A kernel based classifier on a Riemannian manifold
Jean-Michel Loubes, Bruno Pelletier

HAL Id: hal-00460629
https://hal.archives-ouvertes.fr/hal-00460629
Submitted on 1 Mar 2010

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A KERNEL-BASED CLASSIFIER ON A RIEmannIAN MANIFOLD

Jean-Michel LOUBES and Bruno PELLETIER*

Institut de Mathématiques et de Modélisation de Montpellier
UMR CNRS 5149, Equipe de Probabilités et Statistique
Université Montpellier II, CC 051
Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

loubes,pelletier@math.univ-montp2.fr

Abstract

Let $X$ be a random variable taking values in a compact Riemannian manifold without boundary, and let $Y$ be a discrete random variable valued in $\{0; 1\}$ which represents a classification label. We introduce a kernel rule for classification on the manifold based on $n$ independent copies of $(X, Y)$. Under mild assumptions on the bandwidth sequence, it is shown that this kernel rule is consistent in the sense that its probability of error converges to the Bayes risk with probability one.

Index Terms — Classification, Kernel rule, Bayes risk, Consistency.


*Corresponding author.
1 Introduction

In many experiments, the intrinsic structure of the collected data is no longer Euclidean; instead, the observations are points of a given Riemannian manifold. For instance the sphere is the sample space in axial and directional statistics (Fisher et al, 1993; Mardia and Jupp, 2000; Watson, 1983). Three-dimensional rotations or rigid transformations are considered in medical image analysis and high level computer vision (see e.g. Pennec, 2006 and the references therein). Other examples of manifolds encountered in statistical applications include the Stiefel manifold (i.e., the space of $k$-frames in $\mathbb{R}^m$) and the Grassman manifold $G_{k,m-k}$ (i.e., the space of $k$-dimensional hyperplanes in $\mathbb{R}^m$) thoroughly studied by Chikuse (2003), or the manifold of shapes characterized by a corpus of landmarks (Dryden and Mardia, 1998; Kendall et al, 1999; Le and Kendall, 1993; Mardia and Patrangenaru, 2005; Small, 1996).

The aim of the present paper is to generalize the Euclidean kernel rule for the classification of observations to the situation where the data belong to a Riemannian manifold. Stimulated by multiple applications, there is presently a growing literature on statistical inference on manifolds, including the estimation of location parameters (Bhattacharya and Patrangenaru, 2003, 2005), density and regression estimation (Hendriks, 1990; Hendriks et al, 1993; Lee and Ruymgaart, 1996; Pelletier, 2005, 2006), and goodness-of-fit tests (see Jupp (2005) for recent results and further references). However, few is known about classification on a manifold. Indeed, parametric methods are considered in El Khattabi and Streit (1996) and Hayakawa (1997) in the context of directional statistics, i.e. on the sphere, and to the best of our knowledge, no results are available for the nonparametric classification of observations on
a general manifold.

Classification consists in predicting the unknown label $Y \in \{0, 1\}$ of an observation $X \in \mathcal{X}$. It is also called discrimination or supervised classification, this latter terminology being frequently used in the machine learning community, and we will simply use the term classification for short. The observation $X$ as well as its label $Y$ are assumed to be random so that the frequency of outcome of particular pairs is described by the distribution of $(X, Y)$. In practice, the classification procedure is performed by a classifier or classification rule, which in mathematical terms is defined as a function $g : \mathcal{X} \to \{0, 1\}$. The performance of a given classifier $g$ may be quantified by its probability of error $L(g)$ defined by

$$L(g) = P(g(X) \neq Y),$$

an error occurring whenever $g(X) \neq Y$. It is well known (see e.g., Devroye et al, 1996 for a recent exposition) that the minimum of $L(g)$ over all possible classifiers $g$ is achieved by the Bayes rule given by

$$g^*(x) = \begin{cases} 0 & \text{if } P(Y = 0|X = x) \geq P(Y = 1|X = x) \\ 1 & \text{otherwise.} \end{cases}$$

(1.1)

In this sense, the Bayes rule is the optimal decision. However, it depends on the unknown distribution of the pair $(X, Y)$, and for this reason, the Bayes classifier cannot be constructed in practice. Therefore, we shall consider an empirical classifier $g_n$ based on $n$ independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of $(X, Y)$. Following Devroye et al (1996), the classifier $g_n$ will be called strongly consistent if its probability of error

$$L(g_n) = P(g_n(X) \neq Y|(X_1, Y_1), \ldots, (X_n, Y_n))$$
is such that
\[ \lim_{n \to \infty} L(g_n) = L(g^*) \] with probability one.

