Fréchet means of topological summaries: ATMCS

Sayan Mukherjee

Duke University

Joint work with:
K. Turner (University of Chicago), Y. Milyeko and J. Harer (Duke University),

July 1, 2012

Stochastics in TDA

1. Consistency/finite sample bounds – Convergence of topological inferences from random samples;

Stochastics in TDA

1. Consistency/finite sample bounds – Convergence of topological inferences from random samples;
2. Random topology – Topology of stochastic processes;

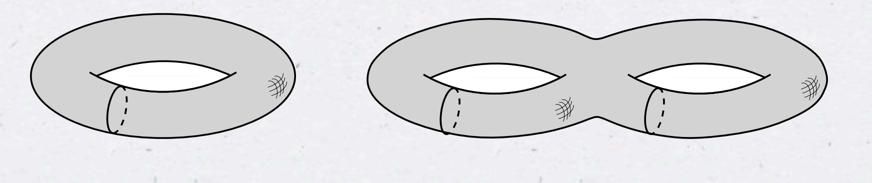
Stochastics in TDA

1. Consistency/finite sample bounds – Convergence of topological inferences from random samples;
2. Random topology – Topology of stochastic processes;
3. Robust topological summaries – Topological summaries robust to stochastic perturbations.

Our approach

1. Think of topological summaries as random variables
2. Use these summaries for inference.

Motivating example



Statistical inference

$D_1 \equiv X_1, \ldots, X_m \overset{iid}{\sim} f_{\theta_1}$: m point samples from a torus $\mathcal{O}_1$
$D_2 \equiv Y_1, \ldots, Y_n \overset{iid}{\sim} f_{\theta_2}$: n point samples from a double torus $\mathcal{O}_2$
Z drawn form either $\mathcal{O}_1$ or $\mathcal{O}_2$

Statistical inference

$D_1 \equiv X_1, ..., X_m \overset{iid}{\sim} f_{\theta_1}$: m point samples from a torus $\mathcal{O}_1$
$D_2 \equiv Y_1, ..., Y_n \overset{iid}{\sim} f_{\theta_2}$: n point samples from a double torus $\mathcal{O}_2$
Z drawn form either $\mathcal{O}_1$ or $\mathcal{O}_2$

Estimate $\hat{\theta}_1$ and $\hat{\theta}_2$ and then use Bayes' rule

$$\hat{p}(\mathcal{O}_1 \mid z, D_1, D_2) = \frac{\hat{p}(z \mid \mathcal{O}_1)\pi(\mathcal{O}_1)}{\hat{p}(z, D_1, D_2)} = \frac{f(z; \hat{\theta}_1)\pi(\mathcal{O}_1)}{f(z; \hat{\theta}_1)\pi(\mathcal{O}_1) + f(z; \hat{\theta}_2)\pi(\mathcal{O}_2)}.$$

Inference with summary statistics

For D_1 and D_2 compute summary statistics (Betti numbers)
$S_1 \equiv B_1, ..., B_m$: m summaries of the torus
$S_2 \equiv C_1, ..., C_n$: n summaries of the double torus
and W a summary from a draw of either $\mathcal{O}_1$ or $\mathcal{O}_2$

Inference with summary statistics

For D_1 and D_2 compute summary statistics (Betti numbers)
$S_1 \equiv B_1, ..., B_m$: m summaries of the torus
$S_2 \equiv C_1, ..., C_n$: n summaries of the double torus
and W a summary from a draw of either $\mathcal{O}_1$ or $\mathcal{O}_2$

$$\hat{p}(\mathcal{O}_1 \mid w, S_1, S_2) = \frac{\tilde{f}(w;\hat{\theta}_1)\pi(\mathcal{O}_1)}{\tilde{f}(w;\hat{\theta}_1)\pi(\mathcal{O}_1) + \tilde{f}(w;\hat{\theta}_2)\pi(\mathcal{O}_2)}.$$

Statistical inference with topological summaries

Natural questions on topological summaries

1. Can we define probability measures on summaries ?

Statistical inference with topological summaries

Natural questions on topological summaries

1. Can we define probability measures on summaries ?
2. Can we define(Fréchet) means and variances ?

Statistical inference with topological summaries

Natural questions on topological summaries

1. Can we define probability measures on summaries ?
2. Can we define(Fréchet) means and variances ?
3. Relation between sampling distribution on data and distribution on the topological summary ?

Statistical inference with topological summaries

Natural questions on topological summaries

1. Can we define probability measures on summaries ?
2. Can we define(Fréchet) means and variances ?
3. Relation between sampling distribution on data and distribution on the topological summary ?
4. Computing (Fréchet) means and variances ?

