Topology of Random Čech complexes in Thermodynamic Regime

Akshay Goel

Kyushu University

October 21, 2019

Outline

- **PART 1:** Introduction
- **PART 2:** Prior Work & Our Contribution
- **PART 3:** Idea of Proof
- **PART 4:** Extension

Random Topology = Probability + Algebraic Topology

- Emerging research area known as random topology = theoretical results that characterize the asymptotic behavior of topological properties of random objects.
- In additions to the mathematical value, such results are also motivated by many issues in manifold learning and topological data analysis.
- One aspect of random topology is the study of random geometric complexes and their topological properties such as Betti numbers.
- In this talk, we concentrate on a typical type of random geometric complexes known as Čech complexes.

Motivation from manifold learning

Assumption in manifold learning: Given data points as realizations of i.i.d. random variables $\{X_1, X_2, \ldots\}$ supported on an unknown non-linear, smooth and compact manifold (of intrinsic dimension $<<$ ambient dimension).

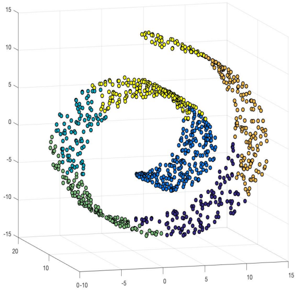


Figure: Swissroll Dataset

After the Manifold Assumption...

Instructive to estimate the topological properties of the unknown manifold- 'Homology Inference'.

*k*th Betti number:
Rank of *k*th homology group

- 0th Betti number β_0:
  Number of connected components
- $(k \geq 1)$th Betti number β_k:
  Number of k-dimensional holes

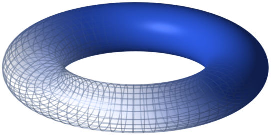


Figure: 2D Torus - $\beta_0 = \beta_2 = 1$; $\beta_1 = 2$

One of the first results in this regard...

Let $\{X_1, \ldots, X_n\}$ be drawn uniformly and independently from the unknown compact manifold $\mathcal{M} \subset \mathbb{R}^N$. Let $B(X_i, r) :=$ closed Euclidean ball in ambient dimension. Then the estimator

$$\hat{\mathcal{M}} := \bigcup_{i=1}^{n} B(X_i, r),$$

has same Betti numbers as of unknown manifold with high probability under certain conditions on n and r (Niyogi, Smale, and Weinberger 2008).

Note

- $\beta_k(\hat{\mathcal{M}}) = 0$ for all $k \geq$ ambient dimension;
- $\beta_k(\hat{\mathcal{M}}) = \beta_k$ of random Čech complex of radius r, constructed on $\{X_1, \ldots, X_n\}$ (follows by Nerve Theorem).

Čech complex

Definition 1.1.

Let $(\mathcal{A}, \rho) :=$ metric space & $\mathfrak{X}_n = \{x_1, x_2, \ldots, x_n\} :=$ finite set of points in $\mathcal{A}$. For any $r > 0$, *the Čech complex of radius r* is

$$\mathcal{C}(\mathfrak{X}_n, r, \rho) = \left\{ \sigma \subset \mathfrak{X}_n \colon \bigcap_{x \in \sigma} B_\rho(x, r) \neq \emptyset \right\},$$

where $B_\rho(x, r) = \{y \in \mathcal{A} \colon \rho(x, y) \leq r\}$.

- $\sigma \in \mathcal{C}(\mathfrak{X}_n, r, \rho)$ is *k-simplex*, if $|\sigma| = k + 1$.
- Dimension of $\mathcal{C}(\mathfrak{X}_n, r, \rho) := (\max_{\sigma \in \mathcal{C}(\mathfrak{X}_n, r, \rho)} |\sigma|) - 1$.

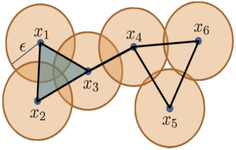
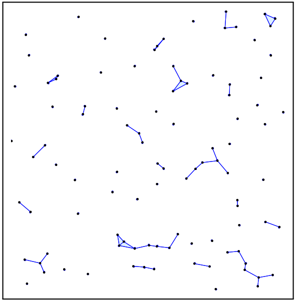


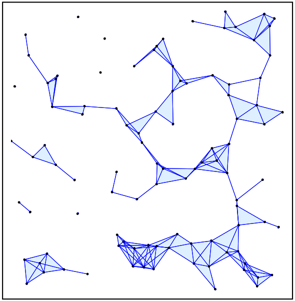
Figure: Čech complex: $\beta_0 = \beta_1 = 1$.

