Communications in Mathematics and Statistics (2022) 10:599–621
https://doi.org/10.1007/s40304-020-00237-0

Nonparametric estimation for stationary and strongly mixing processes on Riemannian manifolds

Amour T. Gbaguidi Amoussou[1] · **Freedath Djibril Moussa**[2] · **Carlos Ogouyandjou**[1] · **Mamadou Abdoul Diop**[3]

Received: 6 November 2019 / Revised: 27 April 2020 / Accepted: 23 December 2020 /

Abstract

In this paper, nonparametric estimation for a stationary strongly mixing and manifold-valued process (X_j) is considered. In this non-Euclidean and not necessarily i.i.d setting, we propose kernel density estimators of the joint probability density function, of the conditional probability density functions and of the conditional expectations of functionals of X_j given the past behavior of the process. We prove the strong consistency of these estimators under sufficient conditions, and we illustrate their performance through simulation studies and real data analysis.

Keywords Riemannian manifolds · Nonparametric estimation · Kernel density estimation · Stationary and strongly mixing processes · Strong consistency

Mathematics Subject Classification 58C35 · 62G05 · 62G20 · 62M09

✉ Freedath Djibril Moussa
freedath.djibril@imsp-uac.org

Amour T. Gbaguidi Amoussou
amour.gbaguidi@imsp-uac.org

Carlos Ogouyandjou
ogouyandjou@imsp-uac.org

Mamadou Abdoul Diop
mamadou-abdoul.diop@ugb.edu.sn

[1] IMSP, Université d'Abomey-Calavi (UAC), Dangbo, Benin

[2] FAST, Université d'Abomey-Calavi, Abomey-Calavi, Benin

[3] Université Gaston Berger, Dakar, Senegal

1 Introduction

Nonparametric estimation is a basic approach in statistical inference and widely used for Euclidean data when no assumption is made on the family to which the data generating process belongs. But in many situations, the data are in a Riemannian manifold or are best modeled as points of a Riemannian manifold which is not an Euclidean space [1,4]. A survey on data on manifolds and nonparametric tools for this kind of data (medical images, meteorological data, directional data, similarity shape data, digital camera images, etc.) can be found in [16]. Many efforts have been devoted to investigating kernel density estimation on Riemannian manifold. Pelletier proposed a kernel density estimator [18] and a nonparametric estimator of the regression function [17] on Riemannian manifolds. Henry et al. [9] gave the asymptotic distribution and uniform consistency of Pelletier's density estimator and proposed an illustration based on real directional paleomagnetism data. Chevallier et al. [5] studied Pelletier's estimator on Siegel spaces (a type of Riemannian manifold) and gave an application to radar observations.

All those works considered i.i.d random objects for the nonparametric inference. But non-i.i.d situations are commonly encountered in statistical applications. Masry [13] established the strong consistency of nonparametric estimators of conditional probability density and conditional expectations for real-valued stationary and strongly mixing processes.

In this paper, we are interested in density estimation for stationary and strongly mixing processes on Riemannian manifolds. We propose strongly consistent estimators of conditional density function and functional conditional expectations that take account the geometric structure of sample space, extending the work of Masry [13] to Riemannian manifolds.

The rest of the paper is organized as follows. In Sect. 2, we review the relevant results related to Riemannian manifold and strongly mixing stationary process on Riemannian manifold. Section 3 is devoted to the new geometric nonparametric density estimator, conditional density estimator and functional conditional expectations estimator and their consistency properties along with the proofs.

The numerical studies on the finite sample performance of the proposed kernel density estimator based on simulated data and real data are presented in Sect. 4. The conclusion and some research perspectives are given in Sect. 5.

2 Preliminaries

2.1 Some notions from Riemannian Geometry

In this section, we introduce some fundamental definitions and properties of Riemannian manifolds. For more details, see [3,20,24].

Definition 2.1 A Riemannian metric g on a differentiable manifold M is a smoothly chosen inner product $g_p : \top_p M \times \top_p M \to \mathbb{R}$ on each of the tangent spaces $\top_p M$ of M with $p \in M$. In other words, for each $p \in M$, $g = g_p$, satisfies

1. $g(u, v) = g(v, u)$ for all $u, v \in \top_p M$;
2. $g(u, u) \geq 0$ for all $u \in \top_p M$;
3. $g(u, u) = 0$ if and only if $u = 0$.

We note $\langle u, v\rangle := g(u, v)$ for all $u, v \in \top_p M$.

Definition 2.2 A differentiable manifold M equipped with a Riemannian metric g is called Riemannian manifold and is denoted by (M, g).

Definition 2.3 Let $p \in M$ and $(x^1, \dots, x^n)$ be a local system of coordinates of p. We define the volume form of a Riemannian metric to be top-dimensional form $\mathrm{d}v_g$ which is given in local coordinates by

$$\mathrm{d}v_g(p) = \sqrt{\det g}\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n,$$

where $(\partial_{x^1}, \dots, \partial_{x^n})$ is a positively oriented basis of $\top_p M$ with $\partial_{x^i} := \frac{\partial}{\partial x^i}$ for all $i \in \{1, \dots, n\}$, $\det g$ is the determinant of the matrix $(g_{ij})_{i,j=1,\dots,n}$, $g_{ij} = g_p(\partial_{x^i}, \partial_{x^j})$ and $\mathrm{d}x^i$ is the linear form on $\top_p M$ defined by

$$\mathrm{d}x^i\left(\partial_{x^j}\right) = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if not} \end{cases} .$$

We set the volume of (M, g) to be

$$v_g(M) = \int_M \mathrm{d}v_g(x).$$

Definition 2.4 Let (M, g) be a Riemannian manifold of dimension d. A smooth curve $\gamma : \mathbb{R} \supset I \to M$ is called geodesic at $t_0 \in I$, if $\frac{D}{\mathrm{d}t}(\dot{\gamma}) = 0$ at $t_0 \in I$, where $\dot{\gamma}$ is the derivative of the function $\gamma : I \to \mathbb{R}$. If γ is geodesic at all points $t \in I$, it is said to be geodesic.

$\frac{D}{\mathrm{d}t}$ is the covariant derivative (see [3]).

Theorem 2.5 [24] *Let* (M,g) *be a Riemannian manifold,* $p \in M$ *and* $v \in \top_p M$. *There is a positive* ϵ *and a unique geodesic* $\gamma_v : [0,\epsilon] \to M$ *such that* $\gamma_v(0) = p$ *and* $\dot{\gamma}_v(0) = v$.

From the geodesics of a Riemannian manifold M, for all $p \in M$, the exponential map at point p, a diffeomorphism (i.e., a differentiable, bijective map of differentiable inverse) denoted by $\exp_p$, which maps a tangent vector v of an open ball $B(0, r) \subset \top_p M$ centered in 0 to the endpoint $\gamma(1) =: \exp_p(v)$ of the geodesic $\gamma : [0, 1] \to M$ verifying $\gamma(0) = p$, $\dot{\gamma}(0) = v$. Intuitively, the exponential map moves the point p along the geodesic starting from p at speed after covering the length $\|v\|$. Conversely, the inverse of the exponential map $\log_p(\tilde{p}) := \exp_p^{-1}(\tilde{p})$ gives the vector that maps p to $\tilde{p}$. The image by the exponential map of the open ball $B(0, r) \subset \top_p M$, with r less than the injectivity radius at p, is called the geodesic ball of radius $r > 0$ centered in p.

Definition 2.6 A Riemannian manifold M is called geodesically complete if the map $\exp_p$ is defined on $\top_p M$ for all $p \in M$.

Theorem 2.7 (Hopf–Rinow [24]) *Let M be a connected Riemannian manifold. The following are equivalent*:

- *M is complete as a metric space.*
- *M is geodesically complete.*
- *There exists a point $p \in M$ such that the map $\exp_p$ is defined on $\top_p M$.*
- *A subset of M is compact if and only if it is closed and bounded.*

Let M be a complete d-dimensional Riemannian manifold. In this case, for all $p \in M$ the exponential map at p is defined on $\top_p M$, and for any value of t, we have $\exp_p(tv) = \gamma_v(t)$.

2.2 Strongly mixing process on Riemannian manifold

In this section, we recall some useful results on mixingales [12] and we introduce the strong mixing condition for Riemannian manifold-valued processes.

