THE PERSISTENT HOMOLOGY OF RANDOM GEOMETRIC COMPLEXES ON FRACTALS

BENJAMIN SCHWEINHART

Abstract. We prove that 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 \( x_1, \ldots, x_n \) are i.i.d. samples from a \( d \)-Ahlfors regular measure on a metric space, and \( E_\alpha(x_1, \ldots, x_n) \) denotes the \( \alpha \)-weight of the minimum spanning tree on \( x_1, \ldots, x_n \):
\[
E_\alpha(x_1, \ldots, x_n) = \sum_{e \in T(x_1, \ldots, x_n)} |e|^\alpha,
\]
then
\[
E_\alpha(x_1, \ldots, x_n) \approx n^{d-\alpha/\delta}
\]
with high probability as \( n \to \infty \). In particular,
\[
\log \left( E_\alpha(x_1, \ldots, x_n) \right) / \log(n) \to (d - \alpha)/d.
\]
This is a generalization of a result of Steele [50] from the absolutely continuous case to the fractal setting. We also prove analogous results for weighted sums defined in terms of higher dimensional persistent homology.

1. Introduction

The first precise notion of a fractional dimension was proposed by Hausdorff in 1918 [31, 25]. Since then, many other definitions have been put forward, including the box-counting [10] and correlation [29] dimensions. These quantities do not agree in general, but coincide on a class of regular sets. Fractal dimension was popularized by Mandelbrot in the 1970s and 1980s [42, 41], and it has since found a wide range of applications in subjects including medicine [3, 38], ecology [30], materials science [19, 55], and the analysis of large data sets [4, 52]. It is also important in pure mathematics and mathematical physics, in disciplines ranging from dynamics [51] to probability [6].

Date: January 2019.
Recently, there has been a surge of interest in applications of topology, and of persistent homology in particular. Several authors have proposed estimators of fractal dimension defined in terms of minimum spanning trees and higher dimensional persistent homology \cite{50, 53, 45, 40, 43, 1}, and provided empirical evidence that those quantities agreed with classical notions of fractal dimension. In Theorem 8 below, we provide the first rigorous justification for the use of minimum spanning trees and higher dimensional persistent homology to estimate fractal dimension.

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

\[
E_n^i (x_1, \ldots, x_n) = \sum_{I \in PH_i(x_1, \ldots, x_n)} |I|^\alpha,
\]

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

\[
E_0^\alpha (x_1, \ldots, x_n) = \sum_{e \in T(x_1, \ldots, x_n)} |e|^\alpha,
\]

then

\[
E_\alpha (x_1, \ldots, x_n) = E_0^\alpha (x_1, \ldots, x_n)
\]

where persistent homology is taken with respect to the Vietoris–Rips complex. In 1988, Steele \cite{50} proved the following celebrated result.

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

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

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

Steele wrote \cite{50}:

One feature of Theorem\[1\] that should be noted is that if \( \mu \) has bounded support and \( \mu \) is singular with respect to Lebesgue measure, then we
have with probability one that \( E_0^\alpha (x_1, \ldots, x_n) = o \left( n^{(d-\alpha)/d} \right) \). Part of the appeal of this observation is the indication that the length of the minimum spanning tree is a measure of the \textit{dimension} of the support of the distribution. This suggests that the asymptotic behavior of the minimum spanning tree might be a useful adjunct to the concept of dimension in the modeling applications and analysis of fractals; see, e.g., [42].

However, despite many subsequent sharper and more general results for non-singular measures [2, 35, 56], little is known about the asymptotic properties of random minimum spanning trees for fractal measures. As far as we know, the only previous result toward that end is that of Kozma, Lotker and Stupp [36], who proved that if \( \mu \) is a \( d \)-Ahlfors regular measure with connected support, then the length of the longest edge of a minimum spanning tree on \( n \) i.i.d. points sampled from \( \mu \) is \( \approx (\log (n)/n)^{1/d} \), where the symbol \( \approx \) 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 analogous asymptotics hold for the alpha-weight of a minimum spanning tree, which we prove here in Theorem 3.

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 [8, 9, 54, 21, 5]. In 2018, we [47] proved results about the asymptotics of \( E_\alpha^0 (x_1, \ldots, x_n) \) 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 [20] 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 [53] proposed using the asymptotics of \( E_\alpha^0 (x_1, \ldots, x_n) \) for negative \( \alpha \) to estimate the generalized Hausdorff dimensions. The PhD thesis of Robins, which was arguably one of the first publications in the field of topological data analysis, studied the scaling of Betti numbers of fractals and proved results for the 0-dimensional persistent homology of disconnected sets [45]. In joint work with Robert MacPherson, we proposed a dimension for probability distributions of geometric objects based on persistent homology in 2012 [40]. 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
Figure 1. Two sets of fractional dimension, and their \( \epsilon \)-neighborhoods: (a) a modified Sierpiński triangle and (b) a branched polymer. Their complex geometry is reflected by their persistent homology.

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 8).

In the extremal setting, Kozma, Lotker and Stupp 37 defined a minimum spanning tree dimension for a metric space \( M \) in terms of the behavior of \( E_0^0(Y) \) 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 persistent and established hypotheses under which it agrees with the upper box dimension 46. In the course of this work, we investigated extremal questions about the number of persistent homology intervals of a set of \( n \) points; these questions are also important in the probabilistic context, as we describe below.

We are currently working on a separate manuscript which compares the practical performance of the persistent homology dimension defined in Definition 7 below to classical techniques for estimating fractal dimension, such as box-counting and the estimation of the correlation dimension. Preliminary results indicate that the persistent homology dimension (for \( i = 0 \)) provides a dimension estimate with lower variance than that of the correlation dimension. 34
1.1. Our Results. We prove analogues of the theorem of Steele [50] for probability measures defined on sets of fractional dimension that satisfy a certain regularity condition:

Definition 2. A probability measure $\mu$ supported on a metric space $X$ is $d$-Ahlfors regular if there exist positive real numbers $c$ and $\delta_0$ so that

$$\frac{1}{c} \delta^d \leq \mu (B_\delta (x)) \leq c \delta^d$$

for all $x \in X$ and $\delta < \delta_0$, where $B_\delta (x)$ denotes the open ball of radius $\delta$ centered at $x$.

Ahlfors regularity is a common hypothesis when studying analysis on fractals [18, 7, 39]. Examples 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 $\mu$ 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 somewhat weaker than Ahlfors regularity. In particular, the proofs of our upper bounds only require that $M_\delta (\mu) = O (\delta^{-d})$, where $M_\delta (\mu)$ is the maximal number of disjoint balls of radius $\delta$ centered at points of $\text{supp} \mu$. 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 $\mu$.

Our main result is:

**Theorem 3.** Let $\mu$ be a $d$-Ahlfors regular measure on a metric space, and let $\{x_n\}_{n \in \mathbb{N}}$ be i.i.d. samples from $\mu$. If $0 < \alpha < d$, then

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

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 minimum spanning trees (rather than persistent homology) in Section 3. The special case where $\mu$ is a measure on Euclidean space is also a consequence of Theorem 4 below.
As we noted in our earlier paper [46], proving asymptotic 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 minimum spanning tree on $n$ points always has $n - 1$ edges, a set of $n$ points may have trivial $PH_i$ for all $i > 0$, and there exist families of finite metric spaces for which the number of persistent homology intervals grows faster than linearly in the number of points.

To prove upper bounds for the asymptotics of $E_\alpha^i$ 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 [46, 28], but are considered somewhat pathological. As far as we know, the Upper Bound Theorem [49] 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 $\mathbb{R}^m$:

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

For the Vietoris–Rips complex of points in Euclidean space, we [46] showed that

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

by modifying an argument of Goff [28].

A different extremal question arises in the process of proving lower bounds for $E_\alpha^i$. In particular, a subset $\mathbb{R}^m$ must have dimension above a certain non-triviality constant $\gamma_m^i$ (defined in Section 6.1) to guarantee the existence of subsets with non-trivial $i$-dimensional persistent homology. We showed that $\gamma_m^i < m - 1/2$ in our previous paper [46].

