Reconstructing measures on shapes: an optimal transport perspective

Vincent Divol
Inria Saclay - Datashape team
Laboratoire Mathématiques d'Orsay
vincentdivol.github.io

January 5, 2021

Séminaire Palaisien

The minimax framework

- Model assumption: $X_1, \ldots, X_n \sim \mu \in \mathcal{P}$, where $\mathcal{P}$ is a set of probability measures on $\mathbb{R}^D$.
- $\theta : \mathcal{P} \to E$ to estimate.
- Estimator: $\hat{\theta}_n : (\mathbb{R}^D)^n \to E$
- Loss function: $\mathcal{L} : E \times E \to [0, \infty]$ pseudo-metric.
- Minimax risk: $\mathcal{R}_n(\theta; \mathcal{P}; \mathcal{L}) := \inf_{\hat{\theta}_n} \sup_{\mu \in \mathcal{P}} \mathbb{E}_{\mu^{\otimes n}} \mathcal{L}(\hat{\theta}_n, \theta(\mu))$

The minimax framework

- Model assumption: $X_1, \ldots, X_n \sim \mu \in \mathcal{P}_D^s$, where $\mathcal{P}_D^s$ is the set of probability measures on $\mathbb{R}^D$ contained in the ball of $H_p^s(\mathbb{R}^D)$ of radius R.
- $\theta(\mu) = f \in L_p(\mathbb{R}^D)$ to estimate.
- Estimator: $\hat{f}_n : (\mathbb{R}^D)^n \to L_p$
- Loss function: $\mathcal{L}(f, g) = \|f - g\|_{L_p}$ pseudo-metric.
- Minimax risk: $\mathcal{R}_n(f; \mathcal{P}_D^s; L_p) := \inf_{\hat{f}_n} \sup_{\mu \in \mathcal{P}_D^s} \mathbb{E}_{\mu^{\otimes n}} \|\hat{f}_n - f\|_{L_p}$

Kernel density estimation on the cube

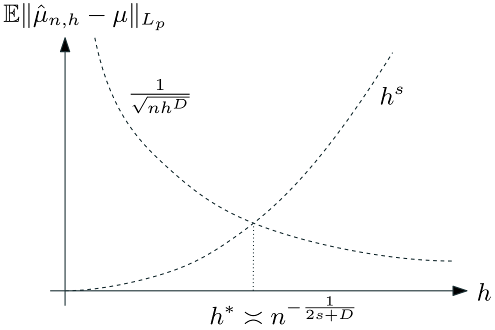
- Let $K : \mathbb{R}^D \to \mathbb{R}$ be a kernel function: $\int K(v)\mathrm{d}v = 1$, and $\int K(v)v^\alpha \mathrm{d}v = 0$ for $0 < |\alpha| \le t =$ order of the kernel.
- Convolution: $K * \nu(x) = \int K(x-y)\mathrm{d}\nu(y)$.
- $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ is the empirical measure, $K_h = h^{-D}K(\cdot/h)$. Let $\hat{\mu}_{n,h} = K_h * \mu_n$.

- **Bias:** Linear operator: $A_h : \phi \mapsto K_h * \phi - \phi$.

$$\|K_h * \mu - \mu\|_{L_p} = \|A_h \mu\|_{L_p} \le \|A_h\|_{H_p^s \to L_p} \|\mu\|_{H_p^s} \lesssim h^s R$$

- **Variance:** $\mathrm{Var}(\hat{\mu}_{n,h}(x)) \le \frac{1}{n}\mathbb{E}|K_h(X-x)|^2 \lesssim \frac{1}{nh^D}$

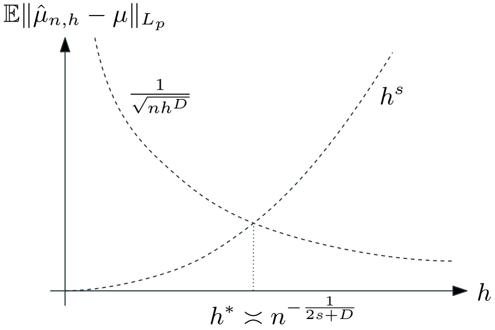
Kernel density estimation on the cube



$$\Longrightarrow \mathbb{E}\|\hat{\mu}_{n,h} - \mu\|_{L_p} \lesssim h^s + \frac{1}{\sqrt{nh^D}}$$