Let $\{F_n, n \in \mathbb{Z}\}$ be a non-decreasing sequence of subsigma algebras of the probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{Z_n\}_{n\in\mathbb{Z}}$ a $\mathbb{R}$-valued process. The sequence $(Z_n, F_n)_{\{n\in\mathbb{Z}\}}$ is called a simple mixingale if Z_n is F_n-measurable, $n \geq 1$, and for some sequences of finite nonnegative constants $c_n, \phi_l, l = 0, 1, \ldots$, where $\phi_l \to 0$ as $l \to \infty$ we have

(a) $\left[\mathbb{E}|\mathbb{E}(Z_n|F_{n-l})|^2\right]^{1/2} \leq \phi_l c_n$, for all $n \geq 1, l \geq 0$,
(b) $\mathbb{E}(Z_n) = 0, n \geq 1$,
(c) $F_n = \{\emptyset, \Omega\}, n \leq 0$

$\phi_l, l = 0, 1, 2, \ldots$ are called mixingale numbers. The following lemma gives sufficient conditions on the sequences (ϕ_l) and (c_n) for the almost sure convergence of $S_n = \sum_{j=1}^n Z_j$.

Lemma 2.8 [12] *Let $(Z_n, F_n)_{n\in\mathbb{Z}}$ be a simple mixingale such that*

$$\sum_{j=1}^{\infty} c_j^2 < \infty \tag{2.1}$$

and for some $\delta > 0$,

$$\sum_{n=1}^{\infty} (\log n)(\log_2 n)^{1+\delta} \phi_n^2 \sum_{j=n}^{\infty} c_j^2 < \infty. \tag{2.2}$$

Then $S_n = \sum_{j=1}^n Z_j$ converges almost surely to a finite limit.

Let (M, g) be a Riemannian manifold. A M-valued random variable X is a measurable map on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into $(M, \mathcal{B}(M))$, where $\mathcal{B}(M)$ is the Borel σ-algebra generated by the open subsets of M.

To each M-valued random variable X, we associate a probability measure $Q = \mathbb{P}_X$ on $\mathcal{B}(M)$ given by $Q(B) = \mathbb{P}(X^{-1}(B))$ for all $B \in \mathcal{B}(M)$.

Generally, Riemannian manifolds do not have a linear structure. Thus, the study of Riemannian manifold-valued random variables relies on the geometric structure of this manifold involving most often the notions of metric, volume form, injectivity radius, volume density, etc. For more details on probabilistic and statistic tools to work on Riemannian manifolds, see [19].

In this paper, we study Riemannian manifold-valued processes. We extend the strong mixing condition for real-valued stationary processes [12] to Riemannian manifold-valued stationary processes as follows:

Definition 2.9 (Strongly mixing process) Let (X_j) be a M-valued process and $\mathcal{F}_i^k$, $(-\infty \le i, k \le \infty)$ be the σ-algebra of events generated by $\{X_j, i \le j \le k\}$. (X_j) is strongly mixing if

$$\sup_{s\in\mathbb{R}} \sup_{A\in\mathcal{F}_{-\infty}^s, B\in\mathcal{F}_{k+s}^{\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| = \alpha(k) \to 0$$

as $k \to \infty$.
$\alpha(k)$ is called the strong mixing coefficient.

Lower and upper bounds of the covariance of $\mathcal{F}_{-\infty}^u$-measurable and $\mathcal{F}_{k+u}^{+\infty}$-measurable $\mathbb{R}$-random variables are given below:

Theorem 2.10 [7] *Let r, s and t be positive reals numbers such that $1/r+1/s+1/t = 1$. Let $u \ge 0$, Y be a $\mathcal{F}_{-\infty}^u$-measurable $\mathbb{R}$-random variable and T a $\mathcal{F}_{k+u}^{\infty}$-measurable $\mathbb{R}$-random variable such that $\mathbb{E}|T|^t < \infty$ and $\mathbb{E}|Y|^s < \infty$. Then, there exists $c \in \mathbb{R}$ such that*

$$|cov(Y, T)| \le c[\alpha(k)]^{1/r}[\mathbb{E}|Y|^s]^{1/s}[\mathbb{E}|T|^t]^{1/t}.$$

2.3 Some useful analysis' results

In this part, we recall some relevant results (Lebesgue's dominated convergence theorem, Toeplitz lemma and Kronecker lemma) that will be used later on.

Theorem 2.11 (Lebesgue's dominated convergence) [10] *Let (X, S, μ) be a measure space and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of (complex-valued) integrable functions defined on X, converging pointwise to a function f. Assume further that for all $x \in X$ and for all $n \in \mathbb{N}$, we have $|f_n(x)| \le g(x)$, where g is a nonnegative integrable function defined on X. Then, f is integrable. Further,*

$$\lim_{n\to\infty} \int_X |f_n(x) - f(x)| \,\mathrm{d}\mu(x) = 0.$$

In particular, we have

$$\lim_{n\to\infty} \int_X f_n(x)\mathrm{d}\mu(x) = \int_X f(x)\mathrm{d}\mu(x).$$

Lemma 2.12 (Toeplitz) [11] *Let* $\{a_{nk}, 1 \leq k \leq k_n, n \geq 1\}$ *be a double array of real numbers such that for any* $k \geq 1$, $\lim_{n\to\infty} a_{nk} = 0$ *and* $\sup_{n\geq 1} \sum_{k=1}^{k_n} |a_{nk}| < \infty$. *Let* $\{x_n, n \geq 1\}$ *be a sequence of real numbers.*

(i) *If* $\lim_{n\to\infty} x_n = 0$, *then* $\lim_{n\to\infty} \sum_{k=1}^{k_n} a_{nk}x_n = 0$.
(ii) *If* $\lim_{n\to\infty} x_n = x \in \mathbb{R}$ *and* $\lim_{n\to\infty} \sum_{k=1}^{k_n} a_{nk} = 1$ *then* $\lim_{n\to\infty} \sum_{k=1}^{k_n} a_{nk}x_n = x$.

The following Kronecker lemma is a corollary of Toeplitz lemma.

Lemma 2.13 (Kronecker) [11] *Let* $\{x_n, n \geq 1\}$ *and* $\{b_n, n \geq 1\}$ *be a increasing sequences of real numbers such that* $b_n > 0$ *for all* $n \geq 1$. *If the series* $\sum_{k=1}^{\infty} x_k/b_k$ *converges, then* $\lim_{n\to\infty} \sum_{k=1}^{n} x_k = 0$.

3 Main results

Let (M, g) be a complete d-dimensional Riemannian manifold and $(X_i)_{i\in\mathbb{Z}}$ be a M-valued stationary process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The aim of this work is to investigate the nonparametric estimation of the finite-dimensional density functions, conditional density functions and conditional expectations of functionals given its past behavior of $(X_i)_{i\in\mathbb{Z}}$. Let $n \in \mathbb{N}^*$. For each integer $m \geq 1$ and integers $0 = i_1 < i_2 < \cdots < i_m < n$, let

$$f(\mathbf{x}, \mathbf{i}_m) = f(x_1, \ldots, x_m; i_1, \ldots, i_m)$$

be the density function of the random vector $\mathbf{X}_0^{1,m} = (X_{i_1}, \ldots, X_{i_m})$, which is assumed to exist.
For any integer s, $1 \leq s < m$, set

$$f(\mathbf{x}_1; \mathbf{i}'_s) = f(x_{j+i_1}, \ldots, x_{j+i_s}; i_1, \ldots, i_s)$$

be the probability density function of $\mathbf{X}_j^{1,s} = (X_{j+i_1}, \ldots, X_{j+i_s})$ and

$$f(\mathbf{x}_2|\mathbf{x}_1) := f(\mathbf{x}_2; \mathbf{i}''_{m-s}|\mathbf{x}_1; \mathbf{i}'_s)$$

be the conditional probability density function of $\mathbf{X}_j^{s+1,m} = (X_{j+i_{s+1}}, \ldots, X_{j+i_m})$ given $\mathbf{X}_j^{1,s} = (X_{j+i_1}, \ldots, X_{j+i_s})$ where $\mathbf{i}''_{m-s} = (i_{s+1}, \ldots, i_m)$. Then

$$f(\mathbf{x}_2|\mathbf{x}_1) = \frac{f(\mathbf{x}; \mathbf{i}_m)}{f(\mathbf{x}_1; \mathbf{i}'_s)},$$

where $\mathbf{x}_1 \in M^s$, $\mathbf{x}_2 \in M^{m-s}$. Let $E = B_{\mathbb{R}}(0, 1)^d = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : 0 \leq x_i \leq 1, i \in \{1, \ldots, d\}\}$ and K_l be a real nonnegative function on $\mathbb{R}^l_+$ for all $l \in \{1, \ldots, m\}$. We make the following assumptions:

H1 (i) For all $l \in \{1, 2, \ldots, m\}$, $\text{supp}K_l = [0, 1]^l$ and $\int_{E^l} K_l(\|u_1\|, \ldots, \|u_l\|)du_1 \ldots du_l = 1$,

(ii) There exists $r \in (2, \infty)$ such that $\int_{E^l} K_l^r(\|u_1\|, \ldots, \|u_l\|)du_1 \ldots du_l < \infty, \forall l$,

H2 $f(\cdot; \mathbf{i}_m)$ and $f(\cdot; \mathbf{i}'_s)$ are continuous functions, respectively, on M^m and M^s (with $M^m = M \times \cdots \times M$, m factors),

H3 The function $R^{(r)}(\cdot) = \mathbb{E}[|q(\mathbf{X}_j^{s+1,m})|^r | \mathbf{X}_j^{1,s} = \cdot]$ is continuous on M^s,

H4 For all p in M, the volume density function θ_p is continuous on M and $\inf_{\tilde{p} \in M} \theta_p(\tilde{p}) = A > 0$.