In the present paper, we focus on the kernel classification rule, which is derived from kernel density estimate, pioneered in Akaike (1954), Parzen (1962) and Rosenblatt (1956). More precisely in a Euclidean space, the kernel rule consists in labeling by 0 a point \( x \) if \( \sum_{i=1}^{n} 1_{\{Y_i=0\}} K((x - X_i)/h_n) \geq \sum_{i=1}^{n} 1_{\{Y_i=1\}} K((x - X_i)/h_n) \), and by 1 otherwise, where the kernel \( K \) is a nonnegative function decreasing with the distance to the origin, and where \( h_n \) is a sequence of smoothing parameters. Using the kernel introduced in Pelletier (2005, 2006), we generalize herein the kernel classification rule to the case of a closed Riemannian manifold and we prove its strong consistency.

The paper is organized as follows. Section 2 introduces the kernel on the manifold defined in Pelletier (2005) as well as some notation. In Section 3, we define the kernel classification rule and prove its strong consistency. For clarity, the proof of our main result, which relies on several auxiliary results, is exposed in Section 4. For materials on differential geometry, we refer to Chavel (1993) and Kobayashi and Nomizu (1969).

## 2 Kernel definition

Let \((M, g)\) be a compact Riemannian manifold without boundary of dimension \(d\). We shall denote by \(d_g\) the Riemannian geodesic distance, and by \(v_g\) the Riemannian volume measure on \(M\). In this section, we define a kernel \(K_h\) on \(M\) with bandwidth parameter \(h\), as in Pelletier (2005), and briefly summarize its main properties.
First of all, let $K : \mathbb{R}_+ \to \mathbb{R}$ be a positive and continuous map such that:

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

(ii) $\text{supp } K = [0; 1],$ 

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

Now for $p$ and $q$ two points of $M$, let $\theta_p(q)$ be the volume density function on $M$ roughly defined by Besse (1978, p. 154):

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

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 $T_p(M)$. Note that 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 (Willmore, 1993, 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 $\theta_p(q) = \theta_q(p)$ (Willmore, 1993, p. 221).

Then we define a kernel $K_h(p, .) : M \to \mathbb{R}_+$ on $M$ by:

$$K_h(p, q) = \frac{1}{\theta_p(q) h^d} K \left( \frac{d_g(q, p)}{h} \right), \quad (2.1)$$

for all $q \in M$. In (2.1), $h$ is the bandwidth or smoothing parameter and we assume that it satisfies the condition

$$h \leq h_0 < inj_g(M), \quad (2.2)$$
for some fixed $h_0$, where $\text{inj}_g(M)$ is the injectivity radius of $M$ [strictly positive since $M$ is compact].

The kernel (2.1) has some interesting properties proved in Pelletier (2005) that we briefly summarize below. First of all, this kernel is a probability density on $M$ with respect to the Riemannian volume measure. Second, if the function $K$ is such that $\int_{R^d} uK(\|u\|)\lambda(du) = 0$, then the kernel is centered on $p$ in the sense that, if a random variable $X$ valued in $M$ has density $K_h(p, \cdot)$ with respect to $\nu_g$, then $p$ is the intrinsic mean of $X$, provided $h$ is small enough. Additionally, when $M$ is $\mathbb{R}^d$, we have $\theta_p(q) = 1$ for all $p, q$, and so $K_h$ reduces to a standard isotropic kernel on $\mathbb{R}^d$ supported by the closed unit Euclidean ball.

In all of the following, we shall assume that the function $K$ is such that

$$\inf_{0 \leq x \leq \frac{1}{2}} K(x) > 0,$$

which implies that the kernel $K_h(p, \cdot)$ takes strictly positive values on the geodesic ball $B_M(p, \frac{h}{2})$ centered at $p$ and of radius $h/2$. This assumption is needed in the proofs of Lemma 4.2 and Lemma 4.4 and is related to the notion of regular kernels on $\mathbb{R}^d$ (see eg., Devroye et al, 1996, Definition 10.1). In this assumption, the scalar $\frac{1}{2}$ is arbitrary. It could be replaced by any real number in the open interval $(0; 1)$, and the particular value of $\frac{1}{2}$ is selected for the sake of simplicity only.

