Probability measures on the space of persistence diagrams

Probability measures on the space of persistence diagrams

Yuriy Mileyko[1], Sayan Mukherjee[2] and John Harer[3]

[1] Departments of Mathematics, Duke University, Durham, NC 27708, USA
[2] Departments of Statistical Science, Computer Science, and Mathematics,
Institute for Genome Sciences & Policy, Duke University, Durham, NC 27708, USA
[3] Department of Mathematics, Computer Science, Electrical and Computer Engineering,
Duke University, Durham, NC 27708, USA

E-mail: yury@math.duke.edu, sayan@stat.duke.edu and 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.

(Some figures in this article are in colour only in the electronic version)

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 [1–3]. 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 [5–10] and non-asymptotic bounds on the convergence or consistency of topological summaries as the number of points increase [11–15]. 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 are examined. The key idea in this paper is that 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 provides 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 bounded continuous function $f : \mathbb{X} \to \mathbb{R}$. Let $\mathbb{X}_a = f^{-1}(-\infty, a]$ denote the sublevel set of f at the threshold a. Inclusions $\mathbb{X}_a \subset \mathbb{X}_b$, $a \leqslant b$, 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 it has only a finite

number of homological critical values and if $H_\ell(\mathbb{X}_a)$ 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

Note 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 denote such an image by $\mathsf{F}_\ell^{a-,b}$. Now, consider the following quotient group:

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

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 *ℓth birth group* at $\mathbb{X}_a$, and we say that a homology class $\alpha \in H_\ell(\mathbb{X}_a)$ 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

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

defined as $\mathsf{g}_\ell^{a,b}([\alpha]) = \left[\mathsf{f}_\ell^{a,b}(\alpha)\right]$, where $\alpha \in H_\ell(\mathbb{X}_a)$, and the square brackets denote the images under the corresponding canonical projections. We set $\mathsf{g}_\ell^{a,b} = 0$ if $b = \sup_{x\in\mathbb{X}} f(x)$, so that each homology class has finite persistence (as defined in section 2.3). The kernel of this map, which we denote by $\mathsf{D}_\ell^{a,b}$, 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 $\alpha \in H_\ell(\mathbb{X}_a)$ *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 $\mathrm{rank}\mathsf{D}_\ell^{a,b} - \mathrm{rank}\mathsf{D}_\ell^{a,b-\delta} = r > 0$ for all sufficiently small $\delta > 0$. Note 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 $\mathsf{b}(\alpha) = a, \mathsf{d}(\alpha) = b$. The *persistence* of α is the difference between the two values, $\mathrm{pers}(\alpha) = \mathsf{d}(\alpha) - \mathsf{b}(\alpha)$. We represent the births and deaths of ℓ-dimensional homology classes by a multiset of points in $\mathbb{R}^2$, the *ℓth persistence diagram* denoted by $\mathrm{Dgm}_\ell(f)$. For each nontrivial birth group B_ℓ^a, the diagram contains points $x_i = (a, b_i)$, where b_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 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 $x \in \mathrm{Dgm}_\ell(f)$, 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, which are defined in the previous section as *finite* multisets of points in a plane.

Definition 1 (Wasserstein distance). *The pth 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

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

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. Note 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 $\Delta = \{ (x, y) \in \mathbb{R}^2 \mid x = y \}$, where each point on the diagonal has infinite multiplicity.*

The pth Wasserstein distance applies naturally to generalized persistence diagrams. We omit the word 'generalized' for the rest of the paper, as these are the only diagrams that we will 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. Note 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*

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

In this paper, we consider only the case $p \geqslant 1$.

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 [18–20] as well as in computer vision [21] and image retrieval [22]. Given probability measures μ, ν with finite *p*th moments on a metric space $(\mathbb{X}, \rho)$, the *p*th Wasserstein distance between μ and ν is defined as follows:

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

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

$$\int_{\mathbb{X}} \rho^p(x_0, x) \, \mathrm{d}\mu(x) < \infty.$$

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 $\mathbb{X}$ 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 $\mathrm{Lip}(f) \leqslant 1$ we have $\mathrm{Pers}_p(d) \leqslant C_{\mathbb{X}}$, where $C_{\mathbb{X}}$ is a constant that depends only on $\mathbb{X}$. We say in this case that $\mathbb{X}$ *implies bounded degree-p total persistence*.

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

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

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 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^*. 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 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 $u_\alpha : D_p \to D_p$ be defined by

$$x \in u_\alpha(d) \iff x \in d \quad \text{and} \quad \text{pers}(x) \geqslant \alpha.$$

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

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

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 analog 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.

Lemma 7 (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 $M_{\text{sup}}^\delta = \limsup_{n \to \infty} |u_\delta(d_n)|$, $M_{\text{inf}}^\delta = \liminf_{n \to \infty} |u_\delta(d_n)|$. Note that $M_{\text{sup}}^\delta < \infty$; otherwise, we could find a subsequence d_{n_k} such that $|u_\delta(d_{n_k})| > k$ so that $W_p(d_{n_k}, d_\emptyset) \geqslant k^{1/p}\delta/2 \to \infty$ as $k \to \infty$. However, $W_p(d_n, d_\emptyset)$ is bounded because d_n is Cauchy.