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 4.** Let $\mu$ be a $d$-Ahlfors regular measure on $\mathbb{R}^m$ with $d > \gamma_m^i$, and let $\{x_n\}_{n \in \mathbb{N}}$ be i.i.d. samples from $\mu$. If there are positive real numbers $D$ and $a$ so that

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

for all finite subsets of $X$, and $0 < \alpha < ad$, then there are real numbers $0 < \zeta < Z$ so that
with high probability, as \( n \to \infty \). In fact, the upper bound holds with probability one.

The upper bound is shown in Proposition 22 and the lower bound in Proposition 33. The following is a corollary.

Corollary 5. Let \( \mu \) be a \( d \)-Ahlfors regular measure on \( \mathbb{R}^2 \), and let \( \{x_n\}_{n \in \mathbb{N}} \) be i.i.d. samples from \( \mu \). If \( 0 < \alpha < d \). If \( d > 1.5 \) and \( 0 < \alpha < d \), and persistent homology is taken of the Čech complex, then

\[
E^i_\alpha (x_1, \ldots, x_n) \approx n^{d-\alpha}
\]

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 6. Let \( \mu \) be a \( d \)-Ahlfors regular measure on \( \mathbb{R}^m \) so that \( d > \gamma_i \), and let \( \{x_n\}_{n \in \mathbb{N}} \) be i.i.d. samples from \( \mu \). If

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

and \( 0 < \alpha < d \), then there are real numbers \( 0 < \lambda < \Lambda \) so that

\[
\lambda n^{d-\alpha} \leq E^i_\alpha (x_1, \ldots, x_n) \leq \Lambda n^{d-\alpha} \log (n)^2
\]

with high probability, as \( n \to \infty \).

The upper and lower bounds are shown in Propositions 27 and 33, 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 a classical fractal dimension, under certain hypotheses.

We define a family of \( PH_i \) dimensions of a measure, one for each real number \( \alpha > 0 \) and \( i \in \mathbb{N} \):
Definition 7.

\[ \dim_{PH}^\alpha (\mu) = \frac{\alpha}{1 - \beta}, \]

where

\[ \beta = \limsup_{n \to \infty} \frac{\log \left( \mathbb{E} \left( E^i_\alpha (x_1, \ldots, x_n) \right) \right)}{\log (n)}. \]

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

\[ \limsup_{n \to \infty} \mathbb{E} \left( E^i_\alpha (x_1, \ldots, x_n) \right) n^{-\frac{k-\alpha}{k}} \]

equals \( \infty \) 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.

Theorem 1 [50] implies that if \( \mu \) is a compactly supported, non-singular probability measure on \( \mathbb{R}^m \), then \( \dim_{PH}^\alpha (\mu) = m \) for \( 0 < \alpha < m \). Similar, the results of Divol and Polonik [20] show that if \( \mu \) is a bounded probability measure on the cube in \( \mathbb{R}^m \), then \( \dim_{PH}^\alpha (\mu) = m \) for \( 0 < \alpha < m \) and \( 0 \leq i < m \).

The following is a corollary of our theorems on the asymptotic behavior of \( E^i_\alpha \):

Theorem 8. If \( \mu \) is a \( d \)-Ahlfors regular measure on a metric space and \( 0 < \alpha < d \) then

\[ \dim_{PH}^\alpha = d. \]

Furthermore, if \( \mu \) is defined on \( \mathbb{R}^m \), \( d > \gamma_i^m \), and

\[ \mathbb{E} \left( |PH (x_1, \ldots, x_n)| \right) = O (n) \quad \text{and} \quad \var{ |PH (x_1, \ldots, x_n)| } /n^2 \to 0, \]

then

\[ \dim_{PH}^\alpha = d. \]

This result is weaker than our main theorems, and it can be proven with weaker hypotheses. For example, the upper bound \( \dim_{PH}^\alpha \leq d \) holds if the hypothesis of \( d \)-Ahlfors regularity is replaced by the requirement that the upper-box dimension of the support of \( \mu \) is less than or equal to \( d \).
1.3. A Conjecture. We conjecture that if the persistent homology of the support of an Ahlfors regular measure is trivial, then the Lebesgue measure can be replaced with the $d$-dimensional Hausdorff measure $\mathcal{H}^d$ in Theorem 1.

Conjecture 9. Let $\mu$ be a $d$-Ahlfors regular measure on a metric space $M$ and let $\{x_n\}_{n \in \mathbb{N}}$ be i.i.d. samples from $\mu$. If $\mathrm{PH}_0(\text{supp } \mu)$ is trivial and $0 < \alpha < d$, then

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

with probability one, where $f(x)$ is the probability density of the absolutely continuous part of $\mu$ with respect to the $d$-dimensional Hausdorff measure $\mathcal{H}^d$ and and $c(\alpha, d)$ is a continuous function of $\alpha$ and $d$.

The support of a fractal measure may have non-trivial persistent homology at infinitesimal length-scales, which could contribute to the asymptotic behavior of the $\alpha$-weighted sums. One might speculate that this could lead to oscillations that would preclude the existence of a limiting constant. However, experimental computations with the Cantor set and Sierpinski triangle are consistent with the existence of a limit (at least for some values of $\alpha$), even though the limit

$$\lim_{\epsilon \to 0} \epsilon^{\alpha-d} \sum_{I \in \mathrm{PH}_0(\text{supp } \mu), |I| > \epsilon} |I|^\alpha$$

does not exist.

2. Preliminaries

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

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

Lemma 10 (Ball-counting Lemma). If $\mu$ is a $d$-Ahlfors regular measure supported on a metric space $X$ then

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

for all $\delta < \delta_0$, where $c$ and $\delta_0$ are the constants given in the definition of Ahlfors regularity.
Proof. Let \( \{x_j\}^{M_\delta(X)}_{j=1} \) be the centers of a maximal set of disjoint balls of radius \( \delta \) centered at points of \( X \).

\[
1 = \mu(X) \\
\geq \sum_{j=1}^{M_\delta(\mu)} \mu(B_\delta(x_j)) \quad \text{by disjointness} \\
\geq \frac{1}{c} \delta^d M_\delta(\mu) \quad \text{by Ahlfors regularity} \\
\implies M_\delta(\mu) \leq c\delta^{-d}.
\]

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

\[
1 = \mu(X) \\
\leq \sum_{j=1}^{M_\delta(X)} \mu(B_{2\delta}(x_j)) \\
\leq c2^d \delta^d M_\delta(X) \quad \text{by Ahlfors regularity} \\
\implies M_\delta(X) \geq \frac{1}{c} 2^{-d} \delta^{-d},
\]

as desired. \( \square \)

We also require the following lemma of Cohen-Steiner et al. [16].

**Lemma 11.** 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 \( \epsilon > 0 \) then

\[
\sum_{j \in J_\epsilon} j^\alpha \leq c^\alpha f(\epsilon) + \alpha \int^{\max J}_{\delta=\epsilon} f(\delta) \delta^{\alpha-1} d\delta
\]

for all \( \alpha > 0 \). Furthermore, if \( |J| \leq f(0) < \infty \) then

\[
\sum_{j \in J} j^\alpha \leq \alpha \int^{\max J}_{\delta=0} f(\delta) \delta^{\alpha-1} d\delta.
\]
For completeness, we reproduce the proof in [16]. \( \sum_{j \in J} j^\alpha \) can be expressed as an integral involving the distributional derivative of \( |J| \). Applying integration by parts yields:

\[
\sum_{j \in J} j^\alpha = \int_{\delta = \epsilon}^{\infty} -\frac{\partial |J_\delta|}{\partial \delta} \delta^\alpha \, d\delta
\]

\[
= \left\lfloor -|J_\delta| \delta^\alpha \right\rfloor_{\delta = \epsilon}^{\infty} + \alpha \int_{\delta = \epsilon}^{\infty} |J_\delta| \delta^{\alpha-1} \, d\delta
\]

\[
= e^\alpha |J_\epsilon| + \alpha \int_{\delta = \epsilon}^{\max J} |J_\delta| \delta^{\alpha-1} \, d\delta
\]

\[
\leq e^\alpha f(\epsilon) + \alpha \int_{\delta = \epsilon}^{\max J} f(\delta) \delta^{\alpha-1} \, d\delta .
\]

2.1. Notation. In the following, \( X \) will denote a metric space and \( x \) will denote a finite point set with an unspecified number of elements. Furthermore, \( x_n \) will be shorthand for a finite point set \( \{x_1, \ldots, x_n\} \subset X \) containing \( n \) points. If the measure \( \mu \) is obvious from the context, \( \{x_j\}_{j \in \mathbb{N}} \) will be a collection of independent random variables with common distribution \( \mu \). 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 for the asymptotic behavior of \( E_i^\alpha(x_1, \ldots, x_n) \) will be to define certain occupancy events that imply the existence of a persistent homology interval (or minimum spanning tree edge) whose length is bounded away from zero.