Limiting Regimes

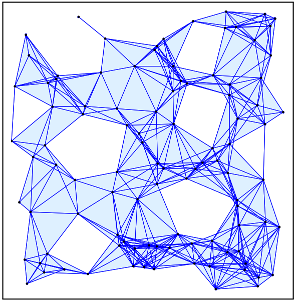
Random geometric complexes - Higher dimensional analogues of Random Geometric Graphs, from which we have



Sparse regime:
$n^{1/m} r_n \to 0$



Thermodynamic regime:
$n^{1/m} r_n \to r \in (0, \infty)$



Dense regime:
$n^{1/m} r_n \to \infty$

Observation

In each regime, topological behavior is significantly different.

For the randomness...

Definition 1.2.

Let $X_i, i \geq 1$ be i.i.d. $\mathbb{R}^N$-valued random variables with common probability density function $f(x)$. Then the union of first n points $\{X_1, X_2, \ldots, X_n\}$ is called a *binomial point process*.

We consider binomial point process in two different yet related settings, on the basis of the support of density function

- **Manifold setting:** support - m-dimensional compact C^1 manifold $\mathcal{M}$, embedded within $\mathbb{R}^N (N > m)$.
- **Euclidean setting:** support - $\mathbb{R}^N$.

For the randomness...

- In stochastic geometry, the standard technique to prove results for binomial point processes is to consider their Poissonized versions.
- Suppose for given $n > 0$, T_n is a Poisson random variable with parameter n and independent of $\{X_i, i \geq 1\}$.
- Then the point process $\mathcal{P}_n := \{X_1, \ldots, X_{T_n}\}$ is called a Poissonized version of the binomial point process.
- It is a Poisson point process on $\mathbb{R}^N$ with intensity function nf. It is denoted by $\mathcal{P}_n$ and $\mathcal{Q}_n$ in Euclidean and manifold setting, respectively.

For the randomness...

Definition 1.3.

A point process $\mathcal{P}$ on $\mathbb{R}^N$ is said to be a *Poisson point process* with intensity function $\lambda(x)$, denoted by $\mathcal{P}(\lambda(x))$, if it satisfies the following two conditions

(i) for any bounded Borel set $B \subset \mathbb{R}^N$, the random variable $\mathcal{P}(B)$ counting the number of points in B has Poisson distribution with parameter $\lambda(B) := \int_B \lambda(x)dx$, i.e.,

$$\mathbb{P}(\mathcal{P}(B) = k) = e^{-\lambda(B)} \frac{\lambda(B)^k}{k!}, \quad k = 0, 1, \ldots;$$

(ii) for disjoint bounded Borel sets $B_1, B_2, \ldots, B_k$, the random variables $\mathcal{P}(B_1)$, $\mathcal{P}(B_2), \ldots, \mathcal{P}(B_k)$ are independent.

Dense Regime

Theorem 1.4 (Manifold setting; Bobrowski and Mukherjee 2015, Theorem 4.9).

Assume κ to be bounded, measurable and supported on a m-dimensional closed manifold $\mathcal{M} \subset \mathbb{R}^N$, where $m < N$.Let $\kappa_{min} := \inf_{z \in \mathcal{M}} \kappa(z) > 0$ and $n^{1/m} r_n \geq C(\log n)^{1/m}$.

(a) *If $C > (\omega_m \kappa_{min})^{-m}$, then as $n \to \infty$,*

$$\mathbb{P}(\beta_k(\mathcal{Q}_n, r_n) = \beta_k(\mathcal{M}), \text{ for all } 0 \leq k \leq m) \to 1.$$

(b) *If $C > 2^m (\omega_m \kappa_{min})^{-m}$, then almost surely there exists $n_0 > 0$ (which is random), such that for all $n > n_0$,*

$$\beta_k(\mathcal{Q}_n, r_n) = \beta_k(\mathcal{M}), \text{ for all } 0 \leq k \leq m,$$

where $\omega_m = Leb^m(B(0,1))$.

Dense Regime

Remarks

On comparing with the result of Niyogi et al., Theorem 1.4

- holds for larger class of probability density functions;
- require less prior information about the manifold;
- is stronger in the sense that the convergence is shown to occur almost surely.

Thermodynamic Regime

- We do not know what happens in the thermodynamic regime clearly since basic questions such as law of large numbers and central limit theorems are not entirely understood yet.
- In our work, we answer the question of law of large numbers in thermodynamic regime.

