WEIGHTED PERSISTENT HOMOLOGY SUMS OF RANDOM ČECH COMPLEXES

BENJAMIN SCHWEINHART

ABSTRACT. We study the asymptotic behavior of random variables of the form

$$E_\alpha^i\left(x_1,\ldots,x_n\right)=\sum_{(b,d)\in PH_i(x_1,\ldots,x_n)}\left(d-b\right)^\alpha$$

where $\left\{x_j\right\}_{j\in\mathbb{N}}$ are i.i.d. samples from a probability measure on a triangulable metric space, and $PH_i\left(x_1,\ldots,x_n\right)$ denotes the i-dimensional reduced persistent homology of the Čech complex of $\left\{x_1,\ldots,x_n\right\}$. These quantities are a higher-dimensional generalization of the α-weighted sum of a minimal spanning tree; we seek to prove analogues of the theorems of Steele [16] and Aldous and Steele [2] in this context.

As a special case of our main theorem, we show that if $\left\{x_j\right\}_{j\in\mathbb{N}}$ are distributed independently and uniformly on the m-dimensional Euclidean sphere, $\alpha<m$, and $0\leq i<n$, then there are real numbers γ and Γ so that

$$\gamma\leq\lim_{n\to\infty}n^{-\frac{m-\alpha}{m}}E_i^\alpha\left(x_1,\ldots,x_n\right)\leq\Gamma$$

in probability. More generally, we prove results about the asymptotics of the expectation of E_α^i for points sampled from a locally bounded probability measure on a space that is the bi-Lipschitz image of an $m-$dimensional Euclidean simplicial complex.

1. Introduction

We are interested in random variables of the form

$$E_\alpha^i\left(x_1,\ldots,x_n\right)=\sum_{(b,d)\in PH_i(x_1,\ldots,x_n)}\left(d-b\right)^\alpha$$

where $\left\{x_j\right\}_{j\in\mathbb{N}}$ are independent samples drawn from a probability measure on a triangulable metric space, and $PH_i\left(x_1,\ldots,x_n\right)$ denotes the i-dimensional reduced persistent homology of the Čech complex of $\left\{x_1,\ldots,x_n\right\}$. The special case $i=0$ is,

Date: July 2018.

under a different guise, already the subject of an expansive literature in probabilistic combinatorics; $E^0_\alpha(\mathbf{x})$ gives the α-weight of the minimal spanning tree on a finite subset of a metric space $\mathbf{x}$, $T(\mathbf{x})$:

$$E^0_\alpha(\mathbf{x}) = 2^{-\alpha} \sum_{e \in T(\mathbf{x})} |e|^\alpha$$

In 1988, Steele [16] showed the following:

Theorem 1 (Steele)**.** *Let μ is a compactly supported probability distribution on $\mathbb{R}^m$, and let $\{x_n\}_{n\in\mathbb{N}}$ be i.i.d. samples from μ. If $\alpha < m$,*

$$\lim_{n\to\infty} n^{-\frac{m-\alpha}{m}} E^0_\alpha(x_1,\ldots,x_n) \to c(\alpha,m) \int_{\mathbb{R}^d} f(x)^{(m-\alpha)/m}$$

with probability one, where $f(x)$ is the probability density of the absolutely continuous part of μ, and $c(\alpha,m)$ is a positive constant that depends only on α and m.

In 1992, Aldous and Steele [2] showed that if $\{x_i\}_{i\in\mathbb{N}}$ sampled independently from the uniform distribution on the unit cube in $\mathbb{R}^m$, then

$$\lim_{n\to\infty} E^m_\alpha(x_1,\ldots,x_n) \to c(d,d)$$

in the L^2 sense. Under the same hypotheses, Kesten and Lee proved the following central limit theorem in 1996 [12]:

$$\frac{E^0_\alpha(X_1,\ldots,X_n) - \mathbb{E}\left(E^0_\alpha(X_1,\ldots,X_n)\right)}{n^{m-2\alpha}2d} \to N\left(0,\sigma^2_{\alpha,d}\right)$$

in distribution, for any $\alpha > 0$. Here, we take the first step toward a higher-dimensional generalization of these celebrated results.

Another special case of $E^i_\alpha(\mathbf{x})$ — $\alpha = 1$ — gives the total lifetime persistence of $\mathbf{x}$. Random variables of the form $E^i_1(\mathbf{x})$ have been investigated by Hiraoka and Shirai [11] in the context of Linial—Meshulam processes. They showed that if X is sampled from the m-Linial—Meshulam process then

$$\mathbb{E}\left(E^{m-1}_1(X)\right) \in O\left(n^{m-1}\right)$$

which is a higher-dimensional generalization of Frieze's $\zeta(3)$-theorem for Erdós—Rényi random graphs [10]. Also, Adams et al. [1] studied the behavior of the lifetime persistence of random measures on Euclidean space, performing computational experiments and conjecturing the existence of a limit function capturing finer properties of the persistent homology.

The properties of $E_{\alpha}^{i}(\mathbf{x})$ for general i and n have until now, as far as we know, not been studied in a probabilistic context (see the note at the end of the introduction). However, some work has been done in the extremal context. In 2010, Cohen-Steiner et al. [6] showed that if M is the bi-Lipschitz image of an m-dimensional simplicial complex and $\alpha > m$, then $E_i^{\alpha}(X)$ is uniformly bounded for $X \subset M$. We use their results to prove the upper bounds in Section 2. Furthermore, in our previous paper [13] we related the upper box dimension of a subset X of a metric space to the behavior of $E_{\alpha}^{i}(Y)$ for extremal subsets $Y \subset X$. We will say more about the relation of this to the present work in Section 1.2.

1.1. **Our Results.** The following are special cases of our main theorem:

Theorem 2. *Let $\left\{x_j\right\}_{j\in\mathbb{N}}$ be be distributed independently and uniformly on the S^n. If $\alpha < m$, $0 \leq i < n$, and persistent homology is taken with respect to the intrinsic metric on S^n,*

$$\gamma \leq \lim_{n\to\infty} n^{-\frac{m-\alpha}{m}} E_i^{\alpha}\left(x_1,\ldots,x_n\right) \leq \Gamma$$

in probability, where γ and Γ are constants that depend on μ and α.