- Choose $h \simeq n^{-\frac{1}{2s+D}}$: $\quad \mathcal{R}_n(f, \mathcal{P}_D^s, \|\cdot\|_{L_p}) \asymp n^{-\frac{s}{2s+D}}$

Kernel density estimation on the cube

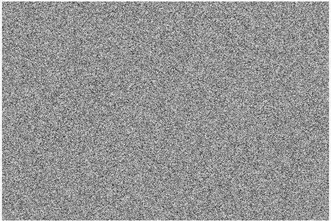


$$\Longrightarrow \mathbb{E}\|\hat{\mu}_{n,h} - \mu\|_{L_p} \lesssim h^s + \frac{1}{\sqrt{nh^D}}$$

- Choose $h \simeq n^{-\frac{1}{2s+D}}$: $\quad \mathcal{R}_n(f, \mathcal{P}_D^s, \|\cdot\|_{L_p}) \asymp n^{-\frac{s}{2s+D}}$

Structural assumptions on the signal

n observations $X_1, \ldots, X_n$ in $\mathbb{R}^D$ with $n \ll D$.



White noise

Key assumption

There is a low d-dimensional structure underlying the observations $\mathbb{X}_n$.

$\rightarrow$ What assumptions should be made on the structure? (Sparsity, single/multi-index model, constraints on the shape of the support)

Structural assumptions on the signal

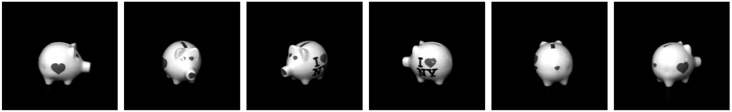


Figure 1: Sober & al. 2020

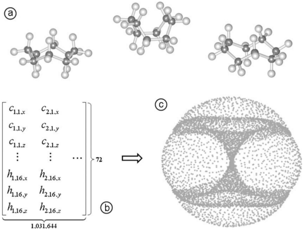


Figure 2: Martin & al. 2010

The delicate choice of the loss function

Problem

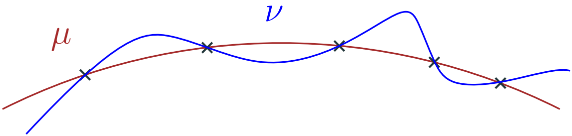
Observe: $X_1, \dots, X_n \sim \mu$ where μ is supported on M is some (unknown) d-dimensional $\mathcal{C}^k$-manifold and μ has a density of regularity s (model $\mathcal{P}_d^{s,k}$). How to estimate μ?

Choice of the loss: Pointwise $\rightarrow$ [Berenfeld Hoffmann 2020, Pelletier 2005, Kerkyacharian & al. 2012] L_2? Total variation? Hellinger? Kullback-Leibler?

Theorem (A negative result)

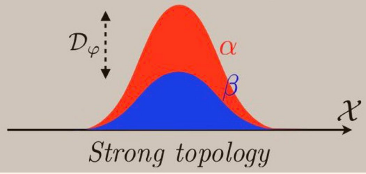
For all $n \geq 1$,

$$\mathcal{R}_n(\mu, \mathcal{P}_d^{s,k}, \mathrm{TV}) \geq 1/2.$$



Csiszár divergences:

$$\mathcal{D}_\varphi(\alpha|\beta) \stackrel{\text{def.}}{=} \int_{\mathcal{X}} \varphi\left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}\right) \mathrm{d}\beta$$

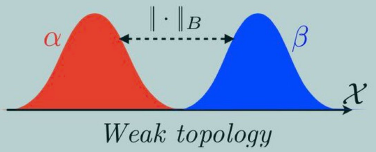


φ convex, $\varphi(1) = 0$.

$$\varphi(r) = \begin{cases} r\log(r) & \to \text{KL} \\ |r-1| & \to \text{TV} \\ |\sqrt{r}-1|^2 & \to \text{Hellinger} \\ \dots \end{cases}$$

