The self-similar evolution of stationary point processes via persistent homology

Daniel Spitz

Journal Club on Topological Data Analysis

December 8th, 2020

Based on: DS, Anna Wienhard, arXiv:2012.05751

Motivation

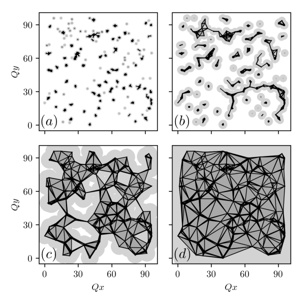
Point clouds from samples of functions $\psi_\omega(t) : \Lambda \to \mathbb{C}$, $\Lambda \subset \mathbb{R}^2$:

$$X_{\bar{\nu},\omega}(t) := |\psi_\omega(t)|^{-1}[0,\bar{\nu}] \subset \Lambda. \quad (1)$$

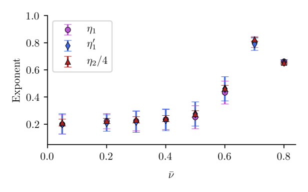
Scaling ansatz for density of persistence diagram measures:

$$\tilde{\mathfrak{p}}(t,\Lambda)(b,d) = (t/t')^{-\eta_2}\tilde{\mathfrak{p}}(t',\Lambda)((t/t')^{-\eta_1}b,(t/t')^{-\eta_1'}d). \quad (2)$$

Alpha complexes of different radii



[Reprinted from Spitz *et al.*, arXiv:2001.02616]



[Reprinted from Spitz *et al.*, arXiv:2001.02616]

Found $\eta_2/\eta_1 \simeq 4$ in $n = 2$. Recent simulations show $\eta_2/\eta_1 \simeq 5$ in $n = 3$. Geometric explanation?

Contents

Contents

Random measures and point processes

$\mathcal{M}^n$: Space of boundedly finite measures on $\mathbb{R}^n$
$\mathcal{N}^n$: Integer-valued ones
$(\Omega, \mathcal{E}, \mathbb{P})$: Probability space

Random measure is a measurable map $\mu : (\Omega, \mathcal{E}, \mathbb{P}) \to (\mathcal{M}^n, \mathcal{B}(\mathcal{M}^n))$
Point process is a measurable map $\xi : (\Omega, \mathcal{E}, \mathbb{P}) \to (\mathcal{N}^n, \mathcal{B}(\mathcal{N}^n))$

To any realization ξ_ω, $\omega \in \Omega$, there exists countable set of *atoms* $\{x_1, x_2, \dots\} \subset \mathbb{R}^n$ such that $\forall\, A \in \mathcal{B}^n$:

$$\xi_\omega(A) = \sum_i \delta_{x_i}(A). \tag{3}$$

Point process *simple*, if a.s. all its atoms are distinct.
Point process *has all finite moments*, if $\forall\, k \geq 1$, $A \in \mathcal{B}_b^n$: $\mathbb{E}[\xi(A)^k] < \infty$.

For simple point process ξ define point clouds:

$$X_{\xi_\omega}(A) = \left\{ x_i \;\middle|\; \xi_\omega(A) = \sum_i \delta_{x_i}(A) \right\}. \tag{4}$$

Filtered simplicial complexes and persistent homology

$X \subset \mathbb{R}^n$ point cloud. *Čech complexes*: $\check{C}_r(X) := \{\sigma \subseteq X \mid \bigcap_{x\in\sigma} B_r(x) \neq \emptyset\}$.