Another motivation

- Besides this motivation from manifold learning, our motivation also comes from stochastic geometry.
- The estimator in the result of Niyogi et. al. is a special case of 'Boolean model', whose geometric properties such as volume and surface area have been well studied.
- Therefore, the next natural question arises about its topological features.
- Furthermore, in stochastic geometry, weak and strong laws of large numbers have been established for a general class of local statistics, however, Betti numbers do not belong to that class.
- Thus, the study of Betti numbers need further development.

Part 2

Prior Work & Our Contribution

Setting

Description

- **Manifold Setting:**
  - $\mathcal{M} \subset \mathbb{R}^N$- m-dimensional compact C^1 manifold, where $m < N$.
  - r_n is a sequence of positive real numbers such that $n^{1/m} r_n \to r \in (0, \infty)$-Thermodynamic regime.
  - $\mathcal{Q}_n$- Poisson point process with intensity function $n\kappa$.
  - $\mathcal{C}(\mathcal{Q}_n, r_n)$- random Čech complex in this setting.
- **Euclidean Setting:**
  - $n^{1/N} r_n \to r \in (0, \infty)$.
  - $\mathcal{P}_n$- Poisson point process with intensity function nf.
  - $\mathcal{C}(\mathcal{P}_n, r_n, \rho)$- random Čech complex in this setting, with the general metric ρ, satisfying certain properties.

Aim: To establish the limiting behavior of the random variables $\beta_k(\mathcal{Q}_n, r_n)$ and $\beta_k(\mathcal{P}_n, r_n, \rho)$ in the thermodynamic regime.

Literature

Bobrowski and Mukherjee 2015, Theorem 4.3

Assume $\mathcal{M} \subset \mathbb{R}^N$ to be closed and smooth manifold, and κ to be bounded and measurable. Then for $1 \leq k \leq m-1$, as $n \to \infty$ with $n^{1/m} r_n \to r \in (0, \infty)$, there exists constants $c_1, c_2 > 0$ such that

$$c_1 n \leq \mathbb{E}[\beta_k(\mathcal{Q}_n, r_n)] \leq c_2 n.$$

Yogeshwaran, Subag, and Adler 2017; Trinh 2017, Corollary 1.4

Assume the support of $f(x)$ is compact and convex and that

$$0 < \inf_{x \in supp(f)} f(x) \leq \sup_{x \in supp(f)} f(x) < \infty.$$

Assume further that f is Riemann integrable. Then for $1 \leq k \leq N-1$, as $n \to \infty$ with $n^{1/N} r_n \to r \in (0, \infty)$,

$$\frac{\beta_k(\mathcal{P}_n, r_n)}{n} \to \int_{\mathbb{R}^N} \hat{\beta}_k^{(N)}(f(x), r) dx \text{ a.s.}$$

Limiting constant: $\hat{\beta}_k^{(N)}(\lambda, r)$

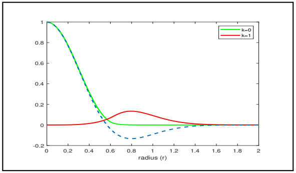
Notation

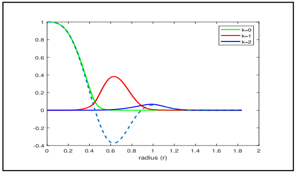
- $\mathcal{P}(\lambda) :=$ homogenous Poisson point process on $\mathbb{R}^N$ with constant intensity function $\lambda > 0$.
- $\mathcal{P}_L(\lambda) :=$ restriction of $\mathcal{P}(\lambda)$ on the window $W_L = (\frac{-L^{1/N}}{2}, \frac{L^{1/N}}{2}]^N$, where $L > 0$. It can also be denoted by $\mathcal{P}(\lambda)|_{W_L}$.

Yogeshwaran, Subag, and Adler 2017, Theorem 3.5

Then for $1 \leq k \leq (N-1)$, as $L \to \infty$,

$$\frac{\beta_k(\mathcal{P}_L(\lambda), r)}{L} \to \hat{\beta}_k^{(N)}(\lambda, r) \text{ a.s.}$$





Limiting constant contd...

Explicit formula for $\hat{\beta}_k$, except for $k = 0$, is unknown- (OPEN PROBLEM).

Trinh 2017, Lemma 2.3

(i) *Scaling Property:* For any $\theta > 0$,

$$\hat{\beta}_k^{(N)}(\lambda, r) = \frac{1}{\theta}\hat{\beta}_k^{(N)}\left(\lambda\theta, \frac{r}{\theta^{1/N}}\right).$$

(ii) *Continuity and Positivity:* $\hat{\beta}_k^{(N)}(\lambda, r)$ is a continuous function in both λ and r. If $\lambda, r > 0$ then $\hat{\beta}_k^{(N)}(\lambda, r) > 0$.