If \( A \) and \( B \) are sets define

\[
\delta(A, B) = \begin{cases} 
0 & A \cap B = \emptyset \\
1 & A \cap B \neq \emptyset .
\end{cases}
\]

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

\[
\Xi(x, A, \mathcal{B}) = \begin{cases} 
1 & \delta(A, x) = 0 \quad \text{and} \quad \delta(B, x) = 1 \quad \forall B \in \mathcal{B} \\
0 & \text{otherwise}
\end{cases}
\]

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

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

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

$$\Xi(x_n, A, B),$$

where $B$ is a collection of at least $r$ sets, and $x_n$ is a collection of $n$ independent random variables with common distribution $\nu$ satisfying

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

If the above conditions on $\nu$ and the number of sets in $B$ hold with equality, we say that $\Xi(x_n, A, B)$ is an $n, p, q, r$-**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 12.** Let $r, a > 0$, and $0 < p, q < 1$. Also, for each $n \in \mathbb{N}$ let $X^n_1, \ldots, X^n_{\lfloor an \rfloor}$ be disjoint $n, p, q, r$-uniform occupancy events. If $Y_n = \frac{1}{n} \sum_{j=1}^{\lfloor an \rfloor} X^n_j$, then

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

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

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

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

by inclusion-exclusion. Therefore

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

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

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

$$\lim_{n \to \infty} \mathbb{E}(X^n_j X^n_k) = e^{-2q} \left( 1 - e^{-p} \right)^{2r}.$$
It follows that
\[
\lim_{n \to \infty} \text{Cov}
\left(X^n_j, X^n_k\right) = \lim_{n \to \infty} \left(\mathbb{E}
\left(X^n_j X^n_k\right) - \mathbb{E}
\left(X^n_j\right) \mathbb{E}
\left(X^n_k\right)\right) = 0.
\]

Therefore
\[
\text{Var}
\left(Y_n\right) = \frac{1}{n^2} \left(\sum_{j=1}^{\lfloor an \rfloor} \text{Var}
\left(X^n_j\right) + 2 \sum_{j=1}^{\lfloor an \rfloor} \sum_{i=1}^{j-1} \text{Cov}
\left(X^n_j, X^n_k\right)\right)
\sim \frac{a}{n} \text{Var}
\left(X^n_1\right) + a \frac{n^2 - n}{n^2} \text{Cov}
\left(X^n_1, X^n_2\right)
\leq \frac{a}{n} + a \left(1 - \frac{1}{n}\right) \text{Cov}
\left(X^n_1, X^n_2\right)
\]
also converges to 0 as \(n\) goes to \(\infty\).

Let \(\epsilon > 0\) and \(0 < \rho < 1\). Choose \(N\) sufficiently large so that
\[
\left|\mathbb{E}
\left(Y_n\right) - \gamma\right| < \epsilon/2 \quad \text{and} \quad \text{Var}
\left(Y_n\right) < \frac{\epsilon^2 \rho}{4}
\]
for all \(n > N\). If \(n > N\),
\[
\mathbb{P}
\left(|Y_n - \gamma| > \epsilon\right) \leq \mathbb{P}
\left(|Y_n - \mathbb{E}
\left(Y_n\right)| > \epsilon/2\right)
\leq \mathbb{P}
\left(|Y_n - \mathbb{E}
\left(Y_n\right)| > \frac{1}{\sqrt{\rho}} \sqrt{\text{Var}
\left(Y_n\right)}\right)
\leq \rho
\]
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 13.** Let \(r, a > 0\), \(0 < p, q < 1\), and \(s_n \geq \lfloor an \rfloor\) for all \(n \in \mathbb{N}\). Also, for each \(n \in \mathbb{N}\) let \(X^n_1, \ldots, X^n_{s_n}\) be disjoint \(n, p, q, r\)-bounded occupancy events. Under these hypotheses, there is a \(\gamma > 0\) so that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{s_n} X^n_j \geq \gamma
\]
in probability.
Proof. Let $a_0 = \min \left(a, \frac{1}{(p + q)}\right)$, $1 \leq j \leq \lfloor a_0 n \rfloor$, and

$$X^n_j = \Xi \left(x_n, A^n_j, B^n_j\right).$$

As $\nu$ is non-atomic, so we can find a subset $\hat{B}$ of each set $B \in B^n_j$ so that $\nu \left(\hat{B}\right) = p/n$. Let

$$\hat{B}^n_j = \left\{ \hat{B} : B \in B^n_j \right\} \quad \text{and} \quad D_n = \bigcup_{j=1}^{\lfloor a_0 n \rfloor} \bigcup_{\hat{B} \in \hat{B}^n_j} \hat{B}.$$

Similarly, there are disjoint sets $\hat{A}^n_1, \ldots, \hat{A}^n_{\lfloor a_0 n \rfloor}$ so that $A^n_j \subseteq \hat{A}^n_j \subseteq D_n^c$ and $\nu \left(\hat{A}^n_j\right) = q/n$ for $j = 1, \ldots, \lfloor a_0 n \rfloor$, where we have used that $a_0 (p + q) \leq 1$. Let

$$\hat{X}_j^n = \Xi \left(x_n, \hat{A}_j^n, \hat{B}_j^n\right).$$

By construction, $X_j^n = 1 \implies \hat{X}_j^n = 1$ so $\frac{1}{n} \sum_{j=1}^{\lfloor a_0 n \rfloor} X_j^n$ 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. \qed

3. The Proof for Minimum Spanning Trees

If $x$ is a finite metric space, $T(x)$ denote the minimum spanning tree on $x$, and let $p(x, \epsilon)$ be the number of edges of $T(x)$ of length greater than $\epsilon$. Also, let $G_{x, \epsilon}$ be graph the with vertex set $x$ 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 on $x$). The following is a corollary of Kruskal’s algorithm.

**Lemma 14.**

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

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

3.1. Proof of the Upper Bound. Our strategy to prove an upper bound for the asymptotics of $E^0_m(x_n)$ is to control the number of edges in $T(x_n)$ of length greater than $\delta$ in terms of the maximal number of disjoint balls of radius $\delta/2$ centered at points of $x_n$. The approach is similar to that in our earlier papers [46, 47].
Lemma 15. 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 $\delta > 0$, where $M_\delta(X)$ was defined in the previous section. Then

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

for all finite subsets $x$ of $X$ and all $\delta > 0$.

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

$$p(x, \delta) = b_0(G_{x, \delta}) - 1 \leq |y| - 1 \leq D (\delta/2)^{-d} = 2^{-d} D \delta^{-d}.$$ 

We prove an extremal upper bound for $E_0^0(x_n)$ that, when combined with Lemma 10, implies the upper bound for our main theorem on minimum spanning trees.