ℓth persistence diagram of $H_\ell(\mathcal{C}(X))$ defined as finite multiset

$$\mathrm{Dgm}_\ell(X) := \left\{ (b_i, d_i) \,\middle|\, H_\ell(\mathcal{C}(X)) \cong \bigoplus_{i=1}^{k} I(b_i, d_i) \right\}. \tag{5}$$

Lemma

For ξ simple point process on $\mathbb{R}^n$, $D \circ X_\xi$ defines point process on space of pers. diags. $\mathscr{D}$ endowed with the Wasserstein metric W_p, for any $p > n$, via

$$\omega \mapsto \left(A \mapsto \sum_{x \in D_A(X_{\xi_\omega}(A))} \delta_x \right), \qquad D_A : X \mapsto \bigcup_{\ell=0}^{n-1} \mathrm{Dgm}_\ell(X) \tag{6}$$

for all $\omega \in \Omega$, $A \in \mathcal{B}_b^n$. Equivalently, $D_A \circ X_{\xi_\cdot}(A) : \Omega \to \mathscr{D}$ is a random variable for each $A \in \mathcal{B}_b^n$.

Bounded total persistence

M: Triangulable, compact metric space. K: Simp. complex giving triangulation ϑ of M, $\mathrm{diam}(\sigma) := \max_{x,y\in\sigma} d(\vartheta(x), \vartheta(y))$,

$$\mathrm{mesh}(K) := \max_{\sigma\in K} \mathrm{diam}(\sigma), \qquad N(r) := \min_{\mathrm{mesh}(K)\leq r} \#K \tag{7}$$

Degree-k total persistence of filtration of sublevel sets of f:

$$\mathrm{Pers}_k(f) := \sum_{\mathrm{pers}(x)>t} \mathrm{pers}(x)^k. \tag{8}$$

Proposition (Cohen-Steiner *et al.*, Found. Comp. Math. 10(2), 2010)

Assume that size of smallest triangulation of M grows polynomially with one over mesh, i.e. there exist C_0, m, *such that* $N(r) \leq C_0/r^m$ *for all* $r > 0$. *Let* $\delta > 0$ *and* $k = m + \delta$. *Then,*

$$\mathrm{pers}(f) \leq \frac{m + 2\delta}{\delta} C_0\,\mathrm{Lip}(f)^m \mathrm{Amp}(f)^\delta, \tag{9}$$

where $\mathrm{Amp}(f) := \max_{x\in M} f(x) - \min_{y\in M} f(y)$.

M implies bounded degree-k total persistence, if $\exists C_M > 0$: $\mathrm{Pers}_k(f) \leq C_M$ for every tame function $f : M \to \mathbb{R}$ with $\mathrm{Lip}(f) \leq 1$.

Example: Compact Riemannian n-manifolds have bounded degree-n total pers.

Ergodicity of point processes

Random measure μ *stationary*, if $\forall\, x \in \mathbb{R}^n$ the probability distributions of μ and $\theta_x\mu$ coincide ($(\theta_x\mu)(A) := \mu(A+x)$)

Stationary random measure μ *ergodic*, if $\mathbb{P}(\mu \in \mathcal{A}) \in \{0,1\}$ for all Borel sets $\mathcal{A}$ of $\mathcal{M}^n$ that are invariant under translation by x for all $x \in \mathbb{R}^n$.

$\{A_k\} \subset \mathcal{B}_b^n$ called a *convex averaging sequence* if

$$\text{(i) each } A_k \text{ is convex,} \qquad \text{(ii) } A_k \subseteq A_{k+1} \text{ for all } k, \qquad \text{(iii) } \lim_{k\to\infty} r(A_k) = \infty, \tag{10}$$

where $r(A) := \sup\{r \,|\, A \text{ contains a ball of radius } r\}$.

Proposition (Daley & Vere-Jones 2007)

$\{A_k\}$ convex averaging sequence. When random measure ξ on $\mathbb{R}^d$ is stationary, ergodic and has finite expectation measure with

$$m = \mathbb{E}[\xi([0,1]^n)] = m\,\lambda_n([0,1]^n), \tag{11}$$

then $\xi(A_k)/\lambda_n(A_k) \to m$ almost surely as $k \to \infty$.