3 Kernel classification rule

In this section, we define a kernel classification rule using the kernel (2.1) and establish its consistency. To this aim, let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be $n$
independent copies of a pair of random variables \((X, Y)\) valued in \(M \times \{0; 1\}\). Then we define the \textit{kernel classification rule} \(g^0_n : M \to \{0; 1\}\) by:

\[
g^0_n(p) = \begin{cases} 
0, & \text{if } \sum_{i=1}^{n} 1\{Y_i = 0\}K_{h_n}(p, X_i) \geq \sum_{i=1}^{n} 1\{Y_i = 1\}K_{h_n}(p, X_i), \\
1, & \text{otherwise}, 
\end{cases} 
\]

for all \(p \in M\), and where \(K_{h_n}\) is a kernel on \(M\) of the form given by (2.1) with bandwidth sequence \(h_n\).

As in the Introduction, \(L(g^*)\) will denote the probability of error of the Bayes rule \(g^*\) defined by (1.1), and the classification error probability of the kernel rule will be denoted by \(L(g^0_n)\), i.e.,

\[
L(g^0_n) = \mathbb{P}(g^0_n(X) \neq Y|(X_1, Y_1), \ldots, (X_n, Y_n)).
\]

We are now in a position to state our main result.

\textbf{Theorem 3.1} Suppose that \(h_n \to 0\) and \(nh_n^{2d} \to \infty\). Then

\[
\lim_{n \to \infty} L(g^0_n) = L(g^*)
\]

with probability one.

\textbf{Remark} Theorem 3.1 states that the kernel classification rule (3.1) is strongly consistent. As exposed in the Introduction, the application field of this type of result is vast, including automatic labelling of shapes, medical images, and signals, for instance. However, the practical implementation of this kernel rule exceeds the scope of the present paper.
4 Proofs

The proof of Theorem 3.1 is given in paragraph 4.3 and relies on several auxiliary results. One first Lemma on the metric entropy of the manifold is proved in paragraph 4.1. Auxiliary Lemmas concerning the classification rule are demonstrated in paragraph 4.2.

4.1 Covering number

Let us first recall that the $\rho$-covering number of a subset $S$ of a metric space is defined as the smallest number of open balls of radius $\rho$ whose union covers $S$. The logarithm of the $\rho$-covering number is generally called the metric entropy of $S$.

**Lemma 4.1** Let $(M, g)$ be a compact Riemannian manifold without boundary of dimension $d$. Let $\delta$ be the infimum of the sectional curvatures of $M$ and let $\mathcal{N}(\rho)$ be the $\rho$-covering number of $M$. If $\rho$ is such that

$$0 < \rho < \min \left\{ \text{inj}_g(M), \frac{\pi}{\sqrt{\delta}}, 2\pi \right\},$$

where $\text{inj}_g(M)$ is the injectivity radius of $M$, and where we have set $\frac{\pi}{\sqrt{\delta}} = +\infty$ whenever $\delta \leq 0$, then

$$\mathcal{N}(\rho) \leq Vol_g(M) \frac{d}{c_{d-1}} \left( \frac{\pi}{2} \right)^{d-1} \left( \frac{\rho}{2} \right)^{-d}$$

where $c_{d-1}$ is the volume of the $(d-1)$-dimensional unit sphere in $\mathbb{R}^d$, and where $Vol_g(M)$ denotes the volume of $M$.

**Proof** Consider a maximal set of points $\{p_i; i \geq 1\}$ such that $d_g(p_i, p_j) > \rho$ for all $i \neq j$. Then $M \subset \cup_{i \geq 1} B_M(p_i, \rho)$ otherwise there would exist a point $p$ on $M$ such that $p_i, d_g(p, p_i) > \rho$ for all points $p_i$, which is not possible by the
definition of the set \( \{(p_i); \; i \geq 1\} \). Furthermore, since \( M \) is compact, there exists an integer \( N \) such that, after sorting the \( p_i \)'s, we have

\[
M \subset \bigcup_{i=1}^{N} B_M(p_i, \rho).
\]

But \( \bigcup_{i=1}^{N} B_M(p_i, \rho/2) \subset M \), and \( B_M(p_i, \rho/2) \cap B_M(p_j, \rho/2) = \emptyset \) whenever \( i \neq j \). As a consequence, we obtain that

\[
\sum_{i=1}^{N} v_g(B_M(p_i, \rho/2)) \leq Vol_g(M),
\]

where \( Vol_g(M) \) is the volume of \( M \). By the Günther-Bishop volume comparison Theorem (Chavel, 1993, Theo. 3.7), we have

\[
v_g(B_M(p_i, \rho/2)) \geq V_\delta(\rho/2), \quad \forall i = 1, \ldots, N,
\]

where \( V_\delta(\rho/2) \) is the volume of the ball of radius \( \rho/2 \) in the space of constant sectional curvature \( \delta \), i.e., the \( d \)-sphere of constant sectional curvature \( \delta \) when \( \delta > 0 \); \( \mathbb{R}^d \) when \( \delta = 0 \); and the hyperbolic space of constant sectional \( \delta \) when \( \delta < 0 \). Reporting the inequality (4.2) in (4.1), we obtain the inequality

\[
N \leq \frac{Vol_g(M)}{V_\delta(\rho/2)},
\]

from which it follows that

\[
N(\rho) \leq \frac{Vol_g(M)}{V_\delta(\rho/2)}, \quad (4.3)
\]

by the definition of the \( \rho \)-covering number.

