THE PERSISTENT HOMOLOGY OF RANDOM GEOMETRIC COMPLEXES ON FRACTALS

BENJAMIN SCHWEINHART

Abstract. We prove that the the fractal dimension of a metric space equipped with an Ahlfors regular measure can be recovered from the persistent homology of random samples. Our main result is that if $x_1, \dots x_n$ are i.i.d. samples from a d-Ahlfors regular measure on a metric space, and $E_\alpha(x_1, \dots, x_n)$ denotes the α-weight of the minimal spanning tree on $x_1, \dots, x_n$:

$$E_\alpha(x_1, \dots, x_n) = \sum_{e \in T(x_1, \dots, x_n)} |e|^\alpha$$

then

$$E_\alpha(x_1, \dots, x_n) \approx n^{\frac{d-\alpha}{d}}$$

with high probability as $n \to \infty$. In particular,

$$\log\left(E_\alpha(x_1, \dots, x_n)\right) / \log(n) \longrightarrow (d - \alpha)/d$$

This is a generalization of a result of Steele [44] from the absolutely continuous case to the fractal setting. We also prove analogous results for weighted sums defined in terms higher dimensional persistent homology.

1. Introduction

Fractal dimension has a wide range of applications in disciplines including medicine [3, 32], ecology [26], materials science [17, 48], and the analysis of large data sets [4, 45]. Several authors have proposed estimators of fractal dimension defined in terms of minimal spanning trees and higher dimensional persistent homology [44, 46, 39, 34, 37, 1], and provided empirical evidence that those quantities agreed with classical notions of fractal dimension. In Theorem 5 below, we provide the first rigorous justification for the use of minimal spanning trees and higher dimensional persistent homology to estimate fractal dimension.

Date: November 2018.

To be precise we study the asymptotic behavior of random variables of the form

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

where $\left\{x_j\right\}_{j\in\mathbb{N}}$ are i.i.d. samples from a probability measure μ on a metric space, and $PH_i\left(x_1,\ldots,x_n\right)$ denotes the i-dimensional reduced persistent homology of the Čech or Vietoris—Rips complex of $\left\{x_1,\ldots,x_n\right\}$. Unless otherwise specified, our results apply to the persistent homology of either the Čech or Vietoris—Rips complex, though the constants may differ. The case where $i=0$ and μ is absolutely continuous is already well-studied, under a different guise: if

$$E_{\alpha}\left(x_1,\ldots,x_n\right)=\sum_{e\in T\left(x_1,\ldots,x_n\right)}\left|e\right|^{\alpha}$$

where T is a minimal spanning tree on $x_1,\ldots,x_n$ then

$$E_{\alpha}\left(x_1,\ldots,x_n\right)=E_{\alpha}^{0}\left(x_1,\ldots,x_n\right)$$

if persistent homology is taken with respect to the Vietoris—Rips complex. In 1988, Steele [44] proved the following celebrated result:

Theorem 1 (Steele)**.** *Let μ be a compactly supported probability measure on $\mathbb{R}^m$, $m\geq 2$, and let $\left\{x_n\right\}_{n\in\mathbb{N}}$ be i.i.d. samples from μ. If $0<\alpha<m$,*

$$\lim_{n\to\infty} n^{-\frac{m-\alpha}{m}}E_{\alpha}^{0}\left(x_1,\ldots,x_n\right)\to c\left(\alpha,m\right)\int_{\mathbb{R}^m} f\left(x\right)^{(m-\alpha)/m}\,dx$$

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

Steele wrote:

> One feature of Theorem 1 that should be noted is that if μ has bounded support and μ is singular with respect to Lebesgue measure, then we have with probability one that $E_{\alpha}^{0}\left(x_1,\ldots,x_n\right)=o\left(n^{(d-\alpha)/d}\right)$. Part of the appeal of this observation is the indication that the length of the minimal spanning tree is a measure of the *dimension* of the support of the distribution. This suggests that the asymptotic behavior of the minimal spanning tree might be a useful adjunct to the concept of dimension in the modeling applications and analysis of fractals; see, e.g., [36]. [44]

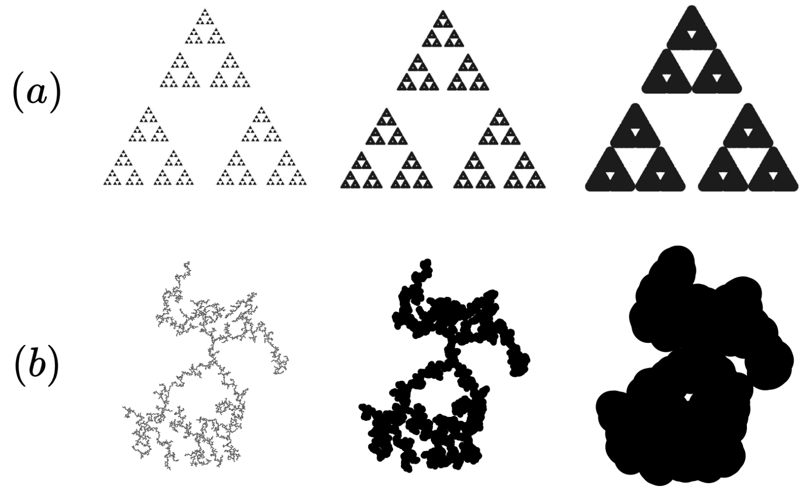

Figure 1. Two sets of fractional dimension, and their ϵ-neighborhoods: (a) a modified Sierpiński triangle and (b) a branched polymer. Their complex geometry is reflected by their persistent homology.

However, despite many subsequent sharper and more general results for non-singular measures [2, 29, 49], little is known about the asymptotic properties of random minimal spanning trees for fractal measures. As far as we know, the only rigorous result in the literature toward this end is that of Kozma, Lotker and Stupp [30], who proved that if μ is a d-Ahlfors regular measure with connected support, then the the length of the longest edge of a minimal spanning tree on n i.i.d. points sampled from μ is $\approx (\log(n)/n)^{1/d}$, where the symbol $\approx$ denotes that the denotes that the ratio between the two quantities is bounded between two positive constants that do not depend on n. They also raised the possibility that analogues asymptotics hold for the alpha-weight of a minimal spanning tree, which we prove here in Theorem 2.

More recently, as the field of stochastic topology has matured, several studies have examined the properties of the higher dimensional persistent homology of random geometric complexes [7, 8, 47, 19, 5]. In 2018, we [41] proved results about the asymptotics of $E_\alpha^i(x_1, \ldots, x_n)$ of i.i.d. samples from a from a locally bounded probability density on the bi-Lipschitz image of a compact m-dimensional simplicial complex. Independently and concurrently, Divol and Polonik [18] showed a sharper analogue of Steele's theorem for the persistent homology of points sampled from bounded, absolutely continuous probability densities on $[0,1]^m$.

A relationship between persistent homology and fractal dimension has been observed in several experimental studies. In 1991, Weygaert, Jones, and Martinez [46] proposed using the asymptotics of $E_\alpha^0(x_1, \ldots, x_n)$ for negative α to estimate the generalized Hausdorff dimensions. The PhD thesis of Robins, which was perhaps the first publication in the field of persistent topology, studied the scaling of Betti numbers of fractals, and proved results for the 0-dimensional persistent homology of disconnected sets [39]. In joint work with Robert MacPherson, we proposed a dimension for probability distributions of geometric objects based on persistent homology in 2012 [34]. Note that the quantities studied in that paper and in the thesis of Robins measure the complexity of a shape rather than the fractional dimension. Most recently, Adams et al. [1] proposed a persistent homology dimension for measures in terms of the asymptotics of $E_i^1(x_1, \ldots, x_n)$. Their computational experiments helped to inspire this work. We study a slightly modified version of their dimension here, and find hypotheses under which it agrees with the Ahlfors dimension (Theorem 5).

In the extremal setting, Kozma, Lotker and Stupp [31] defined a minimal spanning tree dimension for a metric space M in terms of the behavior of $E_\alpha^0(Y)$ as Y ranges over all subsets of M, and proved that it equals the upper box dimension. Earlier this year, we generalized this concept to higher dimensional persistention homology by defining a notion of a PH_i dimension of a metric space, and establishing hypotheses under which it agrees with the upper box dimension [40]. In that work, we also investigated extremal questions about the number of persistent homology intervals of a set of n points; those questions are also important in the probabilistic context, as we describe below.

1.1. **Our Results.** We prove analogues of the theorem of Steele [44] for probability measures defined on sets of fractional dimension that satisfy a certain regularity condition:

Definition 1. *A probability measure μ supported on a metric space X is d-Ahlfors regular if there exist positive real numbers c and r_0 so that*

$$\frac{1}{c} r^d \leq \mu(B_r(x)) \leq c\, r^d \tag{1}$$

for all $x \in X$ and $r < r_0$, where $B_r(x)$ denotes the open ball of radius r centered at x.

Ahlfors regularity is a common hypothesis when studying analysis on fractals [16, 6, 33]. Example of Ahlfors regular measures include the natural measures on the Sierpiński triangle and Cantor set, and, more generally, on any self-similar subset of Euclidean space defined by an iterated function system satisfying the open-set condition. If μ is d-Ahlfors regular on X then it is comparable to the $d-$dimensional Hausdorff measure on X. In particular, d equals the Hausdorff dimension of X. Ahlfors regularity also implies that a host of other fractional dimensions, including the upper and lower box dimensions, coincide and equal d.

The hypotheses we require are actually somewhat weaker than Ahlfors regularity. In particular, the proofs of our upper bounds only require that $M_\delta(\mu) = O\left(\delta^{-d}\right)$, where $M_\delta(\mu)$ is the maximal number of disjoint balls of radius δ centered at points of supp μ; this is stronger than the upper box dimension of the support of μ equaling d, which would imply that $M_\delta(\mu) = O\left(\delta^{-d+\epsilon}\right)$ for all $\epsilon > 0$. Also, the proofs of our lower bounds require that the uniform bounds in Equation 1 are satisfied on a set of positive measure, but not necessarily at every point in the support of μ.

Our main result is:

Theorem 2. *Let μ be a d-Ahlfors regular measure on a metric space, and let $\{x_n\}_{n\in\mathbb{N}}$ be i.i.d. samples from μ. If $0 < \alpha < d$,*

$$E_\alpha(x_1, \ldots, x_n) \approx n^{\frac{d-\alpha}{d}}$$

with high probability as $n \to \infty$, where the symbol $\approx$ denotes that the ratio of the two quantities is bounded between positive constants that do not depend on n.