Statistical inference with topological summaries

Natural questions on topological summaries

1. Can we define probability measures on summaries ?
2. Can we define(Fréchet) means and variances ?
3. Relation between sampling distribution on data and distribution on the topological summary ?
4. Computing (Fréchet) means and variances ?
5. Concentration of the (Fréchet) mean ?

Why summaries ?

1. Inference problem on the data may be hard.

Why summaries ?

1. Inference problem on the data may be hard.
2. Inference problem on data may not be robust.

Why summaries ?

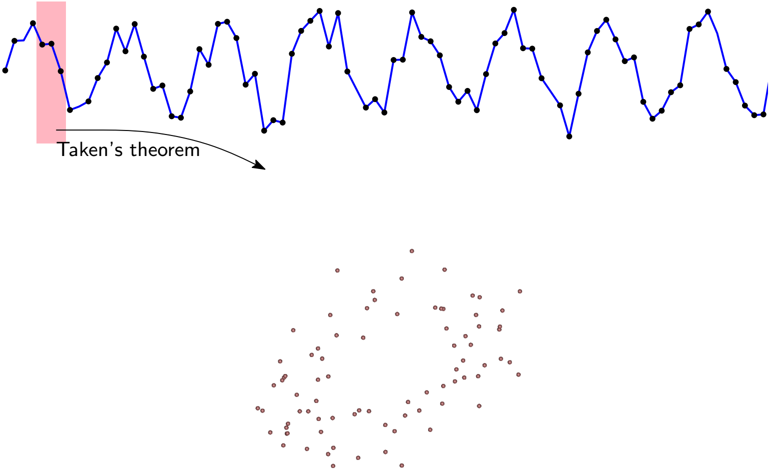
1. Inference problem on the data may be hard.
2. Inference problem on data may not be robust.
3. We are not interested in capturing the exact geometry, certain perturbations or invariants should not matter.

Persistence homology

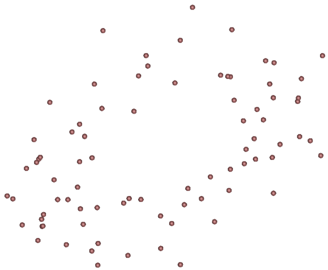
Studies multi-scale features ("holes") of spaces:

1. For a know space gives multi-scale representation of its features.
2. Given a point cloud sample, captures features at different resolution. Separates features from noise.