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

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_i} be two subsequences such that $|u_\delta(d_{n_s})| = M_{\text{sup}}$ and $|u_{\delta_\alpha}(d_{n_i})| = M_{\text{inf}}$. On 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 $\text{pers}(x) \geqslant \delta$, $\text{pers}(\gamma(x)) < \delta_\alpha \Rightarrow \|x - \gamma(x)\|_\infty > \varepsilon/2$. Therefore, $W_p(d_{n_s}, d_{n_i}) > \varepsilon/2$, which is a contradiction. We then set $M_\alpha = M_{\text{sup}} = M_{\text{inf}}$. $\square$
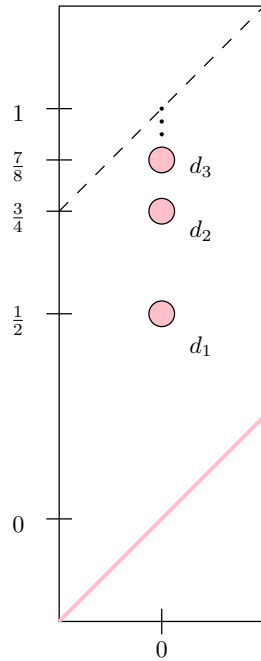
Figure 1. Example of convergence from below. Shown are the first three diagrams from the sequence d_n such that $|d_n| = 1$, $b(x) = 0$ for all $x \in d_n$ and $n \in \mathbb{N}$, and $\text{pers}(x) = 1 - 2^{-n}$ for $x \in d_n$. The sequence of off-diagonal points converges to a single point with persistence 1. It is clear, however, that the 1-upper part of any d_n is empty.

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

Lemma 8. *For any $\alpha > 0$ the sequence d_n^α is Cauchy.*

Proof. Use lemma 7 to choose δ_α. 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 $\gamma : d_n \to d_m$ such that

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

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

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

$\square$

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

Lemma 9 (Persistence-wise convergence). *For any $\alpha > 0 \; \exists \; d^\alpha \in D_p$ such that $\lim_{n \to \infty} W_p(d_n^\alpha, d^\alpha) = 0$; hence, $|d^\alpha| = M_\alpha$ and $u_\alpha(d^\alpha) = d^\alpha$. Moreover, $d^{\alpha_1} \subset d^{\alpha_2}$ if $\alpha_1 > \alpha_2$.*

Proof. Let $\alpha > 0$, $\delta \in (\delta_\alpha, \alpha)$, and let $N > 0$ be such that $\left| d_n^\alpha \right| = |u_\delta(d_n)| = M_\alpha$ for all $n > N$. Let ε be such that $0 < \varepsilon < \delta/2$. Choose a subsequence $d_{n_k}^\alpha$, $k \in \mathbb{N}$, such that $n_1 > N$ and $W_p\left(d_{n_k}^\alpha, d_m^\alpha \right) < 2^{-k}\varepsilon$ for $m \geqslant n_k$. Let $\gamma_k : d_{n_k}^\alpha \to d_{n_{k+1}}^\alpha$ be a bijection realizing the Wasserstein distance $W_p\left(d_{n_k}^\alpha, d_{n_{k+1}}^\alpha \right)$. Note that our choice of ε guarantees that each γ_k maps off diagonal points to off diagonal points. Let $x^1, \ldots, x^{M_\alpha}$ be off-diagonal elements of $d_{n_1}^\alpha$. We can now construct M_α sequences of points in a plane $\left\{ x_k^1 \right\}, \ldots, \left\{ x_k^{m_\alpha} \right\}$, $k \in \mathbb{N}$, such that $x_1^i = x^i$, $i = 1, \ldots, M_\alpha$, and $x_{k+1}^i = \gamma_k\left(x_k^i \right)$. Note that each sequence x_k^i is Cauchy. Indeed, $W_p\left(d_{n_k}^\alpha, d_{n_{k+1}}^\alpha \right) < 2^{-k}\varepsilon$ implies $\| x_k^i - x_l^i \|_\infty < 2^{1-k}\varepsilon$ 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^α be the diagram whose off diagonal elements are $\hat{x}^1, \ldots, \hat{x}^{M_\alpha}$ (note 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^α is the limit of d_n^α 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 γ_k, subsequence $d_{n_k}^\alpha$, or ε.

Let $\varepsilon_0 > 0$ and pick $K > 0$ such that $\forall k > K$ we have $\| x_k^i - \hat{x}^i \| < 0.5\varepsilon_0 M_\alpha^{-1/p}$ and $W_p(d_{n_k}, d_m) < \varepsilon_0/2$, $m \geqslant n_k$. Then, we also have $W_p\left(d_m^\alpha, d^\alpha \right) \leqslant W_p\left(d_m^\alpha, d_{n_k}^\alpha \right) + W_p\left(d_{n_k}^\alpha, d^\alpha \right) < \varepsilon_0/2 + \varepsilon_0/2 = \varepsilon_0$.

