Draft: Distance to the measure induced by a uniform kernel density estimator.

February 7, 2023

Big question: do we gain anything compared to just using the empirical measure?

1 Introduction

1.1 Background

1. In [1], the authors introduce the distance to a measure. Better stability (Wasserstein stability) guarantees with respect to noise and outliers than the classical stability results for distance functions.

2. In [3] the authors propose an approximation scheme for approximating the sublevel sets of the distance to a measure function using power distances and weighted Čech (and Rips) complexes.

3. They use this approximation scheme to compute the sublevel set homology of the distance to the empirical measure.

4. A similar approach using weighted complexes and power distances have been used in [4] for approximating the sublevel set homology of the kernel distance to a measure.

1.2 Goal

1. We want to study kernel density estimators using the disk kernel which can be viewed as a thickening of the Dirac measure.

2. We then consider the measure induced by this KDE and the distance to this measure.

3. Letting the kernel bandwidth approach zero as the number of samples in the input point cloud grows, we want to show that this gives uniformly stable approximations of the sublevel sets of the distance to measure function.

4. We may also relate the disk KDE to the multicover bifiltration ([5], [6], [2] and [7]).

2 Background

2.1 Kernel density estimators

We consider the *disk kernel* $D_h\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ with *bandwidth parameter* $h\mathbb{R}_{>0}$ defined by

$$(x, y) \mapsto \begin{cases} 1 & \|x - y\| \leq h \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

We then define the *normalized disk kernel* $K_h\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ by normalizing the disk kernel such $K_h(x, -)\colon \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ integrates to 1 for all $x \in \mathbb{R}^d$. In other words, K_h is defined as $K_h(x, y) := \frac{\Gamma(\frac{d}{2}+1)}{(h\sqrt{\pi})^d} D_h(x, y)$. Now, given a point cloud $X \subset \mathbb{R}^d$ with $|X| = n$, we define the *disk kernel density estimator (DKDE) $f_{X,h}$ with bandwidth h* on X as the weighted sum

$$f_{X,h} = \frac{1}{n} \sum_{x \in X} K_h(x, -)\colon \mathbb{R}^d \to \mathbb{R}_{\geq 0}.$$

The induced *disk measure on X* is then defined as the probability measure $\mu_{X,h}\colon A \to \int_A f_{X,h}(y)dy$. In other words, for a measurable set $A \subset \mathbb{R}^d$, we have

$$\mu_{X,h}(A) = \frac{\Gamma(\frac{d}{2} + 1)}{n(h\sqrt{\pi})^d} \sum_{x \in X} \mathrm{Vol}(A \cap \bar{B}(x, h)).$$

Given any strictly monotonically function $\alpha\colon \mathbb{Z}_{>0} \longrightarrow \mathbb{R}_{>0}$, we introduce the notation $\mu_{X,\alpha} = \mu_{X,\alpha(n)}$. Put differently, the bandwidth decrease as the number of samples in X increase. For example, we could choose $\alpha(n) = \frac{c}{n}$ for some constant $c \in \mathbb{R}_{>0}$.

2.2 Distance to a measure

Given a probability measure μ on a metric space (M, d_M) and a mass parameter $m \in (0, 1]$, define the *distance to the measure* μ, denoted $d_{\mu,m}\colon M \longrightarrow \mathbb{R}_{\geq 0}$, as

$$d_{\mu,m}(y) = \sqrt{\frac{1}{m}\int_0^m \delta^2_{\mu,t}(y)dt}$$

where $\delta_{\mu,t}\colon M \longrightarrow \mathbb{R}_{\geq 0}$ is defined as $\delta_{\mu,t}(y) = \inf\{r \geq 0 \mid \mu(\bar{B}(y,r)) > t\}$ and $\bar{B}(y,r) = \{m \in M \mid d_M(y,m) \leq r\}$ is the closed ball centred at y of radius r.

Example 2.0.1. The distance to the disk measure $d_{X,\alpha,m} := d_{\mu_{X,\alpha},m}$ is then given by

$$d^2_{X,\alpha,m}(y) = \frac{C(n,d)}{m}\int_0^m \inf\left\{r \geq 0 \,\middle|\, \sum_{x\in X} \mathrm{Vol}(\bar{B}(y,r) \cap \bar{B}(x,\alpha(n))) \geq t\right\}^2 dt$$

where $C(n,d) = \frac{\Gamma(\frac{d}{2}+1)}{n(\alpha(n)\sqrt{\pi})^d}$.

One of the most important properties of the distance to a measure function is the stability it enjoys with respect to the Wasserstein distance.