Proposition 16. 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 $0 < \alpha < d$, then there exists a $D_\alpha > 0$ so that

$$E_\alpha^0(x_n) \leq D_\alpha n^{\frac{d-\alpha}{d}}$$

for all $n$ and all collections $x_n$ of $n$ points in $X$. Furthermore, there exists a $D_d > 0$ so that

$$E_d^0(x_n) \leq D_d \log(n)$$

for all $n$ and all collections $x_n$ of $n$ points in $X$.

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 minimum spanning tree on \( n \) points has \( n - 1 \) lemma implies that 

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

Applying Lemma \[\text{[II]}\] to the set of edge lengths of the minimum spanning tree on \( x_n \) yields 

\[
E_\alpha^0(x_n) = \sum_{e \in T(x_n)} |e|^\alpha \leq \alpha \int_{\delta=0}^{1} f(\delta) \delta^{\alpha-1} \, d\delta
\]

\[
= (n - 1) \int_{\delta=0}^{\kappa} \alpha \delta^{\alpha-1} \, d\delta + \alpha 2^{-d}D \int_{\delta=\kappa}^{1} \delta^{\alpha-d-1} \, d\delta
\]

\[
= (n - 1) \left[ \delta^{\alpha} \right]_{\delta=0}^{\kappa} - \frac{\alpha}{d-\alpha} 2^{-d}D \left[ \delta^{\alpha-d} \right]_{\delta=\kappa}^{1}
\]

\[
= (n - 1) \kappa^{\alpha} + \frac{\alpha}{d-\alpha} 2^{-d}D \left( \kappa^{\alpha-d} - 1 \right)
\]

\[
= 2^{\alpha}D^{\frac{\alpha}{d}} \left( 1 + D \frac{\alpha}{d-\alpha} \right) \left( n - 1 \right) \frac{d-\alpha}{d} - \frac{\alpha}{d-\alpha} 2^{-d}D
\]

\[
\leq D_\alpha n^{\frac{d-\alpha}{d}}
\]

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. \( \square \)

### 3.2. Proof of the Lower Bound

Our strategy to prove a lower bound for the asymptotics of \( E_\alpha^0(x_n) \) is to define random variables in terms of occupancy patterns of disjoint balls of radius \( 2r \). These random variables will imply the existence of minimum spanning tree edges of length at least \( r \).

Let \( M \) be a metric space and let \( \mu \) 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 \( x \) is a finite subset of \( M \), define

\[
\omega(B, x) = \Xi \left( B_{2r}(x) \setminus B_r(x), \{B_r(x)\} \right)
\]
That is, \( \omega(B, x) = 1 \) if \( x \) intersects \( B_r(x) \) but not the annulus centered at \( x \) with radii \( r \) and \( 2r \).

**Lemma 17.** Let \( B \) be a set of disjoint balls of radius \( 2r \) centered at points of \( M \), and let \( x \) be a finite subset of \( M \). Then

\[
p(x, r) \geq \sum_{B \in B} \omega(B, x) - 1.
\]

**Proof.** This is an immediate consequence of Lemma 14. See Figure 2. \( \square \)

Fix \( n \in \mathbb{N} \) and let \( \epsilon = n^{-1/d} \). Let \( B^n_1, \ldots, B^n_{s_n} \) be a maximal collection of disjoint balls of radius \( 2\epsilon \) centered at points of \( X \), and let \( y^n_j \) be the center of \( B^n_j \) for \( j = 1, \ldots, s_n \). We require one more lemma before proving the lower bound.

**Lemma 18.** There is a positive real number \( \gamma > 0 \) so that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{s_n} \omega(B^n_j, x_n) \geq \gamma
\]

in probability as \( n \to \infty \).
**Proof.** Let 

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

where \( c \) is the constant appearing in the definition of Ahlfors regularity. By that definition,

\[ \mu \left( B_e \left( y_j^n \right) \right) \geq p \epsilon^d = \frac{p}{n} \]

and

\[ \mu \left( B_j^n \setminus B_e \left( y_j^n \right) \right) \leq c \left( 2 \epsilon \right)^d - \frac{1}{c} \epsilon^d = \frac{q}{n}. \]

Also, Lemma 10 implies that

\[ s_n \geq \frac{1}{c} 2^{-d} 2 \epsilon^{-d} = \frac{1}{c} 2^{-2d} n. \]

Therefore, the occupancy events \( \omega \left( B_{n_1}^n, x_n \right), \ldots, \omega \left( B_{s_n}^n, x_n \right) \) satisfy the hypotheses of Lemma 13, which immediately implies the desired result. \( \square \)

The lower bound in our main theorem on minimum spanning trees follows quickly.

**Proposition 19.** Let \( \mu \) be a \( d \)-Ahlfors regular measure on a metric space \( M \). If \( \{ x_j \}_{j \in \mathbb{N}} \) are i.i.d. samples from \( \mu \), and \( \gamma \) is as given in the previous lemma, then

\[ \lim_{n \to \infty} n^{-\frac{d \alpha}{d}} E_0^\alpha (x_n) \geq \gamma \]

in probability.

**Proof.** We have that

\[ \lim_{n \to \infty} n^{-\frac{d \alpha}{d}} E_0^\alpha (x_n) \geq \lim_{n \to \infty} n^{-\frac{d \alpha}{d}} n^{-\alpha/d} p \left( x_n, n^{-1/d} \right) \]

\[ \geq \lim_{n \to \infty} \frac{1}{n} \left( \sum_{j=1}^{s_n} \omega \left( B_j^n, x_n \right) - 1 \right) \]

by Lemma 17

\[ \geq \gamma \]

by Lemma 18

in probability as \( n \to \infty. \) \( \square \)
4. Persistent Homology

We provide a brief introduction to the persistent homology [22] of a filtration, loosely following [46]. For a more in-depth survey refer to, e.g., [23, 24, 12, 27]. A filtration of topological spaces is a family \( \{X_\epsilon\}_{\epsilon \in I} \) of topological spaces indexed by an ordered set \( I \), with continuous maps \( X_{\epsilon_1} \rightarrow X_{\epsilon_2} \) for all pairs of indices \( \epsilon_1 < \epsilon_2 \). For example, if \( X \) is a subset of a metric space \( M \), the Čech filtration of \( X \), \( X_{\epsilon \in \mathbb{R}^+} \), is the family of \( \epsilon \)-neighborhoods of \( X \), 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 \( (y_1, \ldots, y_n) \in V(Y, \epsilon) \) if \( d(y_i, y_j) < \epsilon \) for \( i, j = 1, \ldots, n \).

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 \( \prod_{\epsilon \in I} H_i(X_\epsilon) \), together with the homomorphisms \( j_{\epsilon_0, \epsilon_1} : H_i(X_{\epsilon_0}) \rightarrow H_i(X_{\epsilon_1}) \) for \( \epsilon_0 < \epsilon_1 \), where \( H_i(X_\epsilon) \) denotes the reduced homology of \( X_\epsilon \) with coefficients in a field. If \( H_i(X_\epsilon) \) is finite dimensional for all \( \epsilon \in I \) — a hypothesis satisfied by all filtrations considered in this paper [11, 13] — the persistent homology module decomposes canonically into a set of interval modules [57, 17]. We denote the collection of these intervals as \( PH_i(X) \); each interval \((b, d) \in PH_i(X)\) corresponds to a homology generator that is “born” at \( \epsilon = b \) and “dies” at \( \epsilon = d \).

If \( x \) 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 \( PH_0(x) \) and the edges of the minimum spanning tree on \( x \). The same is true if persistent homology is taken with respect to the Čech complex and \( x \subset \mathbb{R}^m \), 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 \( PH^\epsilon_i(X) \) denote the set of intervals of \( PH_i(X) \) of length...
greater than \( \epsilon \):

\[
PH^\epsilon_i(X) = \{I \in PH_i(X) : |I| > \epsilon \}.
\]

Also, define

\[
p_i(X, \epsilon) = |PH^\epsilon_i(X)|.
\]

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

\[
d_H(X, Y) = \inf \{\epsilon \geq 0 : Y \subseteq X^\epsilon \text{ and } X \subset Y^\epsilon\}.
\]

Also, let \( d(X, Y) \) be the 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
   \[
   \eta : PH^{2\epsilon}_i(X) \to PH_i(Y)
   \]
   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 \( \delta \geq 0 \).