We provide a proof of this result using the language of minimal spanning trees (rather than persistent homology) in Section 3. The special case where μ is a measure on Euclidean space is also a consequence of Theorem 3 below.

As we noted in our earlier paper [40], proving results for higher dimensional persistent homology is challenging due to extremal questions about the number of persistent homology intervals of a finite point set. While a minimal spanning tree on n points always has $n-1$ edges, a set of n points may have no PH_i for any $i > 0$, and there exist families of point sets for which $|PH_i(x_1, \ldots, x_n)|$ grows faster than n.

To prove upper bounds for the asymptotics of E_α^i for $i > 0$, we require either extremal or probabilistic control of the number of persistent homology intervals of a set of n points. Families of point sets in Euclidean space with more than a linear number of persistent homology intervals exist [40, 25], but are considered somewhat pathological. As far as we know, the Upper Bound Theorem [43] on the number of faces of a neighborly polytope provides the best upper bound for the number of persistent homology intervals of the Čech complex of a finite subset of $\mathbb{R}^m$:

$$\left|PH_i\left(x_1,\ldots,x_n\right)\right| = \begin{cases} i+1 & i < \lfloor \frac{m}{2} \rfloor \\ \lfloor \frac{m+1}{2} \rfloor & i \geq \lfloor \frac{m}{2} \rfloor \end{cases}$$

For the Vietoris—Rips complex of points in Euclidean space, we [40] showed that

$$\left|PH_1\left(x_1,\ldots,x_n\right)\right| = O\left(n\right)$$

by modifying an argument of Goff [25].

A different extremal question arises in the process of proving lower bounds for E_α^i. In particular, a subset $\mathbb{R}^m$ must have dimension above a certain non-triviality constant γ_i^m (defined in Section 6.1) to guarantee the existence of subsets with non-trivial $i-$dimensional persistent homology. We showed that $\gamma_1^m < m - 1/2$ in our previous paper [40].

The proofs of the upper bounds in the next two theorems work for Ahlfors regular measures on arbitrary triangulable metric spaces, but the lower bound requires that the measure is defined on a subset of Euclidean space:

Theorem 3. *Let μ be a d-Ahlfors regular measure on $\mathbb{R}^m$, and let $\left\{x_n\right\}_{n\in\mathbb{N}}$ be i.i.d. samples from μ. If $0<\alpha<d$. If $d>\gamma_i^m$ and*

$$\left|PH_i\left(x_1,\ldots,x_n\right)\right| < Dn^a$$

for some positive real numbers a and D and all finite subsets of X, and $0<\alpha<ad$, then there are real numbers $0<\zeta<Z$ so that

$$\zeta\, n^{\frac{d-\alpha}{d}} \leq E_\alpha^i\left(x_1,\ldots,x_n\right) \leq Z\, n^{\frac{ad-\alpha}{d}}$$

with high probability, as $n\to\infty$. In fact, the upper bound holds with probability one.

The upper bound is shown in Proposition 3, and the lower bound in Proposition 6. The following is a corollary:

Corollary 1. *Let μ be a d-Ahlfors regular measure on $\mathbb{R}^2$, and let $\left\{x_n\right\}_{n\in\mathbb{N}}$ be i.i.d. samples from μ. If $0<\alpha<d$. If $d>1.5$, $0<\alpha<d$, and persistent homology is taken of the Čech complex,*

$$E_\alpha^i\left(x_1,\ldots,x_n\right) \approx n^{\frac{d-\alpha}{d}}$$

in probability, as $n\to\infty$. In fact, the upper bound holds with probability one.

For large i or m, we show better upper bounds for d-Ahlfors regular measures for which the expectation and variance of $|PH_i(x_1,\ldots,x_n)|$ scale linearly and sub-quadratically, respectively. These quantities can be measured in practice, allowing one to determine whether higher dimensional persistent homology would be suitable for dimension estimation in applications.

Theorem 4. *Let μ be a d-Ahlfors regular measure on $\mathbb{R}^m$ so that $d > \gamma_i^m$, and let $\{x_n\}_{n\in\mathbb{N}}$ be i.i.d. samples from μ. If*

$$\mathbb{E}\left(|PH_i(x_1,\ldots,x_n)|\right) = O(n) \quad \text{and} \quad Var\left(|PH_i(x_1,\ldots,x_n)|\right)/n^2 \to 0$$

and $0 < \alpha < d$, there are real numbers $0 < \lambda < \Lambda$ so that

$$\lambda\, n^{\frac{d-\alpha}{d}} \leq E_\alpha^i(x_1,\ldots,x_n) \leq \Lambda\, n^{\frac{d-\alpha}{d}} \log(n)^{\frac{\alpha}{d}}$$

with high probability, as $n \to \infty$.

The upper and lower bounds are shown in Propositions 5 and 6, respectively.

1.2. **Dimension Estimation.** As we noted earlier in the introduction, several authors have proposed to use persistent homology for dimension estimation. Here, we provide the first proof that these methods recover the fractal dimension, under certain hypotheses. Toward that end, we define a family of PH_i dimensions of a measure, one for each real number $\alpha > 0$ and $i \in \mathbb{N}$:

Definition 2.

$$dim_{PH_i^\alpha}(\mu) = \frac{\alpha}{1-\beta}$$

where

$$\beta = \limsup_{n\to\infty} \frac{\log\left(\mathbb{E}\left(E_\alpha^i(x_1,\ldots,x_n)\right)\right)}{\log(n)}$$

That is, $\dim_{PH_i^\alpha}(\mu)$ is the unique real number d so that

$$\limsup_{n\to\infty} \mathbb{E}\left(E_\alpha^i(x_1,\ldots,x_n)\right) n^{-\frac{k-\alpha}{k}}$$

equals ∞ for all $k < d$, and is bounded for $k > d$. The case $\alpha = 1$ is very closely related to the dimension studied by Adams et al. [1], and agrees with it if defined.

The following is a corollary of our theorems on the asymptotic behavior of E_α^i:

Theorem 5. *If μ is a d-Ahlfors regular measure on a metric space and $0 < \alpha < d$ then*

$$dim_{PH_0^\alpha} = d$$

Furthermore, if μ is defined on $\mathbb{R}^m$, $d > \gamma_i^m$, and

$$\mathbb{E}\left(\left|PH_i\left(x_1, \ldots, x_n\right)\right|\right) = O\left(n\right) \quad and \quad Var\left(\left|PH_i\left(x_1, \ldots, x_n\right)\right|\right)/n^2 \to 0$$

then

$$dim_{PH_i^\alpha} = d$$

This result is weaker than our main theorems, and can be proven with weaker hypotheses. For example, the upper bond $\dim_{PH_i^\alpha} \leq d$ holds if the hypothesis of d-Ahlfors regularity is replaced by the requirement that the upper-box dimension of the support of μ is less than or equal to d.

2. Preliminaries

We will use the following two lemmas in our proofs for both minimal spanning trees and higher dimensional persistent homology.

Let X be a metric space, and let $M_\delta\left(X\right)$ be the maximal number of disjoint open balls of radius δ centered at points of X. (The upper and lower box dimensions are defined in terms of the asymptotic properties of $M_\delta\left(X\right)$). If X admits a d-Ahlfors regular measure, we can control the behavior of $M_\delta\left(X\right)$:

Lemma 1 (Ball-counting Lemma)**.** *If μ is a is d-Ahlfors regular measure supported on a metric space X then*

$$\frac{1}{c}2^{-d}\,\delta^{-d} \leq M_\delta\left(X\right) \leq c\,\delta^{-d}$$

for all $\delta < \delta_0$.

Proof. Let $\{x_j\}_{j=1}^{M_\delta(X)}$ be the centers of a maximal set of disjoint balls of radius δ centered at points of X.

$$\begin{aligned} 1 = & \ \mu(X) \\ \geq & \sum_{j=1}^{M_\delta(\mu)} \mu\left(B_\delta(x_j)\right) \\ \geq & \ \frac{1}{c}\delta^d M_\delta(\mu) \\ \implies & \ M_\delta(\mu) \leq c\delta^{-d} \end{aligned}$$

The maximality of $\{B_\delta(x_i)\}_{i=1}^{M_\delta(\mu)}$ implies that the balls of radius 2δ centered at the points $\{x_i\}$ cover X. It follows that

$$\begin{aligned} 1 = & \ \mu(X) \\ \leq & \sum_{j=1}^{M_\delta(X)} \mu\left(B_{2\delta}(x_j)\right) \\ \leq & \ c2^d\delta^d M_\delta(X) \\ \implies & \ M_\delta(X) \geq \frac{1}{c}2^{-d}\delta^{-d} \end{aligned}$$

as desired. □

We also require the following lemma of Cohen-Steiner et al. [14]:

Lemma 2. *Let $J \subset \mathbb{R}^+$ be a set of positive real numbers and let $J_\epsilon = \{j \in J : j > \epsilon\}$. If $|J_\epsilon| \leq f(\epsilon) < \infty$ for all $\epsilon > 0$ and $\alpha > 0$ then*

$$\sum_{j \in J_\epsilon} j^\alpha \leq \epsilon^\alpha f(\epsilon) + \alpha \int_{\delta=\epsilon}^{\max J} f(\delta)\, \delta^{\alpha-1}\, d\delta$$

Furthermore, if $|J| \leq f(0) < \infty$ then

$$\sum_{j \in J} j^\alpha \leq \alpha \int_{\delta=0}^{\max J} f(\delta)\, \delta^{\alpha-1}\, d\delta$$

For completeness, we reproduce the proof in [14]. $\sum_{j\in J_\epsilon} j^\alpha$ can be expressed as an integral involving the distributional derivative of $|J_\epsilon|$. Applying integration by parts yields:

$$\begin{aligned}\sum_{j\in J_\epsilon} j^\alpha &= \int_{\delta=\epsilon}^{\infty} -\frac{\partial \left|J_\delta\right|}{\delta}\delta^\alpha \, d\delta \\ &= \Big[-\left|J_\delta\right|\delta^\alpha\Big]_{\delta=\epsilon}^{\infty} + \alpha\int_{\delta=\epsilon}^{\infty}\left|J_\delta\right|\delta^{\alpha-1}\, d\delta \\ &= \epsilon^\alpha\left|J_\epsilon\right| + \alpha\int_{\delta=\epsilon}^{\max J}\left|J_\delta\right|\delta^{\alpha-1}\, d\delta \\ &\leq \epsilon^\alpha f\left(\epsilon\right) + \alpha\int_{\delta=\epsilon}^{\max J} f\left(\delta\right)\delta^{\alpha-1}\, d\delta\end{aligned}$$

2.1. **Notation.** $\mathbf{x}_n$ will be shorthand for the finite point set $\{x_1,\ldots,x_n\}$ with n points, and $\mathbf{x}$ will denote a finite point set with an unspecified number of elements. If the measure μ is obvious from the context, $\left\{x_j\right\}_{j\in\mathbb{N}}$ will be a collection of independent random variables with common distribution μ. Finally, we will use symbols with the "mathcal" font (i.e. $\mathcal{A},\mathcal{B},\ldots$) for collections of sets.