Theorem 2.1 (Theorem 3.1 and 3.3 in [3])**.** Let μ and ν be probability measures on a metric space M and let $m \in (0, 1]$ be a mass parameter. Then,

$$\|d_{\mu,m} - d_{\nu,m}\|_\infty \leq \frac{1}{\sqrt{m}} W_2(\mu, \nu).$$

Moreover, if M is triangulable, then $\mathrm{Dgm}(d_{\mu,m})$ and $\mathrm{Dgm}(d_{\nu,m})$ are well-defined and

$$d_B(\mathrm{Dgm}(d_{\mu,m}), \mathrm{Dgm}(d_{\nu,m})) \leq \frac{1}{\sqrt{m}} W_2(\mu, \nu).$$

To establish uniform stability on the level of sublevel set homology, we need to find an upper bound of $W_2(\mu_{X,\alpha}, \mu_{X_0,\alpha})$ which is $O(n^{-1})$. Maybe find an upper bound on the distance from each to a common measure ν, then apply the triangle inequality for W_2?

2.2.1 Approximating the sublevel sets of $d_{\mu,m}$

Following the exposition given in [3], we see how the sublevel sets of $d_{\mu,m}$ can be approximated by the union of balls centred at each point in some sample P of $\mathrm{Supp}(\mu)$. Given a subset P of a metric space (M, d_M) and a weight function $w \colon P \to \mathbb{R}$, we define the *power distance* $p(x) \colon M \to \mathbb{R}_{\geq 0}$ associated with (P, w) by letting

$$p(x) = \sqrt{\min_{p \in P} \left(d_M(x, p)^2 + w(p)^2\right)}.$$

Let μ be a probability measure on a metric space M and let $m \in (0, 1]$. Given a subset $P \subseteq M$, define $d^P_{\mu,m}$ to be the power distance associated with $(P, d_{\mu,m})$. That is, $d^P_{\mu,m} = \sqrt{\min_{p \in P} \left(d_M(x, p)^2 + d_{\mu,m}(p)^2\right)}$

Theorem 2.2 (Theorem 4.5 in [3])**.** If $d_H(P, \mathrm{Supp}(\mu)) \leq \epsilon$, then

$$\frac{1}{\sqrt{2}} d_{\mu,m} \leq d^P_{\mu,m} \leq \sqrt{5}(d_{\mu,m} + \epsilon).$$

For the empirical measure $\mu_X = \frac{1}{|X|} \sum_{x \in X} \delta_x$ on some point cloud X one can choose $P = X$ (giving $\epsilon = 0$ in Theorem 2.2) to obtain interleaving guarantees on the sublevel set filtrations of $d_{\mu_X,m}$ and $d^P_{\mu_X,m}$. In the case where $M = \mathbb{R}^d$ with the L_2-norm and $\mu = \mu_X$, the bounds can be slightly improved to $\frac{1}{\sqrt{2}} d_{\mu,m} \leq d^P_{\mu,m} \leq \sqrt{3} d_{\mu,m}$.

Example 2.2.1. Consider the distance $d_{X,\alpha,m}$ to the the disk measure $\mu_{X,\alpha}$ on X with $|X| = n$. Then, $P = X$ is an $\alpha(n)$-sample of $\mathrm{Supp}(\mu_{X,\alpha})$ so we can apply Theorem 2.2 to obtain the inequalities

$$\frac{1}{\sqrt{2}} d_{X,\alpha,m} \leq d^P_{X,\alpha,m} \leq \sqrt{5}(d_{X,\alpha,m} + \alpha(n)).$$

Can we hope for an additive interleaving instead of a multiplicative one?

References

[1] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. "Geometric inference for probability measures". In: *Foundations of Computational Mathematics* 11.6 (2011), pp. 733–751.

[2] Donald R Sheehy. "A Multicover Nerve for Geometric Inference." In: *CCCG*. 2012, pp. 309–314.

[3] Mickael Buchet et al. *Efficient and Robust Persistent Homology for Measures*. 2013. DOI: 10.48550/ARXIV.1306.0039. URL: https://arxiv.org/abs/1306.0039.

[4] Jeff M. Phillips, Bei Wang, and Yan Zheng. *Geometric Inference on Kernel Density Estimates*. 2013. DOI: 10.48550/ARXIV.1307.7760. URL: https://arxiv.org/abs/1307.7760.

[5] Andrew J. Blumberg and Michael Lesnick. *Stability of 2-Parameter Persistent Homology*. 2020. DOI: 10.48550/ARXIV.2010.09628. URL: https://arxiv.org/abs/2010.09628.

[6] René Corbet et al. *Computing the Multicover Bifiltration*. 2021. DOI: 10.48550/ARXIV.2103.07823. URL: https://arxiv.org/abs/2103.07823.

[7] Herbert Edelsbrunner and Georg Osang. "The multi-cover persistence of Euclidean balls". In: *Discrete & Computational Geometry* 65 (2021), pp. 1296–1313.