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: yury@math.duke.edu, sayan@stat.duke.edu, john.harer@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, 10] and non-asymptotic bounds on the convergence or consistency of topological summaries as the number of points increase [11, 12, 13, 14, 15]. 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 [16] 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 $X$ and a bounded continuous function $f : X \to \mathbb{R}$. Let $X_a = f^{-1}(-\infty, a]$ denote the sublevel set of $f$ at the threshold $a$. Inclusions $X_a \subset X_b$, $a \leq b$, induce homomorphisms of the homology groups of sublevel sets:

$$f_{a,b}^\ell : H_\ell(X_a) \to H_\ell(X_b),$$

for each dimension $\ell$. We call a value $c \in \mathbb{R}$ a homological critical value of $f$ if there exists $\ell$ such that $f_{c}^{\ell-\delta,c}$ is not an isomorphism for any $\delta > 0$. We call $f$ tame if has only a finite number of homological critical values and if $H_\ell(X_a)$ are finitely generated for all $a \in \mathbb{R}$ and all dimensions $\ell$. 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 $\text{Im} f_{c}^{\ell-\delta,c} \subset H_\ell(X_b)$ is independent of $\delta > 0$ if $\delta$ is sufficiently small. We shall denote such an image by $F_{c}^{\ell-\delta}$. Now, consider the following quotient group:

$$B_a^{\ell} = H_\ell(X_a)/F_{a}^{\ell-\delta}.$$

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

Let us now consider the map

$$g_{a,b}^{\ell} : B_a^{\ell} \to H_\ell(X_b)/F_{a}^{\ell-\delta}.$$
defined as $g_{\ell}^{a,b}(\alpha) = f_{\ell}^{a,b}(\alpha)$, where $\alpha \in H_\ell(X_a)$, and the square brackets denote the images under the corresponding canonical projections. We set $g_{\ell}^{a,b} = 0$ if $b = \sup_{x \in X} f(x)$, so that each homology class has finite persistence (as defined in Section 2.3). The kernel of this map, which we denote by $D_{\ell}^{a,b}$, captures homology classes that were born at $X_a$ but at $X_b$ are homologous to homology classes born before $X_a$. We call $D_{\ell}^{a,b}$ the death subgroup of $B_{\ell}^a$ at $X_b$, and we say that a homology class $\alpha \in H_\ell(X_a)$ dies entering $X_b$ if $[\alpha] \in D_{\ell}^{a,b}$ but $[\alpha] \notin D_{\ell}^{a,b-\delta}$ for any $\delta > 0$. We also call $b$ a degree-$r$ death value of $B_{\ell}^a$ if $\text{rank} D_{\ell}^{a,b} - \text{rank} D_{\ell}^{a,b-\delta} = r > 0$ 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 $\alpha$ is born at $X_a$ and dies entering $X_b$ we set $b(\alpha) = a$, $d(\alpha) = b$. The persistence of $\alpha$ is the difference between the two values, $\text{pers}(\alpha) = d(\alpha) - b(\alpha)$. We represent the births and deaths of $\ell$-dimensional homology classes by a multiset of points in $\mathbb{R}^2$, the $\ell$-th persistence diagram denoted by $Dgm_{\ell}(f)$. For each nontrivial birth group $B_{\ell}^a$ the diagram contains points $x_i = (a, b_i)$, where $b_i$ are the death values of $B_{\ell}^a$, and the multiplicity of $x_i$ is equal to the degree of the corresponding death value $b_i$. 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 $x \in Dgm_{\ell}(f)$, denoted by $\text{pers}(x)$, 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, which are defined in the previous section as finite multisets of points in a plane:

**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 $\gamma$ 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 $\infty$ 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 extend the definition of a persistence diagram as follows.

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

The $p$-th Wasserstein distance applies naturally to generalized persistence diagrams. We shall omit the word “generalized” for the rest of the paper, as these are the only diagrams that we shall consider.

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

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

and the infimum is attained at $y = \left(\frac{x_1+x_2}{2}, \frac{x_1+x_2}{2}\right)$, where $x = (x_1, x_2)$. Therefore,

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

Recall from [17] the following definition:

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

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

Thus, $\text{Pers}_p(d) = 2^p(W_p(d, d_\emptyset))^p$, 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

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

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

Let us point out that our definition of the $p$-th Wasserstein distance is a modification of the classical concept from probability theory which has applications in the theory of optimal transportation [18, 19, 20] as well as in computer vision [21]
and image retrieval [22]. Given probability measures \( \mu, \nu \) with finite \( p \)-th moments on a metric space \( (X, \rho) \), the \( p \)-th Wasserstein distance between \( \mu \) and \( \nu \) is defined as follows:

\[
W_p(\mu, \nu) = \left( \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times X} \rho^p(x, y) \, d\gamma(x, y) \right)^{\frac{1}{p}},
\]

where \( \Gamma(\mu, \nu) \) is a collection of probability measures on \( X \times X \) whose marginals on the first and second factors are \( \mu \) and \( \nu \), respectively. Requiring finiteness of the \( p \)-th moment of a probability measure \( \mu \) is similar to requiring finiteness of the degree-\( p \) total persistence of a diagram \( d \), and means that for some \( x_0 \in X \) we have

\[
\int_X \rho^p(x_0, x) \, d\mu(x) < \infty.
\]

The crucial difference between the Wasserstein distance for persistence diagrams and the Wasserstein distance for probability measures is due to the unique role of the diagonal in the former case. The result on completeness and separability of \( D_p \) proved in Section 3.1 is analogous to the classical result for the space of probability measure with finite \( p \)-th moment endowed with the Wasserstein distance [23, 24]. We have not considered the case \( p = \infty \), when the Wasserstein distance between persistence diagrams becomes the bottleneck distance, but we suspect that our results will still hold.

We finish this section by stating an important stability result from [17] which shows that under mild assumptions on \( X \) computing a persistence diagram of a tame Lipschitz functions is a continuous map. Suppose that \( X \) is a metric space such that for any persistence diagram \( d \) computed for a Lipschitz function \( f \) with the Lipschitz constant \( \text{Lip}(f) \leq 1 \) we have \( \text{Pers}_p(d) \leq C_X \), where \( C_X \) is a constant that depends only on \( X \). We shall say in this case that \( X \) implies bounded degree-\( k \) total persistence.