2.2. **Occupancy Events.** Our strategy for proving lower bounds will be to define certain "occupancy events" that imply the existence of a persistent homology interval of a certain length.

If A and B are sets define

$$\delta\left(A,B\right) = \begin{cases} 0 & A\cap B = \varnothing \\ 1 & A\cap B \neq \varnothing\end{cases}$$

Also, If A is a set and $\mathcal{B}$ is a collection of sets define the occupancy event

$$\Xi\left(\mathbf{x},A,\mathcal{B}\right) = \begin{cases} 1 & \delta\left(A,\mathbf{x}\right) = 0 \qquad \text{and} \qquad \delta\left(B,\mathbf{x}\right) = 1 \quad \forall\, B\in\mathcal{B} \\ 0 & \text{otherwise}\end{cases}$$

All occupancy events in this paper will satisfy $A\cap C = \varnothing$ for all $B\in\mathcal{B}$, and $B_1\cap B_2=\varnothing$ for all $B_1,B_2\in\mathcal{B}$ so that $B_1\neq B_2$. We say that two occupancy events $\Xi\left(\mathbf{x},A_1,\mathcal{B}\right)$ and $\Xi\left(\mathbf{x},A_1,\mathcal{C}\right)$ are disjoint if

$$\left(A_1\cup\bigcup_{B\in\mathcal{B}} B\right)\cap\left(A_1\cup\bigcup_{C\in\mathcal{C}} C\right) = \varnothing$$

An n, p, q, r-bounded occupancy event is a random variable of the form

$$\Xi\left(\mathbf{x}_n, A, \mathcal{B}\right)$$

where $\mathcal{B}$ is a collection of $\geq r$ sets and $\mathbf{x}_n$ is a set of n independent random variables with common distribution ν satisfying

$$\nu\left(A\right) \leq q/n \quad \text{and} \quad \nu\left(B\right) \geq p/n \; \forall \; B \in \mathcal{B}$$

If the above conditions on $|\mathcal{B}|$ and ν hold with equality, we say that $\Xi\left(\mathbf{x}_n, A, \mathcal{B}\right)$ is a n, p, q, r-uniform occupancy event.

Disjoint n, p, q, r-uniform occupancy events satisfy something akin to a weak law of large numbers as $n \to \infty$:

Lemma 3. *Let $r, a > 0$, and $0 < p, q < 1$. Also let $X_1^n \ldots, X_{\lfloor a\, n \rfloor}^n$ be disjoint n, p, q, r-uniform occupancy events for each $n \in \mathbb{N}$. If $Y_n = \frac{1}{n} \sum_{j=1}^{\lfloor a\, n \rfloor} X_j^n$,*

$$\lim_{n \to \infty} Y_n = \gamma$$

in probability, where $\gamma = a e^{-q} \left(1 - e^{-p}\right)^r$.

Proof. First, we compute the limiting expectation of the events X_j^n as $n \to \infty$:

$$\mathbb{E}\left(X_j^n\right) = \mathbb{P}\left(X_j^n = 1\right) = \left(1 - \frac{q}{n}\right)^n \sum_{j=0}^{r} (-1)^j \binom{r}{j} \left(1 - j \frac{p/n}{1 - q/n}\right)^n$$

by inclusion-exclusion. Therefore

$$\lim_{n \to \infty} \mathbb{E}\left(X_j^n\right) = e^{-q} \sum_{j=0}^{r} (-1)^j \binom{r}{j} e^{-jp} = e^{-q} \left(1 - e^{-p}\right)^r$$

by the binomial theorem, and $\lim_{n \to \infty} \mathbb{E}\left(Y_n\right) = \gamma$ by linearity of expectation.

A similar computation shows that if $j \neq k$,

$$\lim_{n \to \infty} \mathbb{E}\left(X_j^n X_k^n\right) = e^{-2q} \left(1 - e^{-p}\right)^{2r}$$

It follows that

$$\lim_{n \to \infty} \operatorname{Cov}\left(X_j^n, X_k^n\right) = \lim_{n \to \infty} \left(\mathbb{E}\left(X_j^n X_k^n\right) - \mathbb{E}\left(X_j^n\right) \mathbb{E}\left(X_k^n\right)\right) = 0$$

Therefore

$$\begin{aligned}\mathrm{Var}\left(Y_n\right) = &\ \frac{1}{n^2}\left(\sum_{j=1}^{\lfloor a\,n\rfloor}\mathrm{Var}\left(X_j\right) + 2\sum_{j=1}^{\lfloor a\,n\rfloor}\sum_{i=1}^{j-1}\mathrm{Cov}\left(X_j^n, X_k^n\right)\right)\\ \sim &\ \frac{a}{n}\mathrm{Var}\left(X_1^n\right) + a\frac{n^2-n}{n^2}\mathrm{Cov}\left(X_1^n, X_2^n\right)\\ \leq &\ \frac{a}{n} + a\left(1-\frac{1}{n}\right)\mathrm{Cov}\left(X_1^n, X_2^n\right)\end{aligned}$$

also converges to 0 as n goes to ∞.

Let $\epsilon > 0$ and $0 < \rho < 1$. Choose N sufficiently large so that

$$\left|\mathbb{E}\left(Y_n\right) - \gamma\right| < \epsilon/2 \qquad \text{and} \qquad \mathrm{Var}\left(Y_n\right) < \frac{\epsilon^2\rho}{4}$$

for all $n > N$. If $n > N$,

$$\begin{aligned}\mathbb{P}\left(\left|Y_n - \gamma\right| > \epsilon\right) \leq &\ \mathbb{P}\left(\left|Y_n - \mathbb{E}\left(Y_n\right)\right| > \epsilon/2\right)\\ \leq &\ \mathbb{P}\left(\left|Y_n - \mathbb{E}\left(Y_n\right)\right| > \frac{1}{\sqrt{\rho}}\sqrt{\mathrm{Var}\left(Y_n\right)}\right)\\ \leq &\ \rho\end{aligned}$$

by Chebyshev's Inequality. □

The occupancy events we define below will not be uniform, but we can use the previous lemma to bound them:

Lemma 4. *Let $r, a > 0$, $0 < p, q < 1$, and $s_n \geq \lfloor an \rfloor$ for all $n \in \mathbb{N}$. Also, let $X_1^n \ldots, X_{s_n}^n$ be disjoint n, p, q, r-bounded occupancy events for each $n \in \mathbb{N}$. There is a $\gamma > 0$ so that*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{s_n} X_j^n \geq \gamma$$

in probability.

Proof. Let $a_0 = \min\left(a, 1/\left(p+q\right)\right)$, $1 \leq j \leq \lfloor a_0 n \rfloor$, and

$$X_j^n = \Xi\left(\mathbf{x}_n, A_j^n, \mathcal{B}_j^n\right)$$

ν is non-atomic so for each $b \in \mathcal{B}_j^n$ we can find a $\hat{B} \subseteq B$ with $\nu\left(\hat{B}\right) = p/n$ [42]. Let

$$\hat{\mathcal{B}}_j^n = \left\{\hat{B} : B \in \mathcal{B}_j^n\right\} \qquad \text{and} \qquad D_n = \cup_{j=1}^{\lfloor a_0 n \rfloor} \cup_{\hat{B} \in \hat{\mathcal{B}}_n} \hat{B}$$

Similarly, there are disjoint sets $\hat{A}_1^n, \ldots, \hat{A}_{\lfloor a_0 n \rfloor}^n$ so that $A_j^n \subseteq \hat{A}_j^n \subset D_n^c$ and $\nu\left(\hat{A}_j^n\right) = q/n$ for $j = 1 \ldots, \lfloor a_0 n \rfloor$, where we have used that $a_0 (p+q) \leq 1$. Let

$$\hat{X}_j^n = \Xi\left(\mathbf{x}_n, \hat{A}_j^n, \hat{\mathcal{B}}_j^n\right)$$

By construction, $X_j^n = 1 \implies \hat{X}_j^n = 1$ so $\frac{1}{n}\sum_{j=1}^{\lfloor s_n \rfloor} X_j^n$ stochastically dominates $\frac{1}{n}\sum_{j=1}^{\lfloor a_0 n \rfloor} \hat{X}_j^n$. Applying the previous lemma to the latter sum implies the desired result. □

3. The Proof for Minimal Spanning Trees

If $\mathbf{x}$ is a finite metric space, let $p(\mathbf{x}, \epsilon)$ be the number of edges of $T(\mathbf{x})$ of length greater than ϵ. Also, let $G_{\mathbf{x},\epsilon}$ be graph the with vertex set $\mathbf{x}$ so that x_1 and x_2 are connected by an edge if and only if $d(x_1, x_2) < \epsilon$ (this is the one-skeleton of the Vietoris-Rips complex). The following is a corollary of Kruskal's algorithm:

Lemma 5.

$$p(\boldsymbol{x}, \epsilon) = \beta_0\left(G_{\boldsymbol{x},\epsilon}\right) - 1$$

where $\beta_0\left(G_{\boldsymbol{x},\epsilon}\right)$ is the number of connected components of $G_{\boldsymbol{x},\epsilon}$.

3.1. Proof of the Upper Bound.

Our strategy to prove an upper bound for the asymptotics of $E_\alpha(\mathbf{x}_n)$ is to control the number of edges in $T(\mathbf{x}_n)$ of length greater than ϵ in terms of the maximal number of disjoint balls of radius $\epsilon/2$ centered at points of $\mathbf{x}_n$. The approach is similar to that in our earlier papers [40, 41]

Lemma 6. *Let X be a metric space and suppose there are positive real numbers D and d so that*

$$M_\epsilon(X) \leq D\,\epsilon^{-d}$$

for all $\epsilon > 0$. Then

$$p(\boldsymbol{x}, \epsilon) < 2^{-d} D\,\epsilon^{-d}$$

for all finite subsets $\boldsymbol{x}$ of X and all $\epsilon > 0$.