Ergodicity in persistent homology

Definition

ξ stationary simple point process on $\mathbb{R}^n$ with finite expectation measure, $\{A_k\}$ convex averaging sequence. Set for all k:

$$n(X_k) := \# \bigcup_{\ell=0}^{n-1} \mathrm{Dgm}_\ell(X_\xi(A_k)). \qquad (12)$$

Say that ξ is ergodic in persistence if a.s. for any $\epsilon > 0$ $\exists N \in \mathbb{N}$ such that for all $k, l \geq N$,

$$\left| \frac{n(X_k)}{n(X_l)} - \frac{\lambda_n(A_k)}{\lambda_n(A_l)} \right| < \epsilon. \qquad (13)$$

Lemma

ξ simple point process on $\mathbb{R}^n$ having all finite moments. If ξ is stationary and ergodic, then ξ is ergodic in persistence.

Persistence diagram measures

ξ simple point process, $A \in \mathcal{B}_b^n$, define for all $\omega \in \Omega$:

$$\rho_\omega(A) := \sum_{x \in D_A(X_{\xi_\omega}(A))} \delta_x. \tag{14}$$

Defines a point process $\rho_\cdot(A)$ on $\Delta := \{(b, d) \,|\, 0 \leq b < d \leq \infty\}$.

Persistence diagram measure: $\rho_\cdot : \Omega \times \mathcal{B}_b^n \to \mathcal{N}(\Delta)$.

If they exist, its first moment measures define $\mathfrak{p}(A) := \mathbb{E}[\rho_\omega(A)]$.

Persistence diagram expectation measure: $\mathfrak{p} : \mathcal{B}_b^n \to \mathcal{R}(\Delta)$, $A \mapsto \mathfrak{p}(A)$

Lemma

ξ simple stationary and ergodic point process on $\mathbb{R}^n$ having all finite moments. Then the corresponding persistence diagram expectation measure exists.

Geometric quantities

Number of persistent homology classes:

$$n_\omega(A) := \int_\Delta \rho_\omega(A)(\mathrm{d}x) = \rho_\omega(A)(\Delta) \tag{15}$$

Expectation: $\mathfrak{n}(A) := \int_\Delta \mathfrak{p}(A)(\mathrm{d}x) = \mathfrak{p}(A)(\Delta)$

Let $q > 0$. *Degree-q persistence*:

$$l_{q,\omega}(A) := \left[\frac{1}{n_\omega(A)} \int_\Delta \mathrm{pers}(x)^q \, \rho_\omega(A)(\mathrm{d}x)\right]^{1/q} \tag{16}$$

Expectation: $\mathfrak{l}_q(A) := \left[\frac{1}{\mathfrak{n}(A)} \int_\Delta \mathrm{pers}(x)^q \, \mathfrak{p}(A)(\mathrm{d}x)\right]^{1/q}$

Maximum death:

$$d_{\max,\omega}(A) := \max\{d \mid (b,d) \in D_A(X_{\rho_\omega}(A))\} = \lim_{p\to\infty} \left[\int_\Delta d(x)^p \, \rho_\omega(A)(\mathrm{d}x)\right]^{1/p} \tag{17}$$

Expectation: $\mathfrak{d}_{\max}(A) := \lim_{p\to\infty} \left[\int_\Delta d(x)^p \, \mathfrak{p}(A)(\mathrm{d}x)\right]^{1/p}$

Geometric quantities

Proposition

$\mathfrak{p}$ *persistence diagram expectation measure,* $A \in \mathcal{B}_b^n$*. Find*

$$\mathfrak{n}(A) = \mathbb{E}[n_\omega(A)], \qquad \mathfrak{d}_{\max}(A) = \mathbb{E}[d_{\max,\omega}(A)]. \tag{18}$$

If $\mathfrak{p}(A)$ *is boundedly finite, then* $\mathfrak{n}(A) < \infty$*. If* $\operatorname{supp}(\mathfrak{p}(A)) \subset \Delta$ *is bounded, then* $\mathfrak{d}_{\max}(A) < \infty$*.*