Remark 3.1 Assumption **H4** was used by Henry [9] without the continuity of the volume density function.

Let $\left(h_j\right)_{j \in \mathbb{N}^*}$ be a sequence of positive numbers such that $h_j \longrightarrow 0$ as $j \longrightarrow \infty$ and put

$$K_{l,j}(x) = \frac{1}{h_j^{dl}} K_l \left(\frac{x}{h_j}\right).$$

We suppose that the bandwidth satisfies the following condition:

$$h_j \leq h_0 \tag{3.1}$$

for some fixed h_0 such that $0 < h_0 < \text{inj}_g(M)$, where $\text{inj}_g(M)$ denotes the injectivity radius of the Riemannian manifold (M, g) (see [2]).

3.1 The proposed Kernel density estimator

On the basis of a single realization $(X_i)_{i=1}^n$, we estimate $f(\mathbf{x}; \mathbf{i}_m)$ by

$$\hat{f}_n(\mathbf{x}; \mathbf{i}_m) = \frac{1}{n - i_m} \sum_{j=1}^{n-i_m} \frac{K_{m,j}\left(d(x_1, X_{j+i_1}), \ldots, d(x_m, X_{j+i_m})\right)}{\theta_{x_1}(X_{j+i_1}) \times \cdots \times \theta_{x_m}(X_{j+i_m})}. \tag{3.2}$$

Remark 3.2 – Our density estimator defined by (3.2) is an extension of Masry's (see [13]) density estimator to Riemannian manifold case.

– In Euclidian space $\mathbb{R}^d$, $\theta_p(\tilde{p}) = 1$ for all p and $\tilde{p}$ in M. Then our density estimator function takes the form:

$$\hat{f}_n^*(\mathbf{x}; \mathbf{i}_m) = \frac{1}{n - i_m} \sum_{j=1}^{n-i_m} K_{m,j}\left(\|x_1 - X_{j+i_1}\|, \ldots, \|x_m - X_{j+i_m}\|\right)$$

for $\mathbf{x} \in E^m$.

The function $\hat{f}_n(\cdot; \mathbf{i}_m)$ is a probability density function as stated in the following lemma.

Lemma 3.3 *The estimator* $\hat{f}_n$ *defined by* (3.2) *is a probability density on* M^m*.*

Proof We have $\hat{f}_n(\mathbf{x}; \mathbf{i}_m) \geq 0$ for all $\mathbf{x} \in M^m$. Therefore,

$$\int_{M^m} \hat{f}_n(x_1, \ldots, x_m; i_1, \ldots, i_m) \mathrm{d}v_g(x_1) \cdots \mathrm{d}v_g(x_m) = \frac{1}{n - i_m} \sum_{j=1}^{n-i_m} \left(\int_{\prod_{k=1}^m B\left(X_{j+i_k, h_j}\right)} \frac{K_{m,j}\left(d(x_1, X_{j+i_1}), \ldots, d(x_m, X_{j+i_m})\right)}{\theta_{x_1}(X_{j+i_1}) \cdots \theta_{x_m}(X_{j+i_m})} \mathrm{d}v_g(x_1) \cdots \mathrm{d}v_g(x_m) \right)$$

where v_g is the Riemannian measure on M. Let $\left(U_{X_{j+i_k}}, \exp^{-1}_{X_{j+i_k}}\right)$ be the exponential chart at X_{j+i_k} for all k, with normal coordinates $x^1, \ldots, x^d$. Under the condition (3.1) on the bandwidths, $\prod_{k=1}^m B\left(X_{j+i_k}, h_j\right) \subset \prod_{k=1}^m U_{X_{j+i_k}}$. Let $B(h_j)$ be the ball of radius h_j in $U_{X_{j+i_k}}$, i.e., $B(h_j) = \exp^{-1}_{X_{j+i_k}}\left(B\left(X_{j+i_k}, h_j\right)\right)$. Then by (3.2) we get

$$\begin{aligned}
&\int_{M^m} \hat{f}_n(x_1, \ldots, x_m; i_1, \ldots, i_m) \mathrm{d}v_g(x_1) \cdots \mathrm{d}v_g(x_m) \\
&= \frac{1}{n - i_m} \sum_{j=1}^{n-i_m} \int_{E^m} K_m\left(\|u_1\|, \ldots, \|u_m\|\right) \\
&\times \frac{\sqrt{\det g(u_1)} \times \cdots \times \sqrt{\det g(u_m)}}{\theta_{\exp_{X_{j+i_1}}(u_1)}(X_{j+i_1}) \times \cdots \times \theta_{\exp_{X_{j+i_m}}(u_m)}(X_{j+i_m})} du \\
&= \frac{1}{n - i_m} \sum_{j=1}^{n-i_m} \int_{E^m} K_m\left(\|u_1\|, \ldots, \|u_m\|\right) du_1 \cdots du_m \\
&= \int_{E^m} K_m\left(\|u_1\|, \ldots, \|u_m\|\right) du_1 \ldots du_m. \qquad (3.3)
\end{aligned}$$

Using **H1** and (3.3), we obtain

$$\int_{M^m} \hat{f}_n(x_1, \ldots, x_m; i_1, \ldots, i_m) \mathrm{d}v_g(x_1) \cdots \mathrm{d}v_g(x_m) = 1.$$

□

3.2 Strong consistency of the density estimator

Now, we investigate the asymptotic behavior of the nonparametric estimator of the finite-dimensional density functions and conditional density functions of $(X_i)_{i \in \mathbb{Z}}$. We start by the two following results that are useful to establish the strong consistency of the considered estimators.

Theorem 3.4 *Let* $(X_j)_{j\in\mathbb{Z}}$ *be a* M*-valued stationary and strongly mixing process,*

$$\boldsymbol{X}_j^{1,m} = (X_{j+i_1}, \ldots, X_{j+i_m}), \quad 0 = i_1 < i_2 < \cdots < i_m,$$

and (ϕ_j) *a sequence of* $\mathbb{R}$*-valued Borel measurable functions on* M^m*. Set*

$$Y_j = \phi_j\left(\boldsymbol{X}_j^{1,m}\right) - \mathbb{E}\left[\phi_j\left(\boldsymbol{X}_j^{1,m}\right)\right] \quad \text{and} \quad S_n = \sum_{j=1}^{n} Y_j.$$

If

$$\sum_{j=1}^{\infty}\{\mathbb{E}|Y_j|^r\}^{2/r} < \infty, \tag{3.4}$$

and, for some $\delta > 0$,

$$\sum_{k=1}^{\infty}[\log(k+i_m)][\log_2(k+i_m)]^{1+\delta}[\alpha(k)]^{1-2/r} \times \sum_{j=k+i_m}^{\infty}\{\mathbb{E}|Y_j|^r\}^{2/r} < \infty, \tag{3.5}$$

then S_n *converges almost surely to a finite limit as* $n \to \infty$*.*

Proof Let $F_n = \begin{cases} \mathcal{F}_1^{n+i_m}, & \text{if} \quad n \geq 1 \\ \{\emptyset, \Omega\}, & \text{if} \quad n \leq 0 \end{cases}$.

We shall show that (Y_n, F_n) is a mixingale. Y_n is F_n-measurable for $n \geq 1$. Let $l \in \mathbb{N}$.