Proof. Let $\mathbf{x} \subset X$, $\epsilon > 0$, and let $\mathbf{y}$ be the centers of a maximal set of disjoint balls of radius $\epsilon/2$ centered at points of $\mathbf{x}$. The maximality of $\mathbf{y}$ implies that for every $x \in \mathbf{x}$ there exists a $y \in \mathbf{y}$ so that $d(x,y) < \epsilon$. In particular, every connected component of $G_{\mathbf{x},\epsilon}$ has a vertex that is an element of $\mathbf{y}$. Therefore

$$\begin{aligned} p(\mathbf{x},\epsilon) &= b_0\left(G_{\mathbf{x},\epsilon}\right) - 1 \\ &\leq |\mathbf{y}| - 1 \\ &\leq D(\epsilon/2)^{-d} \\ &= 2^{-d} D \epsilon^{-d} \end{aligned}$$

□

We prove an extremal bound on $E_\alpha(\mathbf{x}_n)$ that, when combined with Lemma 1, implies the upper bound for our main theorem on minimal spanning trees:

Proposition 1. *Let X be a metric space and suppose there are positive real numbers D and d so that*

$$M_\delta(X) \leq D\,\delta^{-d}$$

for all $\delta > 0$. If $0 < \alpha < d$, there exists a $D_\alpha > 0$ so that

$$E_\alpha(x_1,\ldots,x_n) \leq D_\alpha\, n^{\frac{d-\alpha}{d}}$$

for all n and all $x_1,\ldots,x_n \subset X$. Furthermore, there exist a $D_d > 0$ so that

$$E_d(x_1,\ldots,x_n) \leq D_d \log(n)$$

for all n and all $x_1,\ldots,x_n \subset X$.

Proof. Rescale X if necessary so that its diameter is less than 1, and let

$$\kappa = \frac{1}{2}\left(\frac{D}{n-1}\right)^{1/d}$$

The previous lemma together with the fact that a minimal spanning tree on n points has $n-1$ lemma implies that $p(\{\mathbf{x}_n\},\epsilon) \leq f(\epsilon)$ where

$$f(\epsilon) = \min\left(n-1, 2^{-d} D \epsilon^{-d}\right) = \begin{cases} n-1 & \epsilon \leq \kappa \\ 2^{-d} D \epsilon^{-d} & \epsilon \geq \kappa \end{cases}$$

Applying Lemma 2 to the set of edge lengths of the minimal spanning tree on $\mathbf{x}_n$ yields

$$\begin{aligned}
E_\alpha(\mathbf{x}_n) &= \sum_{e \in T(\mathbf{x}_n)} |e|^\alpha \\
&\leq \alpha \int_{\delta=0}^{1} f(\delta)\,\delta^{\alpha-1}\,d\delta \\
&= (n-1)\int_{\delta=0}^{\kappa} \alpha\delta^{\alpha-1}\,d\delta + \alpha 2^{-d} D \int_{\delta=\kappa}^{1} \delta^{\alpha-d-1}\,d\delta \\
&= (n-1)\left[\delta^\alpha\right]_{\delta=0}^{\kappa} - \frac{\alpha}{d-\alpha} 2^{-d} D \left[\delta^{\alpha-d}\right]_{\delta=\kappa}^{1} \\
&= (n-1)\kappa^\alpha + \frac{\alpha}{d-\alpha} 2^{-d} D \left(\kappa^{\alpha-d} - 1\right) \\
&= 2^\alpha D^{\frac{\alpha}{d}} \left(1 + D\frac{\alpha}{d-\alpha}\right)(n-1)^{\frac{d-\alpha}{d}} - \frac{\alpha}{d-\alpha} 2^{-d} D \\
&\leq D_\alpha n^{\frac{d-\alpha}{d}}
\end{aligned}$$

where

$$D_\alpha = 2^\alpha D^{\frac{\alpha}{d}} \left(1 + D\frac{\alpha}{d-\alpha}\right)$$

The result for $\alpha = d$ follows from a similar computation. □

3.2. **Proof of the Lower Bound.** Our strategy to prove a lower bound for the asymptotics of $E_\alpha(\mathbf{x}_n)$ is to define a random variables based on occupancy patterns of disjoint balls of radius $2r$ that imply the existence of minimal spanning tree edges of length at least r.

Let M be a metric space and let μ be a d-Ahlfors regular measure with support M. If B is a ball of radius $2r$ centered at a point $x \in M$ and $\mathbf{x}$ is a finite subset of M, define

$$\omega(B, \mathbf{x}) = \Xi\left(B_{2r}(x) \setminus B_r(x), \{B_r(x)\}\right)$$

That is, $\omega(B, \mathbf{x}) = 1$ if $\mathbf{x}$ intersects $B_r(x)$ but not the annulus centered at x with radii r and $2r$.

Lemma 7. *Let $\mathcal{B}$ be a set of disjoint balls of radius $2r$ centered at points of M, and let $\boldsymbol{x}$ be a finite subset of M.*

$$p(\boldsymbol{x}, r) \geq \sum_{B \in \mathcal{B}} \omega(B, \boldsymbol{x}) - 1$$

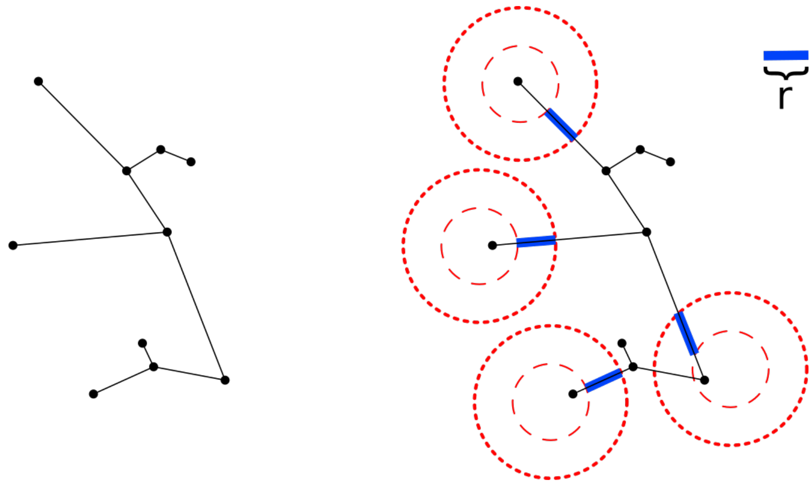

Figure 2. The red balls on the right all satisfy $\omega(B, \mathbf{x}) = 1$, which guarantees that the minimal spanning tree on the left has at least three edges whose length exceeds r.

Proof. This is an immediate consequence of Lemma 5. See Figure 2. □

Fix $n \in \mathbb{N}$ and let $\epsilon = n^{-1/d}$. Let $B_1^n, \ldots, B_{s_n}^n$ be a maximal collection of disjoint balls of radius 2ϵ centered at points of X, and let y_j^n be the center of B_j^n for $j = 1, \ldots, s_n$. We require one more lemma before proving the lower bound:

Lemma 8. *There is a positive real number $\gamma > 0$ so that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{s_n} \omega\left(B_j^n, \boldsymbol{x}_n\right) \geq \gamma$$

in probability as $n \to \infty$.

Proof. Let

$$p = \frac{1}{c} \quad \text{and} \quad q = 2^d c - \frac{1}{c}$$

By the definition of Ahlfors regularity

$$\mu\left(B_\epsilon\left(y_j^n\right)\right) \geq p\epsilon^d = \frac{p}{n}$$

and

$$\mu\left(B_j^n \setminus B_\epsilon\left(y_j^n\right)\right) \leq c\,(2\epsilon)^d - \frac{1}{c}\epsilon^d = \frac{q}{n}$$

Also, Lemma 1 implies that

$$s_n \geq \frac{1}{c}2^{-d}2\epsilon^{-d} = \frac{1}{c}2^{-2d}n$$

Therefore, the occupancy events $\omega\left(B_1^n, \mathbf{x}_n\right), \ldots, \omega\left(B_{s_n}^n, \mathbf{x}_n\right)$ satisfy the hypotheses of Lemma 4, which immediately implies the desired result. □

The lower bound in our main theorem on minimal spanning trees follows quickly:

Proposition 2. *Let μ be a d-Ahlfors regular measure on a metric space M. If $\left\{x_j\right\}_{j\in\mathbb{N}}$ are i.i.d. samples from μ, and γ is as given in the previous lemma,*

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

in probability.

Proof. We have that

$$\begin{aligned}
\lim_{n\to\infty} n^{-\frac{d-\alpha}{d}} E_\alpha^i\left(\mathbf{x}_n\right) \geq\ & \lim_{n\to\infty} n^{-\frac{d-\alpha}{d}} n^{-\alpha/d} p\left(\mathbf{x}_n, n^{-1/d}\right) \\
\geq\ & \lim_{n\to\infty} \frac{1}{n}\left(\sum_{j=1}^{s_n} \omega\left(B_j^n, \mathbf{x}_n\right) - 1\right) && \text{by Lemma 7} \\
\geq\ & \gamma && \text{by Lemma 8}
\end{aligned}$$

in probability, as $n \to \infty$. □

4. Persistent Homology

We provide a brief introduction to persistent homology [20] of a filtration, loosely following [40]. For a more in-depth survey refer to, e.g., [21, 22, 10, 24] A **filtration** of topological spaces is a family $\{X_\alpha\}_{\alpha\in I}$ of topological spaces indexed by an ordered set I, with continuous maps $X_{\alpha_1} \to X_{\alpha_2}$ for $\alpha_1 < \alpha_2$. For example, if X is a subset of a metric space M, the Čech filtration of X, is the family of ϵ-neighborhoods of X:

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

for $\epsilon > 0$, together with inclusion maps $X_{\epsilon_1} \hookrightarrow X_{\epsilon_2}$ for $\epsilon_1 < \epsilon_2$. Another common construction is the Vietoris—Rips complex: if Y is a metric space, let $V(Y, \epsilon)$ be the simplicial complex defined by

$$(y_1, \dots, y_n) \in V(Y, \epsilon) \quad \text{if} \quad d\left(y_i, y_j\right) < \epsilon \text{ for } i, j = 1, \dots, n$$

The family $\{V(Y, \epsilon)\}_{\epsilon > 0}$ together with inclusion maps for $\epsilon_1 < \epsilon_2$ is a filtration indexed by the positive real numbers. As noted earlier, all of our results apply to both the Čech and Vietoris—Rips complexes, though the constants may differ. We will suppress the dependence of persistent homology on the underlying filtration unless otherwise noted.