Follows from Bobrowski and Oliveira 2017, Proposition 6.1

Exponential decay: For $r \in (0, \infty)$,

$$\hat{\beta}_k^{(N)}(1, r) \leq C_{N,k+1} r^{Nk} \exp(-c_N r^N),$$

where $C_{N,k+1}$ and c_N are constants, depending only on their subscripts.

Our Main Results

Theorem 2.1 (For Manifolds).

Let $\mathcal{M} \subset \mathbb{R}^N$ to be a compact m-dimensional C^1 manifold with $m < N$. Assume that $\kappa(z)$ is a non-negative function, supported on $\mathcal{M}$ and for all $j \in \mathbb{N}$, $\int_{\mathcal{M}} \kappa(z)^j dz < +\infty$. Then as $n \to \infty$ with $n^{1/m} r_n \to r \in (0, \infty)$,

$$\frac{\beta_k(\mathcal{Q}_n, r_n)}{n} \to \int_{\mathcal{M}} \hat{\beta}_k^{(m)}(\kappa(z), r) dz \text{ a.s.},$$

where dz is a volume form on $\mathcal{M}$.

It is worth mentioning the following lemma:

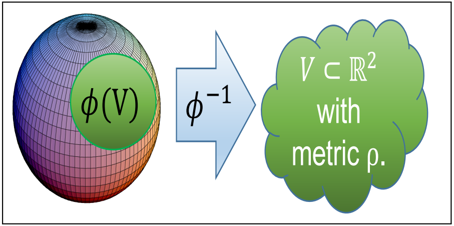
Lemma 2.2.

Under the same assumptions as in Theorem 2.1,

$$\frac{\mathbb{E}[\beta_k(\mathcal{Q}_n, r_n)]}{n} \to \int_{\mathcal{M}} \hat{\beta}_k^{(m)}(\kappa(z), r) dz.$$

A key idea to deal with the manifold setting

- Assume for instance that the support of κ lies entirely in a single chart (V, ϕ), i.e., $\mathrm{supp}(\kappa) \subset \phi(V)$.
- Then $\{X_i = \phi^{-1}(Z_i)\}_{i \geq 1}$ becomes an i.i.d. sequence of random variables on $V \subset \mathbb{R}^m$.
- If we define the metric ρ on V by $\rho(x, y) = \|\phi(x) - \phi(y)\|$, then $\mathcal{C}(\mathcal{Q}_n, r)$ is identical with $\mathcal{C}(\mathcal{P}_n, r, \rho)$, the Čech complex of radius r, constructed on $\mathcal{P}_n \subset \mathbb{R}^m$ using ρ.
- Thus, the problem on a manifold is converted to that on the Euclidean setting with a general metric ρ, which is easier for us to handle.



Main results contd...

Assumptions on the general metric ρ

Let for $x \in \mathbb{R}^N$, $\mathbf{B}_x$ be a positive definite $N \times N$ matrix such that the map $x \mapsto \mathbf{B}_x$ is measurable. Let for $y, z \in \mathbb{R}^N$, $d_x(y,z) := \|\mathbf{B}_x(y-z)\|$. Let $\mathcal{A}$ be a non-empty subset of $\mathbb{R}^N$, equipped with ρ.

(P1) For given $x \in \mathcal{A}$ and $\varepsilon > 0$, there exists $\delta = \delta_{x,\varepsilon} > 0$ such that for $y, z \in \mathcal{A}$, whenever $y, z \in B(x, \delta)$,

$$(1-\varepsilon)d_x(y,z) \leq \rho(y,z) \leq (1+\varepsilon)d_x(y,z).$$

(P2) There exist δ, c and $C > 0$ such that for $y, z \in \mathcal{A}$, whenever $\|y-z\| \leq \delta$,

$$c\|y-z\| \leq \rho(y,z) \leq C\|y-z\|.$$

Main results contd...

Theorem 2.3 (For Euclidean spaces).

Let $(\mathcal{A}, \rho)$ be a metric space, where $\mathcal{A}$ is a Borel subset of $\mathbb{R}^N$ with $Leb^N(\partial\mathcal{A}) = 0$ and the metric ρ satisfies the properties (P1) and (P2). Assume that $f(x)$ is a non-negative function that satisfies for all $j \in \mathbb{N}$, $\int_{\mathbb{R}^N} f(x)^j dx < +\infty$. Then as $n \to \infty$ with $n^{1/N} r_n \to r \in (0, \infty)$,

$$\frac{\beta_k(\mathcal{P}_n, r_n, \rho)}{n} \to \int_{\mathbb{R}^N} \hat{\beta}_k^{(N)} \left(\frac{f(x)}{D(x)}, r \right) D(x) dx \quad \text{a.s.},$$

where $D(x) := \det(\mathbf{B}_x)$.