- If $n - l \leq 0$, $\mathbb{E}(Y_n|F_{n-l}) = 0$ because $\mathbb{E}(Y_n) = 0$ and $F_{n-l} = \{\emptyset, \Omega\}$. Then $\mathbb{E}|\mathbb{E}(Y_n|F_{n-l})|^2 = 0$.
- If $l = 0$, $\mathbb{E}|\mathbb{E}(Y_n|F_{n-l})|^2 = \mathbb{E}|\mathbb{E}(Y_n|F_n)|^2 = \mathbb{E}|Y_n|^2$.
- If $1 \leq l \leq i_m$, we have $n - i_m \leq n - l \leq n - 1$. By the Cauchy–Schwartz inequality, we have

$$\left(\mathbb{E}|\mathbb{E}(Y_n|F_{n-l})|^2\right)^{1/2} \leq \left(\mathbb{E}|Y_n|^2\right)^{1/2}.$$

- If $i_m < l < n$, we have $0 < n - l < n - i_m$ and

$$\left(\mathbb{E}|\mathbb{E}(Y_n|F_{n-l})|^2\right) = cov\left[Y_n, \mathbb{E}(Y_n|F_{n-l})\right].$$

  Since Y_n is $\mathcal{F}_n^{\infty}$-measurable and $\mathbb{E}(Y_n|F_{n-l})$ is $\mathcal{F}_{-\infty}^{n+i_m-1}$-measurable, it follows by Theorem 2.10

$$\left(\mathbb{E}|\mathbb{E}(Y_n|F_{n-l})|^2\right)^{1/2} \leq c[\alpha(l-i_m)]^{1/2-1/r}[\mathbb{E}|Y_n|^r]^{1/r}.$$

Thus, (Y_n, F_n) is a simple mixingale with $c_n = 12[\mathbb{E}|Y_n|^r]^{1/r}$ and

$$\phi_l = \begin{cases} 1, & 1 \le l \le i_m, \\ [\alpha(l - i_m)]^{1/2-1/r}, & i_m < l < n, \\ 0, & n \le l, \end{cases}$$

and, by (3.4) and (3.5), conditions (2.1) and (2.2) of Lemma 2.8 are satisfied. Hence the result. □

Lemma 3.5 *Let **H1**, **H2** and **H4** hold. Then,*

$$\begin{aligned} &\lim_{n\to\infty} h_n^{md(r-1)} \\ &\quad\times \int_{M^m} \left| \frac{K_{m,n}\left(d(x_1, y_1), \ldots, d(x_m, y_m)\right)}{\theta_{x_1}(y_1) \times \cdots \times \theta_{x_m}(y_m)} \right|^r f(y_1, \ldots, y_m) \mathrm{d}v_g(y) \\ &= \frac{f(x_1, \ldots, x_m)}{\theta_{x_1}^{r-1}(x_1) \times \cdots \times \theta_{x_m}^{r-1}(x_m)} \\ &\quad\times \int_{E^m} \left| K_m\left(\|u_1\|, \ldots, \|u_m\|\right)\right|^r du_1 \cdots du_m < \infty. \end{aligned} \tag{3.6}$$

Proof For $k = 1, \ldots, m$, set $y_k = \exp_{x_k}(u_k)$ and $v_k = u_k/h_k$. By **H1-(ii)**, there exists $r > 2$ such that

$$\int_{E^m} \left| K_m\left(\|u_1\|, \ldots, \|u_m\|\right)\right|^r du_1 \cdots du_m < \infty.$$

We have

$$\begin{aligned} &h_n^{md(r-1)} \int_{M^m} \left| \frac{K_{m,n}\left(d(x_1, y_1), \ldots, d(x_m, y_m)\right)}{\theta_{x_1}(y_1) \times \cdots \times \theta_{x_m}(y_m)} \right|^r \\ &\quad\times f(y_1, \ldots, y_m) \mathrm{d}v_g(y_1) \cdots \mathrm{d}v_g(y_m) \\ &= \int_{E^m} \frac{\left| K_m\left(\|v_1\|, \ldots, \|v_m\|\right)\right|^r}{\left|\theta_{x_1}(\exp_{x_1}(h_1 v_1)) \times \cdots \times \theta_{x_m}(\exp_{x_m}(h_n v_m))\right|^{r-1}} \\ &\quad\times f(\exp_{x_1}(h_n v_1), \ldots, \exp_{x_m}(h_n v_m)) \mathrm{d}v_1 \cdots \mathrm{d}v_m. \end{aligned}$$

By the continuity of $\exp_p$ on M for all $p \in M$, Assumptions **H1**-(ii), **H2**, **H4** and Lebesgue's dominated convergence theorem we get the result. □

The results of strong consistency of the proposed estimators are given below.

Theorem 3.6 *Let **H1**, **H2** and **H4** hold. Let $(X_j)_{j\in\mathbb{Z}}$ be a M-valued strongly mixing process. If the strong mixing coefficient $\alpha(k)$ satisfies, for some $\delta > 0$,*

$$\sum_{k=1}^{\infty} [\log(k + i_m)][\log_2(k + i_m)]^{1+\delta}[\alpha(k)]^{1-2/r}$$

$$\times \sum_{j=k+i_m}^{\infty} \frac{1}{j^2 h_j^{2md(1-1/r)}} < \infty \tag{3.7}$$

and the sequence (h_j) *satisfies*

$$\sum_{j=1}^{\infty} \frac{1}{j^2 h_j^{2md(1-1/r)}} < \infty, \tag{3.8}$$

then for almost all $\boldsymbol{x} \in M^m$ *we have*

$$\hat{f}_n(\boldsymbol{x}; \boldsymbol{i}_m) - f(\boldsymbol{x}; \boldsymbol{i}_m) \to 0, \quad n \to \infty$$

almost surely.

Proof We show the consistency result in two steps as follows:
Step 1: We first prove that for almost all $\mathbf{x} \in M^m$ we have

$$\hat{f}_n(\mathbf{x}; \mathbf{i}_m) - \mathbb{E}\left(\hat{f}_n(\mathbf{x}; \mathbf{i}_m)\right) \to 0, \quad n \to \infty$$

almost surely. Using Lemma 3.5, we have

$$\begin{aligned} h_j^{md(r-1)} \mathbb{E} & \left| \frac{K_{m,j}\left(d(x_1, X_{j+i_1}), \ldots, d(x_m, X_{j+i_m})\right)}{\theta_{x_1}(X_{j+i_1}) \times \cdots \times \theta_{x_m}(X_{j+i_m})} \right|^r \\ & \to \frac{f(x_1, \ldots, x_m)}{\theta_{x_1}^{r-1}(x_1) \times \cdots \times \theta_{x_m}^{r-1}(x_m)} \\ & \times \int_{E^m} |K_m(\|u_1\|, \ldots, \|u_m\|)|^r \, du_1 \cdots du_m \end{aligned} \tag{3.9}$$

for almost all x as $j \to \infty$. Fix x and put

$$\begin{aligned} \psi_j(y) &= \frac{1}{j} \frac{K_{m,j}(d(x_1, y_1), \ldots, d(x_m, y_m))}{\theta_{x_1}(y_1) \times \cdots \times \theta_{x_m}(y_m)}, \quad y \in M^m \\ Y_j &= \psi_j(\mathbf{X}_j^{1,n}) - \mathbb{E}\left[\psi_j(\mathbf{X}_j^{1,n})\right]. \end{aligned}$$

By (3.9), there exists a constant β, independent of j, such that

$$[\mathbb{E}|Y_j|^r]^{1/r} \le \frac{1}{j h_j^{md(1-1/r)}} \beta.$$

Hence, by (3.8), we have $\sum_{j=1}^{\infty} [\mathbb{E}|Y_j|^r]^{2/r} < \infty$ and (3.7), (3.8) implying that conditions (3.4) and (3.5) of Theorem 3.4 are satisfied. It follows by Theorem 3.4 that

$S_n = \sum_{j=1}^{n} Y_j$ converges almost surely to a finite limit. Now set $n' = n - i_m$. Then

$$\hat{f}_n(\mathbf{x}; \mathbf{i}_m) - \mathbb{E}\left(\hat{f}_n(\mathbf{x}; \mathbf{i}_m)\right) = \frac{1}{n'} \sum_{j=1}^{n'} jY_j$$

which tends to zero by the Kronecker lemma and the almost sure convergence of S_n. **Step 2:** We show in this second part that for almost all $\mathbf{x} \in M^m$ we have

$$\lim_{n\to\infty} \mathbb{E}\left(\hat{f}_n(\mathbf{x}; \mathbf{i}_m)\right) = f(\mathbf{x}; \mathbf{i}_m).$$

By computing the expectation of $\hat{f}_n(\mathbf{x}; \mathbf{i}_m)$, we have

$$\begin{aligned} \mathbb{E}\left(\hat{f}_n(\mathbf{x}; \mathbf{i}_m)\right) &= \frac{1}{n - i_m} \sum_{j=1}^{n-i_m} \int_{M^m} \frac{K_{m,j}\left(d(x_1, y_1), \ldots, d(x_m, y_m)\right)}{\theta_{x_1}(y_1) \times \cdots \times \theta_{x_m}(y_m)} \\ &\quad \times f(y_1, \ldots, y_m) \mathrm{d}v_g(y) \\ &= \frac{1}{n - i_m} \sum_{j=1}^{n-i_m} \varphi_j(x_1, \ldots, x_m), \end{aligned}$$

where

$$\begin{aligned} \varphi_j(x_1, \ldots, x_m) = \int_{M^m} & \frac{K_{m,j}\left(d(x_1, y_1), \ldots, d(x_m, y_m)\right)}{\theta_{x_1}(y_1) \times \cdots \times \theta_{x_m}(y_m)} \\ & \times f(y_1, \ldots, y_m) \mathrm{d}v_g(y_1) \cdots \mathrm{d}v_g(y_m). \end{aligned}$$