The **persistent homology module** of a filtration is the product $\prod_{\alpha \in I} H_i(X_\alpha)$, together with the homomorphisms $j_{\alpha_0, \alpha_1} : H_i(X_{\alpha_0}) \to H_i(X_{\alpha_1})$ for $\alpha_0 < \alpha_1$, where $H_i(X_\alpha)$ denotes the reduced homology of X_α with coefficients in a field. If $H_i(X_\alpha)$ is finite dimensional for all $\alpha \in I$ — a hypothesis satisfied by all filtrations considered in this paper [9, 11] — the persistent homology module decomposes canonically into a set of interval modules [50, 15]. We denote these intervals $PH_i(X)$; each interval $(b, d) \in PH_i(X)$ corresponds to a homology generator that is "born" at $\alpha = b$ and "dies" at $\alpha = d$.

If $\mathbf{x}$ is a finite metric space and persistent homology is taken with respect to the Vietoris—Rips complex, Kruskal's algorithm implies that there is a length-preserving bijection between intervals of $PH_0(\mathbf{x})$ and the edges of the minimal spanning tree on $\mathbf{x}$. The same is true if if persistent homology is taken with respect to the Čech complex and $\mathbf{x} \subset \mathbb{R}^m$, except that an interval is matched with an edge of twice its length. Note that the Čech complex depends on the ambient metric space.

4.1. **Properties of Persistent Homology.** Let X be a bounded, triangulable metric space. For each $\epsilon > 0$, let $PH_i^\epsilon(X)$ denote the set of intervals of $PH_i(X)$ of length greater than ϵ:

$$PH_i^\epsilon(X) = \{I \in PH_i(X) : |I| > \epsilon\}$$

Also, define

$$p_i(X, \epsilon) = |PH_i^\epsilon(X)|$$

If $X, Y \subset M$, let $d_H(X, Y)$ will denote the Hausdorff distance between X and Y:

$$d_H(X, Y) = \inf_{\epsilon \geq 0} Y \subseteq X_\epsilon \quad \text{and} \quad X \subset Y_\epsilon$$

and $d(X,Y)$ denote infimal distance between pairs of points, one in each set:

$$d(X,Y) = \inf_{x\in X, y\in Y} d(x,y)$$

We use the following properties of persistent homology in our proofs:

(1) **Stability**: If $d_H(X,Y) < \epsilon$, there is an injection
$$\eta : PH_i^{2\epsilon}(X) \to PH_i(Y)$$
so that if $\eta(b_0,d_0) = (b_1,d_1)$ then
$$\max(|b_0 - b_1|, |d_0 - d_1|) < \epsilon$$
In particular,
$$p_i(X, 2\epsilon + \delta) \leq p_i(Y,\delta)$$
for all $\delta \geq 0$. [13, 12]

(2) **Additivity for well-separated sets**: If $X_j \subset M$ for $j = 1\ldots,n$ and
$$d\left(X_j, X_k\right) > \max\left(\operatorname{diam} X_j, \operatorname{diam} X_k\right)\left(1 - \delta_{j,k}\right) \quad \forall j,k$$
then
$$p_i\left(\cup_j X_j, \epsilon\right) \geq \sum_j p_i\left(X_j, \epsilon\right)$$

(3) **Translation invariance**: $PH_i(X) = PH_i(X+t)$ for all $t \in \mathbb{R}^m$.

(4) **Scaling**: For all $\rho > 0$,
$$PH_i(\rho X) = \{(\rho b, \rho d) : (b,d) \in PH_i(X)\}$$

We use property (1) in our proofs of both the upper and lower bounds, and property (2) for our proof of the lower bound. For these results, we also require a non-triviality property (as in Definition 3) and an upper bound on $PH_i(x_1,\ldots,x_n)$ in terms of n.

4.2. **A Lemma.** If X is a subset of a metric space let
$$F_\alpha^i(X,\epsilon) = \sum_{I\in PH_i^\epsilon(X)} |I|^\epsilon$$

We will use the following lemma in the next section:

Lemma 9. *If* $d_H(X,Y) < \epsilon/4$ *then*
$$F_\alpha^i(X,\epsilon) < 2^\alpha F_\alpha^i(Y,\epsilon/2)$$

Proof. By stability, there is an injection

$$\eta : PH_i^{\epsilon}(X) \to PH_i^{\epsilon/2}(Y)$$

satisfying

$$|I| < |\eta(I)| + \epsilon/2 \leq 2|\eta(I)|$$

for all $I \in PH_i^{\epsilon}(X)$.

It follows that

$$\begin{aligned} F_\alpha^i(X,\epsilon) &= \sum_{I \in PH_i^{\epsilon}(X)} |I|^\alpha \\ &< \sum_{I \in PH_i^{\epsilon}(X)} 2^\alpha |\eta(I)|^\alpha \\ &\leq 2^\alpha \sum_{J \in PH_i^{\epsilon/2}(Y)} |J|^\alpha \\ &= 2^\alpha F_\alpha^i(Y,\epsilon/2) \end{aligned}$$

□

5. Upper Bounds

Our strategy to prove an upper bound for the asymptotics of $E_\alpha^i(\mathbf{x}_n)$ is similar to that in Section 3.1: we control to control the number of persistence intervals of length greater than ϵ by approximating $\mathbf{x}_n$ by a set consisting of the centers of disjoint balls of radius $\epsilon/2$ centered at points of $\mathbf{x}_n$.

5.1. Extremal Hypotheses.

First, we prove the upper bound in Theorem 3, which implies the upper bound for for our theorem on minimal spanning trees:

Lemma 10 (Interval Counting Lemma)**.** *If X is a triangulable metric space so that*

$$|PH_i(x_1,\dots,x_n)| < Dn^a$$

for some positive real numbers a and D and all finite subsets $\{x_1,\dots,x_n\}$ of X then

$$p_i(Y,\epsilon) < D'\epsilon^{-ad}$$

for some $D' > 0$, all $Y \subseteq X$, and all $\epsilon > 0$.

Proof. Let $Y \subseteq X$, $\epsilon > 0$, and $\left\{y_j\right\}$ be the centers of a maximal set of disjoint balls of radius $\epsilon/4$ centered at points of Y. The balls of radius $\epsilon/2$ centered at the points $\left\{y_j\right\}$ cover Y so

$$d_H\left(\{y_i\}, Y\right) < \epsilon/2$$

It follows that

$$\begin{aligned} p_i\left(Y, \epsilon\right) \leq\ & p_i\left(\{y_i\}, 0\right) && \text{by stability} \\ \leq\ & D\left|y_i\right|^a && \text{by hypothesis} \\ \leq\ & D M_{\epsilon/4}\left(X\right)^a && \\ \leq\ & D c^a 4^{-a/d} \epsilon^{-ad} && \text{by Lemma 14} \end{aligned}$$

as desired. ☐

Proposition 3. *If X satisfies the hypotheses of the previous lemma and $\alpha < ad$ there exist a $D_\alpha > 0$ so that*

$$E_i^\alpha\left(x_1, \ldots, x_n\right) \leq D_\alpha\left(n^{\frac{ad-\alpha}{d}}\right)$$

for all finite subsets $\{x_1, \ldots, x_n\} \subset X$ and all $n \in \mathbb{N}$. Furthermore there exists a $D_d > 0$ so that

$$E_i^{ad}\left(x_1, \ldots, x_n\right) \leq D_d \log\left(n\right)$$

for all finite subsets $\{x_1, \ldots, x_n\} \subset X$ and all $n \in \mathbb{N}$.

Proof. The proof is nearly identical to that of Proposition 1, and we omit it here. ☐

5.2. **Probabilistic Hypotheses.** While the extremal hypotheses of the previous section allow us to prove the desired upper bound for PH_0, they are inadequate to show a similar upper bound for $i > 0$. Here, we show that hypotheses on the asymptotics of the expectation and variance of the number of PH_i intervals of n imply better asymptotic upper bounds. The idea of the proof is that if the expected number of intervals of n points sampled from X scales linearly with n, we can use that to control the behavior of the intervals of $PH_i\left(X\right)$. With that, we write $PH_i\left(\mathbf{x}_n\right)$ a sum of two terms, one which approximates $PH_i\left(X\right)$ and one which corresponds to "d-dimensional noise" at a certain scale.

First, we require the following lemma, which follows from a standard argument using the union bound; see [38] for a proof:

Lemma 11. *If μ is a probability measure on X and $\left\{B_j\right\}_{j=1}^{l} \subset X$ have the property that $\mu\left(B_j\right) \geq a$ for all j. Then*

$$\mathbb{P}\left(\{\boldsymbol{x}_n\} \cap B_j \neq \varnothing, j = 1 \ldots, l\right) \geq 1 - le^{-an}$$

Lemma 12. *If μ is a d-Ahlfors regular measure with support X then*

$$\mathbb{P}\left(d_H\left(\{\boldsymbol{x}_n\}, X\right) < \epsilon\right) \geq 1 - c\epsilon^{-d} e^{-A_0 \epsilon^d n}$$

for a positive real number A_0 depending only on the constants c and d appearing the definition of Ahlfors regularity.

Proof. Let $\left\{y_1, \ldots, y_{M_{\epsilon/3}(X)}\right\}$ be the centers of a maximal set of disjoint balls of radius $\epsilon/3$ centered at points of X. By the definition of Ahlfors regularity,

$$\mu\left(B_{\epsilon/3}\left(y_j\right)\right) \geq A_0 \epsilon^d$$

for all j, where $A_0 = 3^{-d}\frac{1}{c}$.

The balls of radius $2\epsilon/3$ centered at the points $\left\{y_j\right\}$ cover X so

$$d_H\left(\{y_i\}, X\right) < 2\epsilon/3$$

Therefore, if $\{\mathbf{x}_n\} \cap B_{\epsilon/3}\left(y_j\right) \neq \varnothing$ for $j = 1 \ldots, M_{\epsilon/3}(X)$

$$d_H\left(\{\mathbf{x}_n\}, X\right) < \epsilon/3 + 2\epsilon/3 = \epsilon$$

It follows that

$$\begin{aligned}
\mathbb{P}\left(d_H\left(\{\mathbf{x}_n\}, X\right) < \epsilon\right) \geq \\
& \mathbb{P}\left(\{\mathbf{x}_n\} \cap B_{\epsilon/3}\left(y_j\right) \neq \varnothing \quad \text{for } j = 1 \ldots, M_{\epsilon/3}(X)\right) \\
\geq\ & 1 - M_{\epsilon/3}(X)\, e^{-A_0 \epsilon^d n} && \text{by Lemma 11} \\
\geq\ & 1 - c\epsilon^{-d} e^{-A_0 \epsilon^d n} && \text{by Lemma 1}
\end{aligned}$$

□

Lemma 13. *Let X be a metric space that admits a d-Ahlfors regular measure μ satisfying*

$$\mathbb{E}\left(\left|PH_i\left(\boldsymbol{x}_n\right)\right|\right) = O(n)$$

There are real numbers $A_1, \epsilon_0 > 0$ so that

$$p_i(X, \epsilon) \leq A_1 \epsilon^{-d} \log(1/\epsilon)$$

for all $\epsilon < \epsilon_0$.

