Density Estimation on Manifolds

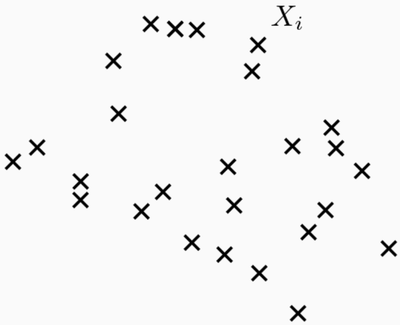
Meeting in Mathematical Statistics, 14th December, 2020

Clément Berenfeld
Joint work with Marc Hoffmann.

Université Paris-Dauphine & PSL, CNRS, CEREMADE.

Introduction

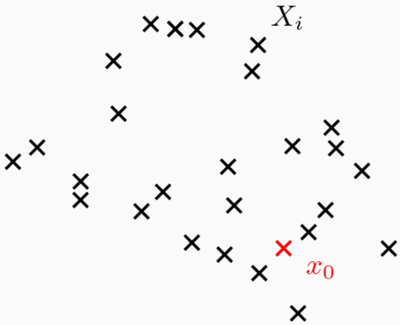
Observations : point cloud $X_1, \ldots, X_n$ in a possibly high-dimensional space $\mathbb{R}^D$.



Introduction

Observations : point cloud $X_1, \dots, X_n$ in a possibly high-dimensional space $\mathbb{R}^D$.

Goal : compute the density of the underlying distribution at a given point x_0.

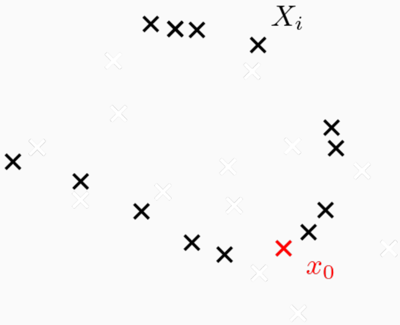


Introduction

Observations : point cloud $X_1, \ldots, X_n$ in a possibly high-dimensional space $\mathbb{R}^D$.

Goal : compute the density of the underlying distribution at a given point x_0.

Assumption : the data actually lie on a low-dimensional, unknown structure.

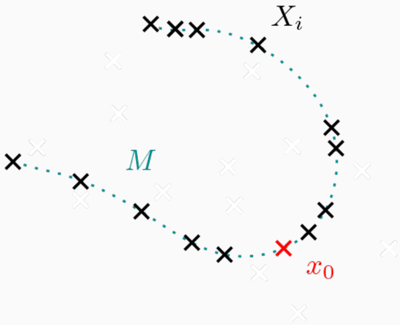


Introduction

Observations : point cloud $X_1, \dots, X_n$ in a possibly high-dimensional space $\mathbb{R}^D$.

Goal : compute the density of the underlying distribution at a given point x_0.

Assumption : the data actually lie on a low-dimensional, <u>unknown</u> structure.



Introduction

When is this structural assumption relevant ?

Introduction

When is this structural assumption relevant ?

- when the data naturally lie on a manifold : geodata, cosmological data, biological data ...

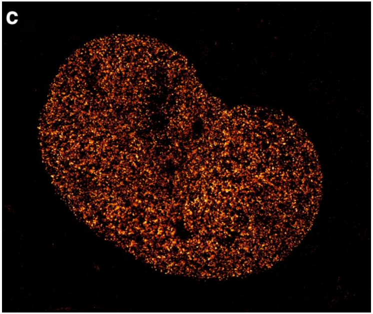


Figure 1: Point cloud obtain using single-molecule localization microscopy on a COS-7 cell. Figure from [Klein et al., 2014].

When is this structural assumption relevant ?

- when the data is high-dimensional but only a small number of parameters is expected to governed the dataset.

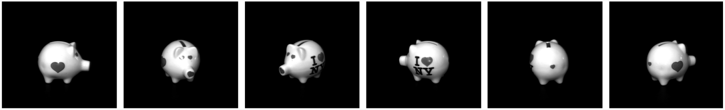


Figure 2: Images from the COIL-20 dataset [Nene et al., 1996] as presented in [Sober et al., 2017]. The ambient dimension is very high (number of pixels) but the data can be parametrized by $SO(3, \mathbb{R})$.

Notations and risk

- Target point $x_0 \in \mathbb{R}^D$;
- Independant data $X_1, \ldots, X_n \in \mathbb{R}^D$ with commom law P;
- P has density f_P and support M_P;
- M_P has dimension d and $x_0 \in M_P$;
- Measure the accuracy of $\widehat{f}(x_0)$ with

$$\mathbb{E}_{P^{\otimes n}}[|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p}.$$

Introduction

What has been done so far ?

What has been done so far ?

- Density estimation on abstract manifolds [Hendriks, 1990, Pelletier, 2005];

Introduction

What has been done so far ?

- Density estimation on abstract manifolds [Hendriks, 1990, Pelletier, 2005];
- Estimation on embedded manifolds with twice differentiable density [Ozakin and Gray, 2009];

What has been done so far ?

- Density estimation on abstract manifolds [Hendriks, 1990, Pelletier, 2005];
- Estimation on embedded manifolds with twice differentiable density [Ozakin and Gray, 2009];
- Adaptive density estimation on abstract manifolds [Kerkyacharian et al., 2012, Cleanthous et al., 2018];

What has been done so far ?

- Density estimation on abstract manifolds [Hendriks, 1990, Pelletier, 2005];
- Estimation on embedded manifolds with twice differentiable density [Ozakin and Gray, 2009];
- Adaptive density estimation on abstract manifolds [Kerkyacharian et al., 2012, Cleanthous et al., 2018];
- Recent developments : strong uniform consistency of KDE on embedded manifolds [Wu and Wu, 2020], lower bounds on Hölder classes for abstract manifolds. [Ki and Park, 2020].