Similarly to the proof of Lemma 3.5, we can show that $\varphi_j(x_1, \ldots, x_m) \to f(x_1, \ldots, x_m)$ for almost all $(x_1, \ldots, x_m) \in M^m$. Moreover, by Toeplitz lemma we have

$$\lim_{n\to\infty} \mathbb{E}\left(\hat{f}_n(\mathbf{x}; \mathbf{i}_m)\right) = f(\mathbf{x}; \mathbf{i}_m).$$

Using **Steps 1–2**, the strong consistency result of $\hat{f}_n(\mathbf{x}; \mathbf{i}_m)$ is obtained. □

Theorem 3.7 *Let **H1**, **H2** and **H4** hold. Let (X_j) be a M-valued strongly mixing process. If the strong mixing coefficient $\alpha(k)$ satisfies, for some $\delta > 0$,*

$$\begin{aligned} & \sum_{k=1}^{\infty} \log(k + i_m)[\log_2(k + i_m)]^{1+\delta} [\alpha(k)]^{1-2/r} \\ & \quad \times \sum_{j=k+i_m}^{\infty} \frac{1}{j^2 h_j^{2md(1-1/r)}} < \infty \end{aligned} \tag{3.10}$$

and the sequence (h_j) *satisfies*

$$\sum_{j=1}^{\infty} \frac{1}{j^2 h_j^{2md(1-1/r)}} < \infty, \tag{3.11}$$

then for almost all $\boldsymbol{x}_1 \in M^s$ *and* $\boldsymbol{x}_2 \in M^{m-s}$ *with* $f(\boldsymbol{x}_1; \boldsymbol{i}'_s) > 0$ *we have*

$$\hat{f}_n(\boldsymbol{x}_2|\boldsymbol{x}_1) - f(\boldsymbol{x}_2|\boldsymbol{x}_1) \to 0, \quad n \to \infty \tag{3.12}$$

almost surely.

Proof Under (3.10) and (3.11), we have Theorem 3.6 so that for almost all $\mathbf{x}$:

$$\hat{f}_n(\mathbf{x}; \mathbf{i}_m) \to f(\mathbf{x}; \mathbf{i}_m), \quad n \to \infty$$

almost surely. (3.10) and (3.11) also imply that

$$\sum_{k=1}^{\infty} [\log(k + i_s)][\log_2(k + i_s)]^{1+\delta} [\alpha(k)]^{1-2/r} \sum_{j=k+i_s}^{\infty} \frac{1}{j^2 h_j^{2sd(1-1/r)}} < \infty$$

and

$$\sum_{j=1}^{\infty} \frac{1}{j^2 h_j^{2sd(1-1/r)}} < \infty. \tag{3.13}$$

By Theorem 3.6, for almost all $\mathbf{x}_1 \in M^s$

$$\hat{f}_n(\mathbf{x}_1; \mathbf{i}_s) \to f(\mathbf{x}_1; \mathbf{i}_s), \quad n \to \infty$$

almost surely. The result follows by using the following identity

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \frac{1}{\hat{b}} \left[(\hat{a} - a) - \frac{a}{b}(\hat{b} - b) \right] \tag{3.14}$$

with $a = f(\mathbf{x}; \mathbf{i}_s), \quad \hat{a} = \hat{f}_n(\mathbf{x}; \mathbf{i}_s), b = f(\mathbf{x}_1; \mathbf{i}'_s), \quad \hat{b} = \hat{f}_n(\mathbf{x}_1; \mathbf{i}'_s)$. □

3.3 Strong consistency of conditional expectations of functionals

We now focus on the nonparametric estimation of functional conditional expectations. Let q be a $\mathbb{R}-$valued Borel measurable function on M^{m-s}. The conditional expectations of $q(\mathbf{X}_j^{s+1,m})$ given $(\mathbf{X}_j^{1,s} = u)$ is denoted by

$$Q(u) = \mathbb{E}\left[q(\mathbf{X}_j^{s+1,m}) | (\mathbf{X}_j^{1,s} = u) \right].$$

Define

$$R(u) = Q(u) f(u; \mathbf{i}'_s)$$

and

$$\hat{R}_n(u) = \frac{1}{n - i_m} \sum_{j=1}^{n-i_m} q(\mathbf{X}_j^{s+1,m}) \frac{K_{s,j}\left(d(x_1, X_{j+i_1}), \ldots, d(x_s, X_{j+i_s})\right)}{\theta_{x_1}(X_{j+i_1}) \times \cdots \times \theta_{x_s}(X_{j+i_s})}.$$

Then we consider the following estimator of Q:

$$\hat{Q}_n(u) = \hat{R}_n(u) / \hat{f}_n(u; \mathbf{i}'_s).$$

We shall first establish the strong consistency of $\hat{R}_n(u)$ and then deduce the asymptotic result on the conditional expectations estimator $\hat{Q}_n(u)$.

Theorem 3.8 *Let* ***H1–H4*** *hold. Let (X_j) be a M-valued strongly mixing process. If the strong mixing coefficient $\alpha(k)$ satisfies, for some $\delta > 0$,*

$$\sum_{k=1}^{\infty} [\log(k + i_s)][\log_2(k + i_s)]^{1+\delta} [\alpha(k)]^{1-2/r} \times \sum_{j=k+i_s}^{\infty} \frac{1}{j^2 h_j^{2sd(1-1/r)}} < \infty \tag{3.15}$$

and the sequence (h_j) satisfies

$$\sum_{j=1}^{\infty} \frac{1}{j^2 h_j^{2sd(1-1/r)}} < \infty; \tag{3.16}$$

then, for almost all $u \in M^s$ with $f(u; \mathbf{i}'_s) > 0$ we have

$$\hat{Q}_n(u) \to Q(u) \tag{3.17}$$

as $n \to \infty$ almost surely.

Proof We prove this theorem by two steps.
Step 1: We show that for almost all $u \in M^s$ we have $\hat{R}_n(u) - \mathbb{E}[\hat{R}_n(u)] \to 0$.
By **H2** and **H3**, the function $u \mapsto R^{(r)}(u) f(u; \mathbf{i}'_s)$ is a bounded and continuous function on M^s. By Lebesgue's dominated convergence theorem, we have for almost all $u \in M^s$,

$$h_j^{dh(r-1)} \mathbb{E} \left| q(\mathbf{X}_j^{s+1,m}) \frac{K_{s,j}\left(d(u_1, X_{j+i_1}), \ldots, d(u_s, X_{j+i_s})\right)}{\theta_{u_1}(X_{j+i_1}) \times \cdots \times \theta_{u_s}(X_{j+i_s})} \right|^r$$

$$= h_j^{dh(r-1)}\mathbb{E}\left\{R^{(r)}(\mathbf{X}_j^{1,s})\left|\frac{K_{s,j}\left(d(u_1,X_{j+i_1}),\ldots,d(u_s,X_{j+i_s})\right)}{\theta_{u_1}(X_{j+i_1})\times\cdots\times\theta_{u_s}(X_{j+i_s})}\right|^r\right\}$$

$$\to \frac{R^{(r)}(u)f(u;\mathbf{i}'_s)}{\theta_{u_1}^{r-1}(u_1)\cdots\theta_{u_m}^{r-1}(u_m)}\int_E^s K^r(\|v_1\|,\ldots,\|v_s\|)\mathrm{d}v_g(v_1)\cdots\mathrm{d}v_g(v_s)$$

as $j\to\infty$. Put

$$\psi_j(y_1,y_2)=\frac{1}{j}q(y_2)\frac{K_{s,j}\left(d(u_1,y_{11}),\ldots,d(u_s,y_{1s})\right)}{\theta_{u_1}(y_{11})\times\cdots\times\theta_{u_s}(y_{1s})},$$

$$Y_j=\psi_j(\mathbf{X}_j^{1,s},\mathbf{X}_j^{s+1,m})-\mathbb{E}\left[\psi_j(\mathbf{X}_j^{1,s},\mathbf{X}_j^{s+1,m})\right],\quad S_n=\sum_{j=1}^{n}Y_j(3.18)$$

Now we proceed as in the first part of the proof of Theorem 3.6 and we have for almost all $u\in M^s$ we have $\hat{R}_n(u)-\mathbb{E}[\hat{R}_n(u)]\to 0$.