Dual norms:

$$\|\alpha - \beta\|_B \stackrel{\text{def.}}{=} \max_{f \in B} \int_{\mathcal{X}} f(x)(\mathrm{d}\alpha(x) - \mathrm{d}\beta(x))$$



$$B = \{f \; ; \; \|f\|_B \leqslant 1\}$$

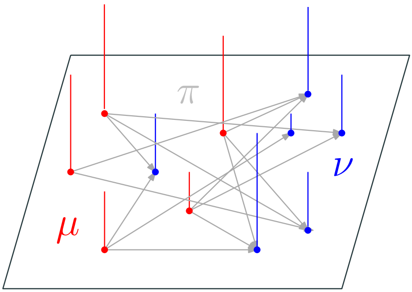
$$\|f\|_B = \begin{cases} \|\nabla f\|_\infty & \to W_1 \\ \|f\|_\infty + \|\nabla f\|_\infty & \to \text{flat} \\ \|\nabla^k f\|_2 & \to \text{MMD} \\ \dots \end{cases}$$

[Gabriel Peyré's Twitter account]

Definition (Wasserstein distance)

- μ, ν probability measures on $\mathbb{R}^D$.
- $\Pi(\mu, \nu)$ set of transport plans between μ and ν.

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\int |x - y|^p \mathrm{d}\pi(x, y) \right)^{1/p}.$$



Definition (Wasserstein distance)

- μ, ν probability measures on $\mathbb{R}^D$.
- $\Pi(\mu, \nu)$ set of transport plans between μ and ν.

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \left(\int |x - y|^p \mathrm{d}\pi(x, y) \right)^{1/p}.$$

Lemma ([Peyre 2018])

Let μ, ν be probability measures on a manifold M, with $\mu, \nu \geq f_{\min} \mathrm{vol}_M$. Then,

$$W_p(\mu, \nu) \leq p^{-1/p} f_{\min}^{1/p-1} \|\mu - \nu\|_{\dot{H}_p^{-1}(M)}.$$

$$\|\mu - \nu\|_{\dot{H}_p^{-1}(M)} = \sup\{(\mu - \nu)(\phi), \ \|\nabla \phi\|_{L_{p^*}(M)} \leq 1\}$$

- If $p = 1$: Kantorovitch-Rubinstein duality formula.
- If $d = 1$: related to the distance between the cdf.

Easy case: Assume that we have access to M.

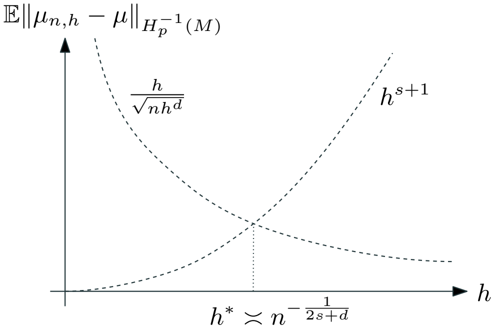
- Let $K : \mathbb{R}^D \to \mathbb{R}$ be a smooth radial function with $\int_{\mathbb{R}^d} K(v)\mathrm{d}v = 1$, of order t.
- Convolution: $K * \nu(x) = \int K(x-y)\mathrm{d}\nu(y)$.
- μ_n is the empirical measure, $K_h = h^{-d}K(\cdot/h)$. Let $\mu_{n,h}$ be the measure having density $K_h * \mu_n$ with respect to vol_M.

- **Bias:** Linear operator: $A_h : \phi \mapsto K_h * \phi - \phi$.

  $$\|K_h * \mu - \mu\|_{H_p^{-1}(M)} = \|A_h \mu\|_{H_p^{-1}(M)} \leq \|A_h\|_{H_p^s(M) \to H_p^{-1}(M)} \|\mu\|_{H_p^s(M)} \lesssim h^{s+1} R$$

- **Variance:** (Harmonic analysis on M) $\mathbb{E}\|\mu_{n,h} - K_h * \mu\|_{H_p^{-1}(M)} \leq \frac{h}{\sqrt{nh^d}}$

Easy case: Assume that we have access to M.



$$\implies \mathbb{E}\|\mu_{n,h} - \mu\|_{H_p^{-1}(M)} \lesssim h^{s+1} + \frac{h}{\sqrt{nh^d}}$$

- Choose $h \simeq n^{-\frac{1}{2s+d}}$: $\quad \mathcal{R}_n(f, \mathcal{P}_d^{s,k}, W_p) \asymp n^{-\frac{s+1}{2s+d}}$

Minimax estimation of the volume measure

- $\mu_{n,h}$ is **NOT** an estimator:

$$\mathrm{d}\mu_{n,h}(x) = (K_h * \mu_n)(x)\mathrm{d}\mathrm{vol}_M(x)$$

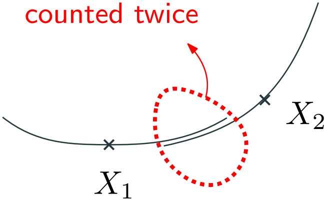
- Idea 1: $\widehat{\mathrm{vol}}_M = d$-dimensional Hausdorff measure on some proxy $\hat{M}$ of M.

