Probability measures on the space of persistence diagrams

Yuriy Mileyko[1], Sayan Mukherjee[2] and John Harer[1]

[1] Departments of Mathematics and Computer Science, Center for Systems Biology, Duke University, 27708, USA

[2] Departments of Statistical Science, Computer Science, and Mathematics, Institute for Genome Sciences & Policy, Duke University, 27708, USA

E-mail: `john.harer@duke.edu,yury@math.duke.edu,sayan@stat.duke.edu`

Abstract. This paper shows that the space of persistence diagrams has properties that allow for the definition of probability measures which support expectations, variances, percentiles and conditional probabilities. This provides a theoretical basis for a statistical treatment of persistence diagrams, for example computing sample averages and sample variances of persistence diagrams. We first prove that the space of persistence diagrams with the Wasserstein metric is complete and separable. We then prove a simple criterion for compactness in this space. These facts allow us to show the existence of the standard statistical objects needed to extend the theory of topological persistence to a much larger set of applications.

1. Introduction

A central idea in topological data analysis (TDA) is to start with point cloud data and compute topological summaries of this data. These summaries should provide useful information about the structure and geometry of the data. The majority of the literature in TDA has focused on the mathematical properties captured by the summaries and the computational issues that arise in obtaining these summaries [1, 2, 3]. This ignores a fundamental aspect of classical data analysis – quantification of the uncertainty, noise, and reproducibility of summaries computed from data. In the framework of statistical inference the objects of study are expectations, variances, and conditional probabilities of these topological summaries. The objective of our paper is to formalize these objects and show that they are well defined.

In this paper we focus on a commonly used topological summary, the persistence diagram [1]. We develop the probability theory needed to define basic statistical objects such as means, variances, and conditional probabilities on the space of persistence diagrams. The following simple problem motivates the theory. Given persistence diagrams from one hundred realizations of point cloud data obtained from one geometric object what is the average diagram and how much do these diagrams vary? The fundamental difficulty in evaluating averages and variances on persistence diagrams is the lack of a clearly defined probability space on persistence diagrams. Statistical inference requires probability spaces with clear definitions of expectations and variances.

In this work we start with the assumption that the point cloud data is generated by a stochastic process with a well defined probability distribution. An example would be n points drawn independently and identically from the uniform distribution on a torus in $\mathbb{R}^3$. Throughout this paper we will refer to a realization of the point cloud data as a point sample – a point sample will typically consist of n points drawn from a geometric object with a specified sampling distribution. We will show that the probability distribution on the point sample induces a probability distribution on persistence diagrams with well defined notions of expectation, variance, percentiles and conditional probabilities. The key challenge in this construction is to show that the space of persistence diagrams is a Polish space – a topological space homeomorphic to a separable complete metric space [4]. We also provide a simple characterization of compactness in the space of persistence diagrams. These two results allow us to define Fréchet expectations and variances as well as conditional probabilities.

Most of the related work on stochastic aspects of topological summaries can be subdivided into two categories: the study of random abstract simplicial complexes generated from stochastic processes [5, 6, 7, 8, 9] and convergence or consistency of topological summaries as the number of points goes to infinity [10, 11, 12, 13]. Neither of these categories are concerned with developing a framework to allow for statistical operations on topological summaries such as persistence diagrams. An effort closer in spirit to our work is developed in Chazal et al [14] where a distance metric between the empirical measure of a point sample and a probability measure is defined and topological summaries of this metric is examined. The key idea in this paper is the metric between measures is more robust than standard distance metrics used in the analysis of point samples. They do not attempt to define probability measures on the topological summaries and define averages and variances.

The paper is structured as follows. In Section 2 we provide an overview of persistent homology and its properties and define the space of persistence diagrams. In Section 3 we prove that the space of persistence diagrams is complete and separable and provide a simple criterion for compactness. Section 4 is devoted to proving existence of Fréchet expectations. We finish by discussing our results in Section 5.

2. Persistent homology

In this section we provide a brief description of persistent homology and persistence diagrams and define the space of persistence diagrams.

2.1. Sublevelset filtration

Let us consider a topological space $\mathbb{X}$ and a continuous function $f : \mathbb{X} \to \mathbb{R}$. For example, $\mathbb{X} = \mathbb{R}^n$ and $f = d_S$, the distance function to a sample $S \subset \mathbb{R}^n$. Let $\mathbb{X}_a = f^{-1}(-\infty, a]$ denote the sublevel set of f at the threshold a. Inclusions $\mathbb{X}_a \subset \mathbb{X}_b$, $a \leq b$, induce homomorphisms of the homology groups of sublevel sets:

$$\mathsf{f}_\ell^{a,b} : \mathsf{H}_\ell(\mathbb{X}_a) \to \mathsf{H}_\ell(\mathbb{X}_b),$$

for each dimension ℓ. We call a value $c \in \mathbb{R}$ a *homological critical value* of f if there exists ℓ such that $\mathsf{f}_\ell^{c-\delta,c}$ is not an isomorphism for any $\delta > 0$. We call f *tame* if has only a finite number of homological critical values and if $\mathsf{H}_\ell(\mathbb{X}_a)$ are finitely generated for all $a \in \mathbb{R}$ and all dimensions ℓ. For the rest of the section, we assume that f is tame and bounded, and that homology groups are defined over field coefficients, e.g. $\mathbb{Z}_2$.

2.2. Birth and death groups

Notice that the assumption of tameness implies that the image $\mathrm{Im}\mathsf{f}_\ell^{a-\delta,b} \subset H_\ell(\mathbb{X}_b)$ is independent of $\delta > 0$ if δ is sufficiently small. We shall denote such an image by $\mathsf{F}_\ell^{a-,b}$. Now, consider the following quotient group:

$$\mathsf{B}_\ell^a = \mathsf{H}_\ell(\mathbb{X}_a)/\mathsf{F}^{a-,a}.$$

This group is the cokernel of $\mathsf{f}_\ell^{a-\delta,a}$ and it captures homology classes which did not exist in sublevel sets preceding $\mathbb{X}_a$. We call this group the *ℓ-th birth group* at $\mathbb{X}_a$, and we say that a homology class $\alpha \in \mathsf{H}_\ell(\mathbb{X}_a)$ is born at $\mathbb{X}_a$ if it represents a nontrivial element $[\alpha] \in \mathsf{B}_\ell^a$, that is, the canonical projection of α is not zero. The tameness assumption implies that there are only a finite number of nontrivial birth groups.

Let us now consider the map

$$\mathsf{g}_\ell^{a,b} : \mathsf{B}_\ell^a \to \mathsf{H}_\ell(\mathbb{X}_b)/\mathsf{F}^{a-,b},$$

such that $\mathsf{g}_\ell^{a,b}([\alpha]) = [\mathsf{f}_\ell^{a,b}(\alpha)]$, where $\alpha \in \mathsf{H}_\ell(\mathbb{X}_a)$, and the square brackets denote the images under the corresponding canonical projections. For technical reasons, we set $\mathsf{g}_\ell^{a,b} = 0$ if $b = \sup_{x\in\mathbb{X}} f(x)$. The kernel of this map, which we denote by $\mathsf{D}_\ell^{a,b}$, captures homology classes that were born at $\mathbb{X}_a$ but at $\mathbb{X}_b$ are homologous to homology classes born before $\mathbb{X}_a$. We call $\mathsf{D}_\ell^{a,b}$ the *death subgroup* of B_ℓ^a at $\mathbb{X}_b$, and we say that a homology class $\alpha \in \mathsf{H}_\ell(\mathbb{X}_a)$ *dies entering* $\mathbb{X}_b$ if $[\alpha] \in \mathsf{D}_\ell^{a,b}$ but $[\alpha] \notin \mathsf{D}_\ell^{a,b-\delta}$ for any $\delta > 0$. We also call b a *degree-r death value* of B_ℓ^a if $\text{rank}\mathsf{D}_\ell^{a,b} - \text{rank}\mathsf{D}_\ell^{a,b-\delta} = r > 0$ for all sufficiently small $\delta > 0$. Notice that the sum of the degrees of all the death values of a birth group is equal to its rank.

2.3. Persistence diagrams

If a homology class α is born at $\mathbb{X}_a$ and dies entering $\mathbb{X}_b$ we set $\mathrm{b}(\alpha) = a, \mathrm{d}(\alpha) = b$. The *persistence* of α is the difference between the two values, $\mathrm{pers}(\alpha) = \mathrm{d}(\alpha) - \mathrm{b}(\alpha)$. We represent the births and deaths of ℓ-dimensional homology classes by a multiset of points in $\mathbb{R}^2$, the *ℓ-th persistence diagram* denoted by $\mathrm{Dgm}_\ell(f)$. For each nontrivial birth group B_ℓ^a the diagram contains points $x_i = (a, d_i)$, where d_i are the death values of B_ℓ^a, and the multiplicity of x_i is equal to the degree of the corresponding death value d_i. Thus, we draw births along the horizontal axis, deaths along the vertical axis, and since deaths happen only after births, all points lie above the diagonal, each point representing the group of homology classes that were born and died at the corresponding values. The diagram also includes points on the diagonal. We can think that such points correspond to trivial homology classes which are born and die at every level. The persistence of a point $x \in \mathrm{Dgm}_\ell(f)$, denoted by $\mathrm{pers}(x)$, is the persistence of the corresponding homology classes, and is equal to the horizontal (or vertical) distance from x to the diagonal.