**Proposition 5** (Wasserstein Stability). If \( X \) is a triangulable, compact metric space that implies bounded degree-\( k \) total persistence for some \( k \geq 1 \) and \( f_1, f_2 : X \to \mathbb{R} \) are tame, Lipschitz functions, then for all dimensions \( \ell \) and \( p \geq k \) we have

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

where \( C = C_X \max\{\text{Lip}(f_1)^k, \text{Lip}(f_2)^k\} \).

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 6. $D_p$ 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 $d^\ast$. Second, we show that $d^\ast$ belongs to $D_p$. Third, we show that $d_n$ converges to $d^\ast$ 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 cardinal number of (off diagonal) points in $d$ counting multiplicities. For $\alpha > 0$ let $u_\alpha : D_p \rightarrow D_p$ be defined by

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

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

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

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

We have introduced the upper and lower parts of persistence diagrams in order to define an analogue of pointwise convergence. Since the $\alpha$-upper part of a diagram has finite total multiplicity for any $\alpha > 0$, it is reasonable to consider convergence of the $\alpha$-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 $\alpha$ 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 $\alpha$ (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.
Lemma 9 (Persistence-wise Separation). Let $\alpha > 0$. Then there exist $M_\alpha \in \mathbb{Z}_+$ and $\delta_\alpha$, $0 < \delta_\alpha < \alpha$, such that for any path $\delta$ 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 $\delta$ with $0 < \delta < \alpha$ let $M^\delta_{\text{sup}} = \limsup_{n \rightarrow \infty} |u_\delta(d_n)|$, $M^\delta_{\text{inf}} = \liminf_{n \rightarrow \infty} |u_\delta(d_n)|$. Notice that $M^\delta_{\text{sup}} < \infty$, otherwise, we could find a subsequence $d_{n_k}$ such that $|u_\delta(d_{n_k})| > k$ so that $W_p(d_{n_k}, d_\emptyset) \geq k^1/\delta/2 \rightarrow \infty$ as $k \rightarrow \infty$. However, $W_p(d_n, d_\emptyset)$ is bounded because $d_n$ is Cauchy.

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

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

Given $\alpha > 0$, let $d^\alpha_n = u_{\delta_\alpha}(d_n), d^\alpha_n$ contain points whose persistence (in the limit) is at least $\alpha$. We also denote $d_{n, \alpha} = l_{\delta_\alpha}(d_n)$.

Lemma 8. For any $\alpha > 0$ the sequence $d^\alpha_n$ is Cauchy.

Proof. Use Lemma 7 to choose $\delta_\alpha$. Let $\delta \in (\delta_\alpha, \alpha)$. Then by Lemma 7 $\exists N > 0$ 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 $\gamma : d_n \rightarrow d_m$ such that

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

This inequality implies that $\gamma$ maps points in $d^\alpha_n$ to points in $d^\alpha_m$, therefore,

$$W_p(d^\alpha_n, d^\alpha_m) \leq \left( \sum_{x \in d^\alpha_n} \|x - \gamma(x)\|_p^p \right)^{1/p} < 2\varepsilon_0 \leq \varepsilon.$$
Proof. Let $\alpha > 0$, $\delta \in (\delta_\alpha, \alpha)$, and let $N > 0$ be such that $|d^\alpha_n| = |u_\delta(d_n)| = M_\alpha$ for all $n > N$. Let $\varepsilon$ be such that $0 < \varepsilon < \delta/2$. Choose a subsequence $d^\alpha_{n_k}$, $k \in \mathbb{N}$, such that $n_1 > N$ and $W_p(d^\alpha_{n_k}, d^\alpha_m) < 2^{-k}\varepsilon$ for $m \geq n_k$. Let $\gamma_k : d^\alpha_{n_k} \to d^\alpha_{n_{k+1}}$ be a bijection realizing the Wasserstein distance $W_p(d^\alpha_{n_k}, d^\alpha_{n_{k+1}})$. Notice that our choice of $\varepsilon$ guarantees that each $\gamma_k$ maps off diagonal points to off diagonal points. Let $x^1, \ldots, x^{M_\alpha}$ be off diagonal elements of $d^\alpha_{n_1}$. We can now construct $M_\alpha$ sequences of points $\{x^1_k\}, \ldots, \{x^{M_\alpha}_k\}$, $k \in \mathbb{N}$, such that $x^i_1 = x^i$, $i = 1, \ldots, M_\alpha$, and $x^i_{k+1} = \gamma_k(x^i_k)$. Notice that each sequence $x^i_k$ is Cauchy. Indeed, $W_p(d^\alpha_{n_k}, d^\alpha_{n_{k+1}}) < 2^{-k}\varepsilon$ implies $\|x^i_k - x^i_l\|_\infty < 2^{-k}\varepsilon$ for all $l > k$, $i = 1, \ldots, M_\alpha$. Taking limits we obtain a collection of points $\hat{x}^1, \ldots, \hat{x}^{M_\alpha}$. Let $d^\alpha$ be the diagram whose off diagonal elements are $\hat{x}^1, \ldots, \hat{x}^{M_\alpha}$ (notice that the multiplicity of a point $x \in d^\alpha$ is equal to the number of sequences whose limit is $x$). We now show that $d^\alpha$ is the limit of $d^\alpha_n$ and hence is unique, which also implies that the collection of limits $\hat{x}^1, \ldots, \hat{x}^{M_\alpha}$ does not depend on the choice of bijections $\gamma_k$, subsequence $d^\alpha_{n_k}$, or $\varepsilon$.

Let $\varepsilon_0 > 0$ and pick $K > 0$ such that $\forall k > K$ we have $\|x^i_k - \hat{x}^i\| < 0.5\varepsilon_0M_\alpha^{-1/p}$ and $W_p(d^\alpha_{n_k}, d_m) < \varepsilon_0/2$, $m \geq n_k$. Then we also have $W_p(d^\alpha_{n_k}, d^\alpha) \leq W_p(d^\alpha_{n_k}, d^\alpha_m) + W_p(d^\alpha_m, d^\alpha) < \varepsilon_0/2 + \varepsilon_0/2 = \varepsilon_0$.