Now we proceed to derive lower bounds on \( V_\delta(\rho/2) \). To this aim, following Chavel (1993, p. 117), the volume \( V_\delta(\rho/2) \) may be evaluated as follows:

\[
V_\delta(\rho/2) = c_{d-1} \int_0^{\rho/2} S^d_\delta(t) dt,
\]
where

\[
S_\delta(t) = \begin{cases} 
\frac{1}{\sqrt{\delta}} \sin(\sqrt{\delta}t), & \text{if } \delta > 0, \\
t, & \text{if } \delta = 0, \\
\frac{1}{\sqrt{-\delta}} \sinh(\sqrt{-\delta}t), & \text{if } \delta < 0,
\end{cases}
\]

and where \( c_{d-1} \) is the volume of the \((d-1)\)-dimensional unit sphere in \( \mathbb{R}^d \).

First of all, observe that, in the case where \( \delta < 0 \), we have \( V_\delta(\rho/2) \geq V_0(\rho/2) \) since \( \sinh(u) \geq u \) for all \( u \geq 0 \). Second, in the case where \( \delta > 0 \), we have \( V_0(\rho/2) \geq V_\delta(\rho/2) \) since \( \frac{1}{\sqrt{\delta}} \sin(\sqrt{\delta}t) \leq t \) for all \( t \geq 0 \). Consequently, it suffices to bound from below \( V_\delta(\rho/2) \) in the case where \( \delta > 0 \).

To this aim, since \( \rho < \frac{\pi}{\sqrt{\delta}} \), we have \( \sqrt{\delta} t \leq \frac{\pi}{2} \) for all \( t \leq \frac{\rho}{2} \). So using the inequality \( \sin u \geq \frac{2}{\pi} u \) for all \( 0 \leq u \leq \frac{\pi}{2} \), we obtain

\[
V_\delta(\rho/2) \geq c_{d-1} \left( \frac{1}{\sqrt{\delta}} \right)^{d-1} \int_0^{\rho/2} \left( \frac{2}{\pi} \sqrt{\delta} t \right)^{d-1} dt
\]

leading to the lower bound

\[
V_\delta(\rho/2) \geq \frac{c_{d-1}}{d} \left( \frac{2}{\pi} \right)^{d-1} \left( \frac{\rho}{2} \right)^d,
\]

which holds for all \( \delta \). Reporting (4.4) in the inequality (4.3) leads to the desired result.

\[ \square \]

4.2 Auxiliary results

Consider the classification rule

\[
g_n(p) = \begin{cases} 
0, & \text{if } \frac{\sum_{i=1}^n 1_{(Y_i=0)} K_{\delta n}(p,X_i)}{nK_{\delta n}(p,X)} \geq \frac{\sum_{i=1}^n 1_{(Y_i=1)} K_{\delta n}(p,X_i)}{nK_{\delta n}(p,X)}, \\
1, & \text{otherwise.}
\end{cases}
\]
Clearly, this classification rule is equivalent to \( g_n^0 \) defined in (3.1). Now we define the function \( \eta_n \) on \( M \) by
\[
\eta_n(p) = \frac{\sum_{j=1}^{n} Y_j K_{h_n}(p, X_j)}{n E_k(p, X)}
\]
and we shall denote by \( \eta(p) \) the conditional probability that \( Y \) is 1 given \( X = p \), i.e.,
\[
\eta(p) = \mathbb{P}\{Y = 1|X = p\} = \mathbb{E}[Y|X = p].
\]
According to Theorem 2.3 in Devroye et al (1996, Chap. 2, p. 17), the theorem will be proved if we show that
\[
\int_M |\eta(p) - \eta_n(p)| \mu(dp) \to 0 \quad \text{with probability one as } n \to \infty, \quad (4.5)
\]
where \( \mu \) is the probability measure of the random variable \( X \).
Lemma 4.2 Let $K_h(p,.)$ be a kernel on $M$ of the form given by (2.1). Let $X$ be a random variable valued in $M$ with probability measure $\mu$. Then there exists a constant $C > 0$ depending only on $K$ and on the geometry of $M$ such that:

$$\sup_{q \in M} \int_M K_h(p, q) \frac{1}{\mathbb{E}K_h(p, X)} \mu(dp) \leq C.$$ 