Lemma 2.4.

Under the same assumptions as in Theorem 2.3,

$$\frac{\mathbb{E}[\beta_k(\mathcal{P}_n, r_n, \rho)]}{n} \to \int_{\mathbb{R}^N} \hat{\beta}_k^{(N)} \left(\frac{f(x)}{D(x)}, r \right) D(x) dx.$$

Part 3

Idea of Proof

Main tools

$S_j(\mathcal{K})$ counts the number of j-simplices in a simplicial complex $\mathcal{K}$.

Lemma 3.1 (Yogeshwaran, Subag, and Adler 2017).

Let $\mathcal{K}, \tilde{\mathcal{K}}$ be two finite simplicial complexes such that $\tilde{\mathcal{K}} \subset \mathcal{K}$. Then for every $k \geq 0$,

$$\left|\beta_k(\mathcal{K}) - \beta_k(\tilde{\mathcal{K}})\right| \leq \sum_{j=k}^{k+1} \left(S_j(\mathcal{K}) - S_j(\tilde{\mathcal{K}})\right).$$

Finite additivity of Betti numbers

Let $\mathcal{K}_1, \mathcal{K}_2 \ldots, \mathcal{K}_n$ be a finite number of disjoint simplicial complexes. Then for all $k \geq 0$,

$$\beta_k\left(\bigcup_{i=1}^{n} \mathcal{K}_i\right) = \sum_{i=1}^{n} \beta_k(\mathcal{K}_i). \tag{1}$$

LLN for simplex counts $S_j(\cdot)$

- Number of j-simplices in $\mathcal{C}(\mathfrak{X}, r, \rho)$ can be written as

$$S_j(\mathfrak{X}, r_n, \rho) = \sum_{\mathcal{Y} \subset \mathfrak{X}} h_{j,r_n,\rho}(\mathcal{Y}), \tag{2}$$

  where $\mathfrak{X} \subset \mathcal{A}$ is finite set and $h_{j,r_n,\rho}(\mathcal{Y})$ is the indicator function which is equal to 1 iff $\mathcal{Y}$ is a j-simplex.
- The above representation implies $S_j(\cdot)$ are local statistics and therefore, their LLN may follow from general theory of local functions due to Penrose 2007; Penrose and Yukich 2003.
- However, we give an elementary proof by calculating the order of the fourth moments.
- For $r \in (0, \infty)$ and $\mathbf{x} = (x_1, x_2, \ldots, x_j) \in (\mathbb{R}^N)^j$, define

$$A_j^{(N)}(r) = \frac{r^{Nj}}{(j+1)!} \int_{(\mathbb{R}^N)^j} h_j(0, \mathbf{x}) d\mathbf{x},$$

  where $h_j(0, \mathbf{x})$ and $d\mathbf{x}$ stand for $h_{j,1,\|\cdot\|}(0, x_1, x_2, \ldots, x_j)$ and $dx_1 \cdots dx_j$ respectively.

LLN for simplex counts $S_j(\cdot)$ contd.

Proposition 3.2.

Assume that $\int_{\mathcal{A}} f(x)^{j+1} dx < +\infty$ *and* $\lim_{n\to\infty} r_n = 0$*. Then*

$$\lim_{n\to\infty} r_n^{-Nj} n^{-(j+1)} \mathbb{E}[S_j(\mathcal{P}_n, r_n, \rho)] = A_j^{(N)}(1) \int_{\mathcal{A}} \frac{f(x)^{j+1}}{D(x)^j} dx.$$

In the thermodynamic regime, Proposition 3.2 is restated as follows.

Corollary 3.3.

Assume that $\int_{\mathcal{A}} f(x)^{j+1} dx < +\infty$ *and* $\lim_{n\to\infty} n^{1/N} r_n = r \in (0,\infty)$*. Then*

$$\lim_{n\to\infty} \frac{\mathbb{E}[S_j(\mathcal{P}_n, r_n, \rho)]}{n} = A_j^{(N)}(r) \int_{\mathcal{A}} \frac{f(x)^{j+1}}{D(x)^j} dx.$$

LLN for simplex counts $S_j(\cdot)$ contd.

Proposition 3.4.