The last statement of the lemma follows from the fact that if $\alpha_1 > \alpha_2$, then points $x \in d^\alpha_{n_2}$ such that $x \notin d^\alpha_{n_1}$ have pers$(x) < \delta_{\alpha_1} < \alpha_1$. Indeed, repeating the above argument with $\alpha = \alpha_2$, $N > 0$ such that $|d^\alpha_n| = |u_\delta(d_n)| = M_{\alpha_2}$ and $|d^\alpha_n| = |u_{\delta_\alpha}(d_n)| = M_{\alpha_2}$, for all $n > N$, where $\delta_\alpha \in (\delta_{\alpha_1}, \alpha_1)$, $\delta_2 \in (\delta_{\alpha_2}, \alpha_2)$, and $\varepsilon > 0$ such that $\varepsilon < \min\{\delta_2/2, (\delta_\alpha - \delta_{\alpha_1})/2\}$, we see that each $\gamma_k : d^\alpha_{n_k} \to d^\alpha_{n_{k+1}}$ maps off diagonal points in $d^\alpha_{n_k}$ to off diagonal points in $d^\alpha_{n_{k+1}}$. Therefore, the collection of limits $\hat{x}^1, \ldots, \hat{x}^{M_\alpha}$ contains the limits that we obtain for the case $\alpha = \alpha_1$. □

Lemma 9 allows us to define $d^* = \cup_{\alpha > 0} d^\alpha$. It is not difficult to show that that $d^* \in D_p$.

**Lemma 10.** $d^* \in D_p$. Furthermore $\lim_{\alpha \to 0} W_p(d^\alpha, d^*) = 0$.

**Proof.** First note that since $d_n$ is Cauchy, there is a constant $C > 0$ such that $\forall n, W_p(d_n, d_0) \leq C$. Let $\alpha > 0$, and let $N > 0$ be such that $\forall n > N$, $W_p(d^\alpha, d^\alpha_n) < 1$. Take any such $n$, then $W_p(d^\alpha, d_0) \leq W_p(d^\alpha, d^\alpha_n) + W_p(d^\alpha_n, d_0) \leq 1 + C$.

Since the right hand side is independent of $\alpha$, we obtain $W_p(d^*, d_0) \leq 1 + C$.

Finally, notice that

$$W_p(d^*, d^*)^p \leq W_p(l_\alpha(d^*), d_0)^p = \sum_{x \in d^*, \text{pers}(x) < \alpha} \left(\frac{\text{pers}(x)}{2}\right)^p \to 0 \text{ as } \alpha \to 0.$$ □

By the triangle inequality $W_p(d^*, d_n) \leq W_p(d^*, d^\alpha) + W_p(d^\alpha, d^\alpha_n) + W_p(d^\alpha_n, d_n)$. The completeness of $D_p$ follows from Lemmas 10, 9 and 11.
Lemma 11. \( \forall \varepsilon > 0, \exists \alpha_0 > 0 \) 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 \( W_p(d_n, d_n) < \varepsilon \).

Proof. We prove the lemma by contradiction. Suppose that \( \exists \varepsilon > 0 \) such that \( \forall \alpha > 0 \exists n_\alpha \in \mathbb{N} \) with \( W_p(d_{n_\alpha,\alpha}, d_\emptyset) \geq \varepsilon \). Take such an \( \varepsilon \). Let \( \{\alpha_i\}_{i \in \mathbb{N}} \) be a sequence of positive values monotonically decreasing to 0. Since \( \alpha_i \to 0, n_{\alpha_i} \to \infty \). 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. For \( i \geq j \), 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
\[
\sum_{x \in d_{n_i,\alpha_i}} \|x - \gamma_i(x)\|_\infty^p = \sum_{x \in d_{n_i,\alpha_i}} \|x - \gamma_i(x)\|_\infty^p + \sum_{x \in d_{n_i,\alpha_i}} \|x - \gamma_i(x)\|_\infty^p < 2\delta^p.
\]
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
\[
\|x - \gamma_i(x)\|_\infty \geq \frac{|\text{pers}(x) - \text{pers}(\gamma_i(x))|}{2} \geq \frac{\delta_{\alpha_j} - \alpha_i}{2} \geq \frac{\alpha_i}{2} \geq \frac{\text{pers}(x)}{2}
\]
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 \( \gamma_i(x) \notin d_{n_k,\alpha_j} \), 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
\[
\sum_{x \in d_{n_i,\alpha_i}} \|x - \hat{\gamma}_i(x)\|_\infty^p = \sum_{x \in d_{n_i,\alpha_i}} \|x - \gamma_i(x)\|_\infty^p + \sum_{x \in d_{n_i,\alpha_i}} \left(\frac{\text{pers}(x)}{2}\right)^p + 
\]
\[
\sum_{x \in d_{n_k,\alpha_j}} \left(\frac{\text{pers}(y)}{2}\right)^p \leq \sum_{x \in d_{n_i,\alpha_i}} \|x - \gamma_i(x)\|_\infty^p + \sum_{x \in d_{n_k,\alpha_j}} \|x - \gamma_i(x)\|_\infty^p + \delta^p
\]
\[
< 2\delta^p + \delta^p = 3\delta^p.
\]
Therefore, \( W_p(d_{n_i,\alpha_i}, d_{n_k,\alpha_j}) < 3\delta \) if \( i \geq l \). Contradiction. \( \square \)

We finish this section by proving separability of \( D_p \).

Theorem 12. \( D_p \) 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,

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

If $d \in D_p$ then $\forall \varepsilon > 0$ we can find $\alpha > 0$ such that $W_p(l_\alpha(d), d_0) < \varepsilon/2$. Then we have $W_p(d, u_\alpha(d)) \leq W_p(l_\alpha(d), d_0) < \varepsilon/2$. Since $\mathbb{Q}^{2|\alpha|}$ is dense in $\mathbb{R}^{2|\alpha|}$, we can find $d_s \in S$ 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 $S = \bigcup_{m=0}^\infty S_m$, where $S_m = \{d \in S \mid |d| = m\}$. 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. We will characterize relatively compact subsets of persistence diagrams. This will require mild conditions which we specify in this subsection.