2. **Additivity for well-separated sets**: If \( X_1, \ldots, X_n \subset M \) and
   \[
d(X_j, X_k) > \max(\text{diam } X_j, \text{diam } X_k)(1 - \delta_{j,k}) \quad \forall j, k
   \]
   then
   \[
p_i(\bigcup_j X_j, \epsilon) \geq \sum_j p_i(X_j, \epsilon).
   \]

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 28) and an upper bound for the number of \( i \)-dimensional persistent homology intervals of a set of \( n \) points.
4.2. **A Lemma.** If $X$ is a metric space, let $F^i_\alpha (X, \epsilon)$ denote the $\alpha$-weighted sum of the persistent homology intervals of $X$ of length greater than $\epsilon$:

$$F^i_\alpha (X, \epsilon) = \sum_{I \in PH^i_\epsilon (X)} |I|^\alpha .$$

We will use the following lemma in the next section.

**Lemma 20.** If $d_H(X, Y) < \epsilon/4$ then

$$F^i_\alpha (X, \epsilon) < 2^\alpha F^i_\alpha (Y, \epsilon/2) .$$

**Proof.** By stability, there is an injection

$$\eta : PH^\epsilon_\epsilon (X) \to PH^{\epsilon/2}_\epsilon (Y)$$

so that

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

for all $I \in PH^\epsilon_\epsilon (X)$.

It follows that

$$F^i_\alpha (X, \epsilon) = \sum_{I \in PH^\epsilon_\epsilon (X)} |I|^\alpha$$

$$< \sum_{I \in PH^\epsilon_\epsilon (X)} 2^\alpha |\eta (I)|^\alpha$$

$$\leq 2^\alpha \sum_{J \in PH^{\epsilon/2}_\epsilon (Y)} |J|^\alpha$$

$$= 2^\alpha F^i_\alpha (Y, \epsilon/2) .$$

\[ \square \]

5. **Upper Bounds**

Our strategy to prove an upper bound for the asymptotics of $E^i_\alpha (x_n)$ is similar to that in Section 3.1: we control the number of persistence intervals of length greater than $\epsilon$ by approximating $x_n$ by a set consisting of the centers of disjoint balls of radius $\epsilon/2$ centered at points of $x_n$. 
5.1. Extremal Hypotheses. First, we prove the upper bound in Theorem 4, which implies the upper bound in our result for measures supported on a subset of $\mathbb{R}^2$.

Lemma 21 (Interval Counting Lemma). If $X$ is a triangulable metric space so that

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

for some positive real numbers $a$ and $D$ and all finite subsets $\{x_1, \ldots, x_n\}$ 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 $\{y_j\}$ 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 $\{y_j\}$ cover $Y$ so

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

It follows that

$$p_i(Y, \epsilon) \leq p_i(\{y_i\}, 0) \leq D|y_i|^a \leq D M_{\epsilon/4}(X)^a \leq Dc^a \epsilon^{-a/d} \epsilon^{-ad}$$

as desired. □

Proposition 22. If $X$ satisfies the hypotheses of the previous lemma and $\alpha < ad$, then there exists a $D_\alpha > 0$ so that

$$E_\alpha^i(x_1, \ldots, x_n) \leq D_\alpha n^{ad-a}$$

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

$$E_{ad}^i(x_1, \ldots, x_n) \leq D_d \log(n)$$

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 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 in Corollary 5, they are inadequate to show a similar upper bound for subsets of higher dimensional Euclidean space. Here, we show that hypotheses on the the expectation and variance of the number of $PH_i$ intervals of a set of $n$ points imply better asymptotic upper bounds. The idea of the proof is to control the behavior of $PH_i(X)$ in terms of the persistent homology of point samples from $X$. With that, we write $PH_i(x_n)$ a sum of two terms, one which approximates $PH_i(X)$ 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 [44] for a proof.

**Lemma 23.** Let $\mu$ be a probability measure on $X$, and $\{B_j\}_{j=1}^l \subset X$ be a collection of balls so that so that $\mu(B_j) \geq a$ for all $j$. Then

$$
\mathbb{P}(x_n \cap B_j \neq \emptyset \text{ for } j = 1, \ldots, l) \geq 1 - le^{-an}.
$$

Next, we apply the previous lemma to control the Hausdorff distance between $X$ and finite samples from an Ahlfors regular measure on $X$.

**Lemma 24.** If $\mu$ is a $d$-Ahlfors regular measure with support $X$ then there exists a positive real number $A_0$ that depends only on the constants $c$ and $d$ appearing in the definition of Ahlfors regularity so that

$$
\mathbb{P}(d_H(\{x_n\}, X) < \epsilon) \geq 1 - c\epsilon^{-d}e^{-A_0\epsilon^d}n
$$

for all $\epsilon > 0$.

**Proof.** Let $\{y_1, \ldots, y_{M_{\epsilon/3}(X)}\}$ 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,

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

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

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

$$
d_H(\{y_i\}, X) < 2\epsilon/3.
$$
Therefore, if \( \{x_n\} \cap B_{\epsilon/3}(y_j) \neq \emptyset \) for \( j = 1, \ldots, M_{\epsilon/3}(X) \)
\[
d_H(\{x_n\}, X) < \epsilon/3 + 2\epsilon/3 = \epsilon.
\]

It follows that
\[
P(\{x_n\}, X) < \epsilon) \geq \P(\{x_n\} \cap B_{\epsilon/3}(y_j) \neq \emptyset \text{ for } j = 1, \ldots, M_{\epsilon/3}(X))
\geq 1 - M_{\epsilon/3}(X) e^{-A_0 \epsilon^d n}
\geq 1 - c\epsilon^{-d} e^{-A_0 \epsilon^d n}
\]
by Lemma 23.

\(\square\)

In the next lemma, we show that if the expected number of persistent homology intervals of \( x_n \) \( X \) is \( O(n) \), then we can control the number of “long” persistent homology intervals of \( X \) itself.

**Lemma 25.** Let \( X \) be a bounded, triangulable metric space that admits a \( d \)-Ahlfors regular measure \( \mu \) satisfying
\[
\E(|PH_i(x_n)|) = O(n).
\]
Then there are positive real numbers \( A_1 \) and \( \epsilon_0 \) so that
\[
p_i(X, \epsilon) \leq A_1 \epsilon^{-d} \log(1/\epsilon)
\]
for all \( \epsilon < \epsilon_0 \).

**Proof.** There are positive real numbers \( D_1 \) and \( N_1 \) so that
\[
\E(|PH_i(x_n)|) \leq D_1/2n
\]
for all \( n > N_1 \). By Markov’s inequality,
\[
P(|PH_i(x_n)| > D_1 n) < 1/2.
\]
Manipulating the inequality in the previous lemma gives that
\[
P(\{x_1, \ldots, x_m(\epsilon)\}, X) < \epsilon/2) \geq 1/2
\]
where
\[ m(\epsilon) = \left\lceil \frac{2^d}{A_0} \epsilon^{-d} \log \left( 2^{d+1} c \epsilon^{-d} \right) \right\rceil. \]

Let \( \epsilon \) be sufficiently small so that \( m(\epsilon) > N_1 \). We have that
\[ \left| \text{PH}_i(x_1,\ldots,x_{m(\epsilon)}) \right| \leq D_1 n \quad \text{and} \quad d_H\left(\{x_1,\ldots,x_{m(\epsilon)}\},X\right) < \epsilon \]
for some finite subset \( x_1,\ldots,x_{m(\epsilon)} \) of \( X \). Therefore, by stability
\[ p_i(X,\epsilon) \leq p_i\left(\{x_1,\ldots,x_{m(\epsilon)}\},0\right) \]
\[ \leq D_1 m(\epsilon) \]
\[ = D_1 \left\lceil \frac{2^d}{A_0} \epsilon^{-d} \log \left( 2^{d+1} c \epsilon^{-d} \right) \right\rceil \]
\[ = O\left( \epsilon^{-d} \log \left( \frac{1}{\epsilon} \right) \right) \]
as \( \epsilon \to 0 \).