Assume that $\int_{\mathcal{A}} f(x)^{4j+1} dx < +\infty$ *and* $\lim_{n\to\infty} n^{1/N} r_n = r \in (0, \infty)$. *Then as* $n \to \infty$,

$$\frac{S_j(\mathcal{P}_n, r_n, \rho)}{n} \to A_j^{(N)}(r) \int_{\mathcal{A}} \frac{f(x)^{j+1}}{D(x)^j} dx \text{ a.s.}$$

The above proposition is proved by using the standard technique:

- Let $\xi_n = S_j - \mathbb{E}[S_j]$. We show that $\mathbb{E}[\xi_n^4] \leq Kn^2$, where K is some positive constant.
- Then by Markov's inequality, $\mathbb{P}(|\xi_n| \geq n\varepsilon) \leq Kn^{-2}\varepsilon^{-4}$.
- Since $\sum n^{-2} < \infty$, by the first Borel–Cantelli lemma, $\mathbb{P}(\limsup_n |n^{-1}\xi_n| \geq \varepsilon) = 0$.
- This means $n^{-1}\xi_n$ converges to zero almost surely.

Sufficient Requirement

To prove results for usual Betti numbers in Euclidean setting, it is sufficient to prove the following proposition

Proposition 3.5.

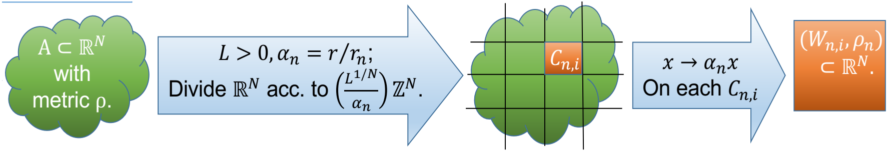
Let $(\mathcal{A}, \rho)$ be a metric space, where $\mathcal{A}$ is a compact subset of $\mathbb{R}^N$ with $Leb^N(\partial\mathcal{A}) = 0$ and the metric ρ satisfies the property (P1). Assume that $f(x)$ is a non negative function on $\mathcal{A}$ and is bounded. Then as $n \to \infty$ with $n^{1/N} r_n \to r \in (0, \infty)$,

$$\frac{\beta_k(\mathcal{P}_n, r_n, \rho)}{n} \to \int_{\mathcal{A}} \hat{\beta}_k^{(N)}\left(\frac{f(x)}{D(x)}, r\right) D(x) dx \;\; \text{a.s.}$$

Here, $\mathcal{P}_n$ is a Poisson point process on $\mathcal{A}$ with intensity function $nf(x)$.

Proof of Proposition 3.5

For the proof, we partition the set $\mathcal{A}$ as follows



- The limiting behavior of $\beta_k(\mathcal{P}_n, r_n, \rho)$ will be estimated by that of $\beta_k\left(\cup_i \mathcal{C}(\mathcal{P}_n|_{C_{n,i}}, r_n, \rho)\right)$
- Consider the map $x \mapsto \alpha_n x$ and let $W_{n,i}$ be the image of $C_{n,i}$. Define a metric on $\alpha_n \mathcal{A}$ as

$$\rho_n(x, y) := \alpha_n \rho\left(x/\alpha_n, y/\alpha_n\right).$$

- Let $\tilde{\mathcal{P}}_n = \alpha_n \mathcal{P}_n$. Then $\tilde{\mathcal{P}}_n$ is a Poisson point process on $\mathbb{R}^N$ with intensity function

$$n/\alpha_n^N f(x/\alpha_n) =: f_n(x).$$

Proof contd.

Then the proof of Proposition follows from the following lemma

Lemma 3.6.

For fixed $L > 0$, as $n \to \infty$,

$$\text{(a)} \quad \frac{1}{n}\sum_i \beta_k(\tilde{\mathcal{P}}_n|_{W_{n,i}}, r, \rho_n) \to \int_{\mathcal{A}} \frac{\mathbb{E}[\beta_k(\mathcal{P}_L(f(x)), r, d_x)]}{L} dx \text{ a.s.,}$$

$$\text{(b)} \quad \frac{1}{n}\sum_i S_j(\tilde{\mathcal{P}}_n|_{W_{n,i}}, r, \rho_n) \to \int_{\mathcal{A}} \frac{\mathbb{E}[S_j(\mathcal{P}_L(f(x)), r, d_x)]}{L} dx \text{ a.s.}$$

As $L \to \infty$,

$$\text{(c)} \quad \int_{\mathcal{A}} \frac{\mathbb{E}[\beta_k(\mathcal{P}_L(f(x)), r, d_x)]}{L} dx \to \int_{\mathcal{A}} \hat{\beta}_k^{(N)}\left(\frac{f(x)}{D(x)}, r\right) D(x) dx,$$