Our contribution :

- Build a model that takes into account the regularity of both the density and the support;
- Get bounds on the minimax risk for this model;
- Adaptive estimation with respect to the intrinsic dimension of the support and to the regularity parameters.

Table of contents

1. A statistical model for sampling on manifolds

Hölder spaces on manifold

A function $g : U \subset \mathbb{R}^d \to \mathbb{R}^k$ is β-Hölder if

- g is m-times differentiable, where $m = \lceil \beta - 1 \rceil$;
- writing $\delta = \beta - m \in (0, 1]$ we have

$$\sup_{u,v \in U} \frac{\| \mathrm{d}^m g(u) - \mathrm{d}^m g(v) \|_{\mathrm{op}}}{\|u - v\|^\delta} < \infty.$$

We call the above quantity the β-Hölder coefficient of g. It is denoted by

$$|g|_\beta.$$

Hölder spaces on manifold

Let M be a submanifold of $\mathbb{R}^D$ of dimension d and take

$$g: \ M \to \mathbb{R}^k.$$

The usual way to define the regularity of g is to see g through a chart $\phi_x : (M, x) \to (\mathbb{R}^d, 0)$

$$g \circ \phi_x^{-1} \ : \ (\mathbb{R}^d, 0) \ \to \ \mathbb{R}^k.$$

Hölder spaces on manifold

Let M be a submanifold of $\mathbb{R}^D$ of dimension d and take

$$g: \; M \to \mathbb{R}^k.$$

The usual way to define the regularity of g is to see g through a chart $\phi_x : (M, x) \to (\mathbb{R}^d, 0)$

$$g \circ \phi_x^{-1} \;:\; (\mathbb{R}^d, 0) \;\to\; \mathbb{R}^k.$$

→ The magnitude of the Hölder coefficients of $g \circ \phi_x^{-1}$ will strongly depend on the choice of ϕ_x

Hölder spaces on manifold

Let M be a submanifold of $\mathbb{R}^D$ of dimension d and take

$$g : \ M \to \mathbb{R}^k.$$

The usual way to define the regularity of g is to see g through a chart $\phi_x : (M, x) \to (\mathbb{R}^d, 0)$

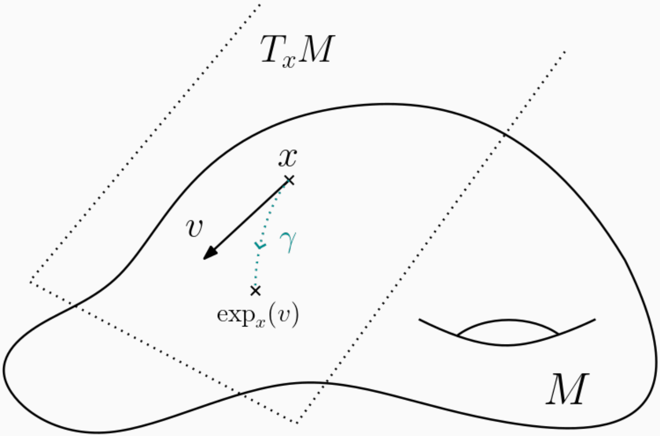
$$g \circ \phi_x^{-1} \ : \ (\mathbb{R}^d, 0) \ \to \ \mathbb{R}^k.$$

→ The magnitude of the Hölder coefficients of $g \circ \phi_x^{-1}$ will strongly depend on the choice of ϕ_x

→ **We need a canonical choice of chart.**

Hölder spaces on manifold

We choose the exponential map $\exp_x : T_xM \to M$ [Triebel, 1987] but other choices are possible [Aamari and Levrard, 2019, Ki and Park, 2020].



Hölder spaces on manifold

We let $\rho > 0$ be a localization parameter.

Definition 1

Let M be a closed submanifold of $\mathbb{R}^D$. We say that $g : M \to \mathbb{R}^k$ is β-Hölder if, for any $x \in M$, the map

$$g \circ \exp_x : B(0, \rho) \to \mathbb{R}^k$$

is β-Hölder.

The β-Hölder coefficient of $g : M \to \mathbb{R}^k$ is simply defined as

$$|g|_\beta = \sup_{x \in M} |g \circ \exp_x|_\beta.$$

Hölder spaces on manifold

This gives us a convenient way to quantify the regularity of a closed submanifold.

Definition 2

We say that M is α-Hölder if the inclusion

$$\iota_M : M \to \mathbb{R}^D$$

is α-Hölder in the sense defined above. We shall write

$$|M|_\alpha := |\iota_M|_\alpha.$$

Volume measure

For a d-submanifold $M \subset \mathbb{R}^D$ we define its volume measure μ_M as

$$\mu_M(A) = \mathcal{H}^d(A \cap M) \quad \forall A \in \mathcal{B}(\mathbb{R}^D)$$

where $\mathcal{H}^d$ is the d-dimensionnal Hausdorff measure.

Volume measure

For a d-submanifold $M \subset \mathbb{R}^D$ we define its volume measure μ_M as

$$\mu_M(A) = \mathcal{H}^d(A \cap M) \quad \forall A \in \mathcal{B}(\mathbb{R}^D)$$

where $\mathcal{H}^d$ is the d-dimensionnal Hausdorff measure.

For $x \in M$, and a smooth function $\psi : M \to \mathbb{R}$ supported on a small neighborhood of x, we have

$$\mu_M(\psi) = \int_{T_x M} \psi(\exp_x(v))\theta_x(\exp_x(v))\,\mathrm{d}v$$

where $\theta_x : (M, x) \to \mathbb{R}^+$ is the **volume density function** of M.