Proof First of all, we have

$$\int_M K_h(p, q) \frac{1}{\mathbb{E}K_h(p, X)} \mu(dp) = \int_{B_M(q, h)} K_h(p, q) \frac{1}{\mathbb{E}K_h(p, X)} \mu(dp).$$

Next, cover the geodesic ball $B_M(q, h)$ by $N_B$ geodesic balls centered at points $p_i$ of $B_M(q, h)$ and of radius $\frac{h}{4}$. Then we start with the following inequality:

$$\int_M K_h(p, q) \frac{1}{\mathbb{E}K_h(p, X)} \mu(dp) \leq \sum_{i=1}^{N_B} \int_{B_M(p_i, h/4)} K_h(p, q) \frac{1}{\mathbb{E}K_h(p, X)} \mu(dp).$$

(4.6)

Now we proceed to bound the two terms in the ratio under the integral above.

First of all, since $K_h(., q)$ is supported by $B_M(q, h)$, we have for all $i = 1, \ldots, N_B$, and all $q \in M$:

$$\sup_{p \in B_M(p_i, \frac{h}{4})} K_h(p, q) \leq \sup_{p \in M} \sup_{q \in B_M(p, h)} K_h(p, q) \leq \left( \sup_{p \in M} \sup_{q \in B_M(p, h)} \theta_p^{-1}(q) \right) \frac{1}{h^d} \sup_{\|x\| \leq h} K\left( \frac{\|x\|}{h} \right) \leq \left( \sup_{p \in M} \sup_{q \in B_M(p, h_0)} \theta_p^{-1}(q) \right) \frac{1}{\pi^d} \sup_{\|x\| \leq 1} K(\|x\|) = C_1 \frac{1}{\pi^d},$$

(4.7)

where we have set

$$C_1 = \left( \sup_{p \in M} \sup_{q \in B_M(p, h_0)} \theta_p^{-1}(q) \right) \sup_{\|x\| \leq 1} K(\|x\|),$$
and where $h_0$ is the constant defined by (2.2).

Second, for all $p \in M$, we have

\[
\mathbb{E}K_h(p, X) = \int_M K_h(p, q) \mu(dq) \\
\geq \int_{B_M(p, h/2)} \theta_p^{-1}(q) \frac{1}{h^d} K \left( \frac{d_g(q, p)}{h} \right) \mu(dq) \\
\geq \left( \inf_{p \in M} \inf_{q \in B_M(p, h/2)} \theta_p^{-1}(q) \right) \frac{1}{h^d} \inf_{q \in B_M(p, h/2)} K \left( \frac{d_g(q, p)}{h} \right) \int_{B_M(p, h/2)} \mu(dq) \\
\geq \left( \inf_{p \in M} \inf_{q \in B_M(p, h_0/2)} \theta_p^{-1}(q) \right) \frac{1}{h^d} \inf_{\|x\| \leq 1/2} K (\|x\|) \int_{B_M(p, h/2)} \mu(dq) \\
= C_2 \frac{1}{h^d} \mu \left( B_M \left( p, \frac{h}{2} \right) \right),
\]

where

\[
C_2 = \left( \inf_{p \in M} \inf_{q \in B_M(p, h_0)} \theta_p^{-1}(q) \right) \inf_{\|x\| \leq 1/2} K (\|x\|).
\]

Now, noting that for all $p \in B_M \left( p_i, \frac{h}{4} \right)$ we have $B_M \left( p_i, \frac{h}{4} \right) \subset B_M(p, h/2)$, we obtain

\[
\mathbb{E}K_h(p, X) \geq C_2 \frac{1}{h^d} \mu \left( B_M(p_i, h/4) \right),
\]

(4.8)

for all $p \in B_M \left( p_i, \frac{h}{4} \right)$.

Reporting (4.7) and (4.8) in (4.6) yields

\[
\int_M \frac{K_h(p, q)}{\mathbb{E}K_h(p, X)} \mu(dp) \leq \sum_{i=1}^{N_B} C_1 \frac{C_2}{C_2} \int_{B_M(p_i, h/4)} \frac{\mu(dp)}{\mu(B_M(p_i, h/4))} \\
= \frac{C_1}{C_2} N_B
\]

for all $q \in M$. Now, applying Lemma 4.1 to $B_M(q, h)$, and since $\text{Vol}_g (B_M(q, h)) = O(h^d)$, where the constant in $O(h^d)$ can be made uniform in $q$ since $M$ is closed, we obtain that there exists a constant $C$ such that $\mathcal{N}_B \leq C$. Hence the Lemma.

\[\square\]

From now on, $\mu$ will denote the probability measure of $X$. 