Next, we use the previous lemma to control \( F^i_\alpha(X,\epsilon) \), the truncated \( \alpha \)-weighted sum defined in Section 4.2.

**Proposition 26.** If \( X \) satisfies the hypotheses of the previous lemma and \( 0 < \alpha < d \), then there exist positive real numbers \( A_2 \) and \( \epsilon_1 \) so that
\[ F^i_\alpha(X,\epsilon) \leq A_2 \epsilon^{\alpha-d} \log \left( \frac{1}{\epsilon} \right) \]
for all \( \epsilon < \epsilon_1 \).

**Proof.** By the previous lemma
\[ p_i(X,\epsilon) \leq f(\epsilon) := A_1 (\epsilon)^{-d} \log \left( \frac{1}{\epsilon} \right) \]
for all \( \epsilon < \epsilon_0 \). Applying Lemma \( \square \) yields
\[ F^i_\alpha(Y,\epsilon) \leq \epsilon^\alpha f(\epsilon) + \alpha \int_{t=\epsilon}^{1} f(t) t^{\alpha-1} \, dt + F^i_\alpha(Y,\epsilon_0). \]
The first term equals
\[ A_1 \epsilon^{d} \left( \log \left( \frac{1}{\epsilon} \right) \right) , \]
which has the desired asymptotics as \( \epsilon \to 0 \). The second term equals
\[
\alpha \int_{t=\epsilon}^{1} A_1 t^{d-1} \log \left( \frac{1}{t} \right) \, dt =
\]
\[
A_1 \left[ -\frac{1}{d-\alpha} t^{d-\alpha} \log \left( \frac{1}{t} \right) - \frac{1}{(d-\alpha)^2} t^{d-\alpha} \right]_\epsilon^1
\]
\[
= A_1 \left( \frac{1}{d-\alpha} \epsilon^{d-\alpha} \log \left( \frac{1}{\epsilon} \right) + \frac{1}{(d-\alpha)^2} \epsilon^{d-\alpha} - \frac{1}{(d-\alpha)^2} \right)
\]
\[
= O \left( \epsilon^{d-\alpha} \log \left( \frac{1}{\epsilon} \right) \right) .
\]
Therefore, \( p_i(X\epsilon) = O \left( \epsilon^{d-\alpha} \log \left( \frac{1}{\epsilon} \right) \right) \), as desired.

\[ \square \]

Finally, we can bootstrap the previous result to control \( E_{i}^\alpha (x_n) \) and prove the desired upper bound.

**Proposition 27.** Let \( \mu \) be a \( d \)-Ahlfors regular measure on a bounded, triangulable metric space. If
\[
\mathbb{E} \left( |PH_x (x_n)| \right) = O(n)
\]
and
\[
\text{Var} \left( |PH_x (x_n)| \right) / n^2 \to 0,
\]
then there is a \( \Lambda > 0 \) so that
\[
E_{i}^\alpha (x_n) \leq \Lambda n^{\frac{d-\alpha}{d}} \log (n)^{\frac{2}{d}}
\]
in probability as \( n \to \infty \).

**Proof.** Let
\[
G_{i}^\alpha (x, \epsilon) = \sum_{I \in PH_i(x) \setminus PH_i^\epsilon(x)} |I|^{\alpha}.
\]
Our strategy is to write
\[
E_{i}^\alpha (x_n) = G_{i}^\alpha (x_n, \epsilon) + F_{i}^\alpha (x_n, \epsilon)
\]
for a well-chosen $\epsilon$. The former term can be interpreted as “noise,” and the latter approximates the persistent homology of the support of $\mu$.

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

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

for all sufficiently large $n$. By Chebyshev’s inequality,

$$
\mathbb{P}
left(\left|\text{PH}_i\left(\mathbf{x}_n\right)\right| > Dn\right) \leq \frac{\mathbb{E}\left(\left|\text{PH}_i\left(\mathbf{x}_n\right)\right|\right)}{D^2 n^2}
$$

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

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

for all $n > M$.

Solving for $\epsilon$ in the expression in Lemma 24 gives that

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

if

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

where $W$ is the Lambert W function. $W(m) \sim \log (m)$ as $m \to \infty$, and $W(m) \leq \log (m)$ for $m \geq e^{33}$. Therefore, there is an $A_3 > 0$ that does not depend on $p$ and an $N_1(p)$ that does depend on $p$ so that

$$
\frac{A_3}{2} n^{-1/d} \log (n)^{1/d} \leq \epsilon(n) \leq A_3 n^{-1/d} \log (n)^{1/d}
$$

for all $n > N_1(p)$.

Choose $N_2(p) > N_1(p)$ to be sufficiently large so that $\epsilon(n) < \epsilon_1$ for all $n > N_2(p)$, where $\epsilon_1$ is given in Proposition 26. Let $n > N_2(p)$ and suppose that $\mathbf{x}_n$ satisfies $\left|\text{PH}_i(\mathbf{x}_n)\right| < Dn$ and $d_H(\mathbf{x}_n, X) < \epsilon(n)/4$ — an event which occurs with probability greater than $1 - p$. Write

$$
E^i_{\alpha}(\mathbf{x}_n) = F^i_{\alpha}(\mathbf{x}_n, \epsilon(n)) + G^i_{\alpha}(\mathbf{x}_n, \epsilon(n)).
$$
We consider the two terms separately.

\[
G_i^\alpha (x_n, \epsilon (n)) \leq D |x_n| \epsilon (n)^\alpha \\
\leq 2^\alpha DA_3^\alpha n^{d-\alpha} \log (n)^{\alpha/d} \\
= A_4 n^{d-\alpha} \log (n)^{\frac{\alpha}{d}},
\]

where \( A_4 = 2^\alpha DA_3^\alpha \) is a positive constant that does not depend on \( n \) or \( p \).

To bound the second term, we apply Lemma [20] to find

\[
F_i^\alpha (x, \epsilon (n)) \leq 2^\alpha F_i^\alpha (X, \frac{\epsilon (n)}{2}) \\
\leq A_2 (\epsilon (n))^{a-d} \log \left( \frac{1}{\epsilon (n)} \right) \quad \text{by Prop. [26]} \\
\leq A_2 A_3^{a-d} n^{d-\alpha} \log (n)^{\frac{d-a}{d}} \log \left( \frac{1}{2A_3 n^{1/d} \log (n)^{-1/d}} \right) \quad \text{by Eqn. [2]} \\
= A_2 A_3^{a-d} n^{d-\alpha} \log (n)^{\frac{d-a}{d}} \left( \frac{1}{d} \log (n) - \log \left( 2A_3 \log (n)^{1/d} \right) \right) \\
\leq \frac{1}{d} A_2 A_3^{a-d} n^{d-\alpha} \log (n)^{\frac{d}{d}} \\
= A_5 n^{d-\alpha} \log (n)^{\frac{d}{d}},
\]

where \( A_5 = \frac{1}{d} A_2 A_3^{a-d} \) is a positive constant that does not depend on \( n \) or \( p \).