Furthermore, there exists a $D \in \mathbb{R}$ so that

$$\lim_{n\to\infty} \frac{1}{\log(n)} E_i^{m}\left(x_1,\ldots,x_n\right) \leq D$$

in probability.

Theorem 3. *Let $\left\{x_j\right\}_{j\in\mathbb{N}}$ be be distributed independently and uniformly on an m-dimensional Euclidean ball. If $\alpha < m$, $0 \leq i < n$,*

$$\gamma \leq \lim_{n\to\infty} n^{-\frac{m-\alpha}{m}} \mathbb{E}\left(E_i^{\alpha}\left(x_1,\ldots,x_n\right)\right) \leq \Gamma$$

where γ and Γ are constants that depend on μ and α. In fact, the lower bound holds in probability.

Furthermore, there exists a $D \in \mathbb{R}$ so that

$$\lim_{n\to\infty} \frac{1}{\log(n)} \mathbb{E}\left(E_i^{m}\left(x_1,\ldots,x_n\right)\right) \leq D$$

We show a stronger result for compactly supported probability measures on $\mathbb{R}^2$ that are locally bounded:

Definition 1. *A probability measure μ on $\mathbb{R}^m$ is **locally bounded** if there is a $A \subset \mathbb{R}^m$ with positive volume and real numbers $a_1 \geq a_0 > 0$ so that*

$$a_0 \, vol(B) \leq \mu(B) \leq a_1 \, vol(B)$$

for all Borel sets $B \subset A$.

Theorem 4. *Let μ is a compactly supported, locally bounded probability measure on $\mathbb{R}^2$, and let $\{x_n\}_{n \in \mathbb{N}}$ be i.i.d. samples from μ. If $\alpha < m$,*

$$\gamma \leq \lim_{n \to \infty} n^{-\frac{m-\alpha}{m}} E_\alpha^1 (x_1, \ldots, x_n) \leq \Gamma$$

in probability. In fact, the upper bound holds with probability one.

Furthermore, there exists a constant D so that

$$\lim_{n \to \infty} \frac{1}{\log(n)} E_2^1 (x_1, \ldots, x_n) \leq D$$

with probability one

More generally, we prove results for locally bounded probability measures on spaces that are the bi-Lipschitz image of a compact, m-dimensional Euclidean simplicial complex:

Definition 2. *Let M be the bi-Lipschitz image of a compact m-dimensional Euclidean simplicial complex Δ_M under a map ϕ_M. A probability measure μ on M is **locally bounded** if there exists a subset $A \subset \Delta_M$ with positive m-dimensional volume, and real numbers $a_1 \geq a_0 > 0$ so that*

$$a_0 \frac{vol(B)}{vol(\Delta_M)} \leq \mu(\phi_M(B)) \leq a_1 \frac{vol(B)}{vol(\Delta_M)}$$

for all Borel sets $B \subseteq A$.

For example, a the uniform measure on a m-dimensional Riemannian manifold is locally bounded, as is any measure that is locally bounded with respect to the Riemannian volume.

While there exist metric spaces M with point sets $\left\{x_j\right\}_{j \in \mathbb{N}}$ so that

$$|PH_i(x_1, \ldots, x_n)| \neq O(n)$$

this is thought to be somewhat pathological behavior [13].

Definition 3. *A probability measure μ on a triangulable metric space has **linear** PH_i **expectation** if*

$$\mathbb{E}\left(\left|PH_i\left(\{x_1,\ldots,x_n\}\right)\right|\right) \in O\left(n\right)$$

*Similarly, μ has **linear** PH_i **variance** if*

$$\mathbb{E}\left(\left(\left|PH_i\left(\{x_1,\ldots,x_n\}\right)\right| - \mathbb{E}\left(\left|PH_i\left(\{x_1,\ldots,x_n\}\right)\right|\right)\right)^2\right) \in O\left(n\right)$$

For example, the uniform measure on a Euclidean ball [8] and any positive, continuous probability density on the Euclidean n-sphere [17] has linear PH_i expectation. It is more difficult to prove that a probability measure has linear PH_i variance. As far as we are aware, this is only known for probability measures on $\mathbb{R}^2$ and the uniform measure on the n-dimensional Euclidean sphere [17] (see Equation 1 and Proposition 3).

Theorem 5. *Let M be the bi-Lipschitz image of an m-dimensional Euclidean simplicial complex, and $0 \leq i < m$. If μ is a locally bounded probability measure on M, there are real numbers $0 < \gamma < \Gamma$ so that*

$$\gamma n^{\frac{m-\alpha}{m}} \leq \mathbb{E}\left(E_\alpha^i\left(x_1,\ldots,x_n\right)\right) \leq \Gamma\,\mathbb{E}\left(\left|PH_i\left(\{x_1,\ldots,x_n\}\right)\right|\right)^{\frac{m-\alpha}{m}}$$

for all sufficiently large n. In particular, if μ has linear PH_i expectation, there is a real number Γ_0 so that

$$\gamma \leq \lim_{n\to\infty} n^{-\frac{m-\alpha}{m}}\,\mathbb{E}\left(E_\alpha^i\left(x_1,\ldots,x_n\right)\right) \leq \Gamma_0$$

The lower bound holds in probability, and the upper bound does if μ has linear PH_i variance.

Furthermore, there exists a real number D so that

$$\mathbb{E}\left(E_n^i\left(x_1,\ldots,x_n\right)\right) \leq D\,\log\left(\mathbb{E}\left(\left|PH_i\left(x_1,\ldots,x_n\right)\right|\right)\right)$$

where analogously sharper statements hold if μ has linear PH_i expectation or variance.

We prove the upper bound in Proposition 2 and the lower bound in Proposition 5.

After completion of this manuscript, we became aware that Divol and Polonik [7] independently and concurrently proved a sharper result for the persistent homology of points sampled from bounded, absolutely continuous probability densities on $[0,1]^m$. We believe this manuscript is still useful in that the proofs are largely self-contained, and the methods are applicable to other situations. In a later paper [14], we use them to study the behavior of $E_\alpha^i(x_1,\ldots,x_n)$ for i.i.d. points sampled from a measure supported on a set of fractional dimension.

