left(f(p)+o(1)\right)^2.
\end{aligned}$$

Now we compute the expectation:

$$\begin{aligned}
\mathbb{E}\left\{\frac{1}{h_n^d}\frac{1}{\theta_{X_1}^2(p)}K^2\left(\frac{d_g(p,X_1)}{h_n}\right)\right\} &= \int_M \frac{1}{h_n^d}\frac{1}{\theta_q^2(p)}K^2\left(\frac{d_g(p,q)}{h_n}\right) f(q)dv_g(q) \\
&= \int_{B_M(p,h_n)} \frac{1}{h_n^d}\frac{1}{\theta_p^2(q)}K^2\left(\frac{d_g(q,p)}{h_n}\right) f(q)dv_g(q) \\
&= \left(f(p)+o(1)\right)\int_{B(h_n)} \frac{1}{h_n^d}\frac{1}{\sqrt{|g_p(x)|}}K^2\left(\frac{\|x\|}{h_n}\right) dx \\
&= f(p)\int_{B(1)} K^2\left(\|u\|\right) du + o(1),
\end{aligned}$$

where we have used the fact that $\sqrt{|g_p(x)|} = 1 + o(1)$ in a normal chart at p. Finally, we obtain:

$$Var\left(f_n(p)\right) = \frac{1}{nh_n^d} f(p)\int_{B(1)} K^2\left(\|u\|\right) du + o\left(\frac{1}{nh_n^d}\right)$$

□

Theorem 3.2 (Pointwise bias and variance of s_n.) *Let s_n be the estimator of s defined by Eq. (2). Assume (K1)-(K4), (A0), (A1), (A3)-(A5) hold. Then the asymptotic pointwise bias and variance of s_n are given by:*

$$\mathbb{E}s_n(p) - s(p) = \frac{1}{2}\left((\nabla^2 s(p))_{ij}\int_{B(1)} K(\|u\|)u^i u^j du\right) h_n^2 + o(h_n^2), \quad (7)$$

$$Var\left(s_n(p)\right) = \frac{1}{nh_n^d} f(p)\varphi(p)\int_{B(1)} K^2\left(\|u\|\right) du + o\left(\frac{1}{nh_n^d}\right). \quad (8)$$

Proof

The proofs of Eq. (7) and Eq. (8) follow the same lines as the proofs of Eq. (5) and Eq. (6) in Theorem 3.1.

Proof of Eq. (7):
We have:

$$\begin{aligned}
\mathbb{E}s_n(p) &= \mathbb{E}Y\frac{1}{h_n^d}\frac{1}{\theta_X(p)}K\left(\frac{d_g(p,X)}{h_n}\right)\\
&= \int_M r(q)\frac{1}{h_n^d}\frac{1}{\theta_q(p)}K\left(\frac{d_g(p,q)}{h_n}\right)f(q)dv_g(q)\\
&= \int_M s(q)\frac{1}{h_n^d}\frac{1}{\theta_q(p)}K\left(\frac{d_g(p,q)}{h_n}\right)dv_g(q).
\end{aligned}$$

Now, by using a covariant Taylor expansion of s around p, the desired result follows.

Proof of Eq. (8):
We have:

$$\begin{aligned}
Var\left(s_n(p)\right) &= \frac{n}{nh_n^d}Var\left(Y\frac{1}{\theta_X(p)}K\left(\frac{d_g(p,X)}{h_n}\right)\right)\\
&= \frac{1}{nh_n^d}\mathbb{E}\left\{Y^2\frac{1}{\theta_X^2(p)}\frac{1}{h_n^d}K^2\left(\frac{d_g(p,X)}{h_n}\right)\right\} - \frac{1}{n}\mathbb{E}\left\{Y\frac{1}{\theta_X(p)}\frac{1}{h_n^d}K\left(\frac{d_g(p,X)}{h_n}\right)\right\}^2\\
&= \frac{1}{nh_n^d}\int_M \varphi(q)\frac{1}{\theta_q^2(p)}\frac{1}{h_n^d}K^2\left(\frac{d_g(p,q)}{h_n}\right)f(q)dv_g(q) - \frac{1}{n}\left(s(p)+o(1)\right)^2,
\end{aligned}$$