Theorem (Aamari-Levrard 2019)

Let π_{X_i} the orthogonal projection on $T_{X_i}M$. There exists estimators $\hat{T}_i$ of $T_{X_i}M$ and $\hat{\Psi}_i$ of Ψ_{X_i} such that, with high probability,
$\forall v \in \hat{T}_i, |v| \leq \varepsilon \asymp \left(\frac{\ln n}{n}\right)^{\frac{1}{d}}$,

$$\angle(\hat{T}_i, T_{X_i}M) \lesssim \left(\frac{\ln n}{n}\right)^{\frac{k-1}{d}} \quad \text{and} \quad |\hat{\Psi}_i(v) - \Psi_{X_i}(\pi_{X_i}(v))| \lesssim \left(\frac{\ln n}{n}\right)^{\frac{k}{d}}$$

$\rightarrow$ $\hat{M} = \bigcup_{i=1}^{n} \hat{\Psi}_i(\hat{T}_i) \cap \mathcal{B}(X_i, \varepsilon)$ is ε^k-close from M... bad idea.



Minimax estimation of the volume measure

- $\hat{\mu}_{n,h}$ is an estimator:

$$\mathrm{d}\hat{\mu}_{n,h}(x) = (K_h * \mu_n)(x)\mathrm{d}\widehat{\mathrm{vol}}_M(x)$$

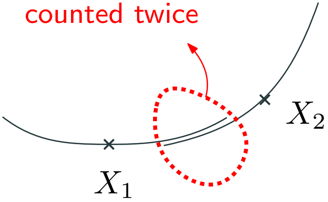
- Idea 1: $\widehat{\mathrm{vol}}_M$ = d-dimensional Hausdorff measure on some proxy $\hat{M}$ of M.

Theorem (Aamari-Levrard 2019)

Let π_{X_i} the orthogonal projection on $T_{X_i}M$. There exists estimators $\hat{T}_i$ of $T_{X_i}M$ and $\hat{\Psi}_i$ of Ψ_{X_i} such that, with high probability,
$\forall v \in \hat{T}_i, |v| \leq \varepsilon \asymp \left(\frac{\ln n}{n}\right)^{\frac{1}{d}}$,

$$\angle(\hat{T}_i, T_{X_i}M) \lesssim \left(\frac{\ln n}{n}\right)^{\frac{k-1}{d}} \quad \text{and} \quad |\hat{\Psi}_i(v) - \Psi_{X_i}(\pi_{X_i}(v))| \lesssim \left(\frac{\ln n}{n}\right)^{\frac{k}{d}}$$

$\rightarrow$ $\hat{M} = \bigcup_{i=1}^{n} \hat{\Psi}_i(\hat{T}_i) \cap \mathcal{B}(X_i, \varepsilon)$ is ε^k-close from M... bad idea.



Minimax estimation of the volume measure

- Idea 2: use an appropriate partition of unity.

1. From $\mathbb{X}_n$, we build a set $\tilde{\mathbb{X}}_n = \{X_1, \ldots, X_J\}$ which is ε-sparse and ε-close from $\mathbb{X}_n$. (Farthest Point Sampling)
2. There exists $\xi_j : \mathcal{B}(X_j, \varepsilon) \to [0,1]$ with $\sum_{j=1}^J \xi_j \equiv 1$ on $M^{\varepsilon/4}$ and $\|\xi_j\|_{\mathcal{C}^1} \lesssim 1$.

$$\int \phi \mathrm{dvol}_M = \int_M \sum_{j=1}^J \xi_j(x)\phi(x)\mathrm{dvol}_M(x) = \sum_{j=1}^J \int_{\Psi_{X_j}(T_{X_j}M)} \xi_j(x)\phi(x)\mathrm{d}x$$

Theorem

Define $\widehat{\mathrm{vol}}_M$ *by* $\int \phi \mathrm{d}\widehat{\mathrm{vol}}_M := \sum_{j=1}^J \int_{\hat{\Psi}_j(\hat{T}_j)} \xi_j(x)\phi(x)\mathrm{d}x$. *Then, w.h.p.*