Statistical model

We introduce some parameters

$$d, \alpha, A \qquad \text{and} \qquad f_{\min}, f_{\max}, \beta, B$$

and define the set $\Sigma^d_{\alpha,\beta}$ of all probability measures P satisfying

Support conditions

Density conditions

Statistical model

We introduce some parameters

$$d, \alpha, A \qquad \text{and} \qquad f_{\min}, f_{\max}, \beta, B$$

and define the set $\Sigma^d_{\alpha,\beta}$ of all probability measures P satisfying

Support conditions

- $\operatorname{supp} P$ is a compact, connected d-submanifold M_P of $\mathbb{R}^D$, containing x_0;

Density conditions

Statistical model

We introduce some parameters

$$d, \alpha, A \qquad \text{and} \qquad f_{\min}, f_{\max}, \beta, B$$

and define the set $\Sigma^d_{\alpha,\beta}$ of all probability measures P satisfying

Support conditions

- $\operatorname{supp} P$ is a compact, connected d-submanifold M_P of $\mathbb{R}^D$, containing x_0;
- M_P is α-Hölder;

Density conditions

Statistical model

We introduce some parameters

$$d, \alpha, A \qquad \text{and} \qquad f_{\min}, f_{\max}, \beta, B$$

and define the set $\Sigma^d_{\alpha,\beta}$ of all probability measures P satisfying

Support conditions

- $\operatorname{supp} P$ is a compact, connected d-submanifold M_P of $\mathbb{R}^D$, containing x_0;
- M_P is α-Hölder;
- $|M_P|_\alpha \leq A$;

Density conditions

Statistical model

We introduce some parameters

$$d, \alpha, A \qquad \text{and} \qquad f_{\min}, f_{\max}, \beta, B$$

and define the set $\Sigma^d_{\alpha,\beta}$ of all probability measures P satisfying

Support conditions

- $\operatorname{supp} P$ is a compact, connected d-submanifold M_P of $\mathbb{R}^D$, containing x_0;
- M_P is α-Hölder;
- $|M_P|_\alpha \leq A$;

Density conditions

- there exists a version of $dP/d\mu_{M_P}$, denoted f_P that satisfies all the below;

Statistical model

We introduce some parameters

$$d, \alpha, A \qquad \text{and} \qquad f_{\min}, f_{\max}, \beta, B$$

and define the set $\Sigma^d_{\alpha,\beta}$ of all probability measures P satisfying

Support conditions

- $\operatorname{supp} P$ is a compact, connected d-submanifold M_P of $\mathbb{R}^D$, containing x_0;
- M_P is α-Hölder;
- $|M_P|_\alpha \leq A$;

Density conditions

- there exists a version of $dP/d\mu_{M_P}$, denoted f_P that satisfies all the below;
- $f_{\min} \leq f_P \leq f_{\max}$;

Statistical model

We introduce some parameters

$$d, \alpha, A \qquad \text{and} \qquad f_{\min}, f_{\max}, \beta, B$$

and define the set $\Sigma^d_{\alpha,\beta}$ of all probability measures P satisfying

Support conditions

- $\operatorname{supp} P$ is a compact, connected d-submanifold M_P of $\mathbb{R}^D$, containing x_0;
- M_P is α-Hölder;
- $|M_P|_\alpha \leq A$;

Density conditions

- there exists a version of $dP/d\mu_{M_P}$, denoted f_P that satisfies all the below;
- $f_{\min} \leq f_P \leq f_{\max}$;
- f_P is β-Hölder;

Statistical model

We introduce some parameters

$$d, \alpha, A \qquad \text{and} \qquad f_{\min}, f_{\max}, \beta, B$$

and define the set $\Sigma^d_{\alpha,\beta}$ of all probability measures P satisfying

Support conditions

- $\operatorname{supp} P$ is a compact, connected d-submanifold M_P of $\mathbb{R}^D$, containing x_0;
- M_P is α-Hölder;
- $|M_P|_\alpha \leq A$;

Density conditions

- there exists a version of $dP/d\mu_{M_P}$, denoted f_P that satisfies all the below;
- $f_{\min} \leq f_P \leq f_{\max}$;
- f_P is β-Hölder;
- $|f_P|_\beta \leq B$.

Reach condition

Are we happy with this model ?

Reach condition

Are we happy with this model ? **Not quite.**

Reach condition

Are we happy with this model ? **Not quite.**

Theorem 1

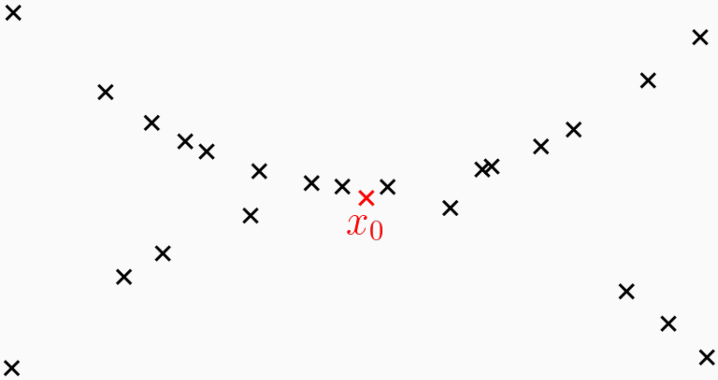
For any estimator $\widehat{f}(x_0)$ we have, for any $n \geq 1$

$$\sup_{P \in \Sigma^d_{\alpha,\beta}} \mathbb{E}_{P^{\otimes n}} [|\widehat{f}(x_0) - f_P(x_0)|^p]^{1/p} \gtrsim 1$$

for whatever value of α, β and d, where $\Sigma^d_{\alpha,\beta}$ is defined as in the slide before.