1.2. PH-dimension.

In [13], we defined a family of persistent homology dimensions for a subset X of a metric space M in terms of the extremal behavior of $E_\alpha^i(Y)$ for subsets $\mathbf{x}$ of X:

$$\dim_{PH}^i(X) = \inf\left\{\alpha : E_\alpha^i(\mathbf{x}) < C\ \forall\, \mathbf{x} \subset X\right\}$$

That is, $E_\alpha^i(\mathbf{x})$ is uniformly bounded for all $\alpha > \dim_{PH}^i(X)$, but not for $\alpha < \dim_{PH}^i(X)$. Note that the persistent homology is taken with respect M. Our results were the first rigorously relating persistent homology to a classically defined fractal dimension, the upper box dimension, but the definition is difficult to compute with in practice. Here, we define a similar notion of fractal dimension for measures on a metric space that may be more computable in practice:

Definition 4. *The PH_i-dimension of a probability measure on a a triangulable metric space is*

$$dim_{PH}^i(\mu) = \sup\left\{\alpha : \limsup_{n\to\infty} \mathbb{E}\left(E_\alpha^i(x_1,\ldots,x_n)\right) = \infty\right\}$$

Clearly, $\dim_{PH}^i(\mu) \leq \dim_{PH}^i(\text{supp}(\mu))$. As a corollary to our main theorem, we show:

Theorem 6. *Let M be the bi-Lipschitz image of a compact m-dimensional Euclidean simplicial complex, and $0 \leq i < m$. If μ is a locally bounded probability measure on M,*

$$dim_{PH}^i(\mu) = m$$

1.3. Persistent Homology.

If X is a bounded subset of a triangulable metric space M, let X_ϵ denote the ϵ-neighborhood of X :

$$X_\epsilon = \{x \in M : d(x,X) < \epsilon\}$$

Also, let $H_i(X)$ be the reduced homology of X, with coefficients in a field k. The **persistent homology** of X is the product $\prod_{\epsilon>0} H_i(X_\epsilon)$, together with the inclusion maps $i_{\epsilon_0,\epsilon_1} : H_i(X_{\epsilon_0}) \to H_i(X_{\epsilon_1})$ for $\epsilon_0 < \epsilon_1$. The structure of persistent homology is captured by a set of intervals, which we refer to as $PH_i(X)$ [18]. These intervals represent how the topology of X_ϵ changes as ϵ increases. Under certain finiteness hypotheses — which are satisfied if X is a finite point set — $PH_i(X)$ is the unique set of intervals so that the rank of $i_{\epsilon_0,\epsilon_1}$ equals the number of intervals containing (ϵ_0, ϵ_1) [5].

If X is finite $PH_i(X)$ is the same as the persistent homology of the Čech complex of X. Note that this depends on the ambient metric space. Here, if "μ is a probability measure on M and $\{x_j\}_{j\in\mathbb{N}}$ are sampled from μ," then $PH_i(x_1,\ldots,x_n)$ is the persistent homology with ambient metric space M. All questions we study here would also be interesting in the context of the Vietoris—Rips Complex.

1.4. **Notation.** In the following, an m-space will be the bi-Lipschitz image of a compact m-dimensional Euclidean simplicial complex. Also, if the measure μ is obvious from the context, $\{x_j\}_{j\in\mathbb{N}}$ will denote a collection of independent random variables with common distribution μ. Also, $\mathbf{x}_n$ will be shorthand for $\{x_1,\ldots,x_n\}$ and $\mathbf{x}$ will denote a finite point set.

2. Upper Bounds

Our strategy to prove an upper bound for the asymptotics of $E_\alpha^i(\{x_1,\ldots,x_n\})$ will be to bound the number and length of the persistent homology intervals in terms of the number of simplices in a triangulation of the ambient metric space. The approach is similar to that in our earlier paper [13].

2.1. **Preliminaries.** We require the following result, which is proven by bounding the number of persistent homology intervals of a triangulable metric space of length greater than δ in terms of the number of simplices in a triangulation of mesh δ:

Proposition 1. *(Cohen-Steiner, Edelsbrunner, Harer, and Mileyko* [6]*) Let M be an m-space. There exists a real number C_0 so that for any $0\le i<m$, $X\subseteq M$, and $\delta>0$,*

$$|\{(b,d)\in PH_i(X): d-b>\delta\}|\le C_0\,\delta^{-m}$$

We use this result to bound $E_\alpha^i(\mathbf{x})$ in terms of the number of PH_i intervals of $\mathbf{x}$:

Lemma 1. *Let M be an m-space, $\alpha<m$, and $i\in\mathbb{N}$. There exists a real number $C_1>0$ so that*

$$E_\alpha^i(X)\le C_1\,|PH_i(X)|^{\frac{m-\alpha}{m}}$$

for all $X\subseteq M$. Furthermore, there exists a real number $D_1>0$ so that

$$E_m^i(X)\le D_1\log(|PH_i(X)|)$$

for all $X \subseteq M$.

Proof. Dilating M by a factor r multiplies $E_\alpha^i(X)$ by r^α, so we may assume without loss of generality that the diameter of M is less than one. Let $n = |PH_i(X)|$ and

$$I_k = \left\{(b,d) \in PH_i(X) : \frac{1}{2^{k+1}} < d - b \leq \frac{1}{2^k}\right\}$$

Also, let C_0 be as in Proposition 1 so

$$|I_k| \leq C_0 2^{mk}$$

The largest C_0 intervals of $PH_i(X)$ each have length less than or equal to 2^0, the next largest $C_0 2^m$ intervals have length less than or equal to 2^{-1}, and so on. It follows that if

$$l = \left\lceil \frac{\log_2(2n/C_0)}{m} \right\rceil$$

then

$$n \leq \sum_{k=0}^{l} C_0 2^{mk}$$

and

$$E_\alpha^i(X) \leq \sum_{k=0}^{l} C_0 2^{mk} \left(\frac{1}{2^k}\right)^\alpha$$

If $\alpha = m$, the previous inequality becomes

$$E_\alpha^i(X) \leq C_0 l = O(\log(n))$$

as desired.