13
Lemma 4.3 If $h_n \to 0$ then

$$
\int_M |\eta(p) - \mathbb{E}\eta_n(p)| \mu(dp) \to 0
$$

as $n \to \infty$.

Proof Let $\varepsilon > 0$. Since $M$ is compact, the set of continuous functions on $M$ is dense in $L^1(M, \mu)$, and so there exists a continuous function $r$ such that

$$
\int_M |\eta(p) - r(p)| \mu(dp) \leq \varepsilon.
$$

First of all, we have

$$
\int_M |\eta(p) - \mathbb{E}\eta_n(p)| \mu(dp)
\leq \int_M |\eta(p) - r(p)| \mu(dp) + \int_M |r(p) - \mathbb{E}\eta_n(p)| \mu(dp)
\leq \varepsilon + \int_M |r(p) - \mathbb{E}\eta_n(p)| \mu(dp). \tag{4.9}
$$

For the second term in the right hand side of (4.9), we may write

$$
\int_M |r(p) - \mathbb{E}\eta_n(p)| \mu(dp)
= \int_M |r(p) - \mathbb{E}\eta_n(p)| \frac{K_{h_n}(p, q)}{\mathbb{E}K_{h_n}(p, X)} \mu(dp) \mu(dq)
\leq \int_M \int_M |r(p) - \eta(q)| \frac{K_{h_n}(p, q)}{\mathbb{E}K_{h_n}(p, X)} \mu(dp) \mu(dq)
\leq \int_M \int_M |r(p) - r(q)| \frac{K_{h_n}(p, q)}{\mathbb{E}K_{h_n}(p, X)} \mu(dp) \mu(dq)
\quad + \int_M \int_M |r(q) - \eta(q)| \frac{K_{h_n}(p, q)}{\mathbb{E}K_{h_n}(p, X)} \mu(dp) \mu(dq). \tag{4.10}
$$

Now we proceed to prove that the two terms in the right hand side of (4.10) are bounded from above by a constant multiple of $\varepsilon$ for all $n$ large enough.
Since the function $r$ is continuous and since $M$ is compact, $r$ is uniformly continuous so there exists $\rho > 0$ such that $|r(q) - r(p)| < \varepsilon$ for all $p$ and $q$ in $M$ with $d_g(p, q) < \rho$. Thus

$$
\int_M \int_M |r(p) - r(q)| \frac{K_{h_n}(p, q)}{E_{K_{h_n}(p, X)}} \mu(dp) \mu(dq)
\leq \int_M \int_{B_M(p, \rho)} |r(q) - r(p)| \frac{K_{h_n}(q, p)}{E_{K_{h_n}(p, X)}} \mu(dp) \mu(dq)
+ \int_M \int_{B_M^c(p, \rho)} |r(q) - r(p)| \frac{K_{h_n}(q, p)}{E_{K_{h_n}(p, X)}} \mu(dp) \mu(dq),
$$

(4.11)

where $B_M(p, \rho)$ and $B_M^c(p, \rho)$ denotes respectively the geodesic ball in $M$ centered at $p$ and of radius $\rho$, and its complement. But for $n$ large enough, $h_n < \rho$ so $B_M(p, h_n) \subset B_M(p, \rho)$. Consequently, the second term in the right hand side of (4.11) vanishes and we obtain

$$
\int_M \int_M |r(p) - r(q)| \frac{K_{h_n}(p, q)}{E_{K_{h_n}(p, X)}} \mu(dp) \mu(dq)
\leq \varepsilon \int_M \int_{B_M(p, h_n)} \frac{K_{h_n}(q, p)}{E_{K_{h_n}(p, X)}} \mu(dp) \mu(dq)
= \varepsilon \int_M \int_{B_M(p, h_n)} \frac{K_{h_n}(q, p)}{E_{K_{h_n}(p, X)}} \mu(dp) \mu(dq)
= \varepsilon \text{Vol}_g(M).
$$

(4.12)

Now for the second term in the right hand side of (4.10), we have

$$
\int_M \int_M |r(q) - \eta(q)| \frac{K_{h_n}(p, q)}{E_{K_{h_n}(p, X)}} \mu(dp) \mu(dq),
\leq \sup_{q \in M} \int_M \frac{K_{h_n}(p, q)}{E_{K_{h_n}(p, X)}} \mu(dp) \int_M |r(q) - \eta(q)| \mu(dq)
\leq C \varepsilon
$$