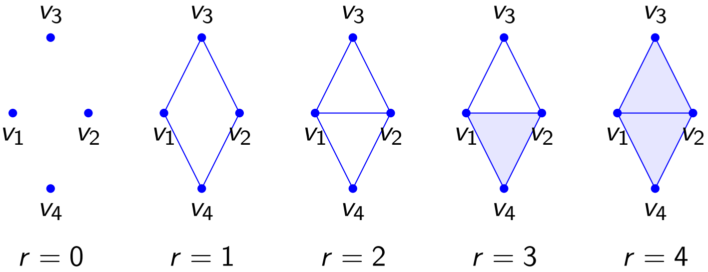
$$\text{(d)} \quad \int_{\mathcal{A}} \frac{\mathbb{E}[S_j(\mathcal{P}_L(f(x)), r, d_x)]}{L} dx \to \int_{\mathcal{A}} \hat{S}_j^{(N)}\left(\frac{f(x)}{D(x)}, r\right) D(x) dx.$$

Part 4
Extension to Persistent Homology

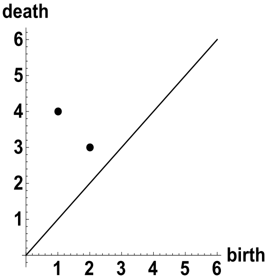
Persistent Homology

- Persistent homology overcomes the issues of noise and sensitivity of the parameter r.
- The main idea behind persistent homology is to consider the whole range of radius r instead of some particular value.
- In other words, persistent homology is defined for a filtration of simplicial complexes.
- *Filtration* of simplicial complexes $:= \{\mathcal{K}_r\}_{r \geq 0}$, such that if $0 \leq r \leq s$ then $\mathcal{K}_r$ is a subcomplex of $\mathcal{K}_s$ and for all $r \geq 0$, $\mathcal{K}_r = \cap_{s>r} \mathcal{K}_s$. Let $\mathbb{K} := \{\mathcal{K}_r\}_{r \geq 0}$.
- *Persistent Betti numbers*$:= \beta_k^{s,t}(\mathbb{K})$ is the number of k-dimensional holes that appear before or at s and still alive at t in the filtration $\mathbb{K}$.
- Persistent homology has unique representation, which is visualized by the kth *persistence diagram*, defined as a multi-subset of $\Delta = \{(x, y) \in \bar{\mathbb{R}}^2 \colon 0 \leq x < y \leq \infty\}$, i.e.,

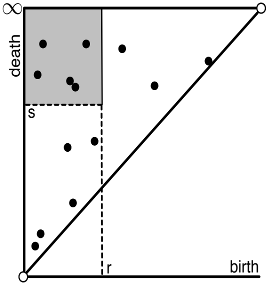
$$\mathrm{Dgm}_k(\mathbb{K}) = \{(b_i, d_i) \in \Delta \colon i = 1, 2 \ldots, p\}.$$


v_3
v_1
v_2
v_4
$r = 0$
$r = 1$
$r = 2$
$r = 3$
$r = 4$

(a)


death
6
5
4
3
2
1
1
2
3
4
5
6
birth

(b)


∞
death
s
r
birth

(c)

Our results for persistent Betti numbers

Theorem 4.1.

(a) Under the same assumptions as in Theorem 2.1, for any $0 \le s \le t < \infty$, *as* $n \to \infty$,

$$\frac{\beta_k^{s,t}\left(\mathcal{C}(n^{\frac{1}{m}}\mathcal{Q}_n)\right)}{n} \to \int_{\mathcal{M}} \hat{\beta}_k^{(m)}(\kappa(z), s, t)\, dz \text{ a.s.}$$

(b) Under the same assumptions as in Theorem 2.3, for any $0 \le s \le t < \infty$, *as* $n \to \infty$,

$$\frac{\beta_k^{s,t}\left(\mathcal{C}(n^{\frac{1}{N}}\mathcal{P}_n, \rho_n)\right)}{n} \to \int_{\mathbb{R}^N} \hat{\beta}_k^{(N)}\left(\frac{f(x)}{D(x)}, s, t\right) D(x)dx \text{ a.s.}$$

The definition of the limiting constant is given by Hiraoka, Shirai, and Trinh 2018.