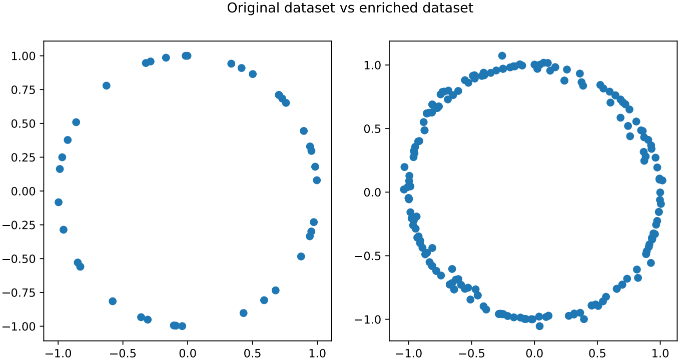
$$W_p\left(\frac{\widehat{\mathrm{vol}}_M}{|\widehat{\mathrm{vol}}_M|}, \frac{\mathrm{vol}_M}{|\mathrm{vol}_M|}\right) \lesssim \left(\frac{\ln n}{n}\right)^{\frac{k}{d}},$$

and this estimator is minimax on $\mathcal{P}_d^{s,k}$.

Toy example

$$\mathrm{d}\hat{\mu}_{n,h}(x) = (K_h * \mu_n)(x)\mathrm{d}\widehat{\mathrm{vol}}_M(x)$$

- $X_1, \ldots, X_n \sim \mu$
- $Y_1, \ldots, Y_N \sim \hat{\mu}_{n,h}$



Thank you for your attention!

Previous results

μ supported on M of dimension d with density f, $\mu_n =$ empirical measure.

- $W_p(\mu_n, \mu) \lesssim n^{-1/d} + n^{-1/(2p)}$ ([Weed, Bach 19], [Singh, Póczos 18], etc.)

If additionally $0 < f_{\min} \leq f \leq f_{\max}$:

- $W_p(\mu_n, \mu) \lesssim (\log n/n)^{1/d}$ [García Trillos & al. 19]
- $W_p(\hat{\mu}_n, \mu) \lesssim n^{-(s+1)/(2s+d)}$ if f is of regularity s and $M = [0,1]^d$ [Weed, Berthet 19]

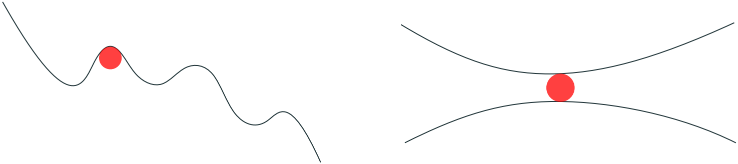
Here:

- Problem 1: M is **not flat**.
- Problem 2: M is **unknown**.

How do we quantify the regularity of a manifold?

Definition

The reach $\tau(M)$ of a manifold $M \subset \mathbb{R}^D$ is the largest radius r such that the projection on the manifold π_M is defined on $\{x \in \mathbb{R}^D,\ d(x, M) < r\}$.



Definition ($\mathcal{C}^k$ model for manifolds [Aamari-Levrard 2019])

$\mathcal{M}_d^k$ is the collection of d-dimensional manifolds $M \subset \mathbb{R}^D$ with $\tau(M) \geq \tau_{\min}$, and for all $x \in M$, M is locally the graph of a function $\Psi_x : T_x M \to \mathbb{R}^D$ with $\|\Psi_x\|_{\mathcal{C}^k} \leq L$.

Definition (Sobolev spaces on manifolds)

Let $M \in \mathcal{M}_d^k$ and $0 \le s \le k-2$ an integer. We let

$$\|f\|_{H_p^s(M)} := \max_{0\le l\le s} \left(\int_M \left\|d^l f(x)\right\|_{\mathrm{op}}^p \mathrm{d}x\right)^{1/p}$$

and $H_p^s(M)$ be the corresponding Banach space.

We let $\mathcal{P}_d^{s,k} = \mathcal{P}_{d,\tau_{\min},L_k,L_s,f_{\min},f_{\max}}^{s,k}$ be the set of probability distributions μ supported on some $M \in \mathcal{M}_{d,\tau_{\min},L_k}^k$, with density $f_{\min} \le f \le f_{\max}$ and $\|f\|_{H_p^s(M)} \le L_s$.

$$\Longrightarrow \text{ What is } \mathcal{R}_n(\mu, \mathcal{P}_d^{s,k}, W_p)?$$