In summary, if \( \Lambda = A_4 + A_5 \) and \( 0 < p < 1 \), then there exists an \( N_2(p) > 0 \) so that

\[
P \left( E_i^\alpha (x_n) \leq \Lambda n^{d-\alpha} \log (n)^{\frac{d}{d}} \right) > 1 - p
\]

for all \( n > N_2(p) \).

\[\Box\]

6. The Lower Bound

While our proofs of the upper bounds work for Ahlfors regular measures on arbitrary metric spaces, here we restrict our attention Ahlfors regular measures on Euclidean space. This will allow us to use the structure of the cubical grid on \( \mathbb{R}^m \).
To prove the lower bound, we modify the approach in our paper on extremal \( PH \)-dimension [46] to work in a probabilistic context. If \( \mu \) is a \( d \)-Ahlfors regular measure on \( \mathbb{R}^m \) and \( \delta > 0 \), let \( C_\delta (\mu) \) be the cubes in the grid of mesh \( \delta \) that intersect the support of \( \mu \). The basic idea is to sub-divide the grid of mesh \( \delta \) so each cube contains \( k^m \) sub-cubes. If \( k \) is chosen carefully, we can find a positive fraction of cubes in \( C_\delta (\mu) \) that contain enough cubes of \( C_{\delta/k} (\mu) \) to guarantee a stable \( PH_i \) class. In fact, we can require that the sub-cubes have probability exceeding a certain threshold. We then control the number of stable \( PH_i \) classes realized by a random sample \( x_n \) with Lemma [13].

First, we define the non-triviality constants \( \gamma_i^m \).

### 6.1. Non-triviality Constants

In previous work [46], 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.

**Definition 28.** 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 \( Y \) is any subset of \( \cup_{x \in X} C(x) \) satisfying

\[
Y \cap C(x) \neq \emptyset \quad \forall \ x \in X,
\]

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 29.** Let $\xi^m(N)$ 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

$$\gamma^m_i = \liminf_{N \to \infty} \frac{\log(\xi^m_i(N))}{\log(N)}.$$ 

$\gamma^m_0 = 0$ 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 [46], we proved that $\gamma^m_1 \leq m - \frac{1}{2}$ if persistent homology is taken with respect to the Čech complex. Note that the previous definition does not include the same restriction on the size as in our previous 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 $C_{\delta,a}(\mu)$ be the set of closed cubes $C$ in the cubic grid of mesh $\delta$ in $\mathbb{R}^m$ centered at the origin that satisfy

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

and let $N_{\delta,a}(\mu) = |C_{\delta,a}(\mu)|$. (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 30.** If $\mu$ is a $d$-Ahlfors regular measure with support $X \subset \mathbb{R}^m$, then there exist real numbers $0 < c_0 \leq c_1 < \infty$ depending on $m$ and the constants $c$ and $d$ appearing in the definitions of Ahlfors regularity so that

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

for all $\delta < \delta_0$, where $\hat{c} = \frac{1}{2^m}$. Similarly, there exist real numbers $0 < c'_0 \leq c'_1 < \infty$ depending on $c$, $d$, and $m$ so that

$$c'_0\delta^{-d} \leq N_{\delta,0}(\mu) \leq c'_1\delta^{-d}$$

for all $\delta < \delta_0$.

**Proof.** Let $C$ be a cube in the grid of mesh $\delta$ that intersects $X$, and $x \in C \cap X$. $\mu(B_{3\delta}(x)) > 1/c\delta^d$ and $B_\delta(x)$ intersects at most $2^m$ cubes in the grid of mesh $\delta$, so at least 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 $C_{\delta,\hat{c}}(\mu)$ is adjacent to at most $3^m$ cubes of $C_{\delta}(\mu)$. It follows that

$$\frac{1}{3^m}N_{\delta,0}(\mu) \leq N_{\delta,\hat{c}}(\mu) \leq N_{\delta,0}(\mu)$$
where the upper bound is trivial. Thus, bounds for $N_{\delta,0}(\mu)$ imply bounds for $N_{\delta,\hat{c}}(\mu)$, and visa versa.

We have that

\[
1 = \mu(X) \\
\leq \sum_{C \in C_{\delta,0}(\mu)} \mu(C) \\
\leq c\delta^d m^{d/2} N_{\delta,0}(\mu) \\
\leq 3^m c\delta^d m^{d/2} N_{\delta,\hat{c}}(\mu) \\
\implies N_{\delta,\hat{c}}(\mu) \geq 3^{-m} m^{-d/2} \delta^{-d}.
\]

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

\[
1 = \mu(X) \\
\geq \frac{1}{3^m} \hat{c}\delta^d N_{\delta,\hat{c}}(\mu) \\
\implies N_{\delta,\hat{c}}(\mu) \leq c 6^m \delta^{-d}.
\]

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

**Lemma 31.** Let $0 < \beta < d$ and let

\[
C_{\delta}^{k,\beta} = \left\{ C \in C_{\delta}(\mu) : D_k(C) > k^\beta \right\}
\]
and

\[ M(\delta, k, \beta) = \left| C_{\delta}^{k, \beta} \right|. \]

Then 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 \( \delta_0 \) be the constants from the previous lemma so \( N_{\delta, 0}(\mu) \leq c_1' \delta^{-d} \) and \( N_{\delta, \epsilon}(\mu) \geq c_0 \delta^{-d} \) for all \( \delta < \delta_1 \).

There are at least \( c_0 k^d \delta^{-d} \) cubes in \( C_{\delta/k, \epsilon}(\mu) \), each of which is either a sub-cube of \( C_{\delta}^{k, \beta} \) or \( C_{\delta, 0}(\mu) \setminus C_{\delta}^{k, \beta} \). A cube in \( C_{\delta}^{k, \beta} \) can contain at most \( k^m \) sub-cubes of \( C_{\delta/k, \epsilon}(\mu) \), and a cube in \( C_{\delta, 0}(\mu) \setminus C_{\delta}^{k, \beta} \) can contain at most \( k^\beta \) sub-cubes of \( C_{\delta/k, \epsilon}(\mu) \). Therefore, \( M(\delta, k, \beta) \) is bounded below by the smallest integer \( a_{k, \delta} \) satisfying

\[ a_{k, \delta} k^m + \left( c_1' \delta^{-d} - a_{k, \delta} \right) k^\beta \geq c_0 k^d \delta^{-d}. \]

Rearranging terms, we have that

\[ a_{k, \delta} = \left\lfloor \frac{\delta^{-d} \left( c_0 k^{d-\beta} - c_1' \right)}{k^{m-\beta} - 1} \right\rfloor. \]

Let

\[ K = \left( \frac{c_1'}{c_0} \right)^{\frac{1}{\beta-\gamma}}, \]

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

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

so

\[ a_{k, \delta} \sim 2c_2 \delta^{-d} \]

as \( \delta \to 0 \). It follows that

\[ M(\delta, k, \beta) \geq a_{k, \delta} > c_2 \delta^{-d} \]

for all sufficiently small \( \delta \), as desired. \( \square \)
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 18.

Lemma 32. If \( \mu \) be a \( d \)-Ahlfors regular measure on \( \mathbb{R}^m \) with \( d > \gamma_i^m \), then there exist positive real numbers \( \epsilon_0 \) and \( \Omega_0 \) so that

\[
\lim_{n \to \infty} \frac{1}{n} p_i \left( x_n, \epsilon_0 n^{-1/d} \right) \geq \Omega_0
\]

in probability.

Proof. Let \( \gamma_i^m < \beta < d \). By the definition of \( \gamma_i^m \) we can find a \( K_0 \) so that \( k^\beta > \xi_i^m (k) \) for all \( k > K_0 \). Let \( k_0 > \min (K, K_0) \), where \( K \) is given in the previous lemma, and let \( \delta_1 \) and \( c_2 \) also be as in the previous lemma. There are only finitely many collections of sub-cubes of \( [k_0]^m \), so there are only finitely many possible stable \( PH_i \) classes of subsets of \( [k_0]^m \). Let \( \epsilon_0 \) 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 \( \{D_1, \ldots, D_s\} \) be a maximal collection of cubes in \( C_{k_0} \) so that \( d \left( D_j, D_k \right) > \delta \sqrt{m} \) for all \( j, k \in \{1, \ldots, s\} \) so that \( j \neq k \). See Figure 5. There is a constant \( 0 < \kappa < 1 \) that depends only on \( d \) so that

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

Let \( l \in \{1, \ldots, s\} \). By the definition of \( \gamma_i^m \), there is a collection of sub-cubes \( B_l \subset D_{k_0} (D_l) \) with a stable \( PH_i \) class. Let

\[
A_l = \hat{B}_{\delta \sqrt{m}} (C) \setminus \cup_{B \in B_l} B
\]
where $\tilde{B}_{\delta \sqrt{m}}(D_j)$ is the union of all cubes in the grid of mesh $\delta/k$ within distance $\delta \sqrt{m}$ of $D_j$. Also, let $\mathcal{B}_l'$ be collection of the interiors of the sets $\mathcal{B}_l$. It follows from property (2) in Section 4.1 that

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

There is a $q > 0$ depending only on $k_0$, $c$, $d$, and $m$ so that

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

for all $l \in \{1, \ldots, s\}$. Also, each $B \in \mathcal{B}_l$ is a cube of width $\delta/k_0$ in $\mathbb{R}^m$ so

$$\mu \left( B \right) \geq \frac{1}{c} \left( \frac{\delta \sqrt{m}}{2k_0} \right)^d = \frac{p}{n},$$

where $p = 2^{-d}k_0^{-d}m^{d/2}/c$. Therefore, $\Xi \left( x_n, A_l, \mathcal{B}_l' \right)$ is a $n, p, q, k^m$-bounded occupancy event for each $l$, and the desired result follows from Lemma 13.