from which the result easily follows. □

Theorem 3.3 (Pointwise bias of r_n.) *Let r_n be the regression estimator defined by Eq. (3). Assume (K1)-(K4) and (A0)-(A4) hold. Then the asymptotic pointwise bias of r_n is given by*

$$\begin{aligned}
\mathbb{E}r_n(p) - r(p) &= \frac{1}{2f(p)}\int_{B(1)} K(\|u\|)u^i u^j du\left(\left(\nabla^2 s(p)\right)_{ij} - r(p)\left(\nabla^2 f(p)\right)_{ij}\right)h_n^2\\
&\quad + o(h_n^2) + O\left(\frac{1}{nh_n^d}\right).
\end{aligned}$$

Proof

Using the decomposition

$$\frac{1}{z} = 1-(z-1)+\frac{(z-1)^2}{z},$$

we obtain that

$$\mathbb{E}r_n(p) = \frac{\mathbb{E}s_n(p)}{\mathbb{E}f_n(p)} - \frac{A}{(\mathbb{E}f_n(p))^2} + \frac{B}{(\mathbb{E}f_n(p))^2}, \tag{9}$$

where

$$\begin{aligned} A &= \mathbb{E}\left\{s_n(p)(f_n(p) - \mathbb{E}f_n(p))\right\}, \\ B &= \mathbb{E}\{(f_n(p) - \mathbb{E}f_n(p))^2 r_n(p)\}. \end{aligned}$$

<u>Expression of A</u> :

Using the independence and the equidistribution of the pairs (X_i, Y_i), we obtain that

$$\begin{aligned} A &= Cov(f_n(p), s_n(p)) \\ &= \frac{1}{nh_n^d}\mathbb{E}\left\{Y\frac{1}{\theta_X^2(p)}\frac{1}{h_n^d}K^2\left(\frac{d_g(p,X)}{h_n}\right)\right\} \\ &\quad -\frac{1}{n}\mathbb{E}\left\{Y\frac{1}{\theta_X(p)}\frac{1}{h_n^d}K\left(\frac{d_g(p,X)}{h_n}\right)\right\}\mathbb{E}\left\{\frac{1}{\theta_X(p)}\frac{1}{h_n^d}K\left(\frac{d_g(p,X)}{h_n}\right)\right\} \\ &\overset{def}{=} \frac{1}{nh_n^d}E_1 - \frac{1}{n}E_2E_3. \end{aligned}$$

Now we compute the three expectations E_1, E_2, and E_3. We have:

$$\begin{aligned} E_1 &= \mathbb{E}\left\{Y\frac{1}{\theta_X^2(p)}\frac{1}{h_n^d}K^2\left(\frac{d_g(p,X)}{h_n}\right)\right\} \\ &= \int_{B_M(p,h_n)} s(q)\frac{1}{\theta_p^2(q)}\frac{1}{h_n^d}K^2\left(\frac{d_g(q,p)}{h_n}\right)dv_g(q) \\ &= (s(p)+o(1))\int_{B_M(p,h_n)}\frac{1}{\theta_p^2(q)}\frac{1}{h_n^d}K^2\left(\frac{d_g(q,p)}{h_n}\right)dv_g(q) \\ &= s(p)\int_{B(1)}K^2(\|u\|)\,du + o(1), \end{aligned}$$

where we have used the fact that $\sqrt{|g_p(x)|} = 1 + o(1)$ in a normal chart at p. Using similar computations, we obtain the following relations:

$$\begin{aligned} E_2 &= \mathbb{E}\left\{Y\frac{1}{\theta_X(p)}\frac{1}{h_n^d}K\left(\frac{d_g(p,X)}{h_n}\right)\right\} \\ &= s(p) + o(1); \\ E_3 &= \mathbb{E}\left\{\frac{1}{\theta_X(p)}\frac{1}{h_n^d}K\left(\frac{d_g(p,X)}{h_n}\right)\right\} \\ &= f(p) + o(1). \end{aligned}$$