Proposition

ξ *stationary, ergodic simple point process on* $\mathbb{R}^n$ *having all finite moments,* $\mathfrak{p}$ *persistence diagram expectation measure from* ξ*,* $\{A_k\}$ *convex averaging sequence,* $\epsilon > 0$*. Find for all* $q > 0$ *and* k *suff. large:*

$$|\mathfrak{l}_q(A_k) - \mathbb{E}[l_{q,\omega}(A_k)]| < \epsilon. \tag{19}$$

Let $A \in \mathcal{B}_b^n$*. If* $\mathfrak{p}(A)$ *is boundedly finite and* $\operatorname{supp}(\mathfrak{p}(A)) \subset \Delta$ *is bounded, then* $\mathfrak{l}_q(A) < \infty$ *for all* $q > 0$*.*

Contents

Functional summaries

Problem: Persistence diagrams do not naturally lead to statistical goals!

$\mathscr{F}(T)$: Collection of functions on compact metric space T, $f : T \to \mathbb{R}$

Functional summary F is a map from the space of persistence diagrams to such a collection of functions, $F : \mathscr{D} \to \mathscr{F}(T)$

F *additive*, if for any two persistence diagrams $D, E \in \mathscr{D}$:
$\mathcal{F}(D + E) = \mathcal{F}(D) + \mathcal{F}(E)$ as sum of multisets

F *uniformly bounded*, if a constant $U < \infty$ exists, such that

$$\sup_{f \in \operatorname{im}(F)} \sup_{s \in T} |f(s)| \leq U \tag{20}$$

Proposition (Berry *et al.*, arXiv:1804.01618)

F uniformly bounded functional summary, $D_i \in \mathscr{D}$ for $i \in \mathbb{N}$, sampled from a probability space $(\mathscr{D}, \mathcal{B}(\mathscr{D}), \mathbb{P}_{\mathscr{D}})$, $f_i := F(D_i)$. If $\operatorname{im}(F)$ *equicontinuous, then a.s. for $n \to \infty$*

$$\sup_{s \in T} \left| \frac{1}{m} \sum_{i=1}^{m} f_i(s) - \mathbb{E}[F(D)(s)] \right| \to 0. \tag{21}$$

Functional summaries and persistence diagram measures

For additive func. summary $\mathcal{A}$, $s \in T$:

$$\mathcal{A}(D_A(X_{\xi_\omega}(A)))(s) = \sum_{x \in D_A(X_{\xi_\omega}(A))} \mathcal{A}(\{x\})(s) = \int_\Delta \mathcal{A}(\{x\})(s)\, \rho_\omega(A)(\mathrm{d}x). \tag{22}$$

$\Longrightarrow$ Persistence diagram measures show up.

Proposition

Assume that pers. diag. expectation measure $\mathfrak{p}$ exists. Then, for any functional summary $\mathcal{F} \in \mathscr{F}(K)$, $s \in K$, $A \in \mathcal{B}_b^n$ have

$$\mathbb{E}\left[\int_\Delta \mathcal{F}(\{x\})(s)\, \rho_\omega(A)(\mathrm{d}x)\right] = \int_\Delta \mathcal{F}(\{x\})(s)\, \mathfrak{p}(A)(\mathrm{d}x). \tag{23}$$

Intensive functional summaries

ξ stationary, ergodic simple point process on $\mathbb{R}^n$, $\mathcal{F} : \mathscr{D} \to \mathscr{F}(T)$ functional summary.

Inspired by statistical mechanics:

$\mathcal{F}$ *ξ-intensive*, if for any convex averaging sequence $\{A_k\}$, $\epsilon > 0$ and k sufficiently large:

$$\lim_{l \to \infty} ||\mathcal{F}(D_l) - \mathcal{F}(D_k)||_\infty < \epsilon, \tag{24}$$

$|| \cdot ||_\infty$: sup-norm, $D_k := \bigcup_{\ell=0}^{n-1} \mathrm{Dgm}_\ell(X_{\xi_\omega}(A_k))$.