The last statement of the lemma follows from the fact that if $\alpha_1 > \alpha_2$, then points $x \in d_n^{\alpha_2}$ such that $x \notin d_n^{\alpha_1}$ have $\mathrm{pers}(x) < \delta_{\alpha_1} < \alpha_1$. Indeed, repeating the above argument with $\alpha = \alpha_2$, $N > 0$ such that $\left| d_n^{\alpha_1} \right| = |u_{\delta_1}(d_n)| = M_{\alpha_1}$ and $\left| d_n^{\alpha_2} \right| = |u_{\delta_2}(d_n)| = M_{\alpha_2}$, for all $n > N$, where $\delta_1 \in (\delta_{\alpha_1}, \alpha_1)$, $\delta_2 \in (\delta_{\alpha_2}, \alpha_2)$, and $\varepsilon > 0$ such that $\varepsilon < \min\{\delta_2/2, (\delta_1 - \delta_{\alpha_1})/2\}$, we see that each $\gamma_k : d_{n_k}^\alpha \to d_{n_{k+1}}^\alpha$ maps off-diagonal points in $d_{n_k}^\alpha$ to off-diagonal points in $d_{n_{k+1}}^{\alpha_1}$. Therefore, the collection of limits $\hat{x}^1, \ldots, \hat{x}^{M_{\alpha_2}}$ contains the limits that we obtain for the case $\alpha = \alpha_1$. $\qquad\square$

Lemma 9 allows us to define $d^* = \cup_{\alpha > 0} d^\alpha$. It is not difficult to show 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 $\forall n$, $W_p(d_n, d_\emptyset) \leqslant C$. Let $\alpha > 0$, and let $N > 0$ be such that $\forall n > N$, $W_p\left(d^\alpha, d_n^\alpha \right) < 1$. Take any such n; then

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

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

Finally, note that

$$W_p(d^\alpha, d^*)^p \leqslant W_p(l_\alpha(d^*), d_\emptyset)^p = \sum_{\substack{x \in d^* \\ \mathrm{pers}(x) < \alpha}} \left(\frac{\mathrm{pers}(x)}{2} \right)^p \to 0 \qquad \text{as} \quad \alpha \to 0.$$

$\qquad\square$

By the triangle inequality $W_p(d^*, d_n) \leqslant W_p(d^*, d^\alpha) + W_p\left(d^\alpha, d_n^\alpha \right) + W_p\left(d_n^\alpha, d_n \right)$. The completeness of D_p follows from lemmas 9–11.

Lemma 11. $\forall \varepsilon > 0$, $\exists \alpha_0 > 0$ *such that* $\forall n \in \mathbb{N}$ *and* $0 < \alpha \leqslant \alpha_0$ *we have* $W_p(d_{n,\alpha}, d_\emptyset) < \varepsilon$ *and hence* $W_p\left(d_n^\alpha, d_n \right) < \varepsilon$.

Proof. We prove the lemma by contradiction. Suppose that $\exists \varepsilon > 0$ such that $\forall \alpha > 0$ $\exists n_\alpha \in \mathbb{N}$ with $W_p(d_{n_\alpha, \alpha}, d_\emptyset) \geqslant \varepsilon$. Take such an ε. 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 $W_p(d_{n_i,\alpha_i}, d_\emptyset) \geqslant \varepsilon$. Let $0 < \delta < \varepsilon/4$, and pick $k \in \mathbb{N}$ such that $W_p(d_{n_k}, d_{n_i}) < \delta$ for all $i \geqslant k$. Now pick $j \geqslant k$ such that $W_p(d_{n_k,\alpha_i}, d_\emptyset) < \delta$ for all $i \geqslant j$. This implies that

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

We now show that this inequality leads to a contradiction. For $i \geqslant j$, 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

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

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

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

where $i \geqslant l$. 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 $i \geqslant l$ we have

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

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

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

Therefore, $W_p(d_{n_i,\alpha_i}, d_{n_k,\alpha_j}) < 3\delta$ if $i \geqslant l$, which is a contradiction. $\qquad\square$

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 $W_p(l_\alpha(d), d_\emptyset) < \varepsilon/2$. Then, we have $W_p(d, u_\alpha(d)) \leqslant W_p(l_\alpha(d), d_\emptyset) < \varepsilon/2$. Since $\mathbb{Q}^{2|u_\alpha(d)|}$ is dense in $\mathbb{R}^{2|u_\alpha(d)|}$, we can find $d_s \in S$ such that $W_p(d_s, u_\alpha(d)) < \varepsilon/2$. Then, $W_p(d, d_s) \leqslant W_p(d, u_\alpha(d)) + W_p(d_s, u_\alpha(d)) < \varepsilon$, which implies that S is dense.

Note that $S = \cup_{m=0}^\infty S_m$, 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. $\qquad\square$
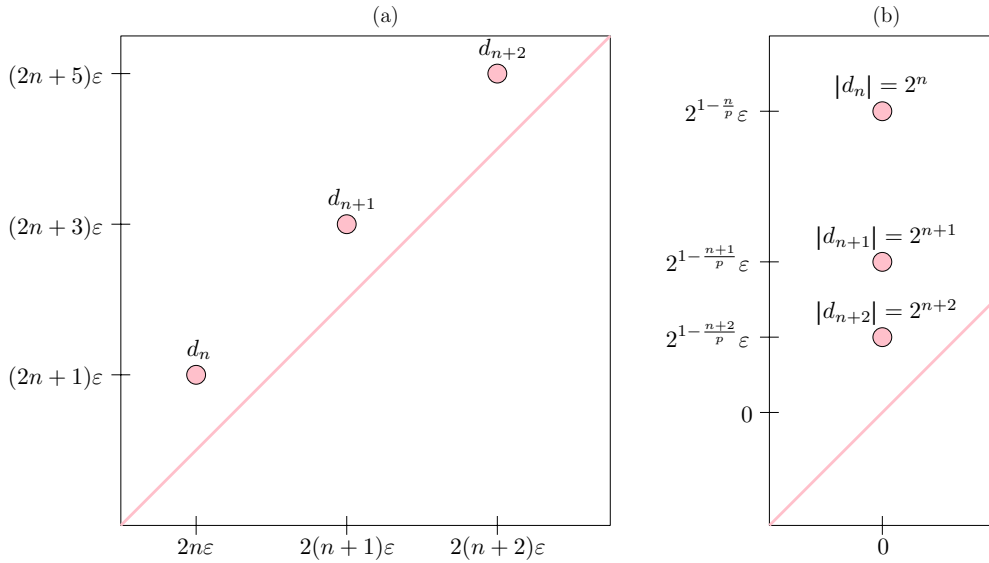
Figure 2. (a) Illustration of three consecutive diagrams from the sequence in example 16. Each point represents a separate diagram with a single off-diagonal point of multiplicity 1. (b) Illustration of three consecutive diagrams from the sequence in example 19. Each point represents a separate diagram with a single off-diagonal point whose multiplicity increases as its persistence decreases.