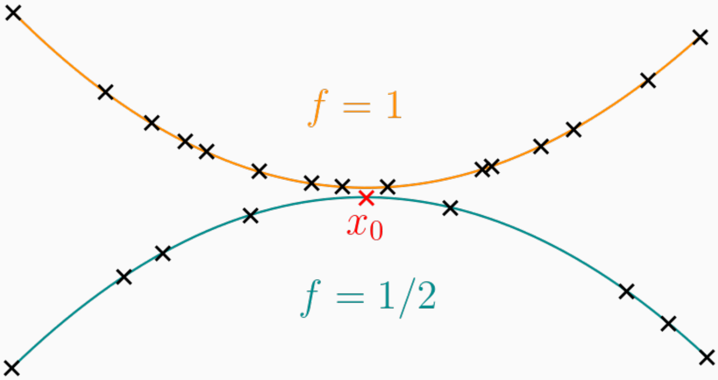
Reach condition

Proof by drawing.



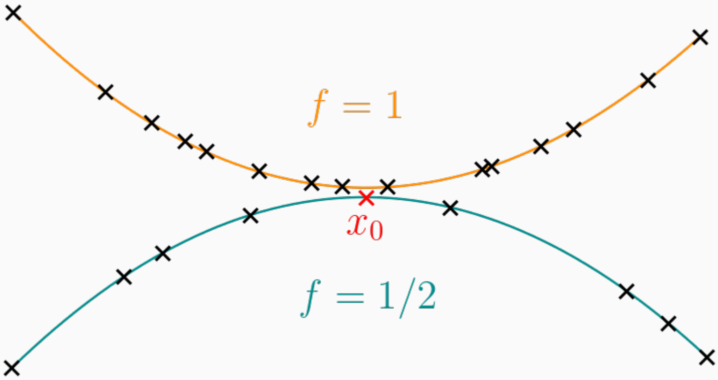
Reach condition

Proof by drawing.



Reach condition

Proof by drawing.



→ we need to avoid manifolds that are close to self-intersect. □

Reach condition

The **reach** of $M \subset \mathbb{R}^D$ is defined as [Federer, 1969]

$$\tau_M = \sup\left\{r \geq 0 \mid \ \mathrm{pr}_M \ \text{ is well defined on } M^{\oplus r}\right\}$$

→ it prevents M from curving too much;

→ it prevents M from self-intersecting.

Reach condition

The **reach** of $M \subset \mathbb{R}^D$ is defined as [Federer, 1969]

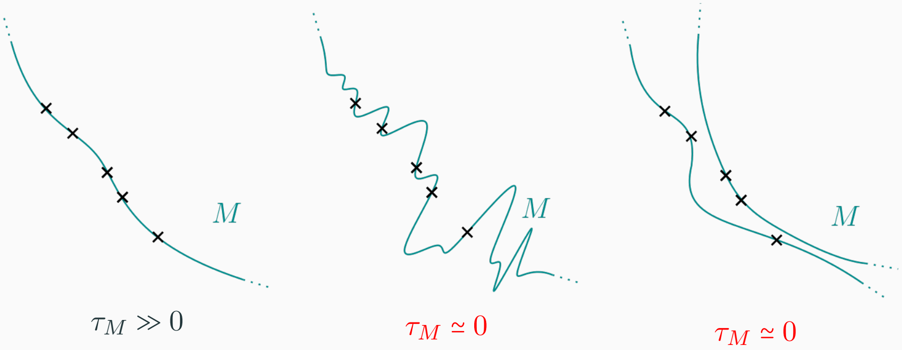
$$\tau_M = \sup \left\{ r \geq 0 \mid \mathrm{pr}_M \text{ is well defined on } M^{\oplus r} \right\}$$

→ it prevents M from curving too much;
→ it prevents M from self-intersecting.

A reach condition is most of the time necessary for getting minimax results in a manifold setting [Niyogi et al., 2008, Genovese et al., 2012, Divol, 2020].

If needed, one can estimate the reach of the support before doing further analysis [Aamari et al., 2019, B. et al., 2020].

Reach condition



A few diagrams of both the local and global constraints imposed by a reach condition.

Statistical model

New definition of the model :

New definition of the model : for parameters

$$d, \tau, \alpha, A \qquad \text{and} \qquad f_{\min}, f_{\max}, \beta, B$$

we define the set $\Sigma^d_{\alpha,\beta}$ of all probability measures P satisfying

Support conditions

- $\operatorname{supp} P$ is a connected, compact d-submanifold M_P of $\mathbb{R}^D$, containing x_0;
- M_P is α-Hölder;
- $|M_P|_\alpha \leq A$;

Density conditions

- there exists a version of $dP/d\mu_{M_P}$, denoted f_P that satisfies all the below;
- $f_{\min} \leq f_P \leq f_{\max}$;
- f_P is β-Hölder;
- $|f_P|_\beta \leq B$.

Statistical model

New definition of the model : for parameters

$$d, \tau, \alpha, A \qquad \text{and} \qquad f_{\min}, f_{\max}, \beta, B$$

we define the set $\Sigma^d_{\alpha,\beta}$ of all probability measures P satisfying

Support conditions

- $\operatorname{supp} P$ is a connected, compact d-submanifold M_P of $\mathbb{R}^D$, containing x_0;
- $\tau_{M_P} \geq \tau$;
- M_P is α-Hölder;
- $|M_P|_\alpha \leq A$;

Density conditions

- there exists a version of $dP/d\mu_{M_P}$, denoted f_P that satisfies all the below;
- $f_{\min} \leq f_P \leq f_{\max}$;
- f_P is β-Hölder;
- $|f_P|_\beta \leq B$.