Definition 13 (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 $\varepsilon$ whose union contains $S$.

Definition 14 (Relative compactness). A subset of a topological space is called relatively compact if its closure is compact.

Proposition 15. In a complete metric space $\mathbb{X}$ a subset $S$ is totally bounded iff it is relatively compact iff every sequence in $S$ has a subsequence convergent in $\mathbb{X}$.

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

Example 16. Consider $S \subset D_p$ consisting of diagrams with a single off diagonal point of multiplicity 1 and 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 $W_p(d_n, d_m) = ((\varepsilon/2)^p + (\varepsilon/2)^p)^{1/p} = 2^{1/p-1}\varepsilon$ for all $n \neq m$. Hence, $d_n$ does not have a convergent subsequence. Thus this set is not relatively compact.

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

Definition 17 (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 $\forall x \in d$ max\{$|b(x)|, |d(x)|\} \leq C$.

We denote $bd(x) = \max\{|b(x)|, |d(x)|\}$.

Definition 18 (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.

These two conditions are not enough to ensure relative compactness as is shown in the following example.
Example 19. Let $\varepsilon > 0$ and $C > \varepsilon$. Consider the set $S = \{ d \mid W_p(d, d_0) \leq \varepsilon \} \cap \{ d \mid b(x) \geq 0 \& d(x) \leq C \ \forall x \in d \}$. For $n \in \mathbb{N}$, let $d_n \in S$ be the diagram consisting of a single off diagonal point $x_n = (0, 2^{1-n/p}\varepsilon)$ with multiplicity $2^n$ (see Figure 2(b)). It is easy to see that for all $n, m \in \mathbb{N}$, $m > n$, we have $W_p(d_n, d_m) \geq \left(2^{n-1}\left(\frac{1}{2}b^{1-m/p}\varepsilon\right)^p\right)^{1/p} = 2^{-1/p}\varepsilon$, as there will be at least $2^{n-1}$ (counting multiplicities) 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 relatively compact.

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

Definition 20. A set $S \subset D_p$ is called uniform if for all $\varepsilon > 0$ there exists $\alpha > 0$ such that $W_p(l_\alpha(d), d_0) \leq \varepsilon$ for all $d \in S$.

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

Theorem 21. 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 $\varepsilon > 0$. Since $S$ is totally bounded it is bounded. 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 $\delta$ which cover $S$. For each $d_n$ we can find a constant $C_n$ such that $bd(x) \leq C_n$ for $x \in d_n$ with $\text{pers}(x) \geq \varepsilon$, and $\text{pers}(x) \leq \varepsilon/4$ for all $x \in d_n$ with $bd(x) > C_n$. Let $C = \max\{C_1, \ldots, C_N\}$. Also, we can find $\alpha > 0$ such that $W_p(l_\alpha(d_n), d_0) \leq \varepsilon/4$ 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 $\text{pers}(x) \geq \varepsilon$ and $bd(x) > C + \varepsilon$. Then for any bijection $\gamma : d \rightarrow d_n$ we have $\|x - \gamma(x)\|_\infty \geq \varepsilon/2 - \varepsilon/8$ which implies that
Suppose that $\gamma$ under $\mathcal{D}$ is bounded, we can also find a constant $M$ such that $\delta_\gamma(t_\delta) \leq M$. Let $l_\delta$ be the subset of the plane corresponding to points whose birth and death are bounded by $C$. Since $l_\delta$ 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 set of all points corresponding to points whose birth and death are bounded by $C$. Since $l_\delta$ 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

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

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

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

\section{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 $\sigma$-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

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

\subsection{Existence of Fréchet expectations}

The idea of the Fréchet expectation and variance \cite{25, 26} was to extend means and variances to general metric spaces. In the case of persistence diagrams the following definition is relevant.

**Definition 22** (Fréchet expectation). Given a probability space $(D_p, \mathcal{B}(D_p), \mathcal{P})$ the quantity

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

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

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

is the Fréchet expectation, also called Fréchet mean.

Being the result of a minimization, the Fréchet mean may be non-unique or empty. There are several results on the existence (and uniqueness) of the Fréchet mean for particular manifolds and distributions \cite{27, 28}. Typically, a local compactness condition as well as convexity and curvature constraints on the metric space are required. It is not clear how to directly apply the results developed in these papers to our setting.

We provide a proof for the existence of the Fréchet expectation under mild regularity conditions on $\mathcal{P}$ specific for the space of persistence diagrams. The main idea here is to show that if $\{d_n\}$ is a sequence which is not off-diagonally birth-death bounded or not uniform but such that $F_{\mathcal{P}}(d_n) \to \text{Var}_\mathcal{P}$, then we can construct a subsequence $d_{nk}$ and subdiagrams $\bar{d}_{nk} \subset d_{nk}$ such that $F_{\mathcal{P}}(\bar{d}_{nk}) \leq F_{\mathcal{P}}(d_{nk}) - \varepsilon$ for some fixed $\varepsilon$, which implies $\text{Var}_\mathcal{P} < \text{Var}_{\mathcal{P}}$, a contradiction. The following lemma provides a crucial component of this idea.

**Lemma 23.** Let $\mathcal{P}$ be a finite measure on $(D_p, \mathcal{B}(D_p))$ with a finite second moment and compact support $S \subset D_p$, and let $\{d_n\} \subset D_p$, $n \in \mathbb{N}$, be a bounded sequence which is not off-diagonally birth-death bounded or/and not uniform. Also let $C_1 > 1$ and $C_2 > 1$ be bounds on $S$ and $d_n$, respectively, that is, $W_p(d, d_\emptyset) \leq C_1$ and $W_p(d_n, d_\emptyset) \leq C_2$. Then
there is $\delta > 0$ (depending only on $d_n$), a subsequence $d_{n_k}$, $k \in \mathbb{N}$, and subdiagrams $d_{n_k}$ such that
\[ \int_S W_p(d_{n_k}, d)^2 d\mathcal{P}(d) \leq \int_S W_p(d_{n_k}, d)^2 d\mathcal{P}(d) - \varepsilon_0 \mathcal{P}(S), \]
where
\[ \varepsilon_0 = (2^{\frac{2}{s}} - 1)(C_1 + C_2)^{2-s} \delta^s, \quad s = \max \{2, p\}. \]