Thus

$$A=\frac{1}{nh_n^d}\left(s(p)\int_{B(1)}K^2\left(\|u\|\right)du+o(1)\right)+\frac{1}{n}\left(s(p)+o(1)\right)\left(f(p)+o(1)\right),$$

which leads to:

$$A=\frac{1}{nh_n^d}s(p)\int_{B(1)}K^2\left(\|u\|\right)du+o\left(\frac{1}{nh_n^d}\right). \tag{10}$$

Expression of B:

Since Y is bounded and since K is positive, $r_n(p)$ is almost surely bounded by a constant C and thus

$$B\leq C\mathbb{E}\left\{\left(f_n(p)-\mathbb{E}f_n(p)\right)^2\right\}.$$

Using similar computations as above, we obtain that

$$B=O\left(\frac{1}{nh_n^d}\right). \tag{11}$$

Now using equations (5), (7), (9), (10), and (11), we have:

$$\begin{aligned}
\mathbb{E}r_n(p) &= \frac{\mathbb{E}s_n(p)}{\mathbb{E}f_n(p)}+O\left(\frac{1}{nh_n^d}\right)\\
&= \frac{s(p)+\frac{1}{2}\left(\nabla^2 s(p)\right)_{ij}\int_{B(1)}K(\|u\|)u^iu^jduh_n^2+o(h_n^2)}{f(p)+\frac{1}{2}\left(\nabla^2 f(p)\right)_{ij}\int_{B(1)}K(\|u\|)u^iu^jduh_n^2+o(h_n^2)}+O\left(\frac{1}{nh_n^d}\right)\\
&= r(p)-\frac{1}{2}\frac{s(p)}{f^2(p)}\left(\nabla^2 f(p)\right)_{ij}\int K(\|u\|)u^iu^jduh_n^2\\
&\quad +\frac{1}{2}\frac{1}{f(p)}\left(\nabla^2 s(p)\right)_{ij}\int_{B(1)}K(\|u\|)u^iu^jduh_n^2+o(h_n^2)+O\left(\frac{1}{nh_n^d}\right)\\
&= r(p)+\frac{1}{2f(p)}\int_{B(1)}K(\|u\|)u^iu^jdu\left(\left(\nabla^2 s(p)\right)_{ij}-r(p)\left(\nabla^2 f(p)\right)_{ij}\right)h_n^2\\
&\quad +o(h_n^2)+O\left(\frac{1}{nh_n^d}\right).
\end{aligned}$$

□

Theorem 3.4 (Pointwise variance of r_n.) *Let r_n be the regression estimator defined by Eq. (3). Assume (K1)-(K4) and (A0)-(A5) hold. Then the asymptotic pointwise variance of r_n is given by:*

$$Var\left(r_n(p)\right)=\frac{1}{nh_n^d}\frac{\varphi(p)-r^2(p)}{f(p)}\int K^2\left(\|u\|\right)du+o\left(\frac{1}{nh_n^d}\right).$$

Proof

We have [26]:

$$\begin{aligned} Var\left(r_n(p)\right) &= \frac{Var\left(s_n(p)\right)}{\left(\mathbb{E}f_n(p)\right)^2} - 4\frac{\mathbb{E}s_n(p)Cov\left(f_n(p), s_n(p)\right)}{\left(\mathbb{E}f_n(p)\right)^3} \\ &\quad +3Var\left(f_n(p)\right)\frac{\left(\mathbb{E}s_n(p)\right)^2}{\left(\mathbb{E}f_n(p)\right)^4} + o\left(\frac{1}{nh_n^d}\right). \end{aligned}$$

By using equations (5) to (8), and the expression for $Cov(f_n(p), s_n(p))$ given in the proof of Theorem 3.3, we obtain the desired result. □

As a corollary of Theorem 3.3 and Theorem 3.4, we obtain the asymptotic pointwise mean squared error.