Otherwise, if $\alpha < m$,

$$\begin{aligned} E_{\alpha}^{i}\left(X\right) \leq & \\ & \sum_{k=0}^{l} C_0 2^{k(m-\alpha)} \\ = & \; C_0 \frac{2^{(m-\alpha)(l+1)} - 1}{2^{m-\alpha} - 1} \\ \leq & \; \frac{C_0}{2^{m-\alpha} - 1} 2^{(m-\alpha)(l+1)} \\ \leq & \; \frac{C_0}{2^{m-\alpha} - 1} 2^{\left(\frac{\log_2(2n/C_0)}{m} + 2\right)(m-\alpha)} \\ = & \; C_1 n^{\frac{m-\alpha}{m}} \end{aligned}$$

where $C_1 = \frac{C_0 4^{m-\alpha}}{2^{m-\alpha}-1}$. □

2.2. **The Upper Bound.** The upper bound in our main theorem now follows immediately from Jensen's inequality, as the function $f\left(x\right) = x^{\frac{m-\alpha}{m}}$ is concave for $0 < \alpha \leq m$:

Proposition 2. *Let M be an m-space, let i be a natural number less than m, and let μ be a locally bounded probability measure on M. For all $0 < \alpha < m$ there exists a real number $C > 0$ so that*

$$\mathbb{E}\left(E_{\alpha}^{i}\left(x_1, \ldots, x_n\right)\right) \leq C\,\mathbb{E}\left(\left|PH_i\left(x_1, \ldots, x_n\right)\right|\right)^{\frac{m-\alpha}{m}}$$

In particular, if μ has linear PH_i expectation and linear PH_i variance then there is a $C' > 0$ so that

$$\lim_{n \to \infty} n^{-\frac{m-\alpha}{m}} E_{\alpha}^{i}\left(x_1, \ldots, x_n\right) \leq C'$$

in probability.

Furthermore, there exists a real number D so that

$$\mathbb{E}\left(E_{m}^{i}\left(x_1, \ldots, x_n\right)\right) \leq D \log\left(\left|PH_i\left(x_1, \ldots, x_n\right)\right|\right)$$

In particular, if μ has linear PH_i expectation and linear PH_i variance then there is a $D' > 0$ so that

$$\lim_{n \to \infty} \frac{1}{\log\left(n\right)} E_{m}^{i}\left(x_1, \ldots, x_n\right) \leq D'$$

in probability.

Proof. Let $\mathbf{x}$ be a finite subset of B, and let C_1 be as in Lemma 1. If $\alpha < m$,

$$\begin{aligned}\mathbb{E}\left(E_\alpha^i(\mathbf{x})\right) \leq & \\ & \mathbb{E}\left(C_1 \left|PH\right|_i(\mathbf{x})^{\frac{m-\alpha}{m}}\right) && \text{by Lemma 1} \\ \leq & \; C_1\mathbb{E}\left(\left|PH\right|_i(\mathbf{x})\right)^{\frac{m-\alpha}{m}} && \text{by Jensen's inequality}\end{aligned}$$

as desired. If μ has linear PH_i expectation and linear PH_i-variance, Chebyshev's Inequality implies that

$$\lim_{n\to\infty} \left|PH_i(x_1,\ldots,x_n)\right|/n \leq C_2$$

in probability, for some $C_2 > 0$, and the desired statement follows from Lemma 1.

The proof for the case $\alpha = m$ is similar. □

2.3. **Sharper Upper Bounds.** Our sharper upper bounds in Theorems 2 and 3 follow from the fact that if $\{x_1,\ldots,x_n\}$ is a finite subset of $\mathbb{R}^m$ of S^m in general position then

$$\left|PH_i(x_1,\ldots,x_n)\right| \leq \left|DT(x_1,\ldots,x_n)\right| \tag{1}$$

where $DT(x_1,\ldots,x_n)$ is the number of simplices of the Delaunay triangulation on $\{x_1,\ldots,x_n\}$. In fact, the Alpha complex is a filtration on the simplices of the Delaunay triangulation that is homotopy equivalent to the ϵ-neighborhood filtration of the points $\{x_1,\ldots,x_n\}$ [9]. This construction is usually defined for points in Euclidean space, but easily extends to points on the m-sphere, in which case the Delaunay triangulation is the spherical convex hull of the points.

Proposition 3. *If B be a bounded subset of $\mathbb{R}^m$*

$$E_\alpha^i(x_1,\ldots,x_n) = O\left(n^{\lfloor\frac{m+1}{2}\rfloor\frac{m-\alpha}{m}}\right)$$

for any general position point set $\{x_1,\ldots,x_n\}$ contained in B.

Proof. The Upper Bound Theorem [15] implies that if $X \subset \mathbb{R}^m$ then

$$\left|(DT)(x_1,\ldots,x_n)\right| = O\left(n^{\lfloor\frac{m+1}{2}\rfloor}\right)$$

The desired statement follows immediately from Lemma 1 and Equation 1 □

3. Lower Bounds

Our strategy to prove lower bounds for the asymptotics of weighted PH-sums is to study collections of sets whose persistent homology obeys a super-additivity property. We define certain "cubical occupancy events" giving rise to such collections, and prove that they occur with positive probability for sets of i.i.d. points drawn from a locally bounded probability measure on an m-space. We bootstrap these results by subdividing a subset of an m-dimensional cube into many small sub-cubes. This bootstrapping argument is similar to the one we used to prove a lower bound for PH_i dimension in our previous paper [13].

In the following, fix $0 \leq i < m$.

