Kernel density estimation for a stochastic process with values in a Riemannian Manifold

Mohamed Abdillahi Isman [1,4] & Salah Khardani [2] & Papa Mbaye [1,3]
Wiem Nefzi [2] & Anne-Françoise Yao [1]

[1] *Laboratoire de Mathématiques Blaise Pascal, CNRS UMR 6620, Campus des Cézeaux, 63171 Aubière Cedex, France. anne.yao@uca.fr, mohamed.abdillahi_isman@doctorant.uca.fr*
[2] *Faculté des Sciences Tunis, Université El-Manar, Laboratoire de Modélisation mathématique, Statistique et Analyse Stochastique M2ϕSAS, Tunisie. khardani_salah@yahoo.fr, nefziwiem24@gmail.com*
[3] *Missioneo, Laboratoire de missioneo (Freeland Group), France. papa.mbaye@missioneo.fr*
[4] *Université de Djibouti, Djibouti. magouleh@gmail.com*

Résumé. Ce travail aborde la problématique de l'estimation de densité pour des observations ayant des valeurs dans une sous-variété riemannienne. Dans ce contexte, Pelletier (2005) a proposé un estimateur de densité à noyau pour des données indépendantes. Nous étudions le comportement asymptotique de l'estimateur de Pelletier lorsque les observations sont générées à partir d'un processus strictement stationnaire α-mixing à valeurs dans cette sous-variété. En particulier, nous donnons les vitesses de convergence en termes d'erreur quadratique moyenne, en probabilité et presque sûrement. Nous établissons également un théorème central limite et illustrons notre propos à travers des simulations ainsi qu'une application à des données réelles.

Mots-clés. Estimateur de densité à noyau, Variétés riemanniennes, Condition de mélange, Processus stochastique, Consistance faible et forte, Théorème central limite.

Abstract. This paper is related to the issue of the density estimation of observations with values in a Riemannian submanifold. In this setting, Pelletier (2005) has proposed a kernel density estimator for independent data. We investigate here the behavior of Pelletier's estimator when the observations are generated from a strictly stationary α-mixing process with values in this submanifold. In particular, we study the pointwise as well as the uniform weak and strong consistency of the estimator. Namely, we give the rate of convergence in mean square error meaning, in probability and almost surely. We also give a central-limit theorem and illustrate our purpose through some simulations and a real data application.

Keywords. : Kernel density estimator, Riemannian manifolds, Mixing condition, Stochastic process, Weak and Strong consistency, Central Limit Theorem.

Motivation

The motivation behind this work is that the probability density estimator is crucial in many fields, including statistics, signal processing, machine learning, etc. The problem of estimating an unknown density using a kernel approach has been widely addressed when the variable of interest lies in a Euclidean space, both for independent data and dependent data. However, in many applications , such as biology, spatial statistics, geology, image analysis, and medicine, the Euclidean assumption about the underlying geometry of the observations fails. An alternative is then to treat the data as lying on some submanifold.

1 Framework and the kernel of interest

Let $(X_t, t \in Z)$ defined on a probability space $(\Omega, \mathcal{A}, P)$ with values on Riemannian submanifold, $(\mathcal{M}, g)$ of R^d $(d \geq 2)$ such that X_t's are dependent and distributed as a random variable, X with an unknown density f on $\mathcal{M}$. We assume that (X_t) satisfies an $\alpha-$mixing condition such that for any integer $n \geq 1$,

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

where $\mathcal{F}_i^k(X)$ is the $\sigma-$field generated by $\{X_i,\ i \leq j \leq k\}$ and $\alpha(n)$ tends to zero.

In this work, we aim to study the behavior of the kernel estimator of Pelletier:

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n^d} \frac{1}{\theta_x(X_i)} K\left(\frac{d(x, X_i)}{h_n}\right),$$

based on observations of the process $(X_t, t \in Z)$, X_i, $i = 1, ..., n.$. Here, h_n is the bandwidth, K is the kernel function, and n is the number of observations. Then, we give some asymptotic results based on the following assumptions. Note that C will represent a positive constant with no impact on the results.