Proof. Find positive real numbers D_1 and N_1 so that

$$\mathbb{E}\left(\left|PH_i\left(\mathbf{x}_n\right)\right|\right) \leq D_1 n$$

for all $n > N_1$. By Markov's inequality,

$$\mathbb{P}\left(\left|PH_i\left(\mathbf{x}_n\right)\right| > 2D_1 n\right) < 1/2$$

Manipulating the inequality in the previous lemma gives that

$$\mathbb{P}\left(d_H\left(\left\{x_1,\ldots,x_{m(\epsilon)}\right\},X\right) < \epsilon/2\right) \geq 1/2$$

where

$$m\left(\epsilon\right) = \lceil \frac{2^d}{A_0}\epsilon^{-d}\log\left(2^{d+1}c\epsilon^{-d}\right)\rceil$$

Let ϵ be sufficiently small so that $m\left(\epsilon\right) > N_1$. We have that

$$\left|PH_i\left(x_1,\ldots,x_{m(\epsilon)}\right)\right| \leq 2D_1 n \qquad \text{and} \qquad d_H\left(\left\{x_1,\ldots,x_{m(\epsilon)}\right\},X\right) < \epsilon$$

for some $\left\{x_1,\ldots,x_{m(\epsilon)}\right\} \subset X$. Therefore, by stability

$$\begin{aligned} p_i\left(X,\epsilon\right) \leq & p_i\left(\left\{x_1,\ldots,x_{m(\epsilon)}\right\},0\right) \\ \leq & \ 2D_1 m\left(\epsilon\right) \\ = & \ \lceil \frac{2^d}{A_0}\epsilon^{-d}\log\left(2^{d+1}c\epsilon^{-d}\right)\rceil \\ = & \ O\left(\epsilon^{-d}\log\left(1/\epsilon\right)\right) \end{aligned}$$

as $\epsilon \to 0$. □

Proposition 4. *If* X *satisfies the hypotheses of the previous lemma and* $0 < \alpha < d$*, then there exist positive real numbers* A_2 *and* ϵ_1 *so that*

$$F_\alpha^i\left(X,\epsilon\right) \leq A_2\epsilon^{\alpha-d}\log\left(1/\epsilon\right)$$

for all $\epsilon < \epsilon_1$.

Proof. By the previous lemma

$$p_i\left(X,\epsilon\right)\leq f\left(\epsilon\right):=A_1\left(\epsilon\right)^{-d}\log\left(\frac{1}{\epsilon}\right)$$

For all $\epsilon<\epsilon_0$. Applying Lemma 2 yields

$$F_\alpha^i\left(Y,\epsilon\right)\leq\epsilon^\alpha f\left(\epsilon\right)+\alpha\int_{t=\epsilon}^{1}f\left(t\right)t^{\alpha-1}\,dt+F_\alpha^i\left(Y,\epsilon_0\right)$$

The first term equals

$$A_1\epsilon^{\alpha-d}\left(\log\left(1/\epsilon\right)\right)$$

which has the desired asymptotics as $\epsilon\to 0$. Evaluating the second term yields

$$\alpha\int_{t=\epsilon}^{1}A_1 t^{\alpha-d-1}\log\left(1/t\right)\,dt=$$

$$\begin{aligned}&A_1\left[-\frac{1}{d-\alpha}t^{\alpha-d}\log\left(1/t\right)-\frac{1}{\left(d-\alpha\right)^2}t^{\alpha-d}\right]_\epsilon^1\\
=\;&A_1\left(\frac{1}{d-\alpha}\epsilon^{\alpha-d}\log\left(1/\epsilon\right)+\frac{1}{\left(d-\alpha\right)^2}\epsilon^{\alpha-d}-\frac{1}{\left(d-\alpha\right)^2}\right)\\
=\;&O\left(\epsilon^{\alpha-d}\log\left(1/\epsilon\right)\right)\end{aligned}$$

□

Proposition 5. *Let μ be a $d-$Ahlfors regular measure on $\mathbb{R}^m$. If*

$$\mathbb{E}\left(\left|PH_i\left(\boldsymbol{x}_n\right)\right|\right)=O\left(n\right)$$

and

$$Var\left(\left|PH_i\left(\boldsymbol{x}_n\right)\right|\right)/n^2\to 0$$

then there is a $\Lambda>0$ so that

$$\lim_{n\to\infty}E_i^\alpha\left(\boldsymbol{x}_n\right)\leq\Lambda n^{\frac{d-\alpha}{d}}\log\left(n\right)^{\frac{\alpha}{d}}$$

in probability.

Proof. Let

$$G_\alpha^i\left(\mathbf{x},\epsilon\right)=\sum_{I\in PH_i\left(\mathbf{x}\right)\setminus PH_i^\epsilon\left(\mathbf{x}\right)}\left|I\right|^\alpha$$

our strategy is to write

$$E_i^{\alpha}\left(\mathbf{x}_n\right)=G_{\alpha}^{i}\left(\mathbf{x}_n,\epsilon\right)+F_{\alpha}^{i}\left(\mathbf{x}_n,\epsilon\right)$$

for a well-chosen ϵ. The former term can be interpreted as "noise" and the latter approximates the persistent homology of the support of μ.

Let $0<p<1$, and let D be a positive real number so that

$$\mathbb{E}\left(\left|PH_i\left(\mathbf{x}_n\right)\right|\right)\leq\left(D/2\right)n$$

for all sufficiently large n. Then

$$\begin{aligned}\mathbb{P}\left(\left|PH_i\left(\mathbf{x}_n\right)\right|>Dn\right)\leq&\\ &\mathbb{P}\left(\Big|\left|PH_i\left(\mathbf{x}_n\right)\right|-\mathbb{E}\left(\left|PH_i\left(\mathbf{x}_n\right)\right|\right)\Big|>Dn/2\right)\\ \leq&\ \operatorname{Var}\left(\left|PH_i\left(\mathbf{x}_n\right)\right|\right)\frac{4}{D^2n^2}\end{aligned}$$

which converges to 0 as $n\to\infty$. Therefore, there is a M so that

$$\mathbb{P}\left(\left|PH_i\left(\mathbf{x}_n\right)\right|>Dn\right)<p/2$$

for all $n>M$.

Solving for ϵ in the expression in Lemma 12 gives that

$$\mathbb{P}\left(d_H\left(\left\{\mathbf{x}_n\right\},X\right)>\epsilon\left(n\right)/4\right)<p/2$$

if

$$\epsilon\left(n\right)=4A_0^{-1/d}n^{-1/d}W\left(\frac{2cA_0n}{p}\right)^{1/d}$$

where W is the Lambert W function. $W\left(m\right)\sim\log\left(m\right)$ as $m\to\infty$, and $W\left(m\right)\leq\log\left(m\right)$ for $m\geq e$. [28] Therefore, there is an $N_1>1/p$ and real number $A_3>0$ so that

$$(1/2)\,A_3n^{-1/d}\log\left(n\right)^{1/d}\leq\epsilon\left(n\right)\leq A_3n^{-1/d}\log\left(n\right)^{1/d}\tag{2}$$

for all $n>N_1$.

Choose $N_2>N_1$ to be sufficiently large so that $\epsilon\left(n\right)<\epsilon_1$ for all $n>N$, where ϵ_1 is given in Proposition 4. Let $n>N_2$ and suppose that $\mathbf{x}_n=\left\{x_1,\ldots,x_n\right\}$ satisfies $\left|\mathbf{x}\right|<Dn$ and $d_H\left(\mathbf{x},X\right)<\epsilon\left(n\right)/4$ — an event which occurs with probability greater than $1-p$. Write

$$E_{\alpha}^{i}\left(\mathbf{x}_n\right)=F_{\alpha}^{i}\left(\mathbf{x}_n,\epsilon\left(n\right)\right)+G_{\alpha}^{i}\left(\mathbf{x}_n,\epsilon\left(n\right)\right)$$

We consider the two terms separately:

$$\begin{aligned} G_\alpha^i\left(\mathbf{x}_n, \epsilon\left(n\right)\right) \leq & \ \left|\mathbf{x}\right| \epsilon\left(n\right)^\alpha \\ \leq & \ 2^\alpha D A_3^\alpha n^{\frac{d-\alpha}{d}} \log\left(n\right)^{\alpha/d} \\ = & \ O\left(n^{\frac{d-\alpha}{d}} \log\left(n\right)^{\frac{\alpha}{d}}\right) \end{aligned}$$

To bound the second term, we apply Lemma 9 to find

$$\begin{aligned} F_\alpha^i\left(\mathbf{x}, \epsilon\left(n\right)\right) \leq & \ 2^\alpha F_\alpha^i\left(X, \epsilon\left(n\right)/2\right) \\ \leq & \ A_2 \epsilon\left(n\right)^{\alpha-d} \log\left(1/\epsilon\left(n\right)\right) && \text{by Proposition 4} \\ \leq & \ A_2 A_3^{\alpha-d} n^{\frac{d-\alpha}{d}} \log\left(n\right)^{-\frac{d-\alpha}{d}} \log\left(1/2 A_3 n^{-1/d} \log\left(n\right)^{1/d}\right) && \text{by Equation 2} \\ = & \ O\left(n^{\frac{d-\alpha}{d}} \log\left(n\right)^{\frac{\alpha}{d}}\right) \end{aligned}$$

□

6. The Lower Bound

While our proof of the upper bounds works for Ahlfors regular measures on arbitrary metric spaces, here we restrict to the case of an Ahlfors regular measure on Euclidean space. This will allow us to use the additional structure of the cubical grid on $\mathbb{R}^m$.

To prove the lower bound, we combine the approach of our paper on extremal PH-dimension [40] with the probabilistic approach in [41]. If μ is a d-Ahlfors regular measure on $\mathbb{R}^m$ and $\delta > 0$, let $\mathcal{C}_\delta\left(\mu\right)$ be the cubes in the grid of mesh δ that intersect the support of μ. The basic idea is to sub-divide the grid of mesh δ so each cube contains k^m sub-cubes. If k is chosen carefully, we can find a positive fraction of cubes of $\mathcal{C}_\delta\left(\mu\right)$ that contain enough cubes of $\mathcal{C}_{\delta/k}\left(\mu\right)$ to guarantee a stable PH_i class. In fact, we can require that the sub-cubes have probability exeeding a certain threshold. We then control the number of stable PH_i classes realized by a random sample $\{x_1, \ldots x_n\}$ with Lemma 4.