3.1. **Super-additivity for Persistent Homology.** Persistent homology does not in general obey a super-additivity property, but we can define a subclass of sets whose persistent homology does. If X and T are subsets of a triangulable metric space and $b < d$, let $M_{X,T}\left(b,d\right)$ be the rank of the homomorphism on homology induced by the inclusion

$$X_b \hookrightarrow X_d \hookrightarrow X_d \cup T_d$$

where X_ϵ denotes the ϵ-neighborhood of X. Note that

$$M_{X,C}\left(b,d\right) \leq N_X\left(b,d\right)$$

where $N_X\left(b,d\right)$ is the number of intervals of $PH_i\left(X\right)$ with birth times less than b and death times greater than d. We will show that if C is an m-dimensional cube and $X \subset C$, then quantities of the form $M_{X,\partial C}\left(d,b\right)$ obey a super-additivity property.

Lemma 2. *Let* $\{C_1,\ldots,C_n\}$ *be* m*-dimensional cubes in* $\mathbb{R}^m$ *so that*

$$C_j \cap C_k \subset \partial C_j \qquad \forall\, j,k \in \{1,\ldots,n\} : j \neq k$$

If $X_j \subset C_j$ *for* $j = 1,\ldots,n$

$$N_{\cup_j X_j}\left(b,d\right) \geq M_{\cup_j X_j, \cup_j \partial C_j}\left(b,d\right) \geq \sum_{j=1}^{n} M_{X_j,\partial C_j}\left(b,d\right)$$

for any $0 \leq b < d$.

Proof. Let $k \in \{1,\ldots,n\}$, $S = \cup_{j=1}^{k-1} X_j$, $T = \cup_{j=1}^{n} \partial C_j$, $X = X_k$, and $C = C_k$. See Figure 1.

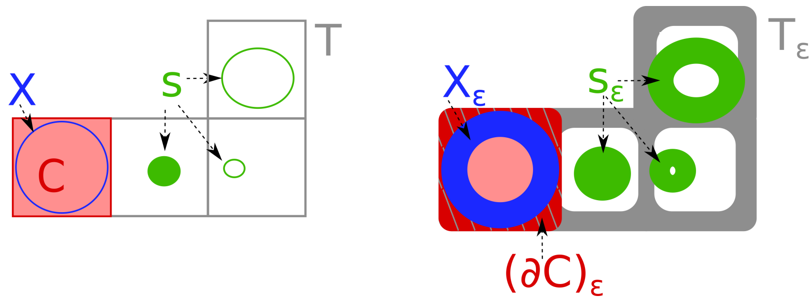

Figure 1. The setup in the proof of Lemma 2.

We consider the cases $i = m - 1$ and $i < m - 1$ separately. If $i = m - 1$, Alexander Duality implies that $N_S(b, d)$ is the number of bounded components of the complement of (S_b) that intersect non-trivially with the complement of S_d. Similarly, $M_{X,C}(b, d)$ is the number of bounded components of the complement of X_b that intersect non-trivially with $\left(X_d \cup (\partial C)_d\right)^c$. Note that all bounded components of $(X_b)^c$ are contained within the interior of C, because C is convex and separates $\mathbb{R}^m$ into two components.

Let Y be a component of the complement of X_b that intersects non-trivially with $\left(X_d \cup (\partial C)_d\right)^c$, and let $y \in Y \cap \left(X_d \cup (\partial C)_d\right)^c$. ∂C separates $\mathbb{R}^m$ into two components so

$$d(y, S) \geq d(y, S \cup T) = d(y, X \cup \partial C) > d$$

Therefore,

$$Y \cap (S_d)^c \supseteq Y \cap (S_d \cup T_d)^c = Y \cap \left(X_d \cup (\partial C)_d\right)^c \neq \varnothing$$

Applying the same argument to each X_j and counting components of the complement yields the desired inequalities.

Otherwise, assume that $i \leq m - 1$. We will show that

$$M_{S \cup X, T}(b, d) \geq M_{S,T}(b, d) + M_{X, \partial C}(b, d)$$

and the desired result will follow by induction. Note that

$$X_\epsilon \cap S_\epsilon \subseteq X_\epsilon \cap (S_\epsilon \cup T_\epsilon) \subseteq (\partial C)_\epsilon$$

for any $\epsilon > 0$. Consider the following commutative diagram of inclusion homomorphisms and Mayer-Vietoris sequences:

$$\begin{array}{ccccc} H_i\left(X_b \cap S_b\right) & \longrightarrow & H_i\left(X_b\right) \oplus H_i\left(S_b\right) & \xrightarrow{\alpha_b+\beta_b} & H_i\left(X_b \cup S_b\right) \\ \downarrow & & \downarrow{\scriptstyle \phi\oplus\psi} & & \downarrow{\scriptstyle \zeta} \\ 0 = H_i\left(\left(\partial C\right)_d\right) & \longrightarrow & H_i\left(X_d \cup \left(\partial C\right)_d\right) \oplus H_i\left(S_d \cup T_d\right) & \xrightarrow{\alpha_d+\beta_d} & H_i\left(X_d \cup S_d \cup T_d\right) \end{array}$$

Observe that $M_{X,\partial C}\left(b,d\right) = \text{rank}\ \phi$, $M_{S,T}\left(b,d\right) = \text{rank}\ \psi$, and $M_{X\cup S,T}\left(b,d\right) = \text{rank}\ \zeta$. It follows that

$$\begin{aligned} M_{X\cup S,T}\left(b,d\right) = & \\ & \text{rank}\,\zeta \\ \geq\ & \text{rank}\ \left(\alpha_d+\beta_d\right)\circ\left(\phi\oplus\psi\right) \\ =\ & \text{rank}\ \left(\phi\oplus\psi\right) && \text{because } H_i\left(\left(\partial C\right)_d\right)=0 \\ =\ & \text{rank}\ \phi + \text{rank}\ \psi \\ =\ & M_{X,\partial C}\left(b,d\right) + M_{S,T}\left(b,d\right) \\ \geq\ & \sum_{j=1}^{k} M_{X_j,\partial C_j}\left(b,d\right) && \text{by induction} \end{aligned}$$

□

3.2. **Occupancy Events.** If B is a subset of an m-space, define the occupancy event

$$\delta\left(B,\mathbf{x}\right) = \begin{cases} 0 & \left|\mathbf{x}\cap B\right| = 0 \\ 1 & \left|\mathbf{x}\cap B\right| > 0 \end{cases}$$