Step 2: We show in this second step that for almost all $u\in M^s$ we have $\lim_{n\to\infty}\mathbb{E}[\hat{R}_n(u)]=R(u)$. Put

$$\phi_j(u_1,\ldots,u_s)=\int_{M^s}Q(x_1,\ldots,x_s)f(x_1,\ldots,x_s)$$
$$\times\frac{K_{s,j}\left(d(u_1,x_1),\ldots,d(u_s,x_s)\right)}{\theta_{x_1}(x_1)\cdots\theta_{u_s}(x_s)}\mathrm{d}v_g(x),$$

we have

$$\mathbb{E}[\hat{R}_n(u)]=\frac{1}{n-i_m}\sum_{k=1}^{n-i_m}\phi_j(u_1,\ldots,u_s).$$

Using Assumption **H3** and Lebesgue's dominated convergence theorem, we have $\phi_j(u_1,\ldots,u_s)\to q(u_1,\ldots,u_s)f(u_1,\ldots,u_s;\mathbf{i}'_s)=R(u)$. Moreover, by Toeplitz lemma we have

$$\lim_{n\to\infty}\mathbb{E}[\hat{R}_n(u)]=R(u).$$

Put $a=R(u)$, $\hat{a}=\hat{R}_n(u)$, $b=f(u;\mathbf{i}'_s)$, $\hat{b}=\hat{f}_n(u;\mathbf{i}'_s)$. By the two previous steps, $\hat{a}-a\to 0$ a.s. and $\hat{b}-b\to 0$ a.s. from Theorem 3.6. Using identity (3.14), we get the result. □

4 Numerical studies

In this section, we report simulations carried out to demonstrate the applicability of our geometric kernel density estimator (3.2) and to investigate its finite sample performance under the statistical software *R* [22]. Let us consider the unit disk $\mathbb{D}^2=$

$\{(x, y) \in \mathbb{R}^2, x^2 + y^2 < 1\}$ of the Euclidean plane, which is not an Euclidean space. Endowed with the Riemannian metric:

$$ds^2_{\mathbb{D}} = 4\frac{\mathrm{d}x^2 + dy^2}{(1 - x^2 - y^2)^2},$$

where x and y are the Cartesian coordinates and $\mathbb{D}^2$ is a model of the two-dimensional hyperbolic geometry $\mathbb{D}^2$ which is a complete manifold called the Poincaré disk. The exponential map $\exp_{\mathbf{0}} : \top_{\mathbf{0}}\mathbb{D}^2 \to \mathbb{D}^2$ and the logarithmic map $\log_{\mathbf{0}} : \mathbb{D}^2 \to \top_{\mathbf{0}}\mathbb{D}^2$ (see [8]) are given for $v \neq \mathbf{0}$ and $y \neq \mathbf{0}$ by:

$$\exp_{\mathbf{0}}(v) = \tanh(\|v\|)\frac{v}{\|v\|},$$
$$\log_{\mathbf{0}}(y) = \tanh^{-1}(\|y\|)\frac{y}{\|y\|},$$

and $\exp_{\mathbf{0}}(\mathbf{0}) = \mathbf{0}$ where $\mathbf{0} = (0, 0)$. On the hyperbolic space $\mathbb{D}^2$, the volume density function θ_p, $p \in \mathbb{D}^2$ is given by

$$\theta_p(q) = \frac{\sinh(r)}{r}, \tag{4.1}$$

where $r = d(p, q) = 2\tanh^{-1}\left|\frac{p-q}{p\bar{q}-1}\right|$ (see [15]). Using the relations (3.2) and (4.1), our density function estimator on $\mathbb{D}^2$ is defined by:

$$\hat{f}_n(\mathbf{x}; \mathbf{i}_m) = \frac{1}{n - i_m}$$
$$\times \sum_{j=1}^{n-i_m} \frac{d\left(x_1, X_{j+i_1}\right) \times \cdots \times d\left(x_m, X_{j+i_m}\right) K_{m,j}\left(d(x_1, X_{j+i_1}), \ldots, d(x_m, X_{j+i_m})\right)}{\sinh\left(d\left(x_1, X_{j+i_1}\right)\right) \times \cdots \times \sinh\left(d\left(x_m, X_{j+i_m}\right)\right)}$$

for $\mathbf{x} \in M^m$. We assume in this section that $h_j = h$ not depend on j. Then

$$\sum_{j=1}^{\infty} \frac{1}{j^2 h_j^{2md(1-1/r)}} = h^{-2md(1-1/r)} \sum_{j=1}^{\infty} \frac{1}{j^2} < \infty.$$

The Gaussian probability distribution function (p.d.f) on the complete Riemannian manifold $\mathbb{D}^2$ is not exactly the p.d.f in the Euclidean space. The main difference is the use of the hyperbolic distance in the exponent and a specific normalization constant which accounts for the underlying geometry (see [23]). The Gaussian probability distribution on Poincaré disk $\mathbb{D}^2$ is defined by

$$h(\mathbf{x}|\mu, \sigma) = \frac{1}{Z(\sigma)} \exp\left\{-\frac{d^2(\mathbf{x}, \mu)}{2\sigma^2}\right\}, \quad \mathbf{x} \in \mathbb{D}^2 \tag{4.2}$$

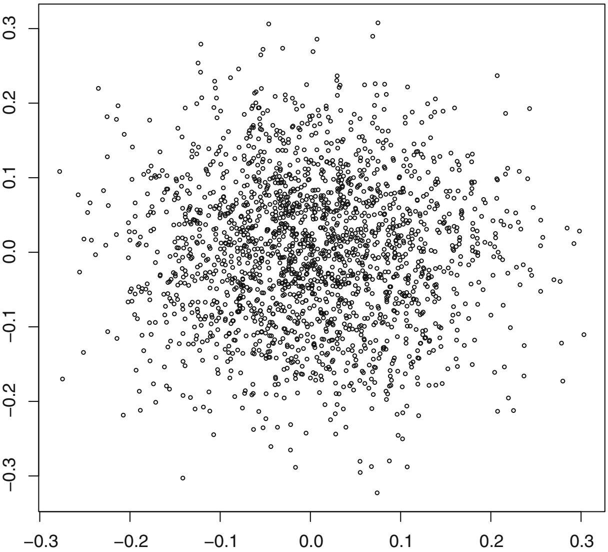


Fig. 1 Sample from Riemannian Gaussian distribution with $\mu = \mathbf{0}$ and $\sigma = 0.2$

where

$$Z(\sigma) = 2\pi\sqrt{\frac{\pi}{2}}\sigma \exp\left(\frac{\sigma^2}{2}\right)\mathrm{erf}\left(\frac{\sigma}{\sqrt{2}}\right)$$

and erf is the Gaussian error function defined by $erf(x) = \frac{1}{\sqrt{\pi}} \int\limits_{-x}^{x} e^{-t^2}\mathrm{d}t$. For more details of the construction of Gaussian sample on $\mathbb{D}^2$, see [15]. Figure 1 provides a plot of a Gaussian random variable sample on the complete Riemannian manifold $\mathbb{D}^2$.

4.1 Simulation 1

Let X be a Gaussian random variable with density probability function $h(\cdot|\mathbf{0}, 0.2)$ defined by (4.2) on Poincaré disk $\mathbb{D}^2$. For fixed $\mathbf{x} = (-0.03, 0.2)$ or $(0.3, 0.1)$, we compute $h(x|\mathbf{0}, 0.2)$. To study the finite sample behavior of the geometric kernel density estimator $\hat{f}_n$, we consider $(X_i)_{i=1}^n$ i.i.d. sample of X for $n = 20, 50, 100, 200,$ 500 and 1000. Since $(X_i)_{i=1}^n$ are i.i.d., the assumptions of Theorem 3.6 are satisfied. For each sample size, we generate 2000 replications and compute $\hat{\theta}_i = \hat{f}_n(\mathbf{x}; \mathbf{i}_1)$ for each sample using the standard kernel $K(x) = \frac{3}{\pi}\left(1 - x^2\right)^2 \mathbf{1}_{x<1}$. Next, we evaluate the mean value obtained over the 2000 replications and the estimated MSE

$$\mathrm{MSE}\left(\hat{\theta}\right) = \frac{1}{2000}\sum_{i=1}^{2000}\left(\hat{\theta}_i - \theta_0\right)^2, \tag{4.3}$$

Table 1 Values of $\hat{f}_n(\mathbf{x})$ and MSE for $\mathbf{x} = (-0.03, 0.2)$, $h(\mathbf{x}|\mu = \mathbf{0}, 0.2) = 0.4795429$