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 ε 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 $\forall x \in d$ $\max\{|b(x)|, |d(x)|\} \leqslant C$.*

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 $S = \{d \mid W_p(d, d_\emptyset) \leqslant \varepsilon\} \cap \{d \mid b(x) \geqslant 0 \ \& \ d(x) \leqslant C \ \forall x \in d\}$. 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 $W_p(d_n, d_m) \geqslant \left(2^{m-1}\left(\frac{1}{2}2^{1-\frac{m}{p}}\varepsilon\right)^p\right)^{\frac{1}{p}} = 2^{-1/p}\varepsilon$, as there will be at least 2^{m-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 $W_p(l_\alpha(d), d_\emptyset) \leqslant \varepsilon$ 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 δ which cover S. For each d_n we can find a constant C_n such that $bd(x) \leqslant C_n$ for $x \in d_n$ with $pers(x) \geqslant \varepsilon$, and $pers(x) \leqslant \varepsilon/4$ 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 $W_p(l_\alpha(d_n), d_\emptyset) \leqslant \varepsilon/4$ 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 $pers(x) \geqslant \varepsilon$ and $bd(x) > C + \varepsilon$. Then, for any bijection $\gamma : d \to d_n$ we have $\|x - \gamma(x)\|_\infty \geqslant \varepsilon/2 - \varepsilon/8$ which implies that $W_p(d, d_n) > \varepsilon/4$. This contradicts $d \in B_n$ 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 and d_t be maximal subdiagrams of $l_{\alpha/2}(d)$ such that $pers(\gamma(x)) < \alpha$ for $x \in d_b$ and $pers(\gamma(x)) \geqslant \alpha$ for $x \in d_t$. If $W_p(d_b, d_\emptyset) > \varepsilon/2$, then

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

where $\gamma(d_b)$ denotes the subdiagram of d_n which coincides with the image of d_b under γ. Since d_b and d_t do not have common off-diagonal points and $l_{\alpha/2}(d)$ the union of d_b and d_t we have $W_p(l_{\alpha/2}(d), d_\emptyset)^p = W_p(d_b, d_\emptyset)^p + W_p(d_t, d_\emptyset)^p$. Thus, if $W_p(d_b, d_\emptyset) \leqslant \varepsilon/2$, then

$W_p(d_t, d_\emptyset) > (\varepsilon^p - 2^{-p}\varepsilon^p)^{\frac{1}{p}} \geqslant \varepsilon/2$. Note also that if $x \in d_t$, then $\|x - \gamma(x)\|_\infty > \alpha/4 \geqslant$ 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, $W_p(d, d_n) \geqslant \varepsilon/4$ which contradicts $d \in B_n$. Consequently, $W_p(l_{\alpha/2}(d), d_\emptyset) \leqslant \varepsilon$ for all $d \in S$ which implies that S is uniform.

We now prove sufficiency. Given $\varepsilon > 0$, 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 bd$(x) \leqslant C$. Since S is bounded, we can also find a constant $M \in \mathbb{N}$ such that $|u_\delta(d)| \leqslant M$ 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 $\|x - x_n\|_\infty \leqslant M^{-1/p}\varepsilon/2$ for some x_n. Let d^* be the diagram consisting of points x_n, $1 \leqslant n \leqslant N$, 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,

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

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

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

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

4.1. Existence of Fréchet expectations

The idea of the Fréchet expectation and variance [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*

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

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

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

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 [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 $F_{\mathcal{P}}(d_n) \to \mathrm{Var}_{\mathcal{P}}$, then we can construct a subsequence d_{n_k} and subdiagrams $\bar{d}_{n_k} \subset d_{n_k}$ such that $F_{\mathcal{P}}(\bar{d}_{n_k}) \leqslant F_{\mathcal{P}}(d_{n_k}) - \varepsilon$ for some fixed ε, which implies $\mathrm{Var}_{\mathcal{P}} < \mathrm{Var}_{\mathcal{P}}$, 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, $W_p(d, d_\emptyset) \leqslant C_1$ and $W_p(d_n, d_\emptyset) \leqslant C_2$. Then, there is $\delta > 0$ (depending only on d_n), a subsequence d_{n_k}, $k \in \mathbb{N}$, and subdiagrams $\bar{d}_{n_k}$ such that*