Also, if $\left\{A_i\right\}_{i=1}^{r}$ and $\left\{B_j\right\}_{j=1}^{s}$, are collections of subsets of M, let

$$\xi\left(\mathbf{x},\left\{A_i\right\},\left\{B_j\right\}\right) = \begin{cases} 1 & \delta\left(A_i,\mathbf{x}\right) = 0 \text{ and } \delta\left(B_j,\mathbf{x}\right) = 1 \qquad \forall\, i,j \\ 0 & \text{otherwise} \end{cases}$$

Lemma 3. *Let μ be a locally bounded probability measure on an m-space M. There exists a real number $V_0 > 0$ so for any $r,s \in \mathbb{N}$ there there exists a real number $\gamma_0 > 0$ so that for any collections of disjoint, congruent cubes$\left\{A_i^k\right\}$ and $\left\{B_j^k\right\}$, for $i \in \left\{1,\ldots,r\right\}$, $j \in \left\{1,\ldots,s\right\}$, and $k \in \left\{1,\ldots,n\right\}$ (for a total of $\left(r+s\right)n$ cubes) with volume*

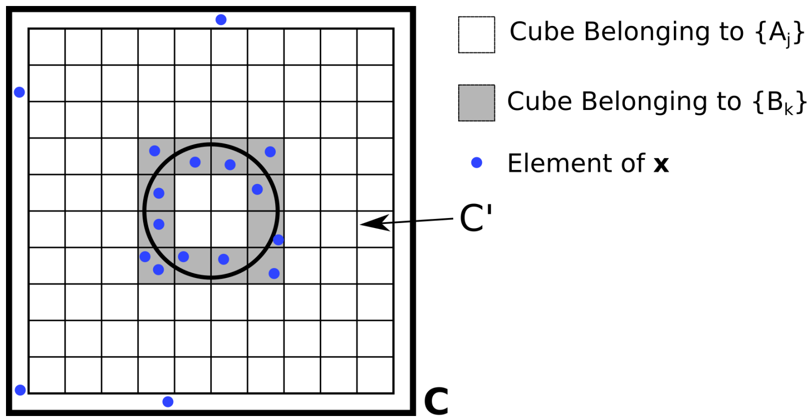

Figure 2. The setup in the proof of Lemma 4.

$$vol\left(A_i^k\right) = vol\left(B_j^k\right) = V_0/n \qquad \forall\, i, j, k$$

then

$$\mathbb{P}\left(\sum_{k=1}^{n} \xi\left(\boldsymbol{x}, \left\{A_i^k\right\}, \left\{B_j^k\right\}\right) \geq s\right) \geq \gamma_0$$

for all sufficiently large n.

Proof. The proof is nearly identical to that of Lemma 3 in [14]. □

Lemma 4. *Let $0 < b < d < 1/6$, and $V_0 > 0$. There exists a $\lambda_0 > 0$ so that if $C \subset \mathbb{R}^m$ is an m-dimensional cube of width R and $\lambda > \lambda_0$, there exist disjoint, congruent cubes $\left\{A_j\right\}$ and $\{B_k\}$ of width $R\,(V_0/\lambda)^{\frac{1}{m}}$ so that*

$$\xi\left(\boldsymbol{x}, \left\{A_j\right\}, \{B_k\}\right) = 1 \implies M_{\boldsymbol{x},\partial C}\left(Rb, Rd\right) > 0$$

Proof. We may assume without loss of generality that $R = 1$ and C is centered at the origin. Let $S^i \subset \mathbb{R}^m$ be an $i-$dimensional sphere of diameter $1/3$ centered at the origin; note that $PH_i\left(S^i\right)$ consists of a single interval $(0, 1/6)$.

Let $\kappa = \min\left(b, 1/6 - d\right)$ and $\Delta_0 = \kappa/\sqrt{m}$.

$$\lim_{\delta \to 0} \delta \lfloor 1/\delta \rfloor = 1$$

so there is a real number $\Delta_1 > 0$ so that $1 - \delta\lfloor 1/\delta \rfloor < \kappa$ for all $\delta < \Delta_1$. Set

$$\lambda_0 = \frac{V_0}{\min\left(\Delta_0, \Delta_1\right)^m}$$

Choose $\lambda > \lambda_0$, set $\delta = \left(V_0/\lambda\right)^{\frac{1}{m}}$, and let C' be the cube of width $\delta\lfloor 1/\delta \rfloor$ centered at the origin. Subdivide C' into $\lfloor 1/\delta \rfloor^m$ sub-cubes of width δ. Call this collection of sub-cubes $\{C_l\}$ and let

$$\left\{A_j\right\} = \left\{c \in \{C_l\} : S^i \cap c = \varnothing\right\} \quad \text{and} \quad \{B_k\} = \left\{c \in \{C_l\} : S^i \cap c \neq \varnothing\right\}$$

See Figure 2 for an illustration.

If $\mathbf{x} \subset C$ and the event $\xi\left(\mathbf{x}, \left\{A_j\right\}, \{B_k\}\right)$ occurs , then

$$d_H\left(\mathbf{x} \cap C', S^i\right) < \kappa$$

where d_H is the Hausdorff distance and we used the fact that the diagonal of an m-dimensional cube of width δ is $\delta\sqrt{m}$. The stability of the bottleneck distance [5] implies that $PH_i\left(\mathbf{x} \cap C'\right)$ includes an interval $\left(\hat{b}, \hat{d}\right)$ so that

$$\hat{b} < \kappa \leq b < d \leq 1/3 - \kappa < \hat{d}$$

In particular,

$$N_{\mathbf{x} \cap C'}\left(b, d\right) > 0$$

By construction,

$$\frac{1}{2} d\left(\mathbf{x} \cap C', C \setminus C'\right) > \frac{1}{2}\left(\frac{1}{6}\sqrt{m}\delta - d\left(C, C'\right)\right) > \frac{1}{6} - \kappa \geq d$$

so the ϵ-neighborhoods of $\mathbf{x} \cap C'$ and $C \setminus C'$ are disjoint for all $\epsilon \leq d$. It follows that the maps on homology induced by the inclusions $\left(\mathbf{x} \cap C'\right)_\epsilon \hookrightarrow \mathbf{x}_\epsilon$ and $\mathbf{x}_\epsilon \hookrightarrow \mathbf{x}_\epsilon \cup \left(\partial C\right)_\epsilon$ are injective for all $\epsilon \leq d$. Therefore, $M_{\mathbf{x}, \partial C}\left(b, d\right) > 0$, as desired. □

3.3. **Proof of the Lower Bound.** In the remainder, let μ be a locally bounded probability measure on an m-space M, let $\left\{x_j\right\}_{j \in \mathbb{N}}$ be i.i.d. samples from μ, and let $\mathbf{x}_n = \left\{x_1, \ldots, x_n\right\}$. Also, let C be as in Lemma 3, and rescale Δ_M if necessary so that C is a unit cube. Finally, let λ_0 be as in Lemma 4.