First, we define the non-triviality constants γ_i^m:

6.1. **Non-triviality.** In previous work [40], we raised the question of how large a subset of the integer lattice can be without having a subset with "stable" i-dimensional persistent homology:

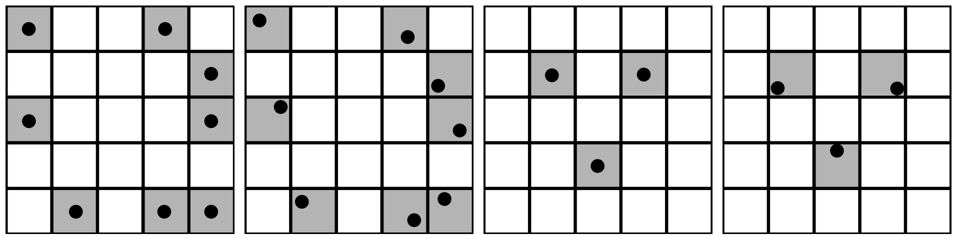

Figure 3. The PH_1 class of the lattice points corresponding to the gray cubes in (a) and (b) is stable — any choice of one point in each cube will result in a set with non-trivial PH_1. The one in (c) and (d) is not. [40]

Definition 3. *For $x \in \mathbb{Z}^m$, let the cube corresponding to x — $C(x)$ — be the Voronoi cell of x in $\mathbb{Z}^m \subset \mathbb{R}^m$. A subset X of $\mathbb{Z}^m$ has a* ***stable*** *i-dimensional persistent homology class if there is a $c > 0$ so that if $Y \subset \cup_{x \in X} C(x)$ satisfies*

$$Y \cap C(x) \neq \varnothing \quad \forall\, x \in X$$

then there is an $I \in PH_i(Y)$ so that $|I| > c$ (see Figure 3). The supremal such c is called the ***size*** *of the stable persistence class.*

Definition 4. *Let $\xi_i^m(N)$ be the size of the largest subset X of $\{1, \ldots, N\}^m \subset \mathbb{Z}^m$ so that no subset Y of X has a stable PH_i-class. Define*

$$\gamma_i^m = \liminf_{N \to \infty} \frac{\log\left(\xi_i^m(N)\right)}{\log(N)}$$

$\gamma_0^m = 0$ for all $m \in \mathbb{N}$: any subset of $\mathbb{Z}^m$ with more than 3^m points has a stable PH_0 class. In [40], we proved that $\gamma_1^m \leq m - \frac{1}{2}$ if persistent homology is taken with respect to the Čech complex. Note that we do not include the same restriction on the size as in that paper.

6.2. **Ahlfors Regular Measures and Box Counting.** Before proceeding to the proof of the lower bound, we prove a lemma about the asymptotics of the number of cubes that intersect the support of a d-Ahlfors regular measure. Let $\mathcal{C}_{\delta,a}(\mu)$ be the set of closed cubes C in the cubic grid of mesh δ in $\mathbb{R}^m$ centered at the origin that intersect X and satisfy

$$\mu(C) \geq a\delta^d$$

and let $N_{\delta,a}(\mu) = \left|\mathcal{C}_{\delta,a}(\mu)\right|$. (The upper and lower box dimensions of a subset of Euclidean space can be defined in terms of the asymptotic properties of $N_{\delta,0}(X)$).

Lemma 14. *If μ is a is d-Ahlfors regular measure with support $X \subset \mathbb{R}^m$, then there exit real numbers $0 < c_0 \leq c_1 < \infty$ depending on c, d, and m so that*

$$c_0 \delta^{-d} \leq N_{\delta,\hat{c}}(\mu) \leq c_1 \delta^{-d}$$

where $\hat{c} = \frac{1}{c 2^m}$ for all $\delta < \delta_0$. Similarly, there exist real numbers $0 < c_0' \leq c_1' < \infty$ depending on c, d, and m

$$c_0' \delta^{-d} \leq N_{\delta,0}(\mu) \leq c_1' \delta^{-d}$$

for all $\delta < \delta_0$.

Proof. Let C be a cube in the grid of mesh δ that intersects X, and $x \in C \cap X$. $\mu(B_\delta(x)) > 1/c\delta^d$ and $B_\delta(x)$ intersects at most 2^m cubes in the grid of mesh δ, so at least one one cube adjacent to C has measure exceeding $\hat{c}\delta^d$ (where two cubes are adjacent if they share at least one point). Also, each cube of $\mathcal{C}_{\delta,a}(\mu)$ is adjacent to at most 3^m cubes of $\mathcal{C}_\delta(\mu)$. It follows that

$$\frac{1}{3^m} N_{\delta,0}(\mu) \leq N_{\delta,\hat{c}}(\mu) \leq N_{\delta,0}(\mu)$$

and the existence of bounds for $N_{\delta,0}(\mu)$ implies bounds for $N_{\delta,\hat{c}}(\mu)$, and visa versa.

We have that

$$\begin{aligned} 1 = & \ \mu(X) \\ \leq & \sum_{C \in \mathcal{C}_\delta(mu)} \mu(C) \\ \leq & \ c\delta^d m^{d/2} N_{\delta,0}(\mu) \\ \leq & \ 3^m c\delta^d m^{d/2} N_{\delta,\hat{c}}(\mu) \\ \implies & \ N_{\delta,\hat{c}}(\mu) \geq 3^{-m} m^{-d/2} \delta^{-d} \end{aligned}$$

For the upper bound, note that the intersection of two cubes may have positive measure, but a cube can share measure with only $3^m - 1$ adjacent cubes. It follows that

$$\begin{aligned} 1 = & \ \mu(X) \\ \geq & \ \frac{1}{3^m} \hat{c}\delta^d N_{\delta,\hat{c}}(\mu) \\ \implies & \ N_{\delta,\hat{c}}(\mu) \leq c\, 6^m \delta^{-d} \end{aligned}$$

□

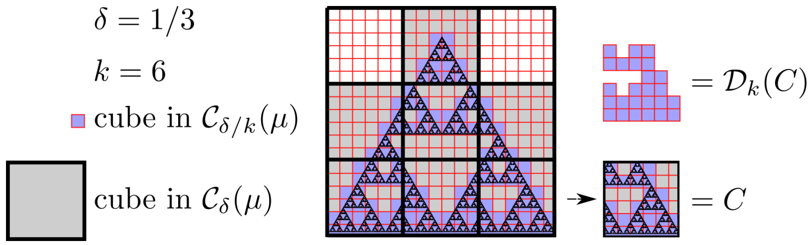

Figure 4. Some of cubes associated with the natural measure μ on the Sierpiński triangle.

For each $k \in \mathbb{N}$, $\delta > 0$, and $C \in \mathcal{C}_\delta(\mu)$, let $\mathcal{D}_k(C)$ be the set of cubes in $\mathcal{C}_{\delta/k,\hat{c}} paren \mu$ that are contained in C, and let $D_k(C) = |\mathcal{D}_k(C)|$. See Figure 4.

Lemma 15. *Let* $0 < \beta < d$ *and let*

$$\mathcal{C}_\delta^{k,\beta} = \left\{ C \in \mathcal{C}_\delta(\mu) : D_k(C) > k^{-\beta} \right\}$$

and

$$M(\delta, k, \beta) = \left| \mathcal{C}_\delta^{k,\beta} \right|$$

There exists a $K > 0$ *so that for all* $k > K$ *there exist* $\delta_1, c_2 > 0$ *so that*

$$M(\delta, k, \beta) > c_2 \delta^{-d}$$

for all $\delta < \delta_1$.

Proof. Let c_0, c_1', and δ be the constants from the previous lemma so $N_\delta(\mu) \leq c_1' \delta^{-d}$ and $N_{\delta,\hat{c}}(\mu) \geq c_0 \delta^{-d}$ for all $\delta < \delta_1$.

$D_k(C) \leq k^m$ for all $C \in \mathcal{C}_\delta^{k,\beta}$ so $N(\delta, k, \beta)$ is bounded below by the smallest integer $a(k, \delta)$ satisfying

$$a(k,\delta) k^m + \left(c_1' \delta^{-d} - a(k,\delta) \right) k^\beta \geq c_0 k^m \delta^{-d}$$

Rearranging terms, we have that

$$a(k,\delta) = \lceil \frac{\delta^{-d} \left(c_0 k^{m-\beta} - c_1' \right)}{k^{m-\beta} - 1} \rceil$$

Let

$$K = \left(\frac{c_1'}{c_0} \right)^{\frac{1}{m-\beta}}$$

so both the numerator and the denominator of the previous expression are positive for $k > K$. Let $k > K$ and set

$$c_2 = \frac{1}{2}\frac{\left(c_0 k^{m-\beta} - c_1'\right)}{k^{m-\beta} - 1}$$

so

$$a\left(k, \delta\right) \sim 2c_2\delta^{-d}$$

as $\delta \to 0$. It follows that

$$N\left(\delta, k, \beta\right) \geq a\left(k, \delta\right) > c_2\delta^{-d}$$

for all sufficiently small δ, as desired. □

6.3. **Proof of the Lower Bound.** We require one more lemma before proving the lower bound. The idea is similar to that of Lemma 8.

Lemma 16. *Let μ be a d-Ahlfors regular measure on $\mathbb{R}^m$ with $d > \gamma_i^m$. There exist positive real numbers ϵ_0 and Ω_0 so that*

$$\lim_{n\to\infty} \frac{1}{n} p_i\left(\boldsymbol{x}_n, \epsilon_0\, n^{-1/d}\right) \geq \Omega_0$$

in probability.

Proof. Let $\gamma_i^m < \beta < d$. By the definition of γ_i^m we can find a K_0 so that $k^\beta > \xi_i^m\left(k\right)$ for all $k > K_0$. Let $k_0 > \min\left(K, K_0\right)$, where K is given in the previous lemma, and let δ_1 and c_2 also be as in the previous lemma. There are only finitely many collections of sub-cubes of $\left[k_0\right]^m$, so there are only finitely many possible stable PH_i classes of subsets of $\left[k_0\right]^m$. Let ϵ_0 be the minimum of the sizes of these stable classes.