Vague convergence of Persistence Diagrams

- By identify persistence diagrams as an integer-valued Radon measures on Δ (shown below), we also discuss their vague convergence. Let $\mathrm{Dgm}_k := \mathrm{Dgm}_k(\mathbb{K})$.

$$\mathrm{Dgm}_k(\mathbb{K}) = \sum_{(b_i,d_i)\in\mathrm{Dgm}_k} \delta_{(b_i,d_i)}.$$

Let $\mathfrak{M}$ be the set of all Radon measures on Δ.

Definition 4.2.

A sequence $\{\mu_n\}_{n\geq 1} \subset \mathfrak{M}$ converges to $\mu \in \mathfrak{M}$ *vaguely* (or *in the vague topology*), denoted by $\mu_n \xrightarrow{v} \mu$, if for every continuous function f with compact support, $\int_\Lambda f d\mu_n$ converges to $\int_\Lambda f d\mu$.

Vague convergence contd.

Theorem 4.3.

(a) Under the same assumptions as in Theorem 2.1, as $n \to \infty$,

$$\frac{\mathrm{Dgm}_k\left(\mathcal{C}(n^{\frac{1}{m}}\mathcal{Q}_n)\right)}{n} \xrightarrow{v} \nu^{(m)}_{k,\kappa} \text{ a.s.},$$

where for $0 \le k \le m-1$ and $A \in \mathcal{R}(\Delta)$,

$$\nu^{(m)}_{k,\kappa}(A) = \int_{\mathcal{M}} \nu^{(m)}_{k,\kappa(z)}(A)dz = \int_{\mathcal{M}} \nu^{(m)}_{k,1}(\kappa(z)^{1/m}A)dz,$$

and for all $k \ge m$, $\nu^{(m)}_{k,\kappa}$ is a null measure.
(b) Under the same assumptions as in Theorem 2.3, as $n \to \infty$,

$$\frac{\mathrm{Dgm}_k\left(\mathcal{C}(n^{\frac{1}{N}}\mathcal{P}_n, \rho_n)\right)}{n} \xrightarrow{v} \nu^{(N)}_{k,f/D} \text{ a.s.}$$

For persistent Betti numbers

Results for persistent Betti numbers can be proved in a similar way as results for usual Betti numbers will be proved. This is because of the following lemma.

Lemma 4.4 (Hiraoka, Shirai, and Trinh 2018).

Let $\mathbb{K} = \{\mathcal{K}_r\}_{r \geq 0}$ *and* $\tilde{\mathbb{K}} = \{\tilde{\mathcal{K}}_r\}_{r \geq 0}$ *be filtrations of Čech complexes such that for all* $r \geq 0$, $\tilde{\mathcal{K}}_r \subset \mathcal{K}_r$. *Then*

$$\left|\beta_k^{s,t}(\mathbb{K}) - \beta_k^{s,t}(\tilde{\mathbb{K}})\right| \leq \sum_{j=k}^{k+1} \left(S_j(\mathcal{K}_t) - S_j(\tilde{\mathcal{K}}_t)\right).$$

Vague convergence of persistence diagrams follows from the results for persistent Betti numbers and the general theory of vague convergence.

References I

Bobrowski, Omer and Sayan Mukherjee (2015). "The topology of probability distributions on manifolds". In: *Probab. Theory Related Fields* 161.3-4, pp. 651–686.

Bobrowski, Omer and Goncalo Oliveira (2017). "Random Čech complexes on Riemannian manifolds". In: *Random Structures & Algorithms*.

Hiraoka, Yasuaki, Tomoyuki Shirai, and Khanh Duy Trinh (2018). "Limit theorems for persistence diagrams". In: *Ann. Appl. Probab.* 28.5, pp. 2740–2780.

Niyogi, Partha, Stephen Smale, and Shmuel Weinberger (2008). "Finding the homology of submanifolds with high confidence from random samples". In: *Discrete Comput. Geom.* 39.1-3, pp. 419–441.

Penrose, Mathew D. (2007). "Laws of large numbers in stochastic geometry with statistical applications". In: *Bernoulli* 13.4, pp. 1124–1150.

References II

Penrose, Mathew D. and J. E. Yukich (2003). "Weak laws of large numbers in geometric probability". In: *Ann. Appl. Probab.* 13.1, pp. 277–303.

Trinh, Khanh Duy (2017). "A remark on the convergence of Betti numbers in the thermodynamic regime". In: *Pac. J. Math. Ind.* 9, Art. 4, 7.

Yogeshwaran, D., Eliran Subag, and Robert J. Adler (2017). "Random geometric complexes in the thermodynamic regime". In: *Probab. Theory Related Fields* 167.1-2, pp. 107–142.

Thank you
for your attention