3.3.1. *The Euclidean Case.* For clarity, we first consider the special case where ϕ_M is the identity map, and μ is a locally bounded probability measure on a compact Euclidean simplicial complex. The argument for the general case contains many of the same elements.

Lemma 5. *Let $0 < b_0 < d_0 < 1/6$. If $n_0 > \lambda_0$, there is a $\gamma_1 > 0$ so that*

$$\lim_{n\to\infty} \frac{1}{n} N_{\boldsymbol{x}_n}\left(\left(\frac{n_0}{n}\right)^{\frac{1}{m}} b_0, \left(\frac{n_0}{n}\right)^{\frac{1}{m}} d_0\right) > \gamma_1$$

in probability.

Proof. Let V_0 be as in Definition 3, and let $r = |A_i|$ and $s = \left|B_j\right|$, where $\{A_i\}$ and $\left\{B_j\right\}$ are as in the previous lemma.

Assuming $n > n_0$, let $\omega = \left(\frac{n_0}{n}\right)^{\frac{1}{m}}$. Subdivide $\mathbb{R}^m$ into cubes of width ω, and let $\{D_l\}_{l=1}^{K_n}$ be the cubes that are fully contained in C. Note that

$$K_n := |\{D_l\}| \approx n/n_0$$

By the previous lemma, there are collections of disjoint, congruent sub-cubes $\left\{A_1^l, \ldots, A_r^l\right\}$ and $\left\{B_1^l, \ldots, B_s^l\right\}$ of width $\omega\left(V_0/n_0\right)^{\frac{1}{m}}$ contained inside each cube D_l so that

$$\xi\left(\mathbf{x}_n, \left\{A_i^l\right\}, \left\{B_j^l\right\}\right) = 1 \implies M_{\mathbf{x}_n \cap D_l, \partial D_l}\left(\omega b_0, \omega d_0\right) > 0$$

Note that

$$\begin{aligned} N_{\mathbf{x}_n}\left(\omega b_0, \omega d_0\right) \geq & \\ & \sum_{l=1}^{K_n} M_{\mathbf{x}_n \cap D_l, \partial D_l}\left(\omega b_0, \omega d_0\right) \qquad \text{by Lemma 2} \\ \geq & \sum_{l=1}^{K_n} \xi\left(\mathbf{x}_n, \left\{A_i^l\right\}, \left\{B_j^l\right\}\right) \end{aligned}$$

Let γ_0 be as in Lemma 3 and $\gamma < \gamma_0/n_0$. Set

$$\delta = \frac{1 + \gamma\frac{n_0}{\gamma_0}}{2} \quad \text{and} \quad \epsilon = \frac{1-\delta}{\delta}$$

so $1/2 < \delta < 1$ and $0 < \epsilon < 1$. Also, find a N so that $K_n > \delta n/n_0$ for all $n > N$. Note that

$$\gamma n = \frac{\gamma_0 \delta n}{n_0}\left(1 - \frac{1-\delta}{\delta}\right) < (1-\epsilon)\,\gamma_0 K_n \tag{2}$$

for all $n > N$. Therefore, if $n > N$,

$$\begin{aligned}
\mathbb{P}\left(N_{\mathbf{x}_n}(\omega b_0, \omega d_0) > \gamma n\right) \geq & \\
& \mathbb{P}\left(\sum_{l=1}^{K_n} \xi\left(\mathbf{x}_n, \left\{A_i^l\right\}, \left\{B_j^l\right\}\right) > \gamma n\right) \\
\geq\ & \mathbb{P}\left(B\left(K_n, \gamma_0\right) > \gamma n\right) && \text{by Lemma 3} \\
\geq\ & \mathbb{P}\left(B\left(K_n, \gamma_0\right) > (1-\epsilon)\,\gamma_0 K_n\right) && \text{by Equation 2}
\end{aligned}$$

which converges to 1 as $n \to \infty$. □

We can now prove the lower bound in the Euclidean setting:

Proposition 4. *There is a $\gamma' > 0$ so that*

$$\lim_{n\to\infty} n^{-\frac{m-\alpha}{m}} E_\alpha^i\left(\boldsymbol{x}_n\right) \geq \gamma'$$

in probability.

Proof. Let $0 < b < d < 1/6$, and let $n_0 > \lambda_0$ and γ_1 be as before. Also, let $\omega = \left(\frac{n_0}{n}\right)^{1/m}$. We have that

$$\begin{aligned}
\lim_{n\to\infty} n^{-\frac{m-\alpha}{m}} E_\alpha^i\left(\mathbf{x}_n\right) \geq & \\
& \lim_{n\to\infty} n^{-\frac{m-\alpha}{m}} \left(\omega d - \omega b\right)^\alpha N_{\mathbf{x}_n}(\omega b, \omega d) \\
=\ & \lim_{n\to\infty} \frac{n_0^{\alpha/m}}{n} (d-b)^\alpha N_{\mathbf{x}_n}(\omega b, \omega d) \\
\geq\ & n_0^{\alpha/m} (d-b)^\alpha \gamma_1 && \text{by Lemma 5} \\
:=\ & \gamma'
\end{aligned}$$

in probability. □

3.3.2. *The General Case.* Before proving the lower bound in our main theorem, we require an interleaving result for the persistent homology of images of bi-Lipschitz maps:

Lemma 6. *Let M_0 and M_1 be metric spaces and let $\psi : M_0 \to M_1$ be L-bilipshitz. If $X \subset M_0$ and $0 \leq b_0 < d_0$*

$$N_X\left(b_0/L, Ld_0\right) \leq N_{\psi(X)}\left(b_0, d_0\right) \leq N_X\left(Lb_0, d_0/L\right)$$

Proof. Fix $i \in \mathbb{N}$, and let $j_{\epsilon_0,\epsilon_1} : X_{\epsilon_0} \hookrightarrow X_{\epsilon_1}$ and $k_{\epsilon_0,\epsilon_1} : \phi\left(X\right)_{\epsilon_0} \hookrightarrow \phi\left(X\right)_{\epsilon_1}$ denote the inclusion maps for $\epsilon_0 \leq \epsilon_1$.