2. Main results

Upper bound

Theorem 2

There exists an estimator $\widehat{f}(x_0)$ *depending on* d, α *and* β *such that*

$$\sup_{P \in \Sigma^d_{\alpha,\beta}} \mathbb{E}_{P^{\otimes n}}\left[|\widehat{f}(x_0) - f_P(x_0)|^p\right]^{1/p} \lesssim n^{-\frac{\beta}{2\beta+d} \wedge \frac{\alpha-1}{2(\alpha-1)+d}}$$

Upper bound

Theorem 2

There exists an estimator $\widehat{f}(x_0)$ *depending on* d, α *and* β *such that*

$$\sup_{P \in \Sigma^d_{\alpha,\beta}} \mathbb{E}_{P^{\otimes n}}\left[|\widehat{f}(x_0) - f_P(x_0)|^p\right]^{1/p} \lesssim n^{-\frac{\beta}{2\beta+d} \wedge \frac{\alpha-1}{2(\alpha-1)+d}}$$

This is achieved using a simple kernel density estimator

$$\widehat{f}_h(x_0) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right).$$

Upper bound

There are two contributions in the bound

$$n^{-\frac{\beta}{2\beta+d} \wedge \frac{\alpha-1}{2(\alpha-1)+d}}$$

Upper bound

There are two contributions in the bound

$$n^{-\frac{\beta}{2\beta+d}\wedge\frac{\alpha-1}{2(\alpha-1)+d}}$$

$\rightarrow \frac{\beta}{2\beta+d}$: classical rate when estimating β-Hölder density.

Upper bound

There are two contributions in the bound

$$n^{-\frac{\beta}{2\beta+d}\wedge\frac{\alpha-1}{2(\alpha-1)+d}}$$

$\rightarrow \frac{\beta}{2\beta+d}$: classical rate when estimating β-Hölder density.

$\rightarrow \frac{\alpha-1}{2(\alpha-1)+d}$: comes from

Upper bound

There are two contributions in the bound

$$n^{-\frac{\beta}{2\beta+d} \wedge \frac{\alpha-1}{2(\alpha-1)+d}}$$

$\rightarrow \frac{\beta}{2\beta+d}$: classical rate when estimating β-Hölder density.

$\rightarrow \frac{\alpha-1}{2(\alpha-1)+d}$: comes from

$$\frac{1}{nh^d} \sum K\left(\frac{X_i - x_0}{h}\right)$$

- Approximating the intrinsic distance $d_{M_P}(X_i, x_0)$ with the euclidean distance $\|X_i - x_0\|$;

Upper bound

There are two contributions in the bound

$$n^{-\frac{\beta}{2\beta+d} \wedge \frac{\alpha-1}{2(\alpha-1)+d}}$$

$\rightarrow \frac{\beta}{2\beta+d}$: classical rate when estimating β-Hölder density.

$\rightarrow \frac{\alpha-1}{2(\alpha-1)+d}$: comes from

$$\frac{1}{nh^d} \sum K\left(\frac{X_i - x_0}{h}\right) \times 1$$

- Approximating the intrinsic distance $d_{M_P}(X_i, x_0)$ with the euclidean distance $\|X_i - x_0\|$;
- Approximating the volume density function θ_{x_0} with 1.

Upper bound

The kernel $K : \mathbb{R}^D \to \mathbb{R}$ we choose only need be

- smooth and compactly supported;
- normalized on all d-dimensional subspace $H \subset \mathbb{R}^D$

$$\int_H K(v)\,\mathrm{d}v = 1.$$

Upper bound

The kernel $K : \mathbb{R}^D \to \mathbb{R}$ we choose only need be

- smooth and compactly supported;
- normalized on all d-dimensional subspace $H \subset \mathbb{R}^D$

$$\int_H K(v)\, \mathrm{d}v = 1.$$

The estimator $\widehat{f}(x_0)$ will depend on

- d : through the normalization h^d, the choice of the kernel K and of its order
- α and β : through the choice of the bandwidth h.

Adaptation

In practice, d, α and β are not known.

- Bandwidth selection via Lepski's method [Lepski, 1992]

$$\widehat{h} = \max\left\{h \in \mathbb{H} \;\middle|\; |\widehat{f}_h(x_0) - \widehat{f}_\eta(x_0)| \le \psi(\eta, h) \quad \forall \eta \le h\right\}$$

where $\mathbb{H}$ is a finite grid of bandwidths and $\psi(\eta, h)$ acts as a proxy of the stochastic deviation of $|\widehat{f}_h(x_0) - \widehat{f}_\eta(x_0)|$.

Adaptation

In practice, d, α and β are not known.

- Bandwidth selection via Lepski's method [Lepski, 1992]

$$\widehat{h} = \max\left\{h \in \mathbb{H} \;\middle|\; |\widehat{f}_h(x_0) - \widehat{f}_\eta(x_0)| \leq \psi(\eta, h) \quad \forall \eta \leq h\right\}$$

  where $\mathbb{H}$ is a finite grid of bandwidths and $\psi(\eta, h)$ acts as a proxy of the stochastic deviation of $|\widehat{f}_h(x_0) - \widehat{f}_\eta(x_0)|$.

- Plug-in of an estimator $\widehat{d}$ of d. There exists a lot of them with very good performance [Kégl, 2003, Farahmand et al., 2007, Kim et al., 2016].

Adaptation

We only ask $\widehat{d}$ to be mildly accurate, meaning

$$\sup_{P \in \Sigma^d_{\alpha,\beta}} P^{\otimes n}(\widehat{d} \neq d) \lesssim n^{-3p/2}.$$

and let $\hat{f}^{\mathrm{ad}}(x_0)$ denote the resulting estimator, built on top of a kernel of order ℓ.