**Proof.** First, consider the case when $d_n$ 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 $\text{pers}(x) \geq \varepsilon$ and $\text{bd}(x) \geq C$. Take $0 < \delta < \varepsilon/4$ and choose $C_0 > 1$ such that for all $d \in S$ we have $\text{bd}(x) \leq C_0$ for $x \in u_{\delta}(d)$. 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 $\text{pers}(x) \geq \varepsilon$ and $\text{bd}(x) \geq C_3$, and let $d_{n_k}$ 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
\[ \sum \|x - \gamma(x)\|_\infty^p \leq W_p(d_{n_k}, d)^p + \delta^p. \]

Notice that $W_p(d_{n_k}, d) \leq W_p(d_{n_k}, d_0) + W_p(d, d_0) \leq C_1 + C_2$, so $\text{bd}(\gamma(x)) > C_0$ for all $x \in d_{n_k}$ with $\text{bd}(x) \geq C_3$. Thus $\gamma(x) \in l_{\delta}(d)$ for $x \in d_{n_k}$ with $\text{bd}(x) \geq C_3$. Hence $\|x - \gamma(x)\|_\infty \geq \text{pers}(x)/2 - \delta/2 > \delta$. Let $\gamma : d_{n_k} \to d$ be the bijection obtained from $\gamma$ by pairing points $\gamma(x)$ such that $\text{pers}(x) \geq \varepsilon$ and $\text{bd}(x) \geq C_3$ to the diagonal. Then we have
\[ \sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p = \sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p + \sum_{x \in d_{n_k} - d_{n_k}} \|x - \gamma(x)\|_\infty^p \]
\[ \geq \sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p + \delta^p \geq \sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p + \delta^p. \]

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
\[ \left( \sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p \right)^{\frac{2}{p}} \geq \left( \sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p \right)^{\frac{2}{p}} + \varepsilon_0, \]
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
\[ W_p(d_{n_k}, d)^2 \geq W_p(d_{n_k}, d)^2 + \varepsilon_0. \]
Therefore
\[
\int_S W_p(\bar{d}_{n_k}, d)^2 d\mathcal{P}(d) \leq \int_S W_p(d_{n_k}, d)^2 d\mathcal{P}(d) - \varepsilon_0 \mathcal{P}(S).
\]

Now, suppose 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 \(W_p(l_\alpha(l_n \cup d_\emptyset)) \geq \varepsilon\). If necessary, decrease the \(\delta\) from the previous case so that \(0 < \delta < \varepsilon/4\) and choose \(\alpha_0\) such that \(W_p(l_{\alpha_0}(d)) \leq \delta\) 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 \(\delta_0 = f^{-1}(M^{-1}C^{-p}d^p)\), and \(\alpha_1 = \min\{\delta_0, \alpha_0, M^{-1/p}\delta\}\). 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 \(d_{n_k} = u_{\alpha_1}(d_{n_k})\). Take \(d \in S\) and let \(\gamma : d_{n_k} \to d\) be a bijection such that
\[
\sum_{x \in d_{n_k}} \|x - \gamma(x)\|^p_\infty \leq W_p(d_{n_k}, d)^p + \delta^p.
\]

Let \(\tilde{\gamma} : d_{n_k} \to d\) be the bijection obtained from \(\gamma\) by pairing points in \(\gamma(l_{\alpha_1}(d_{n_k}))\) to the diagonal. For convenience, let \(s_0 = d_{n_k}\), \(s_1 = \{x \in d_{n_k} | \text{pers}(x) < \alpha_1, \text{pers}(\gamma(x)) < \alpha_0\}\), \(s_2 = \{x \in d_{n_k} | \text{pers}(x) < \alpha_1, \text{pers}(\gamma(x)) \geq \alpha_0\}\). Notice that
\[
\sum_{x \in s_2} \left(\frac{\text{pers}(x)}{2}\right)^p \leq M \frac{\alpha_0^p}{2^p} \leq \frac{\delta^p}{2^p}.
\]

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

Consequently,
\[
W_p(s_1, d_\emptyset) - W_p(\gamma(s_1), d_\emptyset) = \left(\sum_{x \in s_1} \left(\frac{\text{pers}(x)}{2}\right)^p\right)^{1/p} - \left(\sum_{x \in s_1} \left(\frac{\text{pers}(\gamma(x))}{2}\right)^p\right)^{1/p}
\]
\[
\geq \varepsilon \left(1 - \left(\frac{\delta}{2\varepsilon}\right)^p\right)^{1/p} - \delta \geq 2.5\delta,
\]

and thus
\[
\left(\sum_{x \in s_1} \|x - \gamma(x)\|^p_\infty\right)^{1/p} \geq W_p(s_1, \gamma(s_1)) \geq W_p(s_1, d_\emptyset) - W_p(\gamma(s_1), d_\emptyset) \geq 2.5\delta.
\]

Also
\[
\sum_{x \in s_2} \|x - \gamma(x)\|^p_\infty \geq 2^{-p} \sum_{x \in s_2} (\text{pers}(\gamma(x)) - \alpha_1)^p,
\]
and
\[
\sum_{x \in s_2} (\text{pers}(\gamma(x)) - \alpha_1)^p = \sum_{x \in s_2} \left( \text{pers}(\gamma(x))^p - \text{pers}(\gamma(x))^p f \left( \frac{\alpha_1}{\text{pers}(\gamma(x))} \right) \right)
\geq \sum_{x \in s_2} \left( \text{pers}(\gamma(x))^p - C^p f \left( \frac{\alpha_1}{\alpha_0} \right) \right)
\geq \sum_{x \in s_2} \text{pers}(\gamma(x))^p - \delta^p.
\]

Recall that we chose \(\alpha_0\) such that \(W_p(\alpha_0, d_0) \leq \delta\). Therefore,
\[
2^{-p} \sum_{x \in s_1} \text{pers}(\gamma(x))^p \leq W_p(\alpha_0, d_0)^p \leq \delta^p.
\]