Theorem 3.5 (Pointwise mean squared error.) *Let r_n be the regression estimator defined by Eq. (3). Assume (K1)-(K4) and (A0)-(A5) hold. Then the asymptotic quadratic mean of r_n is given by:*

$$\mathbb{E}\left\{r_n(p) - r(p)\right\}^2 = \mu^2(p)h_n^4 + \sigma^2(p)\frac{1}{nh_n^d} + o\left(h_n^4 + \frac{1}{nh_n^d}\right),$$

where

$$\begin{aligned} \mu(p) &= \frac{1}{2f(p)}\int_{B(1)} K(\|u\|)u^i u^j du\left(\left(\nabla^2 s(p)\right)_{ij} - r(p)\left(\nabla^2 f(p)\right)_{ij}\right), \\ \sigma^2(p) &= \frac{1}{nh_n^d}\frac{\varphi(p) - r^2(p)}{f(p)}\int_{B(1)} K^2\left(\|u\|\right)du. \end{aligned}$$

To obtain the mean integrated squared error of r_n:

$$MISE(r_n) = \mathbb{E}\|r_n - r\|_{L^2(M)}^2,$$

we need uniform versions of assumptions (A3)-(A5), i.e.:

(A'3) f is two-times continuously differentiable on M and $\inf_M f > 0$;

(A'4) r is two-times continuously differentiable on M;

(A'5) φ is continuous on M.

As M is compact, the functions r and f are uniformly continuous over M. By inspecting the proofs of the previous results, one easily checks that the o-symbol in Theorem 3.5 is uniform over M, which leads to the following asymptotic expression of $MISE(r_n)$.

Theorem 3.6 (MISE of r_n.) *Let r_n be the regression estimator defined by Eq. (3). Assume (K1)-(K4), (A0)-(A2), and (A'3)-(A'5) hold. Then the asymptotic MISE of r_n is given by:*

$$MISE(r_n) = \int_M \mu^2(p) dv_g(p) h_n^4 + \int_M \sigma^2(p) dv_g(p) \frac{1}{nh_n^d} + o\left(h_n^4 + \frac{1}{nh_n^d}\right),$$

where $\mu(p)$ and $\sigma(p)$ are as in Theorem 3.5.

Corollary 3.7 *With the notations and under the assumptions of the previous theorem, if $h_n \sim n^{-\frac{1}{d+4}}$ then $MISE(r_n) = O\left(n^{-\frac{4}{d+4}}\right)$.*

4 Concluding remarks

Remark 4.1 The assumption of compactness of the manifold is mainly useful for the derivation of the global convergence rate. Note, however, that the compactness of M implies that its injectivity radius is strictly positive. So for the pointwise convergence results, the compactness assumption may be replaced by the assumption that the manifold is complete and has a strictly positive injectivity radius.

Remark 4.2 It is assumed that Y is bounded for the sake of simplicity of the proofs of the asymptotic properties. Nonetheless, it is possible to replace this assumption by the assumption that $\mathbb{E}Y^2 < \infty$ at the price of a stronger condition on the bandwidth sequence, namely that there exists some $\alpha > 0$ such that $n^{1-\alpha} h_n^d \to \infty$ (see [4, 7]).

Remark 4.3 These results show that the estimator r_n of r on (M, g) converges at the same L^2-rate as a kernel regression estimator on $(\mathbb{R}^d, \delta)$, under similar assumptions. The estimator r_n is built on functions on M of the form $\frac{1}{h_n^d} \frac{1}{\theta_{X_i}(p)} K\left(\frac{d_g(p, X_i)}{h_n}\right)$, which are probability density functions on M and may be considered as a kernel on M with bandwidth h_n. It is interesting to note that such a kernel depends on the local geometry of M in a neighborhood of the observation X_i via the volume density function $\theta_{X_i}(p)$. This appears to be necessary for obtaining an estimator which is consistent with kernel estimators on $(\mathbb{R}^d, \delta)$, and which possesses the same properties under a similar bunch of assumptions.

Acknowledgments

The author acknowledges an anonymous referee for interesting comments. This work was supported by the Department of Mathematics of the University of Le Havre, France, and in part by the National Science Foundation under Grant No. OCE-0417748.