Theorem 3

For any $1 \leq d \leq D-1$, $0 \leq \beta \leq \ell$ *and* $2 \leq \alpha \leq \ell + 1$, *we have*

$$\sup_{P \in \Sigma^d_{\alpha,\beta}} \mathbb{E}_{P^{\otimes n}}[|\hat{f}^{\mathrm{ad}}(x_0) - f_P(x_0)|^p]^{1/p} \lesssim \left(\frac{\log n}{n}\right)^{-\frac{\beta}{2\beta+d} \wedge \frac{\alpha-1}{2(\alpha-1)+d}}.$$

Lower bound

Is this rate optimal ?

Lower bound

Is this rate optimal ? **Probably not.**

Lower bound

Is this rate optimal ? **Probably not.**

Theorem 4

For any $\alpha \geq 2$ and $\beta > 0$, we have

$$\inf_{\widehat{f}(x_0)} \sup_{P \in \Sigma^d_{\alpha,\beta}} \mathbb{E}_{P^{\otimes n}} \left[|\widehat{f}(x_0) - f_P(x_0)|^p \right]^{1/p} \gtrsim n^{-\frac{\beta}{2\beta+d} \wedge \frac{2\alpha-2}{d}}$$

where the infimum is taken on all measurable estimators.

Lower bound

Is this rate optimal ? **Probably not.**

Theorem 4

For any $\alpha \geq 2$ and $\beta > 0$, we have

$$\inf_{\widehat{f}(x_0)} \sup_{P \in \Sigma^d_{\alpha,\beta}} \mathbb{E}_{P^{\otimes n}}\left[|\widehat{f}(x_0) - f_P(x_0)|^p\right]^{1/p} \gtrsim n^{-\frac{\beta}{2\beta+d} \wedge \frac{2\alpha-2}{d}}$$

where the infimum is taken on all measurable estimators.

→ we recover the rate $\frac{\beta}{2\beta+d}$

Lower bound

Is this rate optimal ? **Probably not.**

Theorem 4

For any $\alpha \geq 2$ and $\beta > 0$, we have

$$\inf_{\widehat{f}(x_0)} \sup_{P \in \Sigma^d_{\alpha,\beta}} \mathbb{E}_{P^{\otimes n}} \left[|\widehat{f}(x_0) - f_P(x_0)|^p \right]^{1/p} \gtrsim n^{-\frac{\beta}{2\beta+d} \wedge \frac{2\alpha-2}{d}}$$

where the infimum is taken on all measurable estimators.

→ we recover the rate $\frac{\beta}{2\beta+d}$
→ new rate $\frac{2\alpha-2}{d}$

Lower bound

Is this rate optimal ? **Probably not.**

Theorem 4

For any $\alpha \geq 2$ and $\beta > 0$, we have

$$\inf_{\widehat{f}(x_0)} \sup_{P \in \Sigma^d_{\alpha,\beta}} \mathbb{E}_{P^{\otimes n}}\left[|\widehat{f}(x_0) - f_P(x_0)|^p\right]^{1/p} \gtrsim n^{-\frac{\beta}{2\beta+d} \wedge \frac{2\alpha-2}{d}}$$

where the infimum is taken on all measurable estimators.

→ we recover the rate $\frac{\beta}{2\beta+d}$
→ new rate $\frac{2\alpha-2}{d}$
→ matching rates when $\alpha \geq \beta + 1$.

The special case of dimension 1

In dimension 1, the rate in the lower-bound simply becomes

$$\frac{\beta}{2\beta+1} \wedge \frac{2\alpha-2}{1} = \frac{\beta}{2\beta+1}$$

suggesting that α may not have a limiting effect in this case.

In dimension 1, the rate in the lower-bound simply becomes

$$\frac{\beta}{2\beta+1} \wedge \frac{2\alpha-2}{1} = \frac{\beta}{2\beta+1}$$

suggesting that α may not have a limiting effect in this case.

Proposition 1

The volume density function θ_{x_0} of a 1-dimensional submanifold M is identically 1 on M.

The special case of dimension 1

To estimate the intrinsic distance over the support, we use a graph-based distance using a neighborhood graph [Tenenbaum et al., 2000, Arias-Castro and Le Gouic, 2019]. We consider the graph $\mathcal{G}_\varepsilon = (V, E)$ where

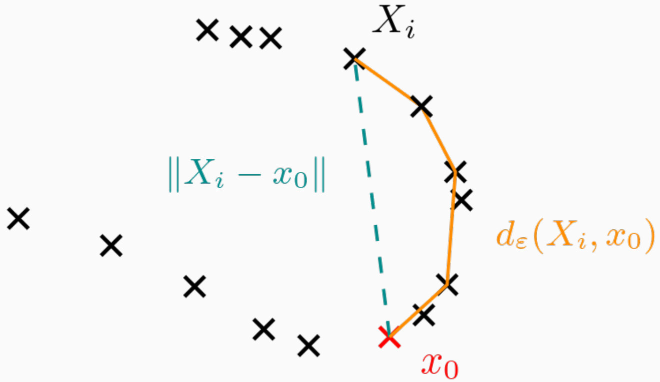
$$V = \{x_0, X_1, \ldots, X_n\} \quad \text{and} \quad E = \{(x, y) \mid \|x - y\| \le \varepsilon\}$$

and define

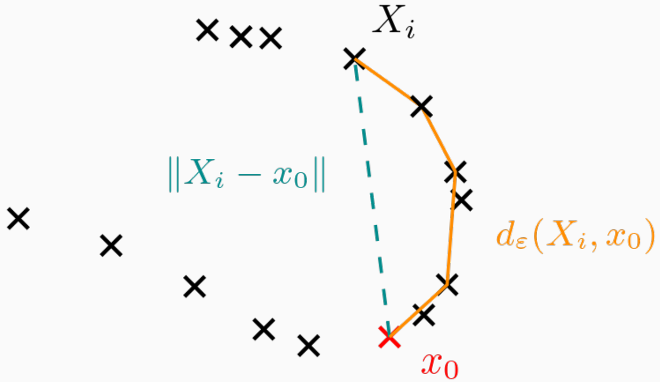
$$d_\varepsilon(x, y) = \text{length of the shortest path in } \mathcal{G}_\varepsilon \text{ from } x \text{ to } y$$

where the length of a path is the sum of the size of its edges.