Lemma

$\mathcal{A} \in \mathscr{A}(K)$ additive functional summary, ξ stationary, ergodic simple point process on $\mathbb{R}^n$, D_k as before. Then, $\mathcal{A}(D_k)/\lambda_n(A_k)$ a.s. defines a ξ-intensive functional summary.

$\Longrightarrow$ Intuition for ergodicity on level of persistence diagrams.

Corollary

$\mathcal{A}(D_k)/n(D_k)$ is a ξ-intensive functional summary

Intensive functional summaries

Proposition

ξ stationary, ergodic simple point process on $\mathbb{R}^n$, $\mathcal{F}$ a ξ-intensive functional summary, $\{A_k\}$ convex averaging sequence. Pick samples $\omega_i \in \Omega$, $i \in \mathbb{N}$, according to the probability distribution $\mathbb{P}$, define

$$D_{k,i} := \bigcup_{\ell=0}^{n-1} \mathrm{Dgm}_\ell(X_{\xi_{\omega_i}}(A_k)). \tag{25}$$

Then for any $j \in \mathbb{N}$, $\epsilon > 0$ and sufficiently large k find a.s.:

$$\sup_{s \in K} \left| \lim_{l \to \infty} \mathcal{F}(D_{l,j})(s) - \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} \mathcal{F}(D_{k,i})(s) \right| < \epsilon. \tag{26}$$

$\Longrightarrow$ For intensive functional summaries ensemble-average and infinite-volume average converge to each other.

Contents

Self-similar scaling

Inspiration: How to describe behavior as it typically appears in quantum-physical time evolutions in persistent-homological quantities?

$(\mathfrak{p}(t))_{t\in(T_0,T_1)}$ family of persistence diagram expectation measures, $0 < T_0 < T_1$. For $t \in (T_0, T_1)$, $A \in \mathcal{B}_b^n$ set $\mathfrak{p}(t, A) := \mathfrak{p}(t)(A)$, $\{A_k\}$ convex averaging sequence.

$(\mathfrak{p}(t))_{t\in(T_0,T_1)}$ *scales self-similarly between times* T_0 *and* T_1 *with exponents* $\eta_1, \eta_2 \in \mathbb{R}$, if for all $t, t' \in (T_0, T_1)$, $B \in \mathcal{B}(\Delta)$, k suff. large:

$$\mathfrak{p}(t, A_k)(B) = (t/t')^{-\eta_2}\mathfrak{p}(t', A_k)((t/t')^{-\eta_1} B), \tag{27}$$

where $\kappa B := \{(\kappa b, \kappa d) \,|\, (b, d) \in B\}$ for $\kappa \in [0, \infty)$.

Intuition based on pers. hom. with length scales as filtration parameter: Any persistence length scale blows up as a power-law t^{η_1}, find e.g.:

$$\mathfrak{n}(t, A_k) = (t/t')^{-\eta_2}\mathfrak{n}(t', A_k), \tag{28a}$$

$$\mathfrak{l}_q(t, A_k) = (t/t')^{\eta_1}\mathfrak{l}_q(t', A_k), \tag{28b}$$

$$\mathfrak{d}_{\max}(t, A_k) = (t/t')^{\eta_1}\mathfrak{d}_{\max}(t', A_k). \tag{28c}$$

Packing relation

Can we find a relation between the occurring exponents based on geometrical arguments?