Assumptions

H1: $K : R^d \to R_+$ is a bounded and continuous map such that

1. K satisfies a Lipschitz condition.
2. *supp* $K = [0; 1]$.
3. $\int K(\|x\|)dx = 1$.
4. $\int xK(\|x\|)dx = 0$ or at least the vector, $\int xK(\|x\|)\,dx$ is orthogonal to span$\{\text{grad} f(x)\}$.
5. $\int \|x\|^2 K(\|x\|)dx < \infty$.

H2: The mixing coefficient of X_i satisfies $\alpha(n) \leq Cn^{-\nu}$ for some $\nu > 2$.
H3: The bandwidth is such that $h_n \to 0$, $nh_n^d \to \infty$ and $h_n < \frac{h^*}{2}$ as $n \to \infty$.
H4: f is bounded, twice continuously differentiable at any $x \in \mathcal{M}$ and $\|\text{Hess} f\,(x)\|_{HS} < \infty$, where $\|\cdot\|_{HS}$ is the Hilbert Schmidt norm (one can also consider the uniform norm of $\text{Hess} f\,(x)$).

H5: $\forall i, j$, the joint density $f_{i,j}$ of (X_i, X_j) exists is such that

$$\sup_j \sup_{u,v \in \mathcal{M}\times\mathcal{M}} |f_{i,j}(u,v) - f(u)f(v)| < M,$$

for some $M > 0$.

Some remarks on these assumptions

Since this work is related to situations where the data are both with values in a Riemannian manifold and dependent, the above assumptions appear as a mixed of conditions in both cases to derive consistency results for kernel density estimators in this setting. Namely,

1. The hypothesis **H2** and **H5** are classical when dealing with the study of the kernel density estimator for dependent data.
2. Assumption **H3** is the Riemannian manifolds counterpart of the classical assumptions on the bandwidth for dependent or independent data.
3. Assumptions **H1** and **H4** are classical regularity conditions on K and f, helpful to control the bias terms (1) obtained using the Taylor expansion:

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

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

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

where grad and Hess denote the gradient and the Hessian operators respectively.

2 Some Theoretical Results

The assumptions stated above allow us to give some theoretical results on the asymptotic behavior of $\widehat{f_n}(x)$ that allow for evaluating the quality of the estimator in terms of bias and variance. Specifically, we study both the weak and strong consistency of $\widehat{f_n}(x)$ in terms of Mean Squared Error(MSE), probability and almost sure convergence.

Weak and Strong Consistency

The bias term

The asymptotic behavior of the bias term has been studied in Pelletier (2005). However, we give here an explicit expression

$$b(x) := E\hat{f}_n(x) - f(x) = \frac{h_n^2}{2}\int_{B(1)} K(\|v\|)\text{Hess} f\,(x)\,(v,v)dv + o(h_n^2) \tag{1}$$

Unlike the bias, all the results below are specific theoretical contributions of this paper.

Asymptotic behavior of the variance of the estimator

Under the assumptions **H1, H2** and **H5**, the variance of $\widehat{f}_n(x)$ is given by

$$V\left(\widehat{f}_n(x)\right) := E\left[\left(\widehat{f}_n(x) - E\widehat{f}_n(x)\right)^2\right] = \frac{1}{nh_n^d}\left[f(x)\int_{B(1)} K^2(\|v\|)dv + o(1)\right]$$

and then

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

Rate of convergence in Mean Squared Error

For each $x \in \mathcal{M}$, we set

$$MSE(x) := E\left((\widehat{f}_n(x) - f(x))^2\right) = b^2(x) + V\left(\widehat{f}_n(x)\right).$$

Under the assumptions **H1** to **H5**,

$$MSE(x) \leq C_{x,f} \times \left(h_n^4 + \frac{1}{nh_n^d}\right),$$

and since $\mathcal{M}$ is compact, and f and K are bounded (see **H1** and **H4**), we have

$$\sup_{x\in\mathcal{M}} MSE(x) \leq C \times \left(h_n^4 + \frac{1}{nh_n^d}\right).$$

An optimal rate in MSE meaning for $\widehat{f}_n(x)$ is deduced from the previous result in the following.