The special case of dimension 1



The special case of dimension 1



$\rightarrow d_\varepsilon$ will be ε^2 close to d_{M_P} with high probability, for a careful choice of ε.

The special case of dimension 1

We introduce

$$\hat{f}^{1\mathrm{D}}_{\varepsilon,h}(x_0) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{d_\varepsilon(X_i, x_0)}{h}\right)$$

where $K : \mathbb{R} \to \mathbb{R}$.

Theorem 5

For any $\beta \geq 0$ and any $\alpha \geq 2$, we have

$$\sup_{P \in \Sigma^1_{\alpha,\beta}} \mathbb{E}_{P^{\otimes n}}[|\hat{f}^{1\mathrm{D}}_{\varepsilon,h}(x_0) - f_P(x_0)|^p]^{1/p} \lesssim n^{-\frac{\beta}{2\beta+1}}$$

when choosing

$$h \simeq n^{-\frac{1}{2\beta+1}} \quad \text{and} \quad \varepsilon \simeq \frac{\log n}{n}.$$

3. Numerical illustrations

Numerical illustrations

We implemented the adaptive estimator $\hat{f}^{\text{ad}}(x_0)$ on two synthetic dataset of intrinsic dimension $d = 1$ and $d = 2$.

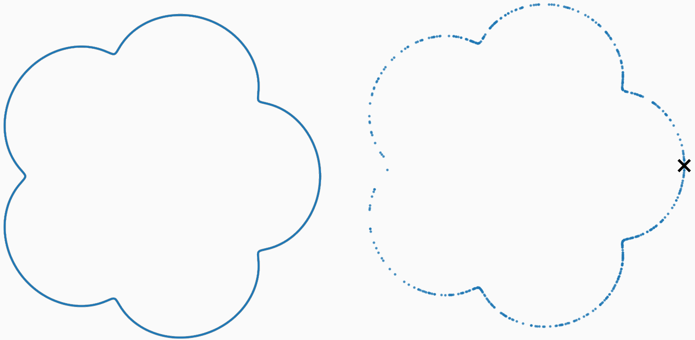


Figure 3: One dimensional synthetic dataset.

Numerical illustrations

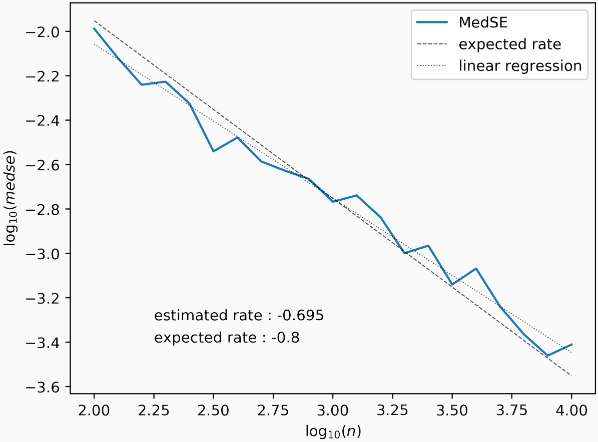


Figure 4: Log-log plot of median square error for $d = 1$. The underlying density had regularity $\beta = 2$. We used from 10^2 to 10^4 numbers of observations, and each experiment was repeated 500 times.

Numerical illustrations

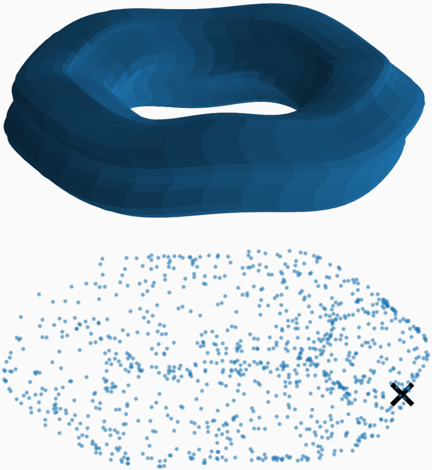


Figure 5: Two dimensional synthetic dataset.

Numerical illustrations

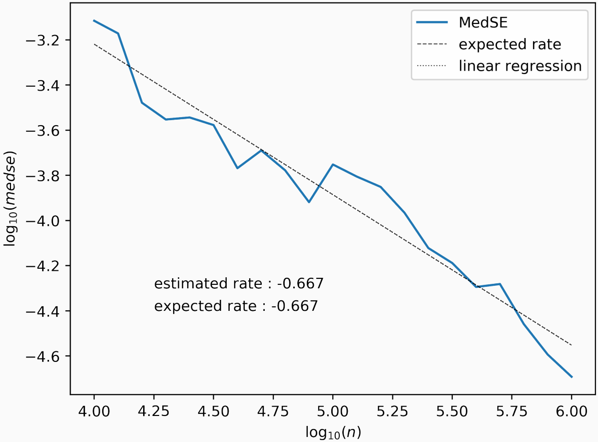


Figure 6: Log-log plot of median square error for $d = 2$. The underlying density had regularity $\beta = 2$. We used from 10^4 to 10^6 numbers of observations, and each experiment was repeated 500 times.