	$\hat{f}_n(\mathbf{x})$	$MSE\left(\hat{f}_n(\mathbf{x})\right)$
$n = 20$	0.4725268	0.001331011
$n = 50$	0.4769959	0.000473344
$n = 100$	0.4783286	0.0002252459
$n = 200$	0.4789286	0.0001136168
$n = 500$	0.4793103	0.00004327218
$n = 1000$	0.4794222	0.00002252942

Table 2 Values of $\hat{f}_n(\mathbf{x})$ and MSE for $\mathbf{x} = (0.3, 0.1)$, $h(\mathbf{x}|\mu = \mathbf{0}, 0.2) = 0.018434481$

	$\hat{f}_n(\mathbf{x})$	$MSE\left(\hat{f}_n(\mathbf{x})\right)$
$n = 20$	0.040968047	0.009645911
$n = 50$	0.036159158	0.004181026
$n = 100$	0.031332491	0.001987973
$n = 300$	0.0244459821	0.0009588741
$n = 500$	0.0227981802	0.0004209544
$n = 1000$	0.0189908596	0.00002017425

where $\theta_0 = h(\mathbf{x}|\mu = \mathbf{0}, 0.2)$ and $\hat{\theta}_i = \hat{f}_n(\mathbf{x}; \mathbf{i}_1)$ for fixed $\mathbf{x} = (-0.03, 0.2)$ or $(0.3, 0.1)$. For $\mathbf{x} = (-0.03, 0.2)$, the true value of the density function is $h(\mathbf{x}|\mu = \mathbf{0}, 0.2) = 0.4795429$ and for $\mathbf{x} = (0.3, 0.1)$, $h(\mathbf{x}|\mu = \mathbf{0}, 0.2) = 0.018434481$. Tables 1 and 2 show the mean value of $\hat{f}_n(\mathbf{x})$ and the MSE value for each sample size, respectively, when $\mathbf{x} = (-0.03, 0.2)$ and $(0.3, 0.1)$. From the results in Tables 1 and 2, we can observe that as the size of the sample increases, $\hat{f}_n(\mathbf{x})$ takes values that are closer and closer to the true value $h(\mathbf{x}|\mathbf{0}, 0.2)$ and the MSE decreases. Hence, our density estimator converges to the true value.

In order to compare the geometric and Euclidean approach for density estimation, we represent the contour of the geometric kernel density function estimate on Riemannian manifold $\mathbb{D}^2$ (with standard kernel K) (Fig. 2) and that of the classical (Euclidean) kernel density estimates on $\mathbb{R}^2$ (with standard $\mathbb{R}^2$-Gaussian kernel) (Fig. 3), evaluated from the same data. The contour plot of the classical kernel density estimate is not symmetric, while the proposed kernel density estimate has the characteristic symmetric shape of Gaussian density (see Fig. 4). Hence, the geometric approach for kernel density estimation yields better results when the data belong to the Riemaniann manifold $\mathbb{D}^2$.

4.2 Simulation 2

We consider the $\mathbb{D}^2$-valued process $\left(\exp_{\mathbf{0}}(Y_t)\right)_{t\in\mathbb{Z}}$ where $(Y_t)_{t\in\mathbb{Z}}$ is the $\mathbb{R}^2$-valued vector autoregressive (VAR) stationary and strongly mixing process defined by

$$Y_t = AY_{t-1} + \epsilon_t \tag{4.4}$$

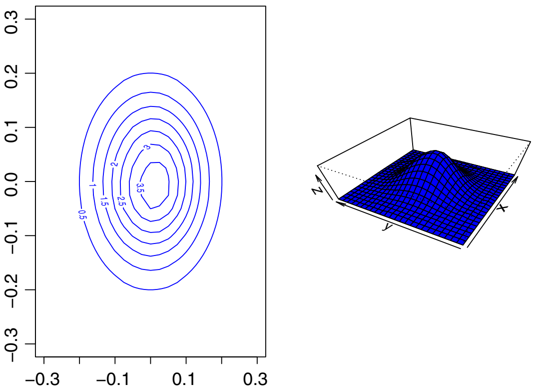


Fig. 2 Geometric density estimation

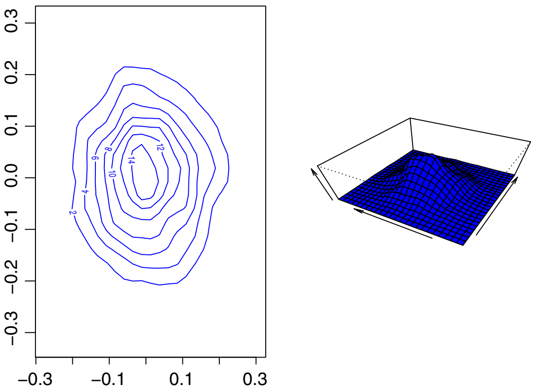


Fig. 3 Euclidean density estimation

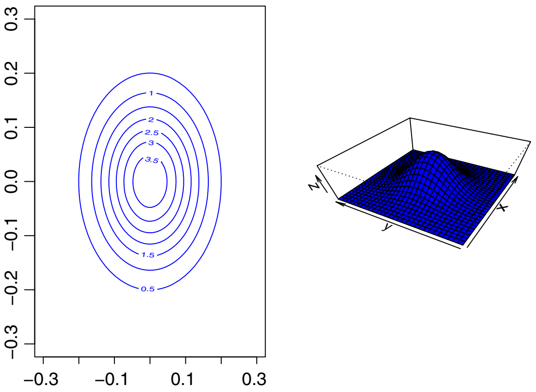


Fig. 4 True probability density $h(\cdot|\mu = \mathbf{0}, 0.2)$

with $A = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}$ and $\epsilon_t \sim \mathcal{N}_2(0, \varSigma)$ are iid, $\varSigma = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$. We set

$$X_t = \exp_{\mathbf{0}}(Y_t)\,. \tag{4.5}$$

Table 3 Values of $\hat{f}_n(\mathbf{x})$ and MSE for $\mathbf{x} = (0.5, 0.6)$, $f(\mathbf{x}) = 0.004369456$

	$\hat{f}_n(\mathbf{x})$	$MSE\left(\hat{f}_n(\mathbf{x})\right)$
$n = 20$	0.005951399	0.00002926904
$n = 50$	0.005279744	0.00001278535
$n = 100$	0.004866229	0.000006562884
$n = 200$	0.004679187	0.000003545061
$n = 500$	0.004506732	0.000001444091
$n = 1000$	0.00437981	0.0000007181330

Table 4 Values of $\hat{f}_n(\mathbf{x})$ and MSE for $\mathbf{x} = (-0.4, 0.7)$, $f(\mathbf{x}) = 0.001921427$

	$\hat{f}_n(\mathbf{x})$	$\mathrm{MSE}\left(\hat{f}_n(\mathbf{x})\right)$
$n = 20$	0.002535025	0.00001714449
$n = 50$	0.00249859	0.000008301609
$n = 100$	0.002317932	0.000004328159
$n = 200$	0.002150694	0.000002291990
$n = 500$	0.002010352	0.0000008698475
$n = 1000$	0.001974886	0.0000004716210

The true density probability function of X_t on the manifold $\mathbb{D}^2$ is defined by

$$f(\mathbf{x}) = \begin{cases} \frac{d(\mathbf{x},0)g(\log_0(\mathbf{x}))}{\sinh(d(\mathbf{x},0))} & \text{if} \quad \mathbf{x} \in \mathbb{D}^2 \backslash \{0\}, \\ 0 & \text{if} \quad \mathbf{x} = 0, \end{cases} \tag{4.6}$$

where g is density probability function of the $\mathbb{R}^2$-process defined by (4.4). We estimate the density probability function of the random variable X_1 which belongs to manifold $\mathbb{D}^2$. Then $m = 1$ and $i_m = 1$. It is well known that (Y_t) is geometrically completely regular (see [14]); then, the mixing coefficients of (X_t) satisfy the assumptions of Theorem 3.6. For the numerical tests, 2000 independent samples $(X_i)_{i=1}^n$ on Poincaré disk are generated under the model (4.5) for $n = 20, 50, 100, 200, 500$ and 1000. For fixed $\mathbf{x} = (0.5, 0.6)$ or $(-0.4, 0.7)$, we compute the mean value of $\hat{f}_n(\mathbf{x}; \mathbf{i}_1)$ and the MSE. The same kernel function as in the previous simulation is used. The obtained results are presented below:

Tables 3 and 4 show that for $\mathbf{x} = (0.5, 0.6)$ and $\mathbf{x} = (-0.4, 0.7)$, the geometric kernel density estimator $\hat{f}_n(\mathbf{x}; \mathbf{i}_1)$ converges to, respectively, the true value $f(\mathbf{x}) = 0.004369456$ and $f(\mathbf{x}) = 0.001921427$ as the sample size increases. Hence, this simulation illustrates the consistency result in the case of stationary and strongly mixing non-i.i.d process.