The proof of the lower bound is now straightforward.

**Proposition 33.** Let $\mu$ be a $d$-Ahlfors regular measure on $\mathbb{R}^m$ with $d > \gamma_i^m$. Then there is an $\Omega > 0$ so that

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

in probability.

**Proof.** It follows immediately from the previous lemma that

$$\lim_{n \to \infty} n^{-\frac{d-\alpha}{d}} E_\alpha^i (x_n) \geq$$

$$\left( \epsilon_0 \frac{1}{n} \right)^\alpha \lim_{n \to \infty} n \left( \epsilon_0 n^{-1/d} \right)^\alpha$$

$$\geq \epsilon_0^\alpha \Omega_0 \quad \text{by Lemma 32}$$

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. *Abel Symposia*, 2019. arXiv:1808.01079.
[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] Vincent Beffara. The dimension of the sle curves. *Annals of Probability*, 2008.
[7] A. Beurling and L. Ahlfors. The boundary correspondence under quasiconformal mappings. *Acta Mathematica*, 1956.
[8] Omer Bobrowski and Matthew Kahle. Topology of random geometric complexes: a survey. *Journal of Applied and Computational Topology*, 2018.
[9] Omer Bobrowski, Matthew Kahle, and Primoz Skraba. Maximally persistent cycles in random geometric complexes. *The Annals of Applied Probability*, 2017.
[10] M. G. Bouligand. Ensembles impropres et nombre dimensionnel. *Bulletin des Sciences Mathématiques*, 1928.
[11] Francesca Cagliari and Claudia Landibc. Finiteness of rank invariants of multidimensional persistent homology groups. *Applied Mathematics Letters*, 2011.
[12] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 2009.
[13] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. *The Structure and Stability of Persistence Modules*. Springer, 2016.
[14] Frédéric Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 2014.
[15] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1), 2007.
[16] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have $l_p$-stable persistence. *Foundations of Computational Mathematics*, 2010.
[17] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and its Applications*, 2015.
[18] Guy David and Stephen Semmes. *Analysis of and on Uniformly Rectifiable Sets*. American Mathematical Society, 1993.
[19] S. Davies and P. Hall. Fractal analysis of surface roughness by using spatial data. *Journal of the Royal Statistical Society Series B*, 1999.
[20] Vincent Divol and Wolfgang Polonik. On the choice of weight functions for linear representations of persistence diagrams. arXiv:1807.03678, July 2018.
[21] Trinh Khanh Duy, Yasuaki Hiraoka, and Tomoyuki Shirai. Limit theorems for persistence diagrams. arXiv:1612.08371, December 2016.
[22] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 2002.
[23] Herbert Edelsbrunner and John Harer. Persistent homology — a survey. *Contemporary Mathematics*, 2008.
[24] Herbert Edelsbrunner and Dmitriy Morozov. Persistent homology: Theory and practice. Technical Report LBNL-6037E, Lawrence Berkeley National Laboratory, 2013.
[25] Gerald A. Edgar. *Classics on Fractals*. Studies in Nonlinearity. Westview Press, 2003.
[26] Kenneth Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, 2014.
[27] Robert Ghrist. Barcodes: the persistent homology of data. *Bulletin of the American Mathematical Society*, 2008.
[28] Michael Goff. Extremal Betti numbers of Vietoris-Rips complexes. *Discrete and Computational Geometry*, 2011.
[29] Peter Grassberger and Itamar Procaccia. Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, 1983.
[30] 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.
[31] Felix Hausdorff. Dimension und äußeres maß. *Mathematische Annalen*, 1918.
[32] Yasuaki Hiraoka and Tomoyuki Shirai. Minimum spanning acycle and lifetime of persistent homology in the Linial—Meshulam process. *Random Structures & Algorithms*, 2017.
[33] Abdolhossein Hoorfar and Mehdi Hassani. Inequalities on the Lambert W function and hyperpower function. *Journal of Inequalities in Pure and Applied Mathematics*, 2008.
[34] Jonothan Jaquette and Benjamin Schweinhart. Comparison of fractal dimension estimation techniques. in preparation.
[35] Harry Kesten and Sungchul Lee. The central limit theorem for weighted minimal spanning trees on random points. *Annals of Applied Probability*, 1996.
[36] G. Kozma, Z. Lotker, and G. Stupp. On the connectivity threshold for general uniform metric spaces. *Information Processing Letters*, 2010.
[37] Gady Kozma, Zvi Lotker, and Gideon Stupp. The minimal spanning tree and the upper box dimension. *Proceedings of the American Mathematical Society*, 2006.
[38] R. Lopes and N. Betrouni. Fractal and multifractal analysis: A review. *Medical Image Analysis*, 2009.
[39] John Mackay and Jeremy Tyson. *Conformal Dimension: Theory and Application*. American Mathematical Society, 2010.
[40] R. D. MacPherson and B. Schweinhart. Measuring shape with topology. *Journal of Mathematical Physics*, 53(7), 2012.
[41] Benoit Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Co., 1982.
[42] Benoît Mandelbrot. *Fractals: Form, Chance and Dimension*. W.H. Freeman & Company, 1977.
[43] Gabriell Máté and Dieter W. Heermann. Persistence intervals of fractals. *Physica A*, 2014.
[44] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete and Computational Geometry*, 2008.
[45] Vanessa Robins. *Computational topology at multiple resolutions: foundations and applications to fractals and dynamics*. PhD thesis, University of Colorado at Boulder, 2000.
[46] Benjamin Schweinhart. Persistent homology and the upper box dimension. arXiv:1802.00533, 2018.
[47] Benjamin Schweinhart. Weighted persistent homology sums of random Čech complexes. arXiv:1807.07054, July 2018.
[48] R. Sikorski. *Real Functions, Vol. 1*. PWN: Warsaw, 1958.
[49] Richard Stalney. The upper bound conjecture and Cohen-Macaulay rings. *Studies in Applied Mathematics*, 1975.
[50] J. Michael Steele. Growth rates of Euclidean minimal spanning trees with power weighted edges. *Annals of Probability*, 1988.
[51] Floris Takens. *Dynamical Systems and Turbulence*, chapter Detecting Strange Attractors in Turbulence, pages 366–381. Springer, 1980.
[52] Caetano Traina, Agma Traina, Leejay Wu, and Christos Faloutsos. Fast feature selection using fractal dimension. *Journal of Information and Data Management*, 2010.
[53] 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.
[54] D. Yogeshwaran, Eliran Subag, and Robert J. Adler. Random geometric complexes in the thermodynamic regime. *Probability Theory and Related Fields*, 2017.
[55] Boming Yu. Analysis of flow in fractal porous media. *Applied Mechanics Reviews*, 2008.
[56] J.E. Yukich. Asymptotics for weighted minimal spanning trees on random points. *Stochastic Processes and their Applications*, 2000.
[57] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:249–274, 2005.