Based on notion of bounded total persistence obtain:

Lemma (Packing lemma)

ξ simple point process on $\mathbb{R}^n$, ρ persistence diagram measure computed from ξ. Then there exists $c > 0$, s.t. for any $\delta > 0$, $\omega \in \Omega$ and $A \in \mathcal{B}_b^n$:

$$n_\omega(A) \leq \frac{c\,(n+2\delta)}{\delta} \frac{d_{\max,\omega}(A)^\delta}{l_{n+\delta,\omega}(A)^{n+\delta}}. \tag{29}$$

Theorem (Packing relation)

$(\xi(t))_{t\in(T_0,T_1)}$ family of stationary, ergodic simple point processes on $\mathbb{R}^n$ having all finite moments, $(\mathfrak{p}(t))_t$ family of persistence diagram expectation measures computed from $(\xi(t))_t$. Assume that the family scales self-similarly between T_0 and T_1 with exponents $\eta_1, \eta_2 \in \mathbb{R}$. Then, if the interval (T_0, T_1) is sufficiently extended find a.s.:

$$\eta_2 = n\eta_1. \tag{30}$$

$\Longrightarrow$ Confirms exponents found in numerical simulations.

Example: Poisson process with power-law scaling intensity

Point processes $(\xi(t))_{t\in[0,\infty)}$ on $\mathbb{R}^n$ form *time-dependent Poisson process*, if there exists $\gamma : [0,\infty) \to (0,\infty)$, s.t. $\forall\, t \in [0,\infty)$,

(i) $\mathbb{E}[\xi(t,A)] = \gamma(t)\lambda_n(A)$,

(ii) for every $m \in \mathbb{N}$ and all pairwise disjoint Borel sets $A_1,\ldots,A_m \in \mathcal{B}^n$ the random variables $\xi(t,A_1),\ldots,\xi(t,A_m)$ are independent.

Proposition

$(\xi(t))_{t\in[0,\infty)}$ *time-dependent Poisson process with intensity function*

$$\gamma(t) = \gamma_0 t^{-n\eta_1}, \tag{31}$$

$\gamma_0 > 0$ *and* $\eta_1 \geq 0$, $\{A_k\}$ *convex averaging sequence. Then the family* $(\mathfrak{p}(t,A_k)/\lambda_n(A_k))_t$ *converges vaguely for* $k \to \infty$ *to a family of Radon measures* $(\mathfrak{P}(t))$ *which scales self-similarly between* 0 *and* ∞ *with exponents* η_1 *and* $n\eta_1$.

Contents

Strong law of large numbers

For the example of the time-dependent Poisson process needed to show the following generalization to the same results for cubes [Hiraoka, Shirai, Trinh, Ann. Appl. Prob. 28(5), 2018].

Theorem (Strong law of large numbers for persistent Betti numbers)

ξ simple point process on $\mathbb{R}^n$ having all finite moments, $\{A_k\}$ convex averaging sequence. If ξ is stationary, then for any $0 \leq r \leq s < \infty$ and $\ell \in \mathbb{Z}$ there exists a constant $\hat{\beta}_\ell^{r,s}$, s.t.

$$\frac{\mathbb{E}[\beta_\ell^{r,s}(\mathcal{C}(X_k))]}{\lambda_n(A_k)} \to \hat{\beta}_\ell^{r,s} \qquad \textit{for } k \to \infty. \tag{32}$$

Additionally, if ξ is ergodic, then a.s.

$$\frac{\beta_\ell^{r,s}(\mathcal{C}(X_k))}{\lambda_n(A_k)} \to \hat{\beta}_\ell^{r,s} \qquad \textit{for } k \to \infty. \tag{33}$$

Proof is lengthy, using techniques from convex and integral geometry.

Contents

Conclusions

- Theory of point processes well-suited to describe persistent homology in a probabilistic setting.
- Ergodicity, functional summaries and persistence diagram measures lead to diverse geometric quantities with well-defined expectations.
- Self-similar scaling approach yields interesting insights into geometry (packing relation).
- Strong law of large numbers for persistent Betti numbers generalized to arbitrary convex averaging sequences.

Further questions

- Can the packing relation be extended to non-ergodic point processes and their persistent homology?
- Do point processes exist, for which correlations do not reveal self-similar scaling but persistence diagram measures do?
- Can we generalize the results to weighted simplicial complexes?