(4.13)

for some constant $C$ by Lemma 4.2.
Finally, reporting (4.13), (4.12), and (4.10) in (4.9) leads to the desired result.

\[\square\]

**Lemma 4.4** There exists a positive constant \(C\) such that

\[
\mathbb{E} \int_M |\eta_n(p) - \mathbb{E}\eta_n(p)| \mu(dp) \leq C \left( \frac{1}{n} \mathcal{N} \left( \frac{h_n}{4} \right) \right)^{\frac{1}{2}}.
\]

**Proof** We have

\[
\mathbb{E} \{ |\eta_n(p) - \mathbb{E}\eta_n(p)| \} \leq \sqrt{\mathbb{E} \{ (\sum_{j=1}^n Y_j K_{h_n}(p,X_j) - \mathbb{E} K_{h_n}(p,X))^2 \} \}^{1/2} \leq \left( \frac{\mathbb{E} \{ (\sum_{j=1}^n Y_j K_{h_n}(p,X_j) - \mathbb{E} K_{h_n}(p,X))^2 \} \}^{1/2} \right),
\]

\[\text{(4.14)}\]

First of all, we have

\[
\mathbb{E} K_{h_n}^2 (p, X) \leq \sup_{q \in B_M(p, h_n)} K_{h_n} (p, q) \mathbb{E} K_{h_n} (p, X) \leq \sup_{\|x\| \leq 1} K(\|x\|) \left( \sup_{p \in M} \sup_{q \in B_M(p, h_0)} \theta_p^{-1}(q) \right) \frac{1}{h_n} \mathbb{E} K_{h_n} (p, X).
\]

Therefore

\[
\frac{\mathbb{E} K_{h_n}^2 (p, X)}{n (\mathbb{E} K_{h_n} (p, X))^2} \leq \frac{C_1}{n h_n^d \mathbb{E} K_{h_n} (p, X)},
\]

where \(C_1 = \sup_{\|x\| \leq 1} K(\|x\|) \left( \sup_{p \in M} \sup_{q \in B_M(p, h_0)} \theta_p^{-1}(q) \right)\).

Now we bound \(\mathbb{E} K_{h_n} (p, X)\) as follows:

\[
\mathbb{E} K_{h_n} (p, X) \geq \frac{1}{h_n^d} \int_{B_M(p, h_n)} \frac{1}{\theta_p(q)} K \left( \frac{d_2(q, p)}{h_n} \right) \mu(dq) \geq \frac{1}{h_n^d} \left( \inf_{p \in M} \inf_{q \in B_M(p, h_0)} \theta_p^{-1}(q) \right) \inf_{\|x\| \leq 1/2} K(\|x\|) \mu \left( B_M(p, h_n/2) \right).
\]
and so
\[ \mathbb{E}K_{h_n}(p, X) \geq C_2 \frac{1}{h_n^d} \mu \left( B_M \left( p, \frac{h_n}{2} \right) \right), \tag{4.16} \]
where \( C_2 = \left( \inf_{p \in M} \inf_{q \in B_M(p, h_0)} \theta_p^{-1}(q) \right) \inf_{\|x\| \leq 1/2} K(\|x\|). \)

From (4.14), (4.15) and (4.16), it follows that
\[ \mathbb{E} \{ |\eta_n(p) - \mathbb{E}\eta_n(p)| \} \leq \frac{C_1}{C_2} \frac{1}{\sqrt{n}} \sqrt{\frac{1}{\mu \left( B_M \left( p, \frac{h_n}{2} \right) \right)}}, \]
for all \( p \in M \), and so
\[ \int_M \mathbb{E} \{ |\eta_n(p) - \mathbb{E}\eta_n(p)| \} \mu(dp) \leq \frac{C_1}{C_2} \sqrt{\text{Vol}(M)} \frac{1}{\sqrt{n}} \left[ \int_M \frac{\mu(dp)}{\mu \left( B_M(p, h_n/2) \right)} \right]^{1/2}, \]
by Cauchy-Schwarz. Now, using a cover of \( M \) by \( N \left( \frac{h_n}{4} \right) \) geodesic balls \( B_M(p_i, \frac{h_n}{4}) \) centered at points \( p_i \) of \( M \) and of radius \( \frac{h_n}{4} \), we obtain that
\[ \int_M \frac{\mu(dp)}{\mu \left( B_M(p, h_n/2) \right)} \leq \sum_{i=1}^{N(h_n/4)} \int_{B_M(p_i, h_n/4)} \frac{\mu(dp)}{\mu \left( B_M(p_i, h_n/4) \right)} = N(h_n/4). \]
Consequently
\[ \int_M \mathbb{E} \{ |\eta_n(p) - \mathbb{E}\eta_n(p)| \} \mu(dp) \leq \frac{C_1}{C_2} \sqrt{\text{Vol}(M)} \left( \frac{1}{n} N \left( \frac{h_n}{4} \right) \right)^{1/2}. \]
\[ \square \]