Let $\delta = n^{-1/d}$ and choose n large enough so that $\delta < \delta_1$. Also, let $\{D_1 \dots D_s\}$ be a maximal collection of cubes in $\mathcal{C}_\delta^{k_0,\beta}$ so that $d\left(D_j, D_k\right) > \delta\sqrt{m}$ for all $j, k \in \{1, \dots, s\}$ so that $j \neq k$. See Figure 5. There is a constant $0 < \kappa < 1$ that depends only on d so that

$$s \geq \kappa N\left(\delta, k, \beta\right) > \kappa c_2 \delta^{-d} = \kappa c_2 n$$

Let $l \in \{1, \dots, s\}$. By the definition of γ_i^m, there is a collection of sub-cubes $\mathcal{B}_l \subset \mathcal{D}_{k_0}\left(D_l\right)$ with a stable PH_i class. Let

$$A_l = \hat{B}_{\delta\sqrt{m}}\left(C\right) \setminus \cup_{B\in\mathcal{B}_l} B$$

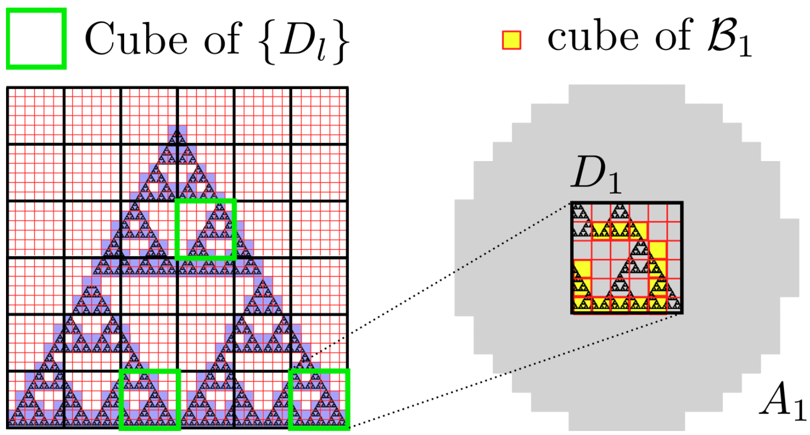

Figure 5. The setup in the proof of Lemma 16.

where $\hat{B}_{\delta\sqrt{m}}\left(D_j\right)$ is the union of all cubes in the grid of mesh δ/k within distance $\delta\sqrt{m}$ of D_j. Also, let $\mathcal{B}'_n$ be collection of the interiors of the sets $\mathcal{B}_l$. It follows from property (2) in Section 4.1 that

$$p_i\left(\mathbf{x}_n, \epsilon_0\, n^{-1/d}\right) \geq \frac{1}{n}\sum_{j=1}^{r} \Xi\left(\mathbf{x}_n, A_l, \mathcal{B}'_l\right)$$

Let $l \in \{1, \ldots, s\}$. There is a $q > 0$ depending only on k_0, c, d, and m so that

$$\mu\left(A_l\right) \leq q\delta^d = \frac{q}{n}$$

Also, each $B \in \mathcal{B}_l$ is an open cube of width δ/k_0 in $\mathbb{R}^m$ so

$$\mu\left(B\right) \geq \frac{1}{c}\left(\frac{\delta\sqrt{m}}{k_0}\right)^d = \frac{p}{n}$$

where $p = k_0^{-d}m^{d/2}/c$. Therefore, $\Xi\left(\mathbf{x}_n, A_l, \mathcal{B}'_l\right)$ is a n, p, q, k^m-bounded occupancy event for each l, and the desired result follows from Lemma 4. ☐

The proof of the lower bound is now straightforward:

Proposition 6. *Let μ be a d-Ahlfors regular measure on $\mathbb{R}^m$ with $d > \gamma_i^m$. There is an $\Omega > 0$ so that*

$$\lim_{n\to\infty} n^{-\frac{d-\alpha}{d}} E_\alpha^i\left(x_1, \ldots, x_n\right) \geq \Omega$$

in probability.

Proof. It follows immediately from the previous lemma that

$$\begin{aligned}
\lim_{n\to\infty} n^{-\frac{d-\alpha}{d}} E^i_\alpha(\mathbf{x}_n) \geq \\
& \lim_{n\to\infty} n^{-\frac{d-\alpha}{d}} p_i\left(\mathbf{x}_n, \epsilon_0 n^{-1/d}\right)\left(\epsilon_0 n^{-1/d}\right)^\alpha \\
&= \epsilon_0^\alpha \lim_{n\to\infty} \frac{1}{n} p_i\left(\mathbf{x}_n, \epsilon_0 n^{-1/d}\right) \\
&\geq \epsilon_0^\alpha \Omega_0 && \text{by Lemma 16} \\
&:= \Omega
\end{aligned}$$

in probability. □

7. Acknowledgments

We would like to thank Henry Adams and Matthew Kahle for interesting discussions. In particular, the computational experiments of Adams et al. [1] helped to inspire this work. Funding for this research was provided by a NSF Mathematical Sciences Postdoctoral Research Fellowship under award number DMS-1606259.

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. arXiv:1808.01079, 2018.

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

[3] James W. Baish and Rakesh K. Jain. Fractals and cancer. *Perspectives in Cancer Research*, 2000.

[4] Daniel Barbará and Ping Chen. Using the fractal dimension to cluster datasets. In *KDD '00 Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2000.

[5] Ulrich Bauer and Florian Pausinger. Persistent Betti numbers of random Čech complexes. arXiv:1801.08376, 2018.

[6] A. Beurling and L. Ahlfors. The boundary correspondence under quasiconformal mappings. *Acta Mathematica*, 1956.

[7] Omer Bobrowski and Matthew Kahle. Topology of random geometric complexes: a survey. *Journal of Applied and Computational Topology*, 2018.

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

[9] Francesca Cagliaria and Claudia Landibc. Finiteness of rank invariants of multidimensional persistent homology groups. *Applied Mathematics Letters*, 2011.

[10] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 2009.
[11] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. *The Structure and Stability of Persistence Modules*. Springer, 2016.
[12] Frédéric Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 2014.
[13] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1), 2007.
[14] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l_p-stable persistence. *Foundations of Computational Mathematics*, 2010.
[15] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and its Applications*, 2015.
[16] Guy David and Stephen Semmes. *Analysis of and on Uniformly Rectifiable Sets*. American Mathematical Soceity, 1993.
[17] S. Davies and P. Hall. Fractal analysis of surface roughness by using spatial data. *Journal of the Royal Statistical Society Series B*, 1999.
[18] Vincent Divol and Wolfgang Polonik. On the choice of weight functions for linear representations of persistence diagrams. arXiv:1807.03678, July 2018.
[19] Trinh Khanh Duy, Yasuaki Hiraoka, and Tomoyuki Shirai. Limit theorems for persistence diagrams. arXiv:1612.08371, December 2016.
[20] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplificitaion. *Discrete and Computational Geometry*, 2002.
[21] Herbert Edelsbrunner and John Harer. Persistent homology — a survey. *Contemporary Mathematics*, 2008.
[22] Herbert Edeslbrunner and Dmitriy Morozov. Persistent homology: Theory and practice. Technical Report LBNL-6037E, Lawrence Berkeley National Laboratory, 2013.
[23] Kenneth Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, 2014.
[24] Robert Ghrist. Barcodes: the persistent homology of data. *Bulletin of the American Mathematical Society*, 2008.
[25] Michael Goff. Extremal Betti numbers of Vietoris-Rips complexes. *Discrete and Computational Geometry*, 2011.
[26] J. M. Halley, S. Hartley, A. S. Kallimanis, W. E. Kunin, J. J. Lennon, and S. P. Sgardelis. Uses and abuses of fractal methodology in ecology. *Ecology Letters*, 2004.
[27] Yasuaki Hiraoka and Tomoyuki Shirai. Minimum spanning acycle and lifetime of persistent homology in the Linial—Meshulam process. *Random Structures & Algorithms*, 2017.
[28] Abdolhossein Hoorfar and Mehdi Hassani. Inequalities on the Lambert W function and hyperpower function. *Journal of Inequalities in Pure and Applied Mathematics*, 2008.
[29] Harry Kesten and Sungchul Lee. The central limit theorem for weighted minimal spanning trees on random points. *Annals of Applied Probability*, 1996.
[30] G. Kozma, Z. Lotker, and G. Stupp. On the connectivity threshold for general uniform metric spaces. *Information Processing Letters*, 2010.
[31] Gady Kozma, Zvi Lotker, and Gideon Stupp. The minimal spanning tree and the upper box dimension. *Proceedings of the American Mathematical Society*, 2006.

[32] R. Lopes and N. Betrouni. Fractal and multifractal analysis: A review. *Medical Image Analysis*, 2009.

[33] John Mackay and Jeremy Tyson. *Conformal Dimension: Theory and Application*. American Mathematical Soceity, 2010.

[34] R. D. MacPherson and B. Schweinhart. Measuring shape with topology. *Journal of Mathematical Physics*, 53(7), 2012.

[35] Benoît Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Co., 1982.

[36] Benoit B Mandelbrot. *Fractals: Form, Chance and Dimension*. W.H.Freeman & Company, 1977.

[37] Gabriell Máté and Dieter W. Heermann. Persistence intervals of fractals. *Physica A*, 2014.

[38] P. Niyogi, S. Smale, and S.Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete and Computational Geometry*, 2008.

[39] Vanessa Robins. *Computational topology at multiple resolutions: foundations and applications to fractals and dynamics*. PhD thesis, University of Colorado at Boulder, 2000.

[40] Benjamin Schweinhart. Persistent homology and the upper box dimension. arXiv:1802.00533, 2018.

[41] Benjamin Schweinhart. Weighted persistent homology sums of random Čech complexes. arXiv:1807.07054, July 2018.

[42] R. Sikorski. *Real Functions, Vol. 1*. PWN: Warsaw, 1958.

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

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

[45] Caetano Traina, Agma Traina, Leejay Wu, and Christos Faloutsos. Fast feature selection using fractal dimension. *Journal of Information and Data Management*, 2010.

[46] Rien van de Weygaert, Bernard J.T. Jones, and Vincent J. Martinez. The minimal spanning tree as an estimator for generalized dimensions. *Physics Letters A*, 1992.

[47] D. Yogeshwaran, Eliran Subag, and Robert J. Adler. Random geometric complexes in the thermodynamic regime. *Probability Theory and Related Fields*, 2017.

[48] Boming Yu. Analysis of flow in fractal porous media. *Applied Mechanics Reviews*, 2008.

[49] J.E. Yukich. Asymptotics for weighted minimal spanning trees on random points. *Stochastic Processes and their Applications*, 2000.

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