4.3 Real data analysis: an application to Canada macroeconomic labor data

In order to illustrate the kernel density estimator for strongly mixing process on Riemannian manifold, we consider the Canada labor data available in the R package $vars$.

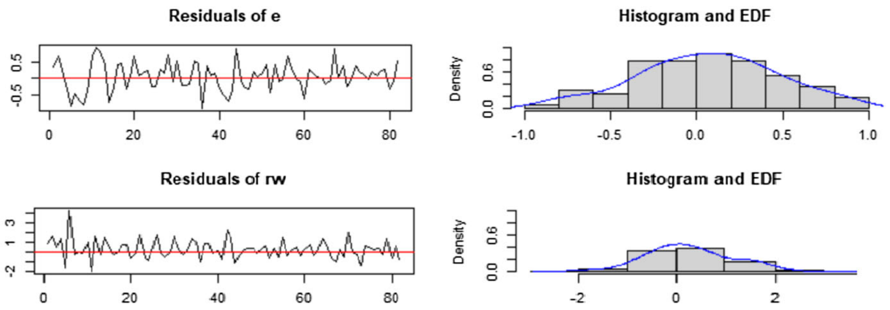
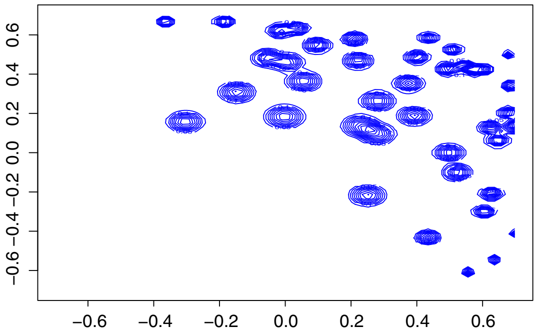


Fig. 5 Residuals diagnostic plots from fitting a VAR(1) to labor data

Fig. 6 Contour density estimation from labor data via diffeomorphism transformation on $\mathbb{D}^2$



The data were deduced from the original time series published by the OECD and was modelled as a VAR in [21]. We used the 84 observations on the vector formed by the variables employment (“e”) and real wage (“rw”). A preliminary analysis is first performed on the data following the methodology in [21]. To get a stationary time series data, we apply the first difference filter to the observations. Then a VAR(1) process $(Y_t)_{t\in\mathbb{Z}}$ defined by $Y_t = AY_{t-1} + \epsilon_t$ with i.i.d. Gaussian errors is fitted to the differenced data by ordinary least squares method and the estimate of A is $\begin{pmatrix} 0.80623004 & 0.01808065 \\ 0.21385420 & 0.67853030 \end{pmatrix}$. The histogram and empirical density function of the residuals are plotted in Fig. 5.

From this preliminary analysis, we can model the differenced data as observations of stationary, strongly mixing and geometrically completely regular VAR(1) process (Y_t) on $\mathbb{R}^2$. Then, considering the same bandwidth and kernel function as in Sect. 4.1, the mixing coefficients of $\left(X_t = \exp_{\mathbf{0}}(Y_t)\right)$ satisfy the assumptions of Theorem 3.6. The geometric kernel density estimate function computed from the sample on $\mathbb{D}^2$ is plotted in Figs. 6 and 7.

5 Conclusion and perspectives

We have discussed the problem of nonparametric density estimator on Riemannian manifold. We have constructed a joint density estimator of a finite-dimensional vector

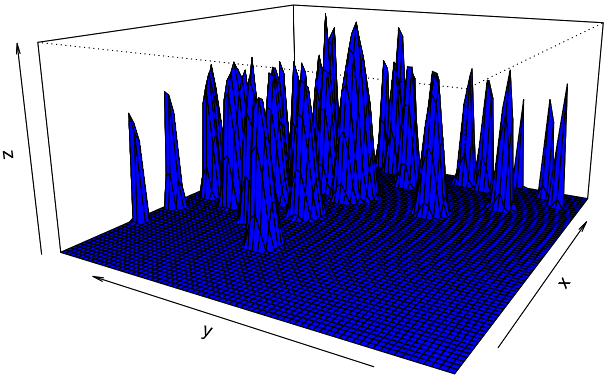


Fig. 7 Density estimation from labor data via diffeomorphism transformation on $\mathbb{D}^2$

of a stationary and strongly mixing process on manifold and have shown that this estimator is strongly consistent. Moreover, a similar result is obtained for conditional density functions and conditional expectations of functionals given its past behavior. Our simulation studies show that the density estimator converges to true density and is more effective than Euclidean kernel density estimator when both methods are applicable. This provides relevant evidence that geometric structure of the sample space has to be taken into account for accurate statistical inference. However, the dimension of the manifold may be high and further investigations would be useful to assess the behavior of the proposed estimator in this setting. One may use a dimension reduction technique like Riemannian principal components analysis [6]. For our estimators of joint density, conditional density and conditional expectations for strongly mixing processes, the almost sure convergence rates remain to be established. Since manifold-valued data and objects are increasingly used in statistics, it would be interesting to study strong mixing processes on other manifolds.

Acknowledgements The authors thank the reviewers for a most careful reading of the manuscript and constructive comments and remarks.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Boothby, W.M.: An Introduction to Differentiable Manifolds and Riemannian geometry. Academic Press, New York (1975)
2. Chavel, I.: Riemannian geometry: a modern introduction. In: Cambridge Tracts in Mathematics vol. 108. Cambridge University Press, Cambridge (1993)
3. Do Carmo, M.: Riemannian Geometry. Birkhauser Boston Inc, Boston (1992)
4. Do Carmo, M.: Geometria Riemaniana, 2nd edn. Proyecto Euclides IMPA, Colchester (1988)
5. Chevallier, E., Forget, T., Barbaresco, F., Angulo, J.: Kernel density estimation on the Siegel space with an application to Radar Processing. Entropy **18**(11), 396 (2016)

6. Dai, X., Muller, H.: Principal component analysis for functional data on Riemannian manifold and spheres. Ann. Stat. **46**(6B), 3334–3361 (2018)
7. Emmanuel, R.: Covariance inequalities for strongly mixing processes. Ann. Prob. Stat. **29**(4), 587–597 (1993)
8. Gamea, O., Bécigneul, G., Hofmann, T.: Hyperbolic neural network. In: Advances in Neutral information Processing Systems, pp. 5350–5360 (2018)
9. Henry, G., Rodriguez, D.: Kernel density estimation on Riemannian manifolds: asymptotic results. J. Math. Imaging Vis. **34**, 235–239 (2009)
10. Kesavan, S.: Measure and Integration. Springer, Berlin (2019)
11. Loeve, M.: Probability Theory, 4th edn. Springer, New York (2017)
12. Masry, E.: Almost sure convergence of recursive density estimators for stationary mixing processes. Stat. Prob. Lett. **5**, 204–207 (1987)
13. Masry, E.: Nonparametric estimation of conditional probability densities and expectations of stationary processes strong consistency and rates. Stoch. Process. Appl. **32**, 109–127 (1989)
14. Mokkade, A.: Mixing properties of ARMA process. Stoch. Process. Appl. **29**(2), 309–315 (1988)
15. Ovinnikov, I.: Poincaré Wasserstein Auntoencoder. In: Third Workshop on Bayesian Deep Learning. Montréal, Canada (2019)
16. Patrangenaru, V., Ellingson, L.: Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis. CRC Press, Boca Raton (2015)
17. Pelletier, B.: Nonparametric regression estimation on closed Riemannian manifolds. J. Nonparametr. Stat. **18**, 57–67 (2006)
18. Pelletier, B.: Kernel density estimation on Riemannian manifolds. Stat. Probab. Lett. **73**(3), 297–304 (2005)
19. Pennec, X.: Probabilities and statistics on Riemannian manifolds: basic tools for geometric measurements. In: Proceedings of Nonlinear Signal and Image Processing (NSIP'99), pp. 194–198 (1999)
20. Petersen, P.: Riemannian geometry 2nd ed. In: Graduate Texts in Mathematics, vol. 171. Springer, New York (2006)
21. Pfaff, B.: Analysis of Integrated and Cointegrated Time series with R. Springer, Berlin (2008)
22. R Development Core Team: R: A language and environment for statistical computing. In: R Foundation for Statistical Computing Vienna Austria. ISBN 3-900051-07-0. http://www.R-project.org (2010)
23. Said, S., Bombrun, L., Berthoumieu, Y.: New Riemannian priors on the univariate normal model. Entropy **16**(7), 4015–4031 (2014)
24. Sakai, T.: Riemannian Geometry, Translations of Mathematical Monographs. American Mathematical Society, Providence, RI (1996)