4. Conclusion

Conclusion

- When estimating pointwise β-Hölder density on α-Hölder support, we obtain minimax bounds

$$n^{-\frac{\beta}{2\beta+d}\wedge\frac{2\alpha-2}{d}} \lesssim R_{\text{minimax}} \lesssim n^{-\frac{\beta}{2\beta+d}\wedge\frac{\alpha-1}{2(\alpha-1)+d}}.$$

- The bounds match whenever $\alpha \geq \beta + 1$ (i.e. when the support is sufficiently smooth with respect to the density).
- In this case, classical KDE is minimax, and we only need to estimate the intrinsic dimension to compute it.
- In the case $d = 1$, α does not impede the speed of estimation, and we can provide a minimax estimator at the price of learning the intrinsic distance of the support.

Thank you for your attention.

Aamari, E., Kim, J., Chazal, F., Michel, B., Rinaldo, A., and Wasserman, L. (2019).
Estimating the reach of a manifold.
Electron. J. Stat., 13(1):1359–1399.

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

Arias-Castro, E. and Le Gouic, T. (2019).
Unconstrained and curvature-constrained shortest-path distances and their approximation.
Discrete & Computational Geometry, 62(1):1–28.

Balakrishnan, S., Rinaldo, A., Sheehy, D., Singh, A., and Wasserman, L. (2012).
Minimax rates for homology inference.
In *Artificial Intelligence and Statistics*, pages 64–72.

Berenfeld, C., Harvey, J., Hoffmann, M., and Shankar, K. (2020).
Estimating the reach of a manifold via its convexity defect function.
arXiv preprint arXiv:2001.08006.

Cleanthous, G., Georgiadis, A., Kerkyacharian, G., Petrushev, P., and Picard, D. (2018).
Kernel and wavelet density estimators on manifolds and more general metric spaces.
arXiv preprint arXiv:1805.04682.

Divol, V. (2020).
Minimax adaptive estimation in manifold inference.
arXiv preprint arXiv:2001.04896.

Eugeciouglu, Ö. and Srinivasan, A. (2000).
Efficient nonparametric density estimation on the sphere with applications in fluid mechanics.
SIAM Journal on Scientific Computing, 22(1):152–176.

Farahmand, A. M., Szepesvári, C., and Audibert, J.-Y. (2007).
Manifold-adaptive dimension estimation.
In *Proceedings of the 24th international conference on Machine learning*, pages 265–272. ACM.

Federer, H. (1959).
Curvature measures.
Transactions of the American Mathematical Society, 93(3):418–491.

Federer, H. (1969).
Geometric measure theory.
Springer.

Gallot, S., Hulin, D., and Lafontaine, J. (2004).
***Riemannian geometry*, volume 3.**
Springer.

Genovese, C., Perone-Pacifico, M., Verdinelli, I., and Wasserman, L. (2012).
Minimax manifold estimation.
Journal of machine learning research, 13(May):1263–1291.

Hall, P., Watson, G., and Cabrera, J. (1987).
Kernel density estimation with spherical data.
Biometrika, 74(4):751–762.

Hendriks, H. (1990).
Nonparametric estimation of a probability density on a riemannian manifold using fourier expansions.
The Annals of Statistics, pages 832–849.

Kégl, B. (2003).
Intrinsic dimension estimation using packing numbers.
In *Advances in neural information processing systems*, pages 697–704.

Kerkyacharian, G., Nickl, R., and Picard, D. (2012).
Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds.
Probability Theory and Related Fields, 153(1-2):363–404.

Ki, D. and Park, B. U. (2020).
Intrinsic hölder classes of density functions on riemannian manifolds and lower bounds to convergence rates.
Statistics & Probability Letters, 169:108959.

Kim, J., Rinaldo, A., and Wasserman, L. (2016).
Minimax rates for estimating the dimension of a manifold.
arXiv preprint arXiv:1605.01011.

Kim, P. T., Koo, J.-Y., and Luo, Z.-M. (2009).
Weyl eigenvalue asymptotics and sharp adaptation on vector bundles.
J. Multivariate Anal., 100(9):1962–1978.

Klein, T., Proppert, S., and Sauer, M. (2014).
Eight years of single-molecule localization microscopy.
Histochemistry and cell biology, 141(6):561–575.

Lee, J. M. (2006).
***Riemannian manifolds: an introduction to curvature*, volume 176.**
Springer Science & Business Media.

Lepskii, O. (1992).
Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates.
Theory of Probability & Its Applications, 36(4):682–697.

Nene, S., Nayar, S., Murase, H., et al. (1996).
Columbia object image library (coil-20), 1996.
Dept. Comput. Sci., Columbia Univ, New York.

Niyogi, P., Smale, S., and Weinberger, S. (2008).
Finding the homology of submanifolds with high confidence from random samples.
Discrete & Computational Geometry, 39(1-3):419–441.

Ozakin, A. and Gray, A. G. (2009).
Submanifold density estimation.
In *Advances in Neural Information Processing Systems*, pages 1375–1382.

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

- Sober, B., Aizenbud, Y., and Levin, D. (2017).
  Approximation of functions over manifolds: A moving least-squares approach.
  arXiv preprint arXiv:1711.00765.
- Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000).
  A global geometric framework for nonlinear dimensionality reduction.
  science, 290(5500):2319–2323.
- Triebel, H. (1987).
  Characterizations of function spaces on a complete Riemannian manifold with bounded geometry.
  Mathematische Nachrichten, 130(1):321–346.

Tsybakov, A. (2008).
Introduction to Nonparametric Estimation.
Springer Series in Statistics. Springer New York.

Wu, H.-T. and Wu, N. (2020).
Strong uniform consistency with rates for kernel density estimators with general kernels on manifolds.
arXiv preprint arXiv:2007.06408.

Yu, B. (1997).
Assouad, Fano, and Le Cam.
In *Festschrift for Lucien Le Cam*, pages 423–435. Springer.