We then have
\[
\sum_{x \in d_{nk}} \|x - \gamma(x)\|_\infty^p = \sum_{x \in s_0} \|x - \gamma(x)\|_\infty^p + \sum_{x \in s_1} \|x - \gamma(x)\|_\infty^p + \sum_{x \in s_2} \|x - \gamma(x)\|_\infty^p
\geq \sum_{x \in s_0} \|x - \gamma(x)\|_\infty^p + 2^{-p} \sum_{x \in s_2} \text{pers}(\gamma(x))^p + (2.5)^p \delta^p - 2^{-p} \delta^p
\geq \sum_{x \in s_0} \|x - \gamma(x)\|_\infty^p + 2^{-p} \sum_{x \in s_2} \text{pers}(\gamma(x))^p +
\]
\[
2^{-p} \sum_{x \in s_2} \text{pers}(\gamma(x))^p + ((2.5)^p - 1 - 2^{-p}) \delta^p
\geq \sum_{x \in d_{nk}} \|x - \gamma(x)\|_\infty^p + \delta^p.
\]

As in the previous case, this implies that
\[
\left( \sum_{x \in d_{nk}} \|x - \gamma(x)\|_\infty^p \right)^{\frac{2}{p}} \geq \left( \sum_{x \in d_{nk}} \|x - \bar{\gamma}(x)\|_\infty^p \right)^{\frac{2}{p}} + \varepsilon_0,
\]
where
\[
\varepsilon_0 = (2^\frac{4}{p} - 1)(C_1 + C_2)^{2-s} \delta^s, \quad s = \max\{2, p\}.
\]

Therefore
\[
W_p(d_{nk}, d)^2 \geq W_p(\bar{d}_{nk}, d)^2 + \varepsilon_0,
\]
and consequently
\[
\int_S W_p(\bar{d}_{nk}, d)^2 d\mathcal{P}(d) \leq \int_S W_p(d_{nk}, d)^2 d\mathcal{P}(d) - \varepsilon_0 \mathcal{P}(S).
\]

We now can prove existence of the Fréchet expectation for probability measures with compact support.
Theorem 24. 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 $\mathbb{P}_\mathcal{P} \neq \emptyset$.

Proof. Let $S \subset D_p$ be the support of $\mathcal{P}$ and let $\{d_n\}_{n=1}^{\infty}$ be a sequence in $D_p$ such that $F_P(d_n) \to \text{Var}_\mathcal{P}$. We shall show that $\{d_n\}$ is bounded, off diagonally birth-death bounded and uniform. By Theorem 21 it is totally bounded. By Proposition 15 $\{d_n\}$ has a subsequence convergent in $D_p$.

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

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

which is a contradiction.

Now assume that $\{d_n\}$ is not off-diagonally birth-death bounded or not uniform. By Lemma 23 we have a subsequence $d_{nk}$ and subdiagrams $\bar{d}_{nk} \subset d_{nk}$ such that

$$\int_S W_p(d_{nk}, d)^2 d\mathcal{P}(d) \leq \int_S W_p(d_{nk}, d)^2 d\mathcal{P}(d) - \varepsilon_0 \mathcal{P}(S).$$

Taking the infimum over $k$ we obtain $\text{Var}_\mathcal{P} \leq \text{Var}_\mathcal{P} - \varepsilon_0 \mathcal{P}(S)$, which is 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 25. Let $\mathbb{X}$ be a Hausdorff topological space, and let $\Sigma$ be a $\sigma$-algebra on $\mathbb{X}$ that contains the topology of $\mathbb{X}$. A measure $\mu$ 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 26. Let $(\mathbb{X}, \rho)$ be a metric space, and let $\Sigma$ be a $\sigma$-algebra on $\mathbb{X}$ that contains the topology of $\mathbb{X}$. A measure $\mu$ 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 shall also need the following lemma.

Lemma 27. Let $\mathcal{P}$ be a tight probability measure on $(D_p, \mathcal{B}(D_p))$ with the rate of decay at infinity $q > \max \{2, p\}$, and let $\{d_n\}_{n=1}^{\infty} \subset D_p$, $n \in \mathbb{N}$, be a bounded sequence. Then for any $\varepsilon > 0$ there are $M \in \mathbb{N}$ and a compact set $S \subset D_p$ such that for any subsequence of subdiagrams $\bar{d}_{nk} \subset d_{nk}$, $k \in \mathbb{N}$, we have

$$\int_{D_p} W_p(d_{nk}, d)^2 d\mathcal{P}(d) < \int_{S \cap B_M(d_0)} W_p(d_{nk}, d)^2 d\mathcal{P}(d) + \frac{\varepsilon}{M^{s-2}},$$

where $s = \max \{2, p\}$ and $B_M(d_0) = \{d \in D_p \mid W_p(d, d_0) \leq M\}$. Moreover, $\mathcal{P}(S \cap B_M(d_0)) > 1 - \varepsilon/4$. 