4.3 Proof of Theorem 3.1

We proceed to demonstrate (4.5), i.e., that
\[ \int_M |\eta(p) - \eta_n(p)| \mu(dp) \to 0 \quad \text{with probability one as } n \to \infty. \]
First of all, we have
\[
\mathbb{E} \int_M |\eta(p) - \eta_n(p)| \mu(dp) \\
\leq \int_M |\eta(p) - \mathbb{E}\eta_n(p)| \mu(dp) + \mathbb{E} \int_M |\eta_n(p) - \mathbb{E}\eta_n(p)| \mu(dp) \\
\leq \int_M |\eta(p) - \mathbb{E}\eta_n(p)| \mu(dp) + C_1 \left( \frac{1}{n} \mathcal{N} \left( \frac{h_n}{4} \right) \right)^\frac{1}{2}
\]
for some positive constant $C_1$ by Lemma 4.4. Since $\mathcal{N} \left( \frac{h_n}{4} \right) = O(\frac{1}{h_n^d})$ by Lemma 4.1, and since $nh_n^{2d} \to \infty$ by assumption, it follows that $nh_n^d \to \infty$ and so
\[
\frac{1}{n} \mathcal{N} \left( \frac{h_n}{4} \right) \to 0 \quad \text{as} \quad n \to \infty.
\]
Next, by applying Lemma 4.3, we obtain
\[
\mathbb{E} \int_M |\eta(p) - \eta_n(p)| \mu(dp) \to 0 \quad \text{as} \quad n \to \infty.
\]

Therefore, (4.5) will be proved if we show that
\[
\int_M |\eta(p) - \eta_n(p)| \mu(dp) - \mathbb{E} \int_M |\eta(p) - \eta_n(p)| \to 0
\]
with probability one as $n \to \infty$. For this purpose, we shall use McDiarmid’s inequality (McDiarmid, 1989) applied to the centered random variable
\[
\int_M |\eta(p) - \eta_n(p)| \mu(dp) - \mathbb{E} \int_M |\eta(p) - \eta_n(p)|.
\]
First of all, keep the data fixed at $(x_1, y_1), \ldots, (x_n, y_n)$ and replace the $i^{th}$ pair $(x_i, y_i)$ by $(\bar{x}_i, \bar{y}_i)$, changing the value of $\eta_n(p)$ to $\bar{\eta}_i(p)$. Then we have
\[
\left| \int_M |\eta_n(p) - \eta(p)| \mu(dp) - |\bar{\eta}_i(p) - \eta(p)| \mu(dp) \right| \leq \int_M |\eta_n(p) - \bar{\eta}_i(p)| \mu(dp) \\
\leq \frac{2}{n} \sup_{p, q \in M} \int_M \frac{K_{h_n}(p, q)}{\mathbb{E}K_{h_n}(p, X)} \mu(dp) \\
\leq \frac{C_1}{n} \mathcal{N} \left( \frac{h_n}{4} \right) \\
\leq \frac{C_2}{nh_n^d}
\]
for some positive constants $C_1$ and $C_2$ by Lemma 4.2 and Lemma 4.1. So, applying McDiarmid’s inequality (McDiarmid, 1989) yields

$$
P \left\{ \int_M |\eta_n(p) - \eta(p)|d\mu(p) \geq \varepsilon \right\} \leq P \left\{ \int_M |\eta_n(p) - \eta(p)|d\mu(p) - E \int_M |\eta_n(p) - \eta(p)|d\mu(p) \geq \frac{\varepsilon}{2} \right\} \leq C \exp \left( -\varepsilon^2 nh_n^{2d} \right).
$$

for all $\varepsilon > 0$. Finally, since $nh_n^{2d} \to +\infty$ by assumption, and using the Borel-Cantelli Lemma, we conclude that

$$
\int_M |\eta(p) - \eta_n(p)| \mu(dp) - E \int_M |\eta(p) - \eta_n(p)| \to 0
$$

with probability one as $n \to \infty$, which proves (4.5), and so the Theorem. □

**References**