The bandwidth which minimizes the function $x \mapsto MSE(x)$, under the assumption **H1** is given by:

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

and the corresponding MSE is

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

where $C_{x,f} = C_0^4 C_1 + C_0^{-d} C_2$ with $C_0 = \left(\frac{dC_2}{4C_1}\right)^{\frac{1}{4+d}}$, $C_1 = \left(\int_{B(1)} \text{Hess} f(x)(v,v)K(\|v\|)dv\right)^2$ and $C_2 = f(x)\int_{B(1)} K^2(\|v\|)dv$.

Rate of Convergence: In Probability and Almost Surely

We now give some pointwise as well as uniform rates of convergence in probability with some additional conditions.

Under the assumptions **H1** to **H5**, if $\alpha(n) \leq Cn^{-\nu}$ with $\nu > 2$ (as stated in **H2**) and if $\frac{nh_n^{d\frac{\nu+1}{\nu-1}}}{\log n} \to \infty$, then for a given $x \in \mathcal{M}$

$$|\widehat{f}_n(x) - f(x)| = O_p\left(h_n^2 + \sqrt{\frac{\log n}{nh_n^d}}\right).$$

Moreover, since $\mathcal{M}$ is compact, if $\nu > d+1$ and $n^{-1}(\log n)h_n^{-d\frac{\nu+d+3}{\nu-d-1}} \to 0$, we have

$$\sup_{x\in\mathcal{M}}|\widehat{f}_n(x) - f(x)| = O_p\left(h_n^2 + \sqrt{\frac{\log n}{nh_n^d}}\right).$$

As previously, with some additional conditions, we get the following almost surely rates of convergence.

Under the assumptions **H1** to **H5**, if $\alpha(n) \leq Cn^{-\nu}$ with $\nu > 3$ and $nh_n^{\frac{d(\nu+1)}{\nu-3}}(\log n)^{-\frac{\nu-1}{\nu-3}}g(n)^{\frac{-2}{\nu-3}} \to \infty$ with $h_n < \frac{h^*}{2}$ and $g(n) = \log n\,(\log\log n)^{1+\epsilon}$ for some small $\epsilon > 0$ then for each given $x \in \mathcal{M}$, we have

$$\widehat{f}_n(x) - f(x) = O_{a.s.}\left(h_n^2 + \sqrt{\frac{\log n}{nh_n^d}}\right)$$

and because $\mathcal{M}$ is compact, we have

$$\sup_{x\in\mathcal{M}}|\widehat{f}_n(x) - f(x)| = O_{a.s.}\left(h_n^2 + \sqrt{\frac{\log n}{nh_n^d}}\right).$$

Asymptotic normality

We now state a Central Limit Theorem (CLT) for each given $x \in \mathcal{M}$.

Under the assumptions **H1**, **H2**, **H3** and **H5** , we have:

$$\sqrt{nh_n^d}\left(\widehat{f}_n(x) - f(x)\right) \longrightarrow \mathcal{N}\left(0, \sigma^2(x)\right),$$

where $\sigma^2(x) = f(x)\int_{B(1)} K^2\left(\|x\|\right)dx$.

3 Numerical results

During the talk, we will illustrate these theoretical results through some simulations and a real data application.

4 Some References

le Brigant, A. and Puechmorel, S. (2019), Approximation of densities on Riemannian manifolds, *Entropy,* 21(1).

Pelletier, B. (2005), Kernel density estimation on Riemannian manifolds, *Statistics & Probability letters*, 73(3):297 - 304.

García-Portugués et al. (2020), On optimal tests for rotational symmetry against new classes of hyperspherical distributions, *Journal of the American Statistical Association*, 115(532):1873 - 1887.

Cleanthous, et al. (2020), Kernel and wavelet density estimators on manifolds and more general metric spaces, *Bernoulli,* 26(3):1832 - 1862.

Berenfeld, C. and Hoffmann, M. (2021), Density estimation on an unknown submanifold, *Electronic Journal of Statistics,* 15(1):2179 - 2223.

Aamari, E. and Levrard, C. (2019), Non-asymptotic rates for manifold, tangent space and curvature estimation, *The Annals of Statistics,* 47(1):177 - 204

.