Proof. Let $C > 0$ and $R > 0$ be such that $\mathcal{P}(B^r(d_0)) \leq Cr^{-q}$, $r \geq R$. Take $M \in \mathbb{N}$ such that $M > R$, $W_p(d_n, d_{\emptyset}) \leq M$ (and hence $W_p(\bar{d}_{n_k}, d_{\emptyset}) \leq M$), $M^{-s} < \varepsilon/(8C)$, and
\[
\frac{(M + 1)^s}{M^q} < \frac{\varepsilon}{16C} \quad \text{and} \quad \sum_{m \geq M} \frac{(2m + 3)^{s-1}}{(m + 1)^q} < \frac{\varepsilon}{16C}.
\]
Denote
\[
B^{m,m+1}(d_0) = B^m(d_0) - B^{m+1}(d_0).
\]
We have
\[
\int_{B^m(d_0)} W_p(\bar{d}_{n_k}, d)^2 d\mathcal{P}(d) \leq \int_{B^m(d_0)} (W_p(\bar{d}_{n_k}, d_0) + W_p(d_0, d))^2 d\mathcal{P}(d),
\]
\[
\leq \int_{B^m(d_0)} (2W_p(d_0, d))^2 d\mathcal{P}(d).
\]
Note that
\[
\int_{B^m(d_0)} (2W_p(d_0, d))^2 d\mathcal{P}(d) = 4 \sum_{m \geq M} \int_{B^{m,m+1}(d_0)} W_p(d_0, d)^2 d\mathcal{P}(d)
\]
\[
\leq 4 \sum_{m \geq M} (m + 1)^2 \left( \mathcal{P}(B^m(d_0)) - \mathcal{P}(B^{m+1}(d_0)) \right).
\]
Denote the right hand side of the above expression by $L$. Then
\[
L = 4 \sum_{m \geq M} ((m + 1)^2 \mathcal{P}(B^m(d_0)) - (m + 2)^2 \mathcal{P}(B^{m+1}(d_0))) + 4 \sum_{m \geq M} (2m + 3) \mathcal{P}(B^{m+1}(d_0)).
\]
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 $S^c = D_p - S$. Then we have
\[
\int_{S^c \cap B_M(d_0)} W_p(\bar{d}_{n_k}, d)^2 d\mathcal{P}(d) \leq \int_{S^c \cap B_M(d_0)} (W_p(\bar{d}_{n_k}, d_0) + W_p(d_0, d))^2 d\mathcal{P}(d),
\]
\[
\leq 4M^2 \mathcal{P}(S^c \cap B_M(d_0)) < \frac{\varepsilon}{2M^{s-2}}.
\]
Combining the two results we get
\[
\int_{D_p} W_p(\bar{d}_{n_k}, d)^2 d\mathcal{P}(d) \leq \int_{S \cap B_M(d_0)} W_p(\bar{d}_{n_k}, d)^2 d\mathcal{P}(d) + \int_{S^c \cap B_M(d_0)} W_p(\bar{d}_{n_k}, d)^2 d\mathcal{P}(d) + \int_{B^M(d_0)} W_p(\bar{d}_{n_k}, d)^2 d\mathcal{P}(d) < \int_{S \cap B_M(d_0)} W_p(\bar{d}_{n_k}, d)^2 d\mathcal{P}(d) + \frac{\varepsilon}{M^{s-2}}.
\]
To prove the last statement of the Lemma, notice that $\mathcal{P}(S) > 1 - \varepsilon/8$ and $\mathcal{P}(B^M(d_0)) < \varepsilon/8$. Since $\mathcal{P}(S) \leq \mathcal{P}(S \cap B_M(d_0)) + \mathcal{P}(B^M(d_0))$ we obtain $\mathcal{P}(S \cap B_M(d_0)) > 1 - \varepsilon/4$. \qed
Now we can prove the following result.

**Theorem 28.** 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 in \( D_p \) such that \( F_\mathcal{P}(d_n) \to \text{Var}_\mathcal{P} \). We shall show that \( \{d_n\} \) is bounded, off diagonally birth-death bounded and uniform. By Theorem 21 it is totally bounded. By Proposition 15 \( \{d_n\} \) has a subsequence convergent in \( D_p \).

First, assume that \( \{d_n\} \) is not bounded. Since \( \mathcal{P} \) is tight, we can find a compact set \( S_0 \subset D_p \) such that \( \mathcal{P}(S_0) \geq 0.5 \). Then \( w_n = \inf_{d \in S_0} W_p(d_n, d) \) is not bounded. Thus, as \( n \to \infty \) we get

\[
F_\mathcal{P}(d_n) = \int_{D_p} W_p^2(d_n, d) d\mathcal{P}(d) \geq \int_{S_0} W_p^2(d_n, d) d\mathcal{P}(d) \geq w_n^2 \mathcal{P}(S_0) \to \infty,
\]

which is a contradiction.

Let us assume now that \( d_n \) is not off-diagonally birth-death bounded or not uniform. Let \( \delta_0 > 0 \) be the \( \delta \) from Lemma 23. Take \( \varepsilon > 0 \) such that \( \varepsilon < (2^{s/2} - 1)2^{1-s}\delta_0^s(1 - \varepsilon/4) \). By Lemma 27 the inequality

\[
\int_{D_p} W_p(\bar{d}_{nk}, d)^2 d\mathcal{P}(d) < \int_{S \cap B_M(d_0)} W_p(\bar{d}_{nk}, d)^2 d\mathcal{P}(d) + \frac{\varepsilon}{M^{s-2}}
\]

holds for the subsequences of subdiagrams \( \bar{d}_{nk} \) from Lemma 23. Now, by Lemma 23 we have

\[
\int_{S \cap B_M(d_0)} W_p(\bar{d}_{nk}, d)^2 d\mathcal{P}(d) \leq \int_{S \cap B_M(d_0)} W_p(d_{nk}, d)^2 d\mathcal{P}(d) - \varepsilon_0 \mathcal{P}(S \cap B_M(d_0)),
\]

where

\[
\varepsilon_0 = \frac{(2^{s/2} - 1)2^{2-s}\delta_0^s}{M^{s-2}}.
\]

By Lemma 27 we have \( \mathcal{P}(S \cap B_M(d_0)) > 1 - \varepsilon/4 \). Therefore,

\[
\int_{D_p} W_p(\bar{d}_{nk}, d)^2 d\mathcal{P}(d) \leq \int_{D_p} W_p(d_{nk}, d)^2 d\mathcal{P}(d) - \frac{\varepsilon_0(1 - \varepsilon/4)}{2}.
\]

Taking the infimum over \( k \) we obtain

\[
\text{Var}_\mathcal{P} \leq \text{Var}_\mathcal{P} - \frac{\varepsilon_0(1 - \varepsilon/4)}{2},
\]

which results in a contradiction. \( \square \)
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 \( \mathcal{P}(\theta | D) \), 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, as considered in [29]. This distribution has a parametric form with parameters \( \theta \) 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 \( F_\theta \)

\[
X_1, \ldots, X_n \overset{iid}{\sim} F_\theta.
\]

This results in a likelihood for the observed point cloud data \( Z \equiv \{X_1, \ldots, X_n\} \)

\[
\text{Lik}(Z; \theta) \equiv f_\theta(Z),
\]

where \( f_\theta \) is the probability density function corresponding to the probability distribution function \( F_\theta \).

We start with the premise that the point cloud data is generated from a probability measure so we have a probability space \((X, \mathcal{B}(X), \mathcal{P}_\theta)\) where \( X \) is a subset of \( \mathbb{R}^d \) (for example a torus), \( \mathcal{B}(X) \) is the Borel \( \sigma \)-algebra on \( X \) and \( \mathcal{P}_\theta \) is the probability measure parameterized by \( \theta \). The observed point cloud data \( Z \equiv \{X_1, \ldots, X_n\} \), where \( X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{P}_\theta \), can be regarded as an element of the probability space \((X^n, \Sigma^n, \mathcal{P}_\theta^n)\), where \( X^n = \prod_{i=1}^n X \), and \( \Sigma^n \) and \( \mathcal{P}_\theta^n \) denote the \( \sigma \)-algebra and probability measure induced by the product structure. Alternatively, \( Z \) can be regarded as a compact subset of \( X \), and we express this formally by defining a map \( h_n : X^n \rightarrow K(X) \),

\[
h_n(X_1, \ldots, X_n) = \{X_1, \ldots, X_n\},
\]

where \( K(X) \) denotes the space of compact subsets of \( X \) endowed with the Hausdorff metric. Suppose now that we have a (continuous) map \( \rho : K(X) \rightarrow \text{Lip}(X) \), where \( \text{Lip}(X) \) denotes the space of Lipschitz functions on \( X \) with the supremum norm. For example, we can take \( \rho(S)(x) = d_S(x) = \inf_{y \in S} \|x - y\| \), the usual distance function. Another choice is to regard \( S \in K(X) \) 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 [16]. Composing these maps and taking the persistence diagram of the resulting function we thus obtain a map \( g : X^n \rightarrow D_p \). 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

\[
\mathcal{P}_D(A) = \mathcal{P}_\theta^n(g^{-1}(A)), \quad \text{for } A \in \mathcal{B}(D_p).
\]
Notice that if \( X \) is triangulable, compact and implies bounded degree-\( k \) total persistence, and if \( \rho \) 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 \( p > k \) and continuous maps between Borel spaces are measurable. Since measurability is a much weaker condition than continuity for Borel spaces, we expect that the induced probability measure on the space of persistence diagrams can be defined in many more general cases.

The probability measure \( P_D \) constructed above is conditioned on the parameter \( \theta \). Suppose that we have a prior distribution of \( \theta \) given by the measure \( \mu \). Then the joint probability measure \( P(D, \theta) \) is given by the product measure

\[
P(D, \theta) = P_D \times \mu.
\]

Bayes’ rule also gives us the conditional measure \( P(\theta \mid D) \):

\[
P(\theta \mid D) \propto P_D \times \mu.
\]

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 \( P_\theta \) with the probability of the topological summary \( P_D \) — this violates the likelihood principle [30]. This idea of a substitution likelihood goes back to Jeffreys [31] and a basic question in TDA is what properties of \( P_\theta \) are preserved by \( P_D \).

### 4.3. An example

Assume we obtain \( m \) point samples from object \( O_1 \) (for example a torus) and \( n \) point samples from an object \( O_2 \) (a double torus). For each point sample we obtain persistence diagrams resulting in two sets of diagrams \( \{x_1, \ldots, x_m\} \subseteq D_p \) for \( O_1 \) and \( \{y_1, \ldots, y_n\} \subseteq D_p \) for \( O_2 \). We are also given a persistence diagram \( z \) that comes from either object but we do not know which one, we would like to assign this diagram to one of the two objects. This is the problem of classification in statistical inference and machine learning. In the following we outline how we can use the results in the previous sections to classify the persistence diagram \( z \). Given the two sets \( \{x_1, \ldots, x_m\} \) and \( \{y_1, \ldots, y_n\} \) can use a variation of kernel density estimation [32, 33] to provide the following density estimates for for diagrams corresponding to object \( O_1 \) and \( O_2 \), respectively

\[
\hat{p}(x \mid O_1) = \frac{1}{n\kappa_\tau} \sum_{i=1}^{m} e^{-W_p^2(x,x_i)/\tau}, \quad \hat{p}(x \mid O_2) = \frac{1}{n\kappa_\tau} \sum_{i=1}^{n} e^{-W_p^2(x,y_i)/\tau},
\]

here \( \tau > 0 \) is the bandwidth parameter that controls the smoothness of the density, and \( \kappa_\tau \) is a normalizing constant. If we assume that the objects have prior probabilities \( \pi_1 = \Pr(O_1) \) and \( \pi_2 = \Pr(O_2) \), we can use Bayes’ rule to compute the posterior probability of membership in class one (the torus) given the diagram \( z \), \( \hat{p}(O_1 \mid z) \)

\[
\hat{p}(O_1 \mid z) = \frac{\hat{p}(z \mid O_1) \pi_1}{\hat{p}(z)} = \frac{\hat{p}(z \mid O_1) \pi_1}{\hat{p}(z \mid O_1) \pi_1 + \hat{p}(z \mid O_2) \pi_2},
\]
note that we do not need to compute the normalization constant $\kappa_\tau$ to compute the posterior probability since $\kappa_\tau$ appears in both the numerator and denominator.

The point of this example is to illustrate that probability distributions on the space of persistence diagrams can be used to make decisions on new observations. Placing persistence diagrams on a probabilistic footing allows for the application of standard ideas and tools in statistical inference including classification, and estimates of variation and means.

### 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 system depending on a parameter $\theta$. Suppose we can obtain samples from the attractors of this system. Then we can try to estimate the distribution of $\theta$ 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 [34] 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.

### Acknowledgements

The authors would like to thank two anonymous referees for their valuable feedback on the manuscript. J.H. and Y.M. are pleased to acknowledge the support of the National Science Foundation under Plant Genome grant No. DBI-08-20624-002 and
the support of the Defense Advance Research Projects Agency under FunBio program (Princeton University Subcontract 00001744). The work of J.H. and S.M. has in part been supported by the National Institute of Health System Biology grant No. 5P50-GM081883-05 and the Air Force Office of Scientific Research grant No. FA9550-10-1-0436. S.M also acknowledges the support of the National Science Foundation under grant No. CCF-1049290.

References