2.4. Wasserstein distance and the space of persistence diagrams

To measure similarities between persistent homology of two functions we use the following definition of a distance between persistence diagrams:

Definition 1 (Wasserstein distance)**.** The p-th Wasserstein distance between two persistence diagrams, d_1 and d_2, is defined as

$$W_p(d_1, d_2) = \left(\inf_\gamma \sum_{x\in d_1} \|x - \gamma(x)\|_\infty^p \right)^{\frac{1}{p}},$$

where γ ranges over all bijections from d_1 to d_2. The set of bijections is nonempty because of the diagonal.

We can now regard a persistence diagram as an element of a metric space – the set of all persistence diagrams endowed with the Wasserstein distance. Unfortunately, this space is not complete, hence not appropriate for statistical inference. Indeed, let $x_n = (0, 2^{-n}) \in \mathbb{R}^2$, $n \in \mathbb{N}$, and let d_n be the persistence diagram containing $x_1, \ldots, x_n$ (each with multiplicity 1). Then

$$W_p(d_n, d_{n+k}) \leq \frac{1}{2^{n+k}},$$

so d_n is Cauchy. It is clear, however, that the number of off-diagonal points in d_n grows to ∞ as $n \to \infty$, so this sequence cannot have a limit in our space. This example suggests that the set of the diagrams forming the space be modified. Notice that the space of all finite sequences endowed with the l_p metric is also not complete for a very similar reason. Hence, we proceed as follows.

Definition 2 (Persistence diagram)**.** A persistence diagram is a countable multiset of points in $\mathbb{R}^2$ along with the diagonal $\Delta = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$, where each point on the diagonal has infinite multiplicity.

While we do not have a notion of a norm of a persistence diagram, we can impose a finiteness condition on the distance to a particular diagram. Let $d_\emptyset$ denote the *empty* persistence diagram, that is, the persistence diagram containing only the diagonal. Notice that

$$\mathrm{pers}(x) = 2 \inf_{y \in \Delta} \|x - y\|_\infty.$$

Therefore,

$$(W_p(d, d_\emptyset))^p = 2^{-p} \sum_{x \in d} (\mathrm{pers}(x))^p.$$

Recall from [15] the following definition:

Definition 3 (Total persistence)**.** The degree-p total persistence of a persistence diagram d is defined as

$$\mathrm{Pers}_p(d) = \sum_{x \in d} (\mathrm{pers}(x))^p.$$

Thus, $\mathrm{Pers}_p(d) = 2^p (W_p(d, d_\emptyset))^p$, and we see that requiring finiteness of the distance to the empty diagram is equivalent to requiring finiteness of total persistence.

Definition 4 (Space of persistence diagrams)**.** We define the space of persistence diagrams as

$$D_p = \{d | W_p(d, d_\emptyset) < \infty\} = \{d | \mathrm{Pers}_p(d) < \infty\}.$$

In this paper we shall consider only the case $p \geq 1$.

We finish this section by stating an important stability result from [15] which shows that under mild assumptions on $\mathbb{X}$ computing a persistence diagram of a tame Lipschitz functions is a continuous map. Suppose that $\mathbb{X}$ is a metric space such that for any persistence diagram d computed for a Lipschitz function f with the Lipschitz constant $\mathrm{Lip}(f) \leq 1$ we have $\mathrm{Pers}_k(d) \leq C_{\mathbb{X}}$, where $C_{\mathbb{X}}$ is a constant that depends only on $\mathbb{X}$. We shall say in this case that $\mathbb{X}$ *implies bounded degree-k total persistence.*

Proposition 1 (Wasserstein Stability). *If $\mathbb{X}$ is a triangulable, compact metric space that implies bounded degree-k total persistence for $k \geq 1$ and $f_1, f_2 : \mathbb{X} \to \mathbb{R}$ are tame, Lipschitz functions, then for all dimensions ℓ and $p \geq k$ we have*

$$W_p(\mathrm{Dgm}_\ell(f_1), \mathrm{Dgm}(_\ell f_2)) \leq C^{\frac{1}{p}} \|f_1 - f_2\|_\infty^{1-\frac{k}{p}},$$

where $C = C_{\mathbb{X}} \max\{\mathrm{Lip}(f_1)^k, \mathrm{Lip}(f_2)^k\}$.

3. Properties of the space of persistence diagrams

Before we define expectations, variances and conditional probabilities for persistence diagrams we need to prove that the space of persistence diagrams has particular properties. This space needs to be a Polish space. We also need to understand what subspaces of D_p are compact.

3.1. Completeness and separability of D_p

We begin by addressing the issue of completeness.

Theorem 1. *D_p is complete in the metric W_p.*

Let $d_n \in D_p$ be a Cauchy sequence. There are three main steps in the proof. First, we show that d_n converges "persistence-wise" (we make this statement precise later) to a diagram d^*. Second, we show that d^* belongs to D_p. Third, we show that d_n converges to d^* in the metric W_p.

Given a persistence diagram $d \in D_p$, we shall use $|d|$ to denote the total multiplicity of d, that is, the number of (off diagonal) points in d counting multiplicities. For $\alpha > 0$ let $u_\alpha : D_p \to D_p$ be defined by

$$x \in u_\alpha(d) \iff x \in d \;\&\; \mathrm{pers}(x) \geq \alpha.$$

The diagram $u_\alpha(d)$ contains only those points in d that have persistence at least α, we call it the *α-upper part of d*. Similarly, we define $l_\alpha : D_p \to D_p$ by:

$$x \in l_\alpha(d) \iff x \in d \;\&\; \mathrm{pers}(x) < \alpha.$$

Thus $l_\alpha(d)$ is the *α-lower part of d* as it contains only those points in d that have persistence less than α.

We have introduced the upper and lower parts of persistence diagrams in order to define an analogue of pointwise convergence. Since the α-upper part of a diagram has finite total multiplicity for any $\alpha > 0$, it is reasonable to consider convergence of the α-upper part of each element of the sequence d_n. If these converged to an element of D_p, the union of such elements over all α would be a natural candidate for the limit of d_n. Unfortunately, the situation is more complicated due to *convergence from below*, when points in $l_\alpha(d_n)$ converge to points with persistence α (see Figure 1). The following lemma is critical as it shows that we can control such behavior because points in d_n start separating according to their persistence as n increases.

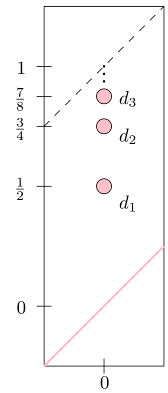

Figure 1. Example of convergence from below. Shown are three first diagrams from the sequence d_n such that $|d_n| = 1$, $\mathrm{b}(x) = 0$ for all for $x \in d_n$ and $n \in \mathbb{N}$, and $\mathrm{pers}(x) = 1 - 2^{-n}$ for $x \in d_n$. The sequence of off diagonal points converges to a single point with persistence 1. It is clear, however, that 1-upper part of any d_n is empty.

Lemma 2 (Persistence-wise Separation)**.** *Let $\alpha > 0$. Then there exist $M_\alpha \in \mathbb{Z}_+$ and δ_α, $0 < \delta_\alpha < \alpha$, such that $\forall \delta$ in the interval $[\delta_\alpha, \alpha)$, eventually $|u_\delta(d_n)| = M_\alpha$; i.e. $\exists N_\delta > 0$ such that $|u_\delta(d_n)| = M_\alpha$ whenever $n > N_\delta$.*

Proof. For each δ with $0 < \delta < \alpha$ let $M^\delta_{sup} = \limsup_{n\to\infty} |u_\delta(d_n)|$, $M^\delta_{inf} = \liminf_{n\to\infty} |u_\delta(d_n)|$. Notice that $M^\delta_{sup} < \infty$, otherwise, we could find a subsequence d_{n_k} such that $|u_\delta(d_{n_k})| > k$ so that $W_p(d_{n_k}, d_\emptyset) \geq k^{1/p}\delta/2 \to \infty$ as $k \to \infty$. However, $W_p(d_n, d_\emptyset)$ is bounded because d_n it is Cauchy.

If $\delta_1 > \delta_2$, $|u_{\delta_1}(d_n)| \leq |u_{\delta_2}(d_n)|$ so $M^{\delta_1}_{sup} \leq M^{\delta_2}_{sup}$ and $M^{\delta_1}_{inf} \leq M^{\delta_2}_{inf}$. Therefore, the limits $\lim_{\delta\to\alpha} M^\delta_{sup} = M_{sup}$ and $\lim_{\delta\to\alpha} M^\delta_{inf} = M_{inf}$ exist. Moreover, for arbitrary $\delta_0 > 0$ the range of values of M^δ_{sup} and M^δ_{inf} for $\delta \geq \delta_0$ is finite, so there is $\delta_\alpha > 0$ such that $M_{sup} = M^\delta_{sup}$ and $M_{inf} = M^\delta_{inf}$ whenever $\delta_\alpha \leq \delta \leq \alpha$.