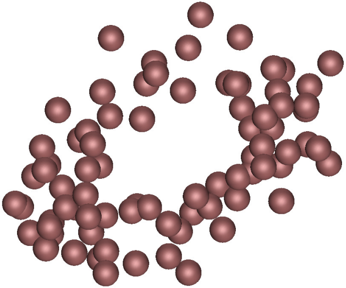
Point cloud data



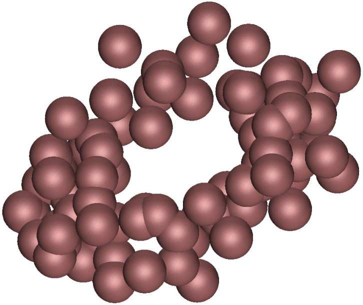
Filtration, $\mathbb{X}_0$



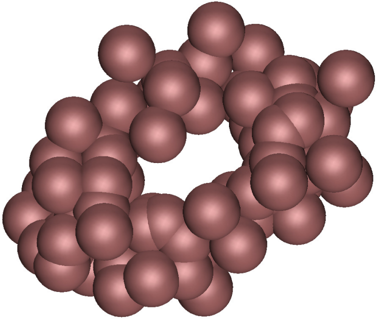
Filtration, $\mathbb{X}_1$



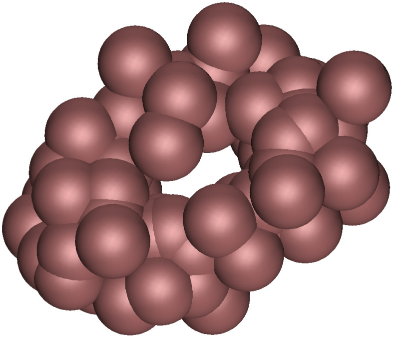
Filtration, $\mathbb{X}_2$



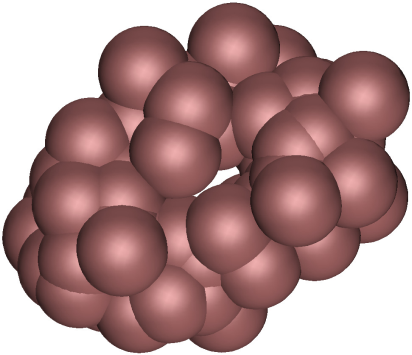
Filtration, $\mathbb{X}_3$



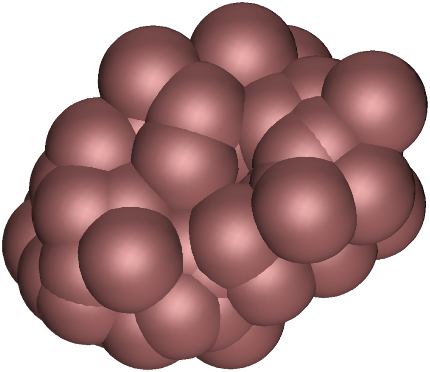
Filtration, $\mathbb{X}_4$



Filtration, $\mathbb{X}_5$

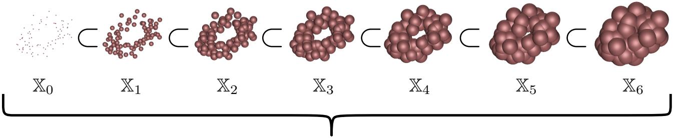


Filtration, $\mathbb{X}_6$



Persistence homology

Construct a filtration

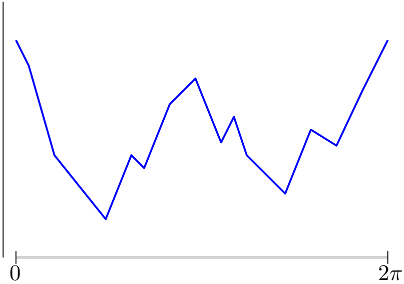


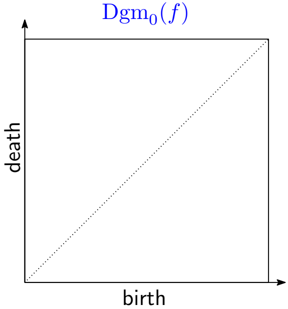
$$H_p(\mathbb{X}_0) \to H_p(\mathbb{X}_1) \to H_p(\mathbb{X}_2) \to H_p(\mathbb{X}_3) \to H_p(\mathbb{X}_4) \to H_p(\mathbb{X}_5) \to H_p(\mathbb{X}_6)$$

Images of linear maps $\phi_p^{i,j} : H_p(\mathbb{X}_i) \to H_p(\mathbb{X}_j)$ induced by inclusion.
Determine when a homology class is born and when it dies.

Persistence diagrams

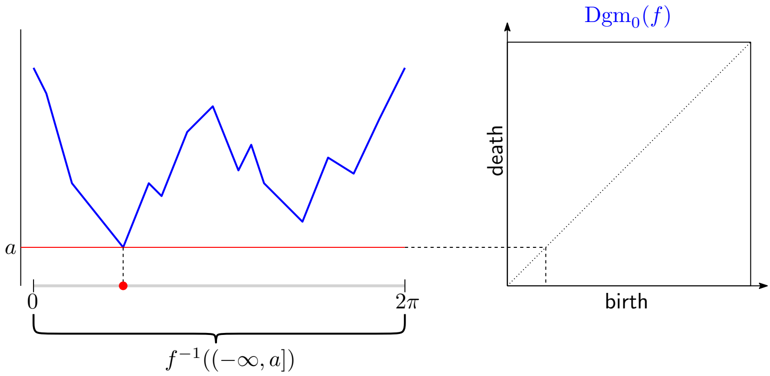
Evolution of homology as birth-death pair.





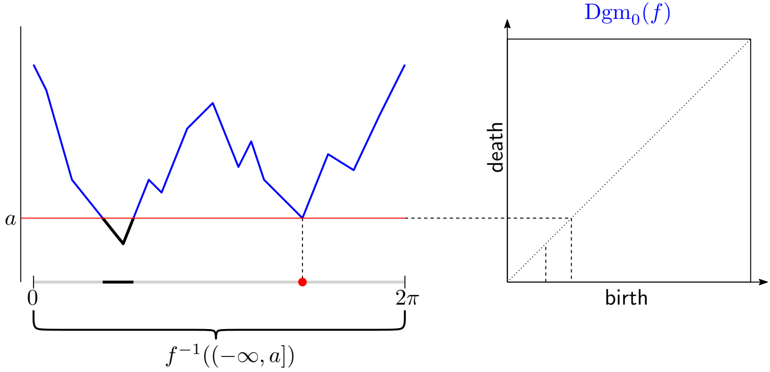
Persistence diagrams

Evolution of homology as birth-death pair.