By the definition of a bi-Lipschitz map

$$\frac{1}{L} d_{M_0}\left(x, y\right) \leq d_{M_1}\left(\psi\left(x\right), \psi\left(y\right)\right) \leq L d_{M_0}\left(x, y\right)$$

for all $x, y \in M_0$. In particular, we have the following inclusions:

$$\psi\left(X_{b_0/L}\right) \hookrightarrow \psi\left(X\right)_{b_0} \hookrightarrow \psi\left(X\right)_{d_0} \hookrightarrow \psi\left(X_{Ld_0}\right)$$

It follows that the rank of map on homology induced by $i_{b_0/L,Ld_0}$ is less than or equal to the rank of the map induced by j_{b_0,d_0} (where we have used that a bi-Lipschitz map is a homeomorphism). Therefore,

$$N_X\left(b_0/L, Ld_0\right) \leq N_{\psi(X)}\left(b_0, d_0\right)$$

The argument for the other inequality is very similar. □

Proposition 5. *Let μ be a locally bounded probability measure on an m-space M and $0 \leq i < m$. There is a $\gamma > 0$ so that*

$$\lim_{n \to \infty} n^{-\frac{m-\alpha}{m}} E_\alpha^i\left(x_1, \ldots x_n\right) > \gamma$$

in probability

Proof. Let L be the bi-Lipschitz constant of ϕ_M, and choose $b, d > 0$ so that

$$L^2 b < d < 1/6$$

Set

$$n_0 = \max\left(\left(d/L - Lb\right)^{-m}, n_0\right)$$

so

$$n_0^{\frac{1}{m}}\left(d/L - Lb\right) \geq 1 \tag{3}$$

Let $\omega = \left(\frac{n_0}{n}\right)^{\frac{1}{m}}$ and $\mathbf{y}_n = \phi_m^{-1}(\mathbf{x}_n)$. Our strategy is to bound $E_\alpha^i(\mathbf{x}_n)$ by applying Lemma 5 to $\mathbf{y}_n$.

First,

$$\begin{aligned} & E_\alpha^i(\mathbf{x}_n) \\ & \quad (\omega(d/L - Lb))^\alpha N_{\mathbf{x}_n}(\omega Lb, \omega d/L) \\ \geq & \ n^{-\alpha/m} N_{\mathbf{x}_n}(\omega Lb, \omega d/L) && \text{by Equation 3} \\ \geq & \ n^{-\alpha/m} N_{\mathbf{y}_n}(\omega b, \omega d) && \text{by Lemma 6} \\ = & \ n^{-\alpha/m} N_{\mathbf{y}_n}(\omega b, \omega d) \end{aligned}$$

Therefore,

$$\lim_{n\to\infty} n^{-\frac{m-\alpha}{m}} E_\alpha^i(\mathbf{x}_n) \geq \lim_{n\to\infty} \frac{1}{n} N_{\mathbf{y}_n}(\omega b, \omega d) > \gamma_1$$

in probability, where $\gamma_1 > 0$ is as given in Lemma 5.

□

References

[1] Henry Adams, Manuchehr Aminian, Elin Farnell, Michael Kirby, Joshua Mirth, Rachel Neville, Chris Peterson, Patrick Shipman, and Clayton Shonkwiler. A fractal dimension for measures via persistent homology. Preprint, 2018.

[2] David Aldous and J. Michael Steele. Asymptotics for Euclidean minimal spanning trees on random points. *Probability Theory and Related Fields*, 1992.

[3] Ulrich Bauer and Florian Pausinger. Persistent Betti numbers of random Čech complexes. arXiv:1801.08376.

[4] Omer Bobrowski, Matthew Kahle, and Primoz Skraba. Maximally persistent cycles in random geometric complexes. *The Annals of Applied Probability*, 2017.

[5] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103120, 2007.

[6] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l_p-stable persistence. *Foundations of Computational Mathematics*, 2010.

[7] Vincent Divol and Wolfgang Polonik. On the choice of weight functions for linear representations of persistence diagrams. arXiv:1807.03678, July 2018.

[8] Rex Dwyer. Higher-dimensional voronoi diagrams in linear expected time. *Discrete and Computational Geometry*, 1991.

[9] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplificitaion. *Discrete and Computational Geometry*, 2002.

[10] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Applied Mathematics*, 1985.

[11] Yasuaki Hiraoka and Tomoyuki Shirai. Minimum spanning acycle and lifetime of persistent homology in the Linial—Meshulam process. *Random Structures & Algorithms*, 2017.

[12] Harry Kesten and Sungchul Lee. The central limit theorem for weighted minimal spanning trees on random points. *Annals of Applied Probability*, 1996.

[13] B. Schweinhart. Persistent homology and the upper box dimension. arXiv:1802.00533.

[14] Benjamin Schweinhart. The persistent homology of random geometric complexes on fractals. arXiv:1808.02196.

[15] Richard Stalney. The upper bound conjecture and Cohen-Macaulay rings. *Studies in Applied Mathematics*, 1975.

[16] J. Michael Steele. Growth rates of Euclidean minimal spanning trees with power weighted edges. *Annals of Probability*, 1988.

[17] Johannes Stemeseder. *Random polytopes with vertices on the sphere*. PhD thesis, Paris-Lodron University of Salzburg, 2014.

[18] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:249–274, 2005.