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

where

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

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 $\mathrm{pers}(x) \geqslant \varepsilon$ and $\mathrm{bd}(x) \geqslant C$. Take $0 < \delta < \varepsilon/4$ and choose $C_0 > 1$ such that for all $d \in S$ we have $\mathrm{bd}(x) \leqslant C_0$ 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 $\mathrm{pers}(x) \geqslant \varepsilon$ and $\mathrm{bd}(x) \geqslant C_3$, 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

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

Note that $W_p(d_{n_k}, d) \leqslant W_p(d_{n_k}, d_\emptyset) + W_p(d, d_\emptyset) \leqslant C_1 + C_2$, so $\mathrm{bd}(\gamma(x)) > C_0$ for all $x \in d_{n_k}$ with $\mathrm{bd}(x) \geqslant C_3$. Thus, $\gamma(x) \in l_\delta(d)$ for $x \in d_{n_k}$ with $\mathrm{bd}(x) \geqslant C_3$. Hence $\|x - \gamma(x)\|_\infty \geqslant \mathrm{pers}(x)/2 - \delta/2 > \delta$. Let $\bar{\gamma} : \bar{d}_{n_k} \to d$ be the bijection obtained from γ by pairing points $\gamma(x)$ such that $\mathrm{pers}(x) \geqslant \varepsilon$ and $\mathrm{bd}(x) \geqslant C_3$ to the diagonal. Then, we have

$$\sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p = \sum_{x \in \bar{d}_{n_k}} \|x - \gamma(x)\|_\infty^p + \sum_{x \in d_{n_k} - \bar{d}_{n_k}} \|x - \gamma(x)\|_\infty^p$$

$$\geqslant \sum_{x \in \bar{d}_{n_k}} \|x - \gamma(x)\|_\infty^p + \delta^p \geqslant \sum_{x \in \bar{d}_{n_k}} \|x - \bar{\gamma}(x)\|_\infty^p + \delta^p.$$

13

Using the inequalities

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

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

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

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

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

where

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

Taking the infima, we obtain

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

Therefore,

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

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 $W_p(l_\alpha(d_n), d_\emptyset) \geqslant \varepsilon$. If necessary, decrease the δ from the previous case so that $0 < \delta < \varepsilon/4$ and choose α_0 such that $W_p(l_{\alpha_0}(d), d_\emptyset) \leqslant \delta$ for all $d \in S$. Take $M \geqslant 1$ and $C > \delta$ such that for all $d \in S$ we have $|u_{\alpha_0}(d)| \leqslant M$ and $\mathrm{pers}(x) \leqslant C$ for $x \in d$. Define $f : [0, 1] \to [0, 1]$ as $f(x) = 1 - (1 - x)^p$. Note that f is a continuous, monotonically increasing function and $f(0) = 0$, $f(1) = 1$. Set $\delta_0 = f^{-1}(M^{-1}C^{-p}\delta^p)$, and $\alpha_1 = \min\{\delta_0\alpha_0, M^{-1/p}\delta\}$. Let d_{n_k} be a subsequence of d_n such that $W_p(l_{\alpha_1}(d_{n_k}), d_\emptyset) \geqslant \varepsilon$, $k \geqslant 1$, 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

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

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 $s_0 = \bar{d}_{n_k}$, $s_1 = \{x \in d_{n_k} \mid \mathrm{pers}(x) < \alpha_1, \mathrm{pers}(\gamma(x)) < \alpha_0\}$, $s_2 = \{x \in d_{n_k} \mid \mathrm{pers}(x) < \alpha_1, \mathrm{pers}(\gamma(x)) \geqslant \alpha_0\}$. Note that

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

Therefore,

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

Consequently,

$$W_p(s_1, d_\emptyset) - W_p(\gamma(s_1), d_\emptyset) = \left(\sum_{x \in s_1} \left(\frac{\mathrm{pers}(x)}{2} \right)^p \right)^{\frac{1}{p}} - \left(\sum_{x \in s_1} \left(\frac{\mathrm{pers}(\gamma(x))}{2} \right)^p \right)^{\frac{1}{p}}$$

$$\geqslant \varepsilon \left(1 - \left(\frac{\delta}{2\varepsilon} \right)^p \right)^{\frac{1}{p}} - \delta \geqslant 2.5\delta,$$

14

and thus

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

Also

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

and

$$\sum_{x \in s_2} (\operatorname{pers}(\gamma(x)) - \alpha_1)^p = \sum_{x \in s_2} \left(\operatorname{pers}(\gamma(x))^p - \operatorname{pers}(\gamma(x))^p f\left(\frac{\alpha_1}{\operatorname{pers}(\gamma(x))} \right) \right)$$

$$\geqslant \sum_{x \in s_2} \left(\operatorname{pers}(\gamma(x))^p - C^p f\left(\frac{\alpha_1}{\alpha_0} \right) \right)$$

$$\geqslant \sum_{x \in s_2} \operatorname{pers}(\gamma(x))^p - \delta^p.$$

Recall that we chose α_0 such that $W_p(l_{\alpha_0}(d), d_\emptyset) \leqslant \delta$. Therefore,

$$2^{-p} \sum_{x \in s_1} \operatorname{pers}(\gamma(x))^p \leqslant W_p(l_{\alpha_0}(d), d_\emptyset)^p \leqslant \delta^p.$$

We then have

$$\sum_{x \in d_{n_k}} \|x - \gamma(x)\|_\infty^p = \sum_{x \in s_0} \|x - \gamma(x)\|_\infty^p + \sum_{x \in s_1} \|x - \gamma(x)\|_\infty^p + \sum_{x \in s_2} \|x - \gamma(x)\|_\infty^p$$

$$\geqslant \sum_{x \in s_0} \|x - \bar{\gamma}(x)\|_\infty^p + 2^{-p} \sum_{x \in s_2} \operatorname{pers}(\gamma(x))^p + (2.5)^p \delta^p - 2^{-p}\delta^p$$

$$\geqslant \sum_{x \in s_0} \|x - \bar{\gamma}(x)\|_\infty^p + 2^{-p} \sum_{x \in s_1} \operatorname{pers}(\gamma(x))^p$$

$$+ 2^{-p} \sum_{x \in s_2} \operatorname{pers}(\gamma(x))^p + ((2.5)^p - 1 - 2^{-p})\delta^p$$

$$\geqslant \sum_{x \in \bar{d}_{n_k}} \|x - \bar{\gamma}(x)\|_\infty^p + \delta^p.$$

As in the previous case, this implies that

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

where

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

Therefore,

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

and consequently

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

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 $\mathbb{E}_\mathcal{P} \neq \emptyset$.*

Proof. Let $S \subset D_p$ be the support of $\mathcal{P}$ and let $\{d_n\}_{n=1}^\infty$ be a sequence in D_p such that $F_\mathcal{P}(d_n) \to \mathrm{Var}_\mathcal{P}$. We 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 obtain

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

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_{n_k} and subdiagrams $\bar{d}_{n_k} \subset d_{n_k}$ such that

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

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

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 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 26. *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 $r \geqslant R$ we have $\mu(B^r(x_0)) \leqslant C r^{-q}$, where $B^r(x_0) = \{x \in \mathbb{X} \mid \rho(x, x_0) \geqslant r\}$.*

We will 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 $\{d_n\} \subset D_p$, $n \in \mathbb{N}$, 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 $\bar{d}_{n_k} \subset d_{n_k}$, $k \in \mathbb{N}$, we have*

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

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

Proof. Let $C > 0$ and $R > 0$ be such that $\mathcal{P}(B^r(d_\emptyset)) \leqslant C r^{-q}$, $r \geqslant R$. Take $M \in \mathbb{N}$ such that $M > R$, $W_p(d_n, d_\emptyset) \leqslant M$ (and hence $W_p(\bar{d}_{n_k}, d_\emptyset) \leqslant M$), $M^{-s} < \varepsilon/(8C)$, and

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

Denote

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

We have

$$\int_{B^M(d_\emptyset)} W_p(\bar{d}_{n_k}, d)^2 \, \mathrm{d}\mathcal{P}(d) \leqslant \int_{B^M(d_\emptyset)} \left(W_p(\bar{d}_{n_k}, d_\emptyset) + W_p(d_\emptyset, d) \right)^2 \, \mathrm{d}\mathcal{P}(d),$$

$$\leqslant \int_{B^M(d_\emptyset)} (2W_p(d_\emptyset, d))^2 \, \mathrm{d}\mathcal{P}(d).$$

Note that

$$\int_{B^M(d_\emptyset)} \left(2W_p(d_\emptyset, d) \right)^2 \, \mathrm{d}\mathcal{P}(d) = 4 \sum_{m \geqslant M} \int_{B^{m,m+1}(d_\emptyset)} W_p(d_\emptyset, d)^2 \, \mathrm{d}\mathcal{P}(d)$$

$$\leqslant 4 \sum_{m \geqslant M} (m+1)^2 (\mathcal{P}(B^m(d_\emptyset)) - \mathcal{P}(B^{m+1}(d_\emptyset))).$$

Denote the right hand side of the above expression by L. Then

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

Finally,

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

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

$$\int_{S^c \cap B_M(d_\emptyset)} W_p(\bar{d}_{n_k}, d)^2 \, \mathrm{d}\mathcal{P}(d) \leqslant \int_{S^c \cap B_M(d_\emptyset)} \left(W_p(\bar{d}_{n_k}, d_\emptyset) + W_p(d_\emptyset, d) \right)^2 \, \mathrm{d}\mathcal{P}(d)$$

$$\leqslant 4M^2 \mathcal{P}(S^c \cap B_M(d_\emptyset)) < \frac{\varepsilon}{2M^{s-2}}.$$

Combining the two results, we obtain

$$\int_{D_p} W_p(\bar{d}_{n_k}, d)^2 \, \mathrm{d}\mathcal{P}(d) \leqslant \int_{S \cap B_M(d_\emptyset)} W_p(\bar{d}_{n_k}, d)^2 \, \mathrm{d}\mathcal{P}(d) + \int_{S^c \cap B_M(d_\emptyset)} W_p(\bar{d}_{n_k}, d)^2 \, \mathrm{d}\mathcal{P}(d)$$

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

To prove the last statement of the lemma, note that $\mathcal{P}(S) > 1 - \varepsilon/8$ and $\mathcal{P}(B^M(d_\emptyset)) < \varepsilon/8$. Since $\mathcal{P}(S) \leqslant \mathcal{P}(S \cap B_M(d_\emptyset)) + \mathcal{P}(B^M(d_\emptyset))$, we obtain $\mathcal{P}(S \cap B_M(d_\emptyset)) > 1 - \varepsilon/4$. □

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 $F_\mathcal{P}(d_n) \to \mathrm{Var}_\mathcal{P}$. We 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 $\mathcal{P}(S_0) \geqslant 0.5$. Then, $w_n = \inf_{d \in S_0} W_p(d_n, d)$ is not bounded. Thus, as $n \to \infty$ we obtain

$$F_{\mathcal{P}}(d_n) = \int_{D_p} W_p^2(d_n, d)\, \mathrm{d}\mathcal{P}(d) \geqslant \int_{S_0} W_p^2(d_n, d)\, \mathrm{d}\mathcal{P}(d) \geqslant w_n^2 \mathcal{P}(S_0) \to \infty,$$

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 δ from lemma 23. Take $\varepsilon > 0$ such that $\varepsilon < (2^{\frac{2}{s}} - 1)2^{1-s}\delta_0^s(1 - \varepsilon/4)$. By lemma 27 the inequality

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

holds for the subsequences of subdiagrams $\bar{d}_{n_k}$ from lemma 23. Now, by lemma 23 we have

$$\int_{S \cap B_M(d_\emptyset)} W_p(\bar{d}_{n_k}, d)^2\, \mathrm{d}\mathcal{P}(d) \leqslant \int_{S \cap B_M(d_\emptyset)} W_p(d_{n_k}, d)^2\, \mathrm{d}\mathcal{P}(d) - \varepsilon_0 \mathcal{P}(S \cap B_M(d_\emptyset)),$$

where

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

By lemma 27 we have $\mathcal{P}(S \cap B_M(d_\emptyset)) > 1 - \varepsilon/4$. Therefore,

$$\int_{D_p} W_p(\bar{d}_{n_k}, d)^2\, \mathrm{d}\mathcal{P}(d) \leqslant \int_{D_p} W_p(d_{n_k}, d)^2\, \mathrm{d}\mathcal{P}(d) - \frac{\varepsilon_0(1 - \varepsilon/4)}{2}.$$

Taking the infimum over k, we obtain

$$\mathrm{Var}_{\mathcal{P}} \leqslant \mathrm{Var}_{\mathcal{P}} - \frac{\varepsilon_0(1 - \varepsilon/4)}{2},$$

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 $\mathcal{P}(\theta \mid D)$, 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, as considered in [29]. This distribution 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:

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

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

$$\mathrm{Lik}(Z; \theta) \equiv f_\theta(Z),$$

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), $\mathcal{B}(\mathbb{X})$ is the Borel σ-algebra on $\mathbb{X}$ and $\mathcal{P}_\theta$ is the probability measure parameterized by θ. The observed point cloud data $Z \equiv \{X_1, \ldots, X_n\}$, where $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathcal{P}_\theta$, 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, respectively. 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 $\rho : K(\mathbb{X}) \to \mathrm{Lip}(\mathbb{X})$, where $\mathrm{Lip}(\mathbb{X})$ denotes the space of Lipschitz functions on $\mathbb{X}$ 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 $S \in K(\mathbb{X})$ 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 [16]. 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

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

Note that if $\mathbb{X}$ is 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 $p > k$ 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 $\mathcal{P}_D$ 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 $\mathcal{P}(\theta \mid D)$:

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

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 [30]. This idea of a substitution likelihood goes back to Jeffreys [31] and a basic question in TDA is what properties of $\mathcal{P}_\theta$ are preserved by $\mathcal{P}_D$.

4.3. An example

Assume that we obtain m point samples from the object $\mathcal{O}_1$ (for example a torus) and n point samples from an object $\mathcal{O}_2$ (a double torus). For each point sample we obtain persistence diagrams resulting in two sets of diagrams $\{x_1, \ldots, x_m\} \in D_p$ for $\mathcal{O}_1$ and $\{y_1, \ldots, y_n\} \in D_p$ for $\mathcal{O}_2$. 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 $\{y_1, \ldots, y_n\}$, we can use a variation of kernel density estimation [32, 33] to provide the following density estimates for diagrams corresponding to the objects $\mathcal{O}_1$ and $\mathcal{O}_2$, respectively:

$$\hat{p}(x|\mathcal{O}_1) = \frac{1}{m\kappa_\tau} \sum_{i=1}^{m} e^{-W_p^2(x,x_i)/\tau}, \qquad \hat{p}(x|\mathcal{O}_2) = \frac{1}{n\kappa_\tau} \sum_{i=1}^{n} e^{-W_p^2(x,y_i)/\tau};$$

here, $\tau > 0$ is the bandwidth parameter that controls the smoothness of the density, and κ_τ is a normalizing constant. If we assume that the objects have prior probabilities $\pi_1 = \Pr(\mathcal{O}_1)$ and $\pi_2 = \Pr(\mathcal{O}_2)$, we can use Bayes' rule to compute the posterior probability of membership in class one (the torus) given the diagram z, $\hat{p}(\mathcal{O}_1 \mid z)$:

$$\hat{p}(\mathcal{O}_1 \mid z) = \frac{\hat{p}(z \mid \mathcal{O}_1)\pi_1}{\hat{p}(z)} = \frac{\hat{p}(z \mid \mathcal{O}_1)\pi_1}{\hat{p}(z \mid \mathcal{O}_1)\pi_1 + \hat{p}(z \mid \mathcal{O}_2)\pi_2};$$

note that we do not need to compute the normalization constant κ_τ to compute the posterior probability since κ_τ 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 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 [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.

Acknowledgments

References

[1] Edelsbrunner H and Harer J 2010 *Computational Topology: An Introduction* (Providence, RI: American Mathematical Society)
[2] de Silva V and Carlsson G 2004 Topological estimation using witness complexes *Symp. on Point-based Graphics* pp 157–66
[3] Edelsbrunner H and Harer J 2008 Persistent homology—a survey *Discrete and Computational Geometry. Twenty Years Later (Contemporary Mathematics* vol 453*)* ed J E Goodman, J Pach and R Pollack (Providence, RI: American Mathematical Society) pp 257–82
[4] Dudley R M 2002 *Real Analysis and Probability* (Cambridge: Cambridge University Press)
[5] Penrose M D 2003 *Random Geometric Graphs* (New York: Oxford University Press)
[6] Penrose M D and Yukich J E 2001 Central limit theorems for some graphs in computational geometry *Ann. Appl. Probab.* **11** 1005–41
[7] Kahle M 2011 Random geometric complexes arXiv:0910.1649
[8] Kahle M 2009 Topology of random clique complexes *Discrete Math.* **309** 1658–71
[9] Lunagómez S, Mukherjee S and Wolpert R 2009 Geometric representations of hypergraphs for prior specification and posterior sampling arXiv:0912.3648
[10] Kahle M and Meckes E 2010 Limit theorems for Betti numbers of random simplicial complexes arXiv:1009.4130v3 [math.PR]
[11] Niyogi P, Smale S and Weinberger S 2008 Finding the homology of submanifolds with high confidence from random samples *Discrete Comput. Geom.* **39** 419–41
[12] Niyogi P, Smale S and Weinberger S 2011 A topological view of unsupervised learning from noisy data *SIAM J. Comput.* **40** 646–63
[13] Chazal F, Cohen-Steiner D and Lieutier A 2009 A sampling theory for compact sets in euclidean space *Discrete Comput. Geom.* **41** 461–79
[14] Bendich P, Mukherjee S and Wang B 2010 Towards stratification learning through homology inference arXiv:1008.3572
[15] Bubenik P, Carlsson G, Kim P T and Luo Z-M 2010 Statistical topology via morse theory, persistence, and nonparametric estimation *Algebraic Methods in Statistics and Probability II* (*Contemporary Mathematics* vol 516) pp 75–92
[16] Chazal F, Cohen-Steiner D and M'erigot Q 2010 Geometric inference for measures based on distance functions http://hal.inria.fr/inria-00383685/
[17] Cohen-Steiner D, Edelsbrunner H, Harer J and Mileyko Y 2010 Lipschitz functions have l_p -stable persistence *Found. Comput. Math.* **10** 127–39
[18] Wasserstein L N 1969 Markov processes over denumerable products of spaces describing large systems of automata *Probl. Inf. Transm.* **5** 47–52
[19] Villani C 2003 *Topics in Optimal Transportation* (Providence, RI: American Mathematical Society)
[20] Kantorovich L V 1942 On the translocation of masses *C. R. (Doklady) Acad. Sci. URSS* **37** 199–201
[21] Peleg S, Werman M and Rom H 1989 A unified approach to the change of resolution: space and gray-level *IEEE Trans. Pattern Anal. Mach. Intell.* **11** 739–42
[22] Rubner Y, Tomasi C and Guibas L J 2000 The earth mover's distance as a metric for image retrieval *Int. J. Comput. Vis.* **40** 99–121
[23] Rachev S T 1991 *Probability Metrics and the Stability of Stochastic Models* (New York: Wiley)
[24] Bolley F 2008 Separability and completeness for the Wasserstein distance *Séminaire de probabilités XLI* (*Lecture Notes in Mathematics* vol 1934) pp 371–7

[25] Fréchet M 1944 L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application á la moyenne d'un élément aléatoire de nature quelconque *Rev. Sci.* **82** 483–512

[26] Fréchet M 1948 Les éléments aléatories de nature quelconque dans un espace distancié *Ann. Inst. Henri Poincare* **10** 215–310

[27] Karcher H 1977 Riemannian center of mass and mollifier smoothing *Commun. Pure Appl. Math.* **30** 509–41

[28] Kendall W S 1990 Probability, convexity, and harmonic maps with small image: I. Uniqueness and fine existence *Proc. London Mathematical Society (Third Series)* vol 61 pp 371–406

[29] Bubenik P and Kim P T 2007 A statistical approach to persistent homology *Homology Homotopy Appl.* **9** 337–62

[30] Berger J O and Wolpert R L 1984 *The Likelihood Principle (IMS Lecture Notes—Monograph Series* vol 6*)* (Beachwood, OH: Institute of Mathematical Statistics)

[31] Jeffreys H 1961 *Theory of Probability* (Oxford: Clarendon)

[32] Rosenblatt M 1956 Remarks on some nonparametric estimates of a density function *Ann. Math. Stat.* **27** 832–7

[33] Parzen E 1962 On estimation of a probability density function and mode *Ann. Math. Stat.* **33** 1065–76

[34] Gamble J and Heo G 2010 Exploring uses of persistent homology for statistical analysis of landmark-based shape data *J. Multivariate Anal.* 2184—99