References

[1] A.L. Besse. *Manifolds all of whose geodesics are closed*. Ergebnisse der Mathematik und ihrer Grenzgebiete 93. Springer, 1978.

[2] R. Bhattacharya and V. Patrangenaru. Nonparametric estimation of location and dispersion on riemannian manifolds. *Journal of Statistical Planning and Inference*, 108:23–35, 2002.

[3] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifold (part i). *Annals of Statistics*, 31(1):1–29, 2003.

[4] D. Bosq and J.-P. Lecoutre. *Théorie de l'estimation fonctionnelle*. Economica, 1987.

[5] I. Chavel. *Riemannian geometry: a modern introduction*, volume 108 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 1993.

[6] Y. Chikuse. Density estimation on the stiefel manifold. *Journal of Multivariate Analysis*, 66:188–206, 1998.

[7] G. Collomb. Quelques propriétés de la méthode du noyau pour l'estimation non paramétrique de la régression en un point fixé. *C. R. Acad. Sc. Paris, Série A*, 285:289–292, 1977.

[8] S. Dabo-Niang and N. Rhomari. Kernel regression when the regressor takes values in metric space. *C. R. Acad. Sci. Paris, Série I*, 336:75–80, 2003.

[9] L. Devroye and L. Gyorfi. *Nonparametric Density Estimation. The L_1-View*. Wiley, New York, 1985.

[10] F. Ferraty, A. Goia, and P. Vieu. Nonparametric regression for mixing functional random variables. *C. R. Acad. Sci. Paris, Série I*, 334:217–220, 2002.

[11] F. Ferraty and P. Vieu. Fractal dimensionality and regression estimation in semi-normed vectorial spaces. *C. R. Acad. Sci. Paris, Série I*, 330:139–142, 2000.

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

[13] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer, 1993.

[14] P. Hall, G.S. Watson, and J. Cabrera. Kernel density estimation with spherical data. *Biometrika*, 74:751–762, 1987.

[15] D. Healy, H. Hendriks, and P. Kim. Spherical deconvolution. *J. Multivariate Anal.*, 67:1–22, 1998.

[16] D. Healy and P.T. Kim. An empirical bayes approach to directional data and efficient computation on the sphere. *Ann. Statis.*, 24:232–254, 1996.

[17] H. Hendriks. Nonparametric estimation of a probability density on a riemannian manifold using fourier expansions. *Ann. Statist.*, 18:832–849, 1990.

[18] H. Hendriks. Application of fast spherical fourier transform to density estimation. *J. Multivariate Anal.*, 84:209–221, 2003.

[19] H. Hendriks, J.H.M. Janssen, and F.H. Ruymgaart. Strong uniform convergence of density estimators on compact euclidean manifolds. *Statis. Probab. Letters*, 16:305–311, 1993.

[20] P.E. Jupp and K.V. Mardia. A unified view of the theory of directional statistics 1975-1988. *Internat. Statist. Rev.*, 57:261–294, 1989.

[21] S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry*, volume 1,2. Wiley, 1969.

[22] J.M. Lee and F.H. Ruymgaart. Nonparametric curve estimation on stiefel manifolds. *Nonparam. Statistics*, 6:57–68, 1996.

[23] K.V. Mardia. *Statistics of Directional Data*. Academic Press, 1972.

[24] B. Pelletier. Kernel density estimation on riemannian manifolds. *Statistics and Probability Letters*, 73(3):297-304, 2005.

[25] M.J. Prentice. Spherical regression on matched pairs of orientation statistics. *J. R. Statist. Soc. B*, 51(2):241–248, 1989.

[26] P. Sarda and P. Vieu. Kernel regression. In M.G. Schimek, editor, *Smoothing and Regression: Approaches, Computation, and Application*, pages 43–70. Wiley, 2000.

[27] A.W. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[28] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman and Hall, London, UK, 1995.

[29] G.S. Watson. *Statistics on Spheres*. Wiley, 1983.

[30] T.J. Willmore. *Riemannian geometry*. Oxford University Press, 1993.