Suppose now that $M_{inf} < M_{sup}$. Take $\delta \in (\delta_\alpha, \alpha)$, and let $\varepsilon = \delta - \delta_\alpha > 0$. Let d_{n_s} and d_{n_i} be two subsequences such that $|u_\delta(d_{n_s})| = M_{sup}$ and $|u_{\delta_\alpha}(d_{n_i})| = M_{inf}$. On the one hand, we can pick $K > 0$ such that $W_p(d_{n_s}, d_{n_i}) < \varepsilon/4$ $\forall s, i > K$. On the other hand, $|u_\delta(d_{n_s})| > |u_{\delta_\alpha}(d_{n_i})|$, which implies that for any bijection $\gamma : d_{n_s} \to d_{n_i}$ there is a point $x \in d_{n_s}$ such that $\mathrm{pers}(x) \geq \delta$, $\mathrm{pers}(\gamma(x)) < \delta_\alpha \Rightarrow \|x - \gamma(x)\|_\infty \geq \varepsilon/2$. Therefore, $W_p(d_{n_s}, d_{n_i}) > \varepsilon/2$, which is a contradiction. We then set $M_\alpha = M_{sup} = M_{inf}$. $\square$

Given $\alpha > 0$, let $d^\alpha_n = u_{\delta_\alpha}(d_n)$, d^α_n contain points whose persistence (in the limit) is at least α.

Lemma 3. *For any $\alpha > 0$ the sequence d^α_n is Cauchy.*

Proof. Lemma 2 implies that $\forall \delta \in (\delta_\alpha, \alpha)$ $\exists N > 0$ such that $\forall n > N$, d_n contains no points with persistence in the range $[\delta_\alpha, \delta)$. Let $\varepsilon > 0$, $\varepsilon_0 = \min\{\varepsilon/2, (\delta - \delta_\alpha)/8\}$. Increase N so that $\forall n, m > N$ we have $W_p(d_n, d_m) < \varepsilon_0$. Then there is a bijection

$\gamma : d_n \to d_m$ such that

$$\left(\sum_{x\in d_n} \|x-\gamma(x)\|_\infty^p\right)^{\frac{1}{p}} < 2\varepsilon_0 \leq \frac{\delta-\delta_\alpha}{4}.$$

This inequality implies that γ maps points in d_n^α to points in d_m^α, therefore,

$$W_p(d_n^\alpha, d_m^\alpha) \leq \left(\sum_{x\in d_n^\alpha} \|x-\gamma(x)\|_\infty^p\right)^{\frac{1}{p}} < 2\varepsilon_0 \leq \varepsilon.$$

□

The following lemma shows that for each persistence level α the sequence d_n^α converges.

Lemma 4 (Persistencewise Convergence)**.** *For any* $\alpha > 0$ $\exists$ $d^\alpha \in D_p$ *such that* $\lim_{n\to\infty} W_p(d_n^\alpha, d^\alpha) = 0$*, and hence* $|d^\alpha| = M_\alpha$*. Moreover,* $d^{\alpha_1} \subset d^{\alpha_2}$ *if* $\alpha_1 > \alpha_2$*.*

Proof. Let $\alpha > 0$, and let K_n^ε denote the minimal number of closed disjoint ℓ_∞ plane disks (i.e. squares) of radius ε needed to cover the off diagonal points of d_n^α. Taking $N > 0$ such that $|d_n^\alpha| = M_\alpha$ and $W_p(d_n^\alpha, d_m^\alpha) < \varepsilon$ for all $n, m > N$, we see that K_n^ε is non-increasing and $\lim_{n\to\infty} K_n^\varepsilon = K^\varepsilon \leq M_\alpha$. Since $K^{\varepsilon_1} \leq K^{\varepsilon_2}$ if $\varepsilon_1 \geq \varepsilon_2$ we can also take $\lim_{\varepsilon\to 0} K^\varepsilon = K \leq M_\alpha$. Notice that $\exists\varepsilon_\alpha$ such that $\forall\varepsilon \in (0, \varepsilon_\alpha)$ $\exists N > 0$ with $K_n^\varepsilon = K$ $\forall n \geq N$. Moreover, we can take $N > 0$ such that if $B_1^\varepsilon, \ldots, B_K^\varepsilon$ are disks of radius ε that cover d_N^α, then they also cover d_n^α for all $n > N$. These observations allow us to construct sequences $\varepsilon_i \searrow 0$, $n_i \nearrow \infty$, and K sequences of disks $B_j^{\varepsilon_i}$, $j = 1, \ldots, K$ such that $B_j^{\varepsilon_i}$ cover d_n^α for $n \geq n_i$, and $B_j^{\varepsilon_{i+1}} \subset B_j^{\varepsilon_i}$, $j = 1, \ldots, K$. Therefore, we obtain K points $x_j = \cap_i B_j^{\varepsilon_i}$, and we assign x_j the multiplicity equal to the number (counting multiplicities) of points of $d_{n_i}^\alpha$ inside $B_j^{\varepsilon_i}$ (notice that this number is well defined).

Let d^α be the persistence diagram consisting of these x_j. Let $\varepsilon > 0$, and take $I > 0$ such that $\varepsilon_i < \varepsilon M_\alpha^{-1/p}$ for all $i > I$. By pairing points of $d_{n_i}^\alpha$ inside $B_j^{\varepsilon_i}$ with x_j we see that $W_p(d_{n_i}^\alpha, d^\alpha) < \varepsilon$, for $i > I$. This shows that $d_n^\alpha \to d^\alpha$.

The last statement of the lemma follows from the fact that if $\alpha_1 > \alpha_2$, then points $x \in d_n^{\alpha_2}$ such that $x \notin d_n^{\alpha_1}$ have $\mathrm{pers}(x) < \delta_{\alpha_1} < \alpha_1$, while for large n points in $d_n^{\alpha_1}$ have persistence at least $\alpha_1 - \delta$, where δ is arbitrarily small. Hence, the K_2 sequences of disks $B_j^{\varepsilon_i}$ covering $d_{n_i}^{\alpha_2}$ (as constructed above) contain K_1 sequences of disks that cover $d_{n_i}^{\alpha_1}$ and also satisfy the requirements of the above construction. □

Lemma 4 allows us to define $d^* = \cup_{\alpha>0} d^\alpha$. It is not difficult to show that that $d^* \in D_p$.

Lemma 5. $d^* \in D_p$*.*

Proof. First note that since d_n is Cauchy, there is a constant $C > 0$ such that $\forall n, W_p(d_n, d_\emptyset) \leq C$. Let $\alpha > 0$, and let $N > 0$ be such that $\forall n > N, W_p(d^\alpha, d_n^\alpha) < 1$. Take any such n, then

$$W_p(d^\alpha, d_\emptyset) \leq W_p(d^\alpha, d_n^\alpha) + W_p(d_n^\alpha, d_\emptyset) \leq 1 + C.$$

Since the right hand side is independent of α, we obtain $W_p(d^*, d_\emptyset) \leq 1 + C$. □

At this point it follows from lemmas 3 and 5 that given $\varepsilon > 0$, we can pick $\alpha > 0$ and $N > 0$ such that $W_p(d^\alpha, d^*) < \frac{\varepsilon}{3}$ and $W_p(d^\alpha, d_n^\alpha) < \frac{\varepsilon}{3}$, for $n > N$. Using the triangle inequality (twice) we see that completeness of D_p follows from the following final result. Let $d_{n,\alpha} = l_{\delta_\alpha}(d_n)$.

Lemma 6. *$\forall \varepsilon > 0$, $\exists \alpha_0 > 0$ such that $\forall n \in \mathbb{N}$ and $0 < \alpha \leq \alpha_0$ we have $W_p(d_{n,\alpha}, d_\emptyset) < \varepsilon$ and hence $W_p(d_n^\alpha, d_n) < \varepsilon$.*

Proof. We prove the lemma by contradiction. Suppose that $\exists \varepsilon > 0$ such that $\forall \alpha > 0$ $\exists n \in \mathbb{N}$ with $W_p(d_{n,\alpha}, d_\emptyset) \geq \varepsilon$. Take such an ε. Let $\{\alpha_i\}_{i\in\mathbb{N}}$ be a sequence of positive values monotonically decreasing to 0. Then we can find a subsequence d_{n_i} such that $W_p(d_{n_i,\alpha_i}, d_\emptyset) \geq \varepsilon$. Let $0 < \delta < \varepsilon/4$, and pick $k \in \mathbb{N}$ such that $W_p(d_{n_k}, d_{n_i}) < \delta$ for all $i \geq k$. Now pick $j \geq k$ such that $W_p(d_{n_k,\alpha_i}, d_\emptyset) < \delta$ for all $i \geq j$. This implies that

$$W_p(d_{n_i,\alpha_i}, d_{n_k,\alpha_j}) \geq W_p(d_{n_i,\alpha_i}, d_\emptyset) - W_p(d_{n_k,\alpha_j}, d_\emptyset) \geq \varepsilon - \delta > 3\delta.$$

We shall now show that this inequality leads to a contradiction. Let $\gamma_i : d_{n_i} \to d_{n_k}$ be a bijection such that

$$\sum_{x\in d_{n_i}} \|x - \gamma_i(x)\|_\infty^p < 2\delta^p.$$

Then we have the same inequality for the part of the sum over points $x \in d_{n_i,\alpha_i}$, that is

$$\sum_{x\in d_{n_i,\alpha_i}} \|x - \gamma_i(x)\|_\infty^p = \sum_{\substack{x\in d_{n_i,\alpha_i} \\ \gamma_i(x)\in d_{n_k,\alpha_j}}} \|x - \gamma_i(x)\|_\infty^p + \sum_{\substack{x\in d_{n_i,\alpha_i} \\ \gamma_i(x)\notin d_{n_k,\alpha_j}}} \|x - \gamma_i(x)\|_\infty^p < 2\delta^p.$$

Notice that $\delta_{\alpha_j} > 0$, so let us pick $l > j$ such that $\delta_{\alpha_j} > 2\alpha_i$ for all $i \geq l$. Then taking $x \in d_{n_i,\alpha_i}$ such that $\gamma_i(x) \notin d_{n_k,\alpha_j}$ we see that

$$\|x - \gamma_i(x)\|_\infty \geq \frac{|\mathrm{pers}(x) - \mathrm{pers}(\gamma_i(x))|}{2} \geq \frac{\delta_{\alpha_j} - \alpha_i}{2} \geq \frac{\alpha_i}{2} \geq \frac{\mathrm{pers}(x)}{2}$$

where $i \geq l$. Let $\hat{\gamma}_i : d_{n_i,\alpha_i} \to d_{n_k,\alpha_j}$ be the bijection such that $\hat{\gamma}_i(x) = \gamma_i(x)$ if $x \in d_{n_i,\alpha_i}$ and $\gamma_i(x) \in d_{n_k,\alpha_j}$, and points $x \in d_{n_i,\alpha_i}$ with $\gamma_i(x) \notin d_{n_k,\alpha_j}$ as well as points $y \in d_{n_k,\alpha_j}$ with $\gamma^{-1}(y) \notin d_{n_i,\alpha_i}$ get mapped to the diagonal. Then for $i \geq l$ we have

$$\sum_{x\in d_{n_i,\alpha_i}} \|x - \hat{\gamma}_i(x)\|_\infty^p = \sum_{\substack{x\in d_{n_i,\alpha_i} \\ \gamma_i(x)\in d_{n_k,\alpha_j}}} \|x - \gamma_i(x)\|_\infty^p + \sum_{\substack{x\in d_{n_i,\alpha_i} \\ \gamma_i(x)\notin d_{n_k,\alpha_j}}} \left(\frac{\mathrm{pers}(x)}{2}\right)^p +$$

$$\sum_{\substack{y\in d_{n_k,\alpha_j} \\ \gamma_i^{-1}(y)\notin d_{n_i,\alpha_i}}} \left(\frac{\mathrm{pers}(y)}{2}\right)^p \leq \sum_{\substack{x\in d_{n_i,\alpha_i} \\ \gamma_i(x)\in d_{n_k,\alpha_j}}} \|x - \gamma_i(x)\|_\infty^p + \sum_{\substack{x\in d_{n_i,\alpha_i} \\ \gamma_i(x)\notin d_{n_k,\alpha_j}}} \|x - \gamma_i(x)\|_\infty^p + \delta^p$$

$$< 2\delta^p + \delta^p = 3\delta^p.$$

Therefore, $W_p(d_{n_i,\alpha_i}, d_{n_k,\alpha_j}) < 3\delta$ if $i \geq l$. Contradiction. □

We finish this section by proving separability of D_p.

Theorem 7. *D_p is separable.*

Proof. Let $S \subset D_p$ be a set of persistence diagrams with finite total multiplicity and such that their points have rational coordinates, that is,

$$S = \{d \in D_p | |d| < \infty \;\&\; x \in \mathbb{Q}^2 \;\forall x \in d\}.$$

If $d \in D_p$ then $\forall \varepsilon > 0$ we can find $\alpha > 0$ such that $W_p(l_\alpha(d), d_\emptyset) < \varepsilon/2$. Then we have $W_p(d, u_\alpha(d)) \leq W_p(l_\alpha(d), d_\emptyset) < \varepsilon/2$. Take $d_s \in S$ such that $W_p(d_s, u_\alpha(d)) < \varepsilon/2$. Then $W_p(d, d_s) \leq W_p(d, u_\alpha(d)) + W_p(d_s, u_\alpha(d)) < \varepsilon$, which implies that S is dense.

Notice that $S = \cup_{m=0}^{\infty} S_m$, where $S_m = \{d \in S | |d| = m\}$. Each S_m is isomorphic to subset of $\mathbb{Q}^{2m}$ and thus is countable. Hence, S is countable. □

3.2. Compactness in D_p

Of a particular interest are subspaces of persistence diagrams which are compact. This will require mild conditions which we specify in this subsection. We first recall a few properties of a compact metric space.

Proposition 2 (Totally bounded)**.** *A subset S of a metric space $\mathbb{X}$ is called totally bounded if $\forall \varepsilon > 0$ there exists a finite collection of open balls in $\mathbb{X}$ of radius ε whose union contains S.*

Proposition 3 (Relative compactness)**.** *In a complete metric space a set is totally bounded if and only if its closure is compact. This is relative compactness.*

Proposition 4. *Relatively compact sets have the property that their infinite subsets have at least one limit point.*

We first state some examples of sets of persistence diagrams that are not compact in D_p. We then define restrictions to a set $S \subset D_p$ that ensure compactness by eliminating such examples.

Example 1. *Consider $S \subset D_p$ consisting of diagrams with a single off diagonal point of persistence exactly $\varepsilon > 0$. Take a sequence $d_n \in S$ such that the birth of the off diagonal point of d_n is equal to $2n\varepsilon$ (see Figure 2(a)). We have $W_p(d_n, d_m) = 2^{1/p-1}\varepsilon$ for all $n \neq m$. Hence, no subsequence of d_n can be Cauchy and d_n does not have a limit point. Proposition 4 then implies this set is not compact.*

We can eliminate this example by imposing one of the following two conditions.

Definition 5 (Birth-death bounded)**.** A set $S \subset D_p$ is called *birth-death bounded*, if there is a constant $C > 0$ such that $\forall d \in S$ and $\forall x \in d$ $\mathrm{bd}(x) = \max\{|\mathrm{b}(x)|, |\mathrm{d}(x)|\} \leq C$.

Definition 6 (Off-diagonally birth-death bounded)**.** A set $S \subset D_p$ is called *off-diagonally birth-death bounded* if $\forall \varepsilon > 0$ $u_\varepsilon(S)$ is birth-death bounded.

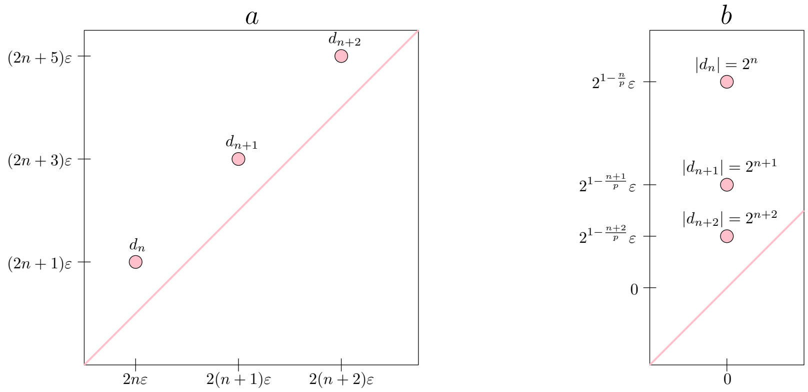

Figure 2. (a) Illustration of three consecutive diagrams from the sequence in Example 1. Each point represents a separate diagram with a single off diagonal point of multiplicity 1. (b) Illustration of three consecutive diagrams from the sequence in Example 2. Each point represents a separate diagram with a single off diagonal point whose multiplicity increases as its persistence decreases.

These two conditions are not enough to ensure compactness as is shown in the following example.

Example 2. *Consider a set $S \subset D_p$ with $W_p(d, d_\emptyset) \leq \varepsilon$, $\varepsilon > 0$, for all $d \in S$, and the death bounded by some constant $C > \varepsilon$. Choose a sequence $d_n \in S$ such that $|d_n| = 2^n$, and $\mathrm{b}(x) = 0$ for all for $x \in d_n$ and $n \in \mathbb{N}$, and $\mathrm{pers}(x) = 2^{1-n/p}\varepsilon$ for all $x \in d_n$ (see Figure 2(b)). It is easy to see that for all $n, m \in \mathbb{N}$, $m > n$, we have $W_p(d_n, d_m) \geq 2^{-1/p}\varepsilon$ as there will be at least 2^{m-1} points of persistence $2^{1-m/p}\varepsilon$ paired to the diagonal. Thus, no subsequence of d_n can be Cauchy and S is not compact.*

To deal with the above case we introduce the following notion:

Definition 7. A set $S \subset D_p$ is called *uniform* if for all $\varepsilon > 0$ there exists $\alpha > 0$ such that $W_p(l_\alpha(d), d_\emptyset) \leq \varepsilon$ for all $d \in S$.

It turns out that excluding cases that fall under the above examples is enough to achieve total boundedness.

Theorem 8. *A set $S \subset D_p$ is totally bounded if and only if it is bounded, off-diagonally birth-death bounded, and uniform.*

Proof. First, we prove the necessary part.

Assume that S is totally bounded, and let $\varepsilon > 0$. Take $0 < \delta < \varepsilon/4$ and let $B_n = B(d_n, \delta)$ for $n = 1, \ldots, N$ be a collection of balls of radius δ which cover S. For each d_n we can find a constant C_n such that $\mathrm{bd}(x) \leq C_n$ for $x \in d_n$ with $\mathrm{pers}(x) \geq \varepsilon$, and $\mathrm{pers}(x) \leq \varepsilon/4$ for all $x \in d_n$ with $\mathrm{bd}(x) > C_n$. Let $C = \max\{C_1, \ldots, C_N\}$. Also, we can find $\alpha > 0$ such that $W_p(l_\alpha(d_n), d_\emptyset) \leq \varepsilon/4$ for $n = 1, \ldots, N$.

We now prove by contradiction that S is off-diagonally birth-death bounded. Suppose that $d \in B_n$ and there is an $x \in d$ such that $\mathrm{pers}(x) \geq \varepsilon$ and $\mathrm{bd}(x) > C + \varepsilon$. Then for any bijection $\gamma : d \to d_n$ we have $\|x - \gamma(x)\|_\infty \geq \varepsilon/2 - \varepsilon/8$ which implies that $W_p(d, d_n) > \varepsilon/4$. This contradicts $W_p(l_\alpha(d_n), d_\emptyset) \leq \varepsilon/4$ and implies $C + \varepsilon$ as a birth-death bound for $u_\varepsilon(S)$.

The proof of the necessity of S being uniform also follows from contradiction. Suppose that $d \in B_n$ and $W_p(l_{\alpha/2}(d), d_\emptyset) > \varepsilon$. Consider a bijection $\gamma : d \to d_n$ and let $(d_b, d_t) \subset d$ be subdiagrams of $l_{\alpha/2}(d)$ such that $\text{pers}(\gamma(x)) < \alpha$ for $x \in d_b$ and $\text{pers}(\gamma(x)) \geq \alpha$ for $x \in d_t$. If $W_p(d_b, d_\emptyset) > \varepsilon/2$ then

$$\left(\sum_{x\in d_b} \|x - \gamma(x)\|_\infty^p\right)^{\frac{1}{p}} \geq W_p(d_b, \gamma(d_b)) \geq W_p(d_b, d_\emptyset) - W_p(\gamma(d_b), d_\emptyset) > \frac{\varepsilon}{2} - \frac{\varepsilon}{4},$$

where $\gamma(d_b)$ denotes the subdiagram of d_n which coincides with the image of d_b under γ. If $W_p(d_b, d_\emptyset) \leq \varepsilon/2$ then $W_p(d_t, d_\emptyset) > (\varepsilon^p - 2^{-p}\varepsilon^p)^{\frac{1}{p}} \geq \varepsilon/2$. Notice also that if $x \in d_t$ then $\|x - \gamma(x)\|_\infty > \alpha/4 \geq \text{pers}(x)/2$. Therefore,

$$\left(\sum_{x\in d_t} \|x - \gamma(x)\|_\infty^p\right)^{\frac{1}{p}} > \left(\sum_{x\in d_t} \left(\frac{\text{pers}(x)}{2}\right)^p\right)^{\frac{1}{p}} = W_p(d_t, d_\emptyset) > \frac{\varepsilon}{2}.$$

Thus, for any bijection $\gamma : d \to d_n$ we have

$$\left(\sum_{x\in d} \|x - \gamma(x)\|_\infty^p\right)^{\frac{1}{p}} > \frac{\varepsilon}{4}.$$

Therefore, $W_p(d, d_n) \geq \varepsilon/4$ which leads to a contradiction. Consequently $W_p(l_{\alpha/2}(d), d_\emptyset) \leq \varepsilon$ for all $d \in S$ which implies that S is uniform.

We now prove sufficiency. Let $\varepsilon > 0$ and let $\delta > 0$ be such that $W_p(l_\delta(d), d_\emptyset) < \varepsilon/2$ $\forall d \in S$. Take C such that for all $d \in S$ and all $x \in u_\delta(d)$ we have $\text{bd}(x) \leq C$. Since S is bounded, we can also find a constant $M \in \mathbb{N}$ such that $|u_\delta(d)| \leq M$ for all $d \in S$. Let $R \subset \mathbb{R}^2$ be the subset of the plane corresponding to points whose birth and death are bounded by C. Since R is a bounded subset of the plane it is also totally bounded, and we can find points $x_1, \ldots, x_N \in R$ such that for any $x \in R$ we have $\|x - x_n\|_\infty \leq M^{-1/p}\varepsilon/2$ for some x_n. Let d^* be the diagram consisting of points x_n, $1 \leq n \leq N$, each with multiplicity M and let $d_1, \ldots, d_L$ with $L = N^{M+1}$ be all subdiagrams of d^*. If $d \in S$ we can find d_n and a bijection $\gamma : u_\delta(d) \to d_n$ such that

$$\left(\sum_{x\in u_\delta(d)} \|x - \gamma(x)\|_\infty^p\right)^{\frac{1}{p}} < \frac{\varepsilon}{2}.$$

Let $\bar{\gamma} : d \to d_n$ be the extension of γ to d obtained by mapping the points in $l_\delta(d)$ to the diagonal. Then

$$\left(\sum_{x\in d} \|x - \bar{\gamma}(x)\|_\infty^p\right)^{\frac{1}{p}} = \left(\sum_{x\in u_\delta(d)} \|x - \gamma(x)\|_\infty^p + \sum_{x\in l_\delta(d)} \|x - \bar{\gamma}(x)\|_\infty^p\right)^{\frac{1}{p}} < 2^{\frac{1}{p}-1}\varepsilon \leq \varepsilon.$$

Therefore $W_p(d, d_n) < \varepsilon$. □

4. Existence of Fréchet expectations

In this section we define expectations and variances on the space of persistence diagrams. To this end we require a probability measure $\mathcal{P}_D$ on $(D_p, \mathcal{B}(D_p))$ where $\mathcal{B}(D_p)$ is the Borel σ-algebra on D_p. Later in this section we will relate the $\mathcal{P}_D$ to the measure $\mathcal{P}_\theta$ from which the data was generated. We will require that the measure $\mathcal{P}_D$ have a finite second moment

$$F_{\mathcal{P}_D}(d) = \int_{D_p} W_p(d,e)^2 \, \mathcal{P}_D(\mathrm{d}e) < \infty.$$

4.1. Existence of Fréchet expectations

We define the Fréchet expectation and variance as follows.

Definition 8 (Fréchet expectation)**.** Given a probability space $(D_p, \mathcal{B}(D_p), \mathcal{P})$ the quantity

$$\mathrm{Var}_{\mathcal{P}} = \inf_{d \in D_p} \left[F_{\mathcal{P}}(d) = \int_{D_p} W_p(d,e)^2 \, \mathcal{P}(\mathrm{d}e) < \infty \right],$$

is the Fréchet variance of $\mathcal{P}$ and the set at which the value is obtained

$$\mathbb{E}_{\mathcal{P}} = \{ d \mid F_{\mathcal{P}}(d) = \mathrm{Var}_{\mathcal{P}} \},$$

is the Fréchet expectation.

This set may in general be empty. We now show that under regularity conditions on $\mathcal{P}$ the Fréchet expectation exists. First, we consider probability measures with compact support.

Theorem 9. *Let $\mathcal{P}$ be a probability measure on $(D_p, \mathcal{B}(D_p))$ with a finite second moment. If $\mathcal{P}$ has compact support then $\mathbb{E}_{\mathcal{P}} \neq \emptyset$.*

Proof. Let $S \subset D_p$ be the support of $\mathcal{P}$ and let $\{d_n\}_{n=1}^{\infty}$ be a sequence such that $F_{\mathcal{P}}(d_n) \to \mathrm{Var}_{\mathcal{P}}$. We shall show that d_n has a convergent subsequence. Suppose the contrary. Then d_n cannot belong to a compact set. By Theorem 8, there are three cases to consider. Each of them, as we shall see, leads to a contradiction.

First, assume that $\{d_n\}$ is not bounded. Then $w_n = W_p(d_n, S) = \inf_{d \in S} W_p(d_n, d)$ is not bounded. Thus, as $n \to \infty$ we get

$$F_{\mathcal{P}}(d_n) = \int_S W_p^2(d_n, d) \mathrm{d}\mathcal{P}(d) \geq w_n^2 \mathcal{P}(S) \to \infty,$$

which is a contradiction.

Now assume that $\{d_n\}$ is not off-diagonally birth-death bounded. Then there exists $0 < \varepsilon < 1$ such that for any $C > 0$ and $N > 0$ there is $n > N$ and $x \in d_n$ satisfying $\mathrm{pers}(x) \geq \varepsilon$ and $\mathrm{bd}(x) \geq C$. Take $0 < \delta < \varepsilon/4$ and choose $C_0, C_1 > 1$ such that for all $d \in S$ we have $W_p(d, d_\emptyset) \leq C_1$ and $\mathrm{bd}(x) \leq C_0$ for $x \in u_\delta(d)$. Also, let $C_2 > 1$ be such that $W_p(d_n, d_\emptyset) \leq C_2$, $n \geq 1$, and set $C_3 = C_0 + C_1 + C_2 + 1$. Let d_{n_k} be a subsequence of d_n such that each d_{n_k} contains a point x with $\mathrm{pers}(x) \geq \varepsilon$ and $\mathrm{bd}(x) \geq C_3$, and let $\bar{d}_{n_k}$ be the subdiagram of d_{n_k} obtained by removing all such points x. Take $d \in S$ and let $\gamma : d_{n_k} \to d$ be a bijection such that

$$\sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p \leq W_p(d_{n_k}, d)^p + \delta^p.$$

Notice that $W_p(d_{n_k}, d) \leq W_p(d_{n_k}, d_\emptyset) + W_p(d, d_\emptyset) \leq C_1 + C_2$, so $\mathrm{bd}(\gamma(x)) > C_0$ for all $x \in d_{n_k}$ with $\mathrm{bd}(x) \geq C_3$. Let $\bar{\gamma} : \bar{d}_{n_k} \to d$ be the bijection obtained from γ by pairing points $\gamma(x)$ such that $\mathrm{pers}(x) \geq \varepsilon$ and $\mathrm{bd}(x) \geq C_3$ to the diagonal. Then we have

$$\begin{aligned}
\sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p &= \sum_{x \in \bar{d}_{n_k}} \|x - \gamma(x)\|_\infty^p + \sum_{x \in d_{n_k} - \bar{d}_{n_k}} \|x - \gamma(x)\|_\infty^p \\
&\geq \sum_{x \in \bar{d}_{n_k}} \|x - \gamma(x)\|_\infty^p + \sum_{x \in d_{n_k} - \bar{d}_{n_k}} \left(\frac{\mathrm{pers}(x) - \delta}{2} \right)^p \\
&\geq \sum_{x \in \bar{d}_{n_k}} \|x - \gamma(x)\|_\infty^p + \sum_{x \in d_{n_k} - \bar{d}_{n_k}} \left(\frac{\mathrm{pers}(\gamma(x))}{2} + \delta \right)^p \\
&\geq \sum_{x \in \bar{d}_{n_k}} \|x - \bar{\gamma}(x)\|_\infty^p + \delta^p.
\end{aligned}$$

Using the inequalities

$$(x + y)^\alpha \geq x^\alpha + y^\alpha,$$

where $x, y \geq 0$, $\alpha \geq 1$, and

$$(x + y)^\beta \geq x^\beta + (2^\beta - 1)c^{\beta - 1} y,$$

where $x, y \in [0, c]$, $\beta \in (0, 1)$, we obtain

$$\left(\sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p \right)^{\frac{2}{p}} \geq \left(\sum_{x \in \bar{d}_{n_k}} \|x - \bar{\gamma}(x)\|_\infty^p \right)^{\frac{2}{p}} + \varepsilon_0,$$

where

$$\varepsilon_0 = (2^{\frac{2}{s}} - 1)(C_1 + C_2)^{2 - s} \delta^s, \quad s = \max\{2, p\}.$$

Taking the infima we obtain

$$W_p(d_{n_k}, d)^2 \geq W_p(\bar{d}_{n_k}, d)^2 + \varepsilon_0.$$

Therefore

$$\int_S W_p(\bar{d}_{n_k}, d)^2 \mathrm{d}\mathcal{P}(d) \leq \int_S W_p(d_{n_k}, d)^2 \mathrm{d}\mathcal{P}(d) - \varepsilon_0 \mathcal{P}(S)$$

Taking the infimum over k we obtain $\mathrm{Var}_\mathcal{P} \leq \mathrm{Var}_\mathcal{P} - \varepsilon_0 \mathcal{P}(S)$ which is a contradiction.

Finally, assume that $\{d_n\}$ is not uniform. Let $\varepsilon > 0$ be such that for any $\alpha > 0$ and $N > 0$ there is $n > N$ such that $W_p(l_\alpha(d_n), d_\emptyset) \geq \varepsilon$. Take $0 < \delta < \varepsilon/4$ and choose

α_0 such that $W_p(l_{\alpha_0}(d), d_\emptyset) \leq \delta$ for all $d \in S$. Take $M \geq 1$ and $C > \delta$ such that for all $d \in S$ we have $|u_{\alpha_0}(d)| \leq M$ and $\text{pers}(x) \leq C$ for $x \in d$. Define $f : [0,1] \to [0,1]$ as $f(x) = 1 - (1-x)^p$. Notice that f is a continuous, monotonically increasing function and $f(0) = 0$, $f(1) = 1$. Set $\delta_0 = f^{-1}\left(M^{-1}C^{-p}\delta^p\right)$, and $\alpha_1 = \min\left\{\delta_0\alpha_0, M^{-1/p}\delta\right\}$. Let d_{n_k} be a subsequence of d_n such that $W_p(l_{\alpha_1}(d_{n_k}), d_\emptyset) \geq \varepsilon$, $k \geq 1$, and let $\bar{d}_{n_k} = u_{\alpha_1}(d_{n_k})$. Take $d \in S$ and let $\gamma : d_{n_k} \to d$ be a bijection such that

$$\sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p \leq W_p(d_{n_k}, d)^p + \delta^p.$$

Let $\bar{\gamma} : \bar{d}_{n_k} \to d$ be the bijection obtained from γ by pairing points in $\gamma(l_{\alpha_1}(d_{n_k}))$ to the diagonal. For convenience, let $s_0 = \{x \in d_{n_k} \mid \text{pers}(x) \geq \alpha_1\}$, $s_1 = \{x \in d_{n_k} \mid \text{pers}(x) < \alpha_1, \text{pers}(\gamma(x)) < \alpha_0\}$, $s_2 = \{x \in d_{n_k} \mid \text{pers}(x) < \alpha_1, \text{pers}(\gamma(x)) \geq \alpha_0\}$. Notice that

$$\sum_{x \in s_2} \left(\frac{\text{pers}(x)}{2}\right)^p \leq M\frac{\alpha_1^p}{2^p} \leq \frac{\delta^p}{2^p}.$$

Therefore

$$\sum_{x \in s_1} \left(\frac{\text{pers}(x)}{2}\right)^p \geq \varepsilon^p - \frac{\delta^p}{2^p}.$$

Consequently,

$$\begin{aligned} W_p(s_1, d_\emptyset) - W_p(\gamma(s_1), d_\emptyset) &= \left(\sum_{x \in s_1} \left(\frac{\text{pers}(x)}{2}\right)^p\right)^{\frac{1}{p}} - \left(\sum_{x \in s_1} \left(\frac{\text{pers}(\gamma(x))}{2}\right)^p\right)^{\frac{1}{p}} \\ &\geq \varepsilon\left(1 - \left(\frac{\delta}{2\varepsilon}\right)^p\right)^{\frac{1}{p}} - \delta \geq 2.5\delta, \end{aligned}$$

and thus

$$\left(\sum_{x \in s_1} \|x - \gamma(x)\|_\infty^p\right)^{\frac{1}{p}} \geq W_p(s_1, \gamma(s_1)) \geq W_p(s_1, d_\emptyset) - W_p(\gamma(s_1), d_\emptyset) \geq 2.5\delta.$$

Also

$$\sum_{x \in s_2} \|x - \gamma(x)\|_\infty^p \geq 2^{-p} \sum_{x \in s_2} (\text{pers}(\gamma(x)) - \alpha_1)^p,$$

and

$$\begin{aligned} \sum_{x \in s_2} (\text{pers}(\gamma(x)) - \alpha_1)^p &= \sum_{x \in s_2} \left(\text{pers}(\gamma(x))^p - \text{pers}(\gamma(x))^p f\left(\frac{\alpha_1}{\text{pers}(\gamma(x))}\right)\right) \\ &\geq \sum_{x \in s_2} \left(\text{pers}(\gamma(x))^p - C^p f\left(\frac{\alpha_1}{\alpha_0}\right)\right) \\ &\geq \sum_{x \in s_2} \text{pers}(\gamma(x))^p - \delta^p. \end{aligned}$$

We then have

$$\begin{aligned}\sum_{x\in d_{n_k}} \|x-\gamma(x)\|_\infty^p &= \sum_{x\in s_0}\|x-\gamma(x)\|_\infty^p + \sum_{x\in s_1}\|x-\gamma(x)\|_\infty^p + \sum_{x\in s_2}\|x-\gamma(x)\|_\infty^p \\ &\geq \sum_{x\in s_0}\|x-\bar\gamma(x)\|_\infty^p + 2^{-p}\sum_{x\in s_2}\mathrm{pers}(\gamma(x))^p + (2.5)^p\delta^p - 2^{-p}\delta^p \\ &\geq \sum_{x\in s_0}\|x-\bar\gamma(x)\|_\infty^p + 2^{-p}\sum_{x\in s_1}\mathrm{pers}(\gamma(x))^p + \\ &\quad 2^{-p}\sum_{x\in s_2}\mathrm{pers}(\gamma(x))^p + ((2.5)^p - 1 - 2^{-p})\delta^p \\ &\geq \sum_{x\in \bar d_{n_k}} \|x-\bar\gamma(x)\|_\infty^p + \delta^p.\end{aligned}$$

As in the previous case, this implies that

$$\left(\sum_{x\in d_{n_k}} \|x-\gamma(x)\|_\infty^p\right)^{\frac{2}{p}} \geq \left(\sum_{x\in \bar d_{n_k}} \|x-\bar\gamma(x)\|_\infty^p\right)^{\frac{2}{p}} + \varepsilon_0,$$

where

$$\varepsilon_0 = (2^{\frac{2}{s}}-1)(C_1+C_2)^{2-s}\delta^s, \quad s=\max\{2,p\}.$$

and C_1 and C_2 are chosen exactly as in the previous case. Therefore

$$W_p(d_{n_k},d)^2 \geq W_p(\bar d_{n_k},d)^2 + \varepsilon_0,$$

and consequently

$$\int_S W_p(\bar d_{n_k},d)^2 \mathrm{d}\mathcal{P}(d) \leq \int_S W_p(d_{n_k},d)^2\mathrm{d}\mathcal{P}(d) - \varepsilon_0\mathcal{P}(S),$$

implying $\mathrm{Var}_{\mathcal{P}} \leq \mathrm{Var}_{\mathcal{P}} - \varepsilon_0\mathcal{P}(S)$ leading to a contradiction. □

Requiring compactness of the support of $\mathcal{P}$ may be too restrictive. A less stringent condition is that the distribution has a particular tail decay for which we need the following definitions.

Definition 9. Let $\mathbb{X}$ be a Hausdorff topological space, and let Σ be a σ-algebra on $\mathbb{X}$ that contains the topology of $\mathbb{X}$. A measure μ on the measurable space $(\mathbb{X},\Sigma)$ is called *inner regular*, or *tight*, if $\forall\varepsilon>0$ there exists a compact set $S\subset\mathbb{X}$ such that $\mu(\mathbb{X}-S)<\varepsilon$.

Definition 10. Let $(\mathbb{X},\rho)$ be a metric space, and let Σ be a σ-algebra on $\mathbb{X}$ that contains the topology of $\mathbb{X}$. A measure μ on the measurable space $(\mathbb{X},\Sigma)$ has *rate of decay at infinity* q if for some (hence for all) $x_0\in\mathbb{X}$ there exist $C>0$ and $R>0$ such that for all $r\geq R$ we have $\mu(B^r(x_0))\leq Cr^{-q}$, where $B^r(x_0)=\{x\in\mathbb{X}\mid\rho(x,x_0)\geq r\}$.

We can now prove the following result.

Theorem 10. *Let $\mathcal{P}$ be a tight probability measure on $(D_p, \mathcal{B}(D_p))$ with the rate of decay at infinity $q > \max\{2, p\}$. Then $\mathbb{E}_{\mathcal{P}} \neq \emptyset$.*

Proof. Let $\{d_n\}_{n=1}^{\infty}$ be a sequence such that $F_{\mathcal{P}}(d_n) \to \mathrm{Var}_{\mathcal{P}}$. The same argument as in Theorem 9 shows that d_n has to be bounded. Let $C > 0$ and $R > 0$ be such that $\mathcal{P}(B^r(d_\emptyset)) \leq Cr^{-q}$, $r \geq R$. Let $\varepsilon > 0$ be arbitrarily small and let $s = \max\{2, p\}$. Take $M \in \mathbb{N}$ such that $M > R$, $W_p(d_n, d_\emptyset) \leq M$ and

$$\frac{(M+1)^s}{M^q} < \frac{\varepsilon}{16C} \text{ and } \sum_{m \geq M} \frac{(2m+3)^{s-1}}{(m+1)^q} < \frac{\varepsilon}{16C}.$$

Denote

$$B^{m,m+1}(d_\emptyset) = B^m(d_\emptyset) - B^{m+1}(d_\emptyset)$$

we then obtain

$$\begin{aligned}\int_{B^M(d_\emptyset)} W_p(d_n, d)^2 \mathrm{d}\mathcal{P}(d) &\leq \int_{B^M(d_\emptyset)} \left(W_p(d_n, d_\emptyset) + W_p(d_\emptyset, d)\right)^2 \mathrm{d}\mathcal{P}(d), \\ &\leq \int_{B^M(d_\emptyset)} \left(2W_p(d_\emptyset, d)\right)^2 \mathrm{d}\mathcal{P}(d).\end{aligned}$$

Note that

$$\begin{aligned}\int_{B^M(d_\emptyset)} \left(2W_p(d_\emptyset, d)\right)^2 \mathrm{d}\mathcal{P}(d) &= 4 \sum_{m \geq M} \int_{B^{m,m+1}(d_\emptyset)} W_p(d_\emptyset, d)^2 \mathrm{d}\mathcal{P}(d) \\ &\leq 4 \sum_{m \geq M} (m+1)^2 \left(\mathcal{P}(B^m(d_\emptyset)) - \mathcal{P}(B^{m+1}(d_\emptyset))\right) = L,\end{aligned}$$

and

$$L = 4 \sum_{m \geq M} \left((m+1)^2 \mathcal{P}(B^m(d_\emptyset)) - (m+2)^2 \mu(B^{m+1}(d_\emptyset))\right) + 4 \sum_{m \geq M} (2m+3) \mathcal{P}(B^{m+1}(d_\emptyset)).$$

Finally

$$L \leq 4C \left(\frac{(M+1)^2}{M^q} + \sum_{m \geq M} \frac{2m+3}{(m+1)^q}\right) < \frac{\varepsilon}{2M^{s-2}}.$$

Now let $S \subset D_p$ be a compact set such that $\mathcal{P}(S^c) < M^{-s}\varepsilon/8$, where $S^c = D_p - S$. Letting $B_M(d_\emptyset) = \{d \in D_p \mid W_p(d, d_\emptyset) \leq M\}$ we have

$$\begin{aligned}\int_{S^c \cap B_M(d_\emptyset)} W_p(d_n, d)^2 \mathrm{d}\mathcal{P}(d) &\leq \int_{S^c \cap B_M(d_\emptyset)} \left(W_p(d_n, d_\emptyset) + W_p(d_\emptyset, d)\right)^2 \mathrm{d}\mathcal{P}(d) \\ &\leq 4M^2 \mathcal{P}(S^c \cap B_M(d_\emptyset)) < \frac{\varepsilon}{2M^{s-2}}.\end{aligned}$$

Combining the two results we get

$$\int_{D_p} W_p(d_n,d)^2 \mathrm{d}\mathcal{P}(d) \leq \int_{S\cap B_M(d_\emptyset)} W_p(d_n,d)^2 \mathrm{d}\mathcal{P}(d) + \int_{S^c\cap B_M(d_\emptyset)} W_p(d_n,d)^2 \mathrm{d}\mathcal{P}(d) + \int_{B^M(d_\emptyset)} W_p(d_n,d)^2 \mathrm{d}\mathcal{P}(d) < \int_{S\cap B_M(d_\emptyset)} W_p(d_n,d)^2 \mathrm{d}\mathcal{P}(d) + \frac{\varepsilon}{M^{s-2}}.$$

Let us assume now that d_n does not have a converging subsequence. Since the above inequality holds for any subsequence d_{n_k} of d_n as well as any subsequence of subdiagrams $\bar{d}_{n_k} \subset d_{n_k}$, the proof of Theorem 9 implies that

$$\int_{D_p} W_p(\bar{d}_{n_k},d)^2 \mathrm{d}\mathcal{P}(d) \leq \int_{S\cap B_M(d_\emptyset)} W_p(d_{n_k},d)^2 \mathrm{d}\mathcal{P}(d) - \varepsilon_0 \mathcal{P}(S) + \frac{\varepsilon}{M^{s-2}},$$

where d_{n_k}, $\bar{d}_{n_k}$ are the same as in the proof of Theorem 9 and

$$\varepsilon_0 = \frac{(2^{\frac{2}{s}}-1)2^{2-s}\delta^s}{M^{s-2}}.$$

Taking $\varepsilon < (2^{\frac{2}{s}}-1)2^{1-s}\delta^s\mathcal{P}(S)$ we have

$$\int_{D_p} W_p(\bar{d}_{n_k},d)^2 \mathrm{d}\mathcal{P}(d) \leq \int_{D_p} W_p(d_{n_k},d)^2 \mathrm{d}\mathcal{P}(d) - \frac{\varepsilon_0\mathcal{P}(S)}{2},$$

Taking the infimum over k we obtain

$$\mathrm{Var}_{\mathcal{P}} \leq \mathrm{Var}_{\mathcal{P}} - \frac{\varepsilon_0\mathcal{P}(S)}{2},$$

which results in a contradiction. □

4.2. The measure $\mathcal{P}_D$ and conditional probabilities

The point of the previous section was to prove that for natural restrictions of a distribution of persistence diagrams $\mathcal{P}_D$ the expected diagram and variance over these diagrams are defined. In this section we first show how a measure on the point samples $\mathcal{P}_\theta$ implies measure on persistence diagrams. We then define joint and conditional measures $\mathcal{P}(D,\theta)$ and $\mathcal{P}(\theta \mid D)$, respectively. We later discuss the relevance of these measures in inference.

From the perspective of a probabilist or statistician there is a stochastic process that generates the point cloud data. For example, a family of distributions on the $(p-1)$-dimensional sphere in $\mathbb{R}^p$ can be the von Mises-Fisher distribution which has a parametric form with parameters θ and recovers the uniform distribution for a particular parameter setting. Our point cloud data may be drawn identically and independently from the von Mises-Fisher distribution F_θ

$$X_1,...,X_n \overset{iid}{\sim} F_\theta.$$

This results in a likelihood for the observed point cloud data $Z \equiv \{X_1, ..., X_n\}$

$$\text{Lik}(Z \mid \theta) = f_\theta(Z),$$

where f_θ is the probability density function corresponding to the probability distribution function F_θ.

We start with the premise that the point cloud data is generated from a probability measure so we have a probability space $(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mathcal{P}_\theta)$ where $\mathbb{X}$ is a subset of $\mathbb{R}^d$ (for example a torus), $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{X}$ and $\mathcal{P}_\theta$ is the probability measure parameterized by θ. The observed point cloud data $Z \equiv \{X_1, ..., X_n\}$, where $X_1, ..., X_n \overset{iid}{\sim} \mathcal{P}_\theta$, can be regarded as an element of the probability space $(\mathbb{X}^n, \Sigma^n, \mathcal{P}_\theta^n)$, where $\mathbb{X}^n = \prod_{i=1}^n \mathbb{X}$, and Σ^n and $\mathcal{P}_\theta^n$ denote the σ-algebra and probability measure induced by the product structure. Alternatively, Z can be regarded as a compact subset of $\mathbb{X}$, and we express this formally by defining a map $h_n : \mathbb{X}^n \to K(\mathbb{X})$, $h_n(X_1, \ldots, X_n) = \{X_1, \ldots, X_n\}$, where $K(\mathbb{X})$ denotes the space of compact subsets of $\mathbb{X}$ endowed with the Hausdorff metric. Suppose now that we have a (continuous) map $\rho : K(\mathbb{X}) \to \text{Lip}(\mathbb{X})$, where $\text{Lip}(\mathbb{X})$ denotes the space of Lipschitz functions on $\mathbb{X}$ with the supnorm. For example, we can take $\rho(S) = d_S$, the usual distance function. Another choice is to regard $S \in K(\mathbb{X})$ as a measure (which in the case of the point cloud data will be an empirical probability measure) and map S to the distance function to this measure as defined in [14]. Composing these maps and taking the persistence diagram of the resulting function we thus obtain a map $g : \mathbb{X}^n \to D_p$. The map g is measurable if for every $A \in \mathcal{B}(D_p)$ the inverse image

$$g^{-1}(A) = \{\omega : g(\omega) \in A\} \in \Sigma^n.$$

Assuming that g is measurable we then have the induced measure $\mathcal{P}_D$ on $(D_p, \mathcal{B}(D_p))$ defined by

$$\mathcal{P}_D(A) = \mathcal{P}_\theta(g^{-1}(A)), \quad \text{for } A \in \mathcal{B}(D_p).$$

Notice that if $\mathbb{X}$ triangulable, compact and implies bounded degree-k total persistence, and if ρ maps point cloud data to only tame functions with bounded Lipschitz constants then the Wasserstein stability result from Section 2 shows that g is, in fact, continuous when $p > k$. Since measurability is a much weaker condition that continuity, we expect that the induced probability measure on the space of persistence diagrams can be defined in many more general cases.

The probability measure $\mathcal{P}_D$ constructed above is conditioned on the parameter θ. Suppose that we have a prior distribution of θ given by the measure μ. Then the joint probability measure $\mathcal{P}(D, \theta)$ is given by the product measure

$$\mathcal{P}(D, \theta) = \mathcal{P}_D \times \mu.$$

Bayes' rule also gives us the conditional measure $\mathcal{P}(\theta|D)$:

$$\mathcal{P}(\theta|D) \propto \mathcal{P}_D \times \mu.$$

Thus, we have the basic building blocks for performing statistical inference on topological summaries such as persistence diagrams. An interesting subtle point about the above conditional probability is that it is not strictly Bayesian since we substitute the likelihood $\mathcal{P}_\theta$ with the probability of the topological summary $\mathcal{P}_D$ – this violates the likelihood principle [16]. This idea of a substitution likelihood goes back to Jeffreys [17] and a basic question in TDA is what properties of $\mathcal{P}_\theta$ are preserved by $\mathcal{P}_D$.

5. Discussion

We have shown that persistence diagrams form a space on which basic statistical objects such as means, variances, and conditional probabilities are well defined. This result is crucial for our ability to perform statistical inference on persistence diagrams and provides a foundation for further integration of TDA methods into the standard statistical framework. For example, we can consider homological estimators based on the Fréchet mean of persistence diagrams, and we might be able to quantify the uncertainty of such an estimator using the Fréchet variance.

Existence of conditional probabilities on persistence diagrams provides a basis for topology based parameter estimators. For example, consider a stochastic dynamical systems depending on a parameter θ. Suppose we can obtain samples from the attractors of this system. Then we can try to estimate the distribution of θ using persistence diagrams of these samples.

We would like to emphasize that our result does not depend on a particular procedure used to compute persistence diagrams. Hence, we are free to choose the best application dependent procedure as long as the resulting map from the sample space to the space of persistence diagrams is measurable (see Section 4.2 for details).

While our result shows a theoretical possibility of performing rigorous statistical inference on persistence diagrams there remain several issues to address. For example, the Fréchet expectation is not unique due to peculiarities of the Wasserstein distance, which complicates standard statistical procedures. Also, we do not yet have an algorithm for computing the Fréchet mean of persistence diagrams. An algorithm for variance decomposition for persistence diagrams was developed in [18] using the Wasserstein distance metric and multidimensional scaling. The framework in this paper may provide a theoretical basis for this procedure. It is also important to better understand the conditions required for measurability of the map from the sample space to the space of persistence diagrams.

References

[1] Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2010.

[2] V. de Silva and G. Carlsson. Topological estimation using witness complexes. *Symposium on Point-Based Graphics*, pages 157–166, 2004.

[3] Herbert Edelsbrunner and John Harer. Persistent homology - a survey. *Contemporary Mathematics*, 453:257–282, 2008.

[4] Richard M. Dudley. *Real analysis and probability*. Chapman & Hall, New York, NY, 1989.

[5] Mathew D. Penrose. *Random Geometric Graphs*. Oxford Univ. Press, New York, NY, 2003.

[6] Mathew D. Penrose and Joseph E. Yukich. Central limit theorems for some graphs in computational geometry. *Ann. Appl. Probab.*, 11(4):1005–1041, 2001.

[7] Matthew Kahle. Random geometric complexes. http://arxiv.org/abs/0910.1649, 2011.

[8] Matthew Kahle. Topology of random clique complexes. *Discrete Math.*, 309(6):1658–1671, 2009.

[9] Simón Lunagómez, Sayan Mukherjee, and Robert Wolpert. Geometric representations of hypergraphs for prior specification and posterior sampling. http://arxiv.org/abs/0912.3648, 2009.

[10] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Computational Geometry*, 39:419–441, 2008.

[11] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. A topological view of unsupervised a topological view of unsupervised learning from noisy data. Manuscript, 2008.

[12] Frédéric Chazal, David Cohen-Steiner, and André Lieutier. A sampling theory for compact sets in euclidean space. *Discrete and Computational Geometry*, 41:461–479, 2009.

[13] Paul Bendich, Sayan Mukherjee, and Bei Wang. Towards stratification learning through homology inference. http://arxiv.org/abs/1008.3572, 2010.

[14] Frédéric Chazal, David Cohen-Steiner, and Quentin M'erigot. Geometric inference for measures based on distance functions. http://hal.inria.fr/inria-00383685/, 2010.

[15] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l_p-stable persistence. *Foundations of Computational Mathematics*, 10:127–139, 2010. 10.1007/s10208-010-9060-6.

[16] James O. Berger and Robert L. Wolpert. *The likelihood principle*. Institute of Mathematical Statistics, 1984.

[17] Harold Jeffreys. *Theory of Probability*. Clarendon Press, 1961.

[18] Jennifer Gamble and Giseon Heo. Exploring uses of persistent homology for statistical analysis of landmark-based shape data. *Journal of Multivariate Analysis*, pages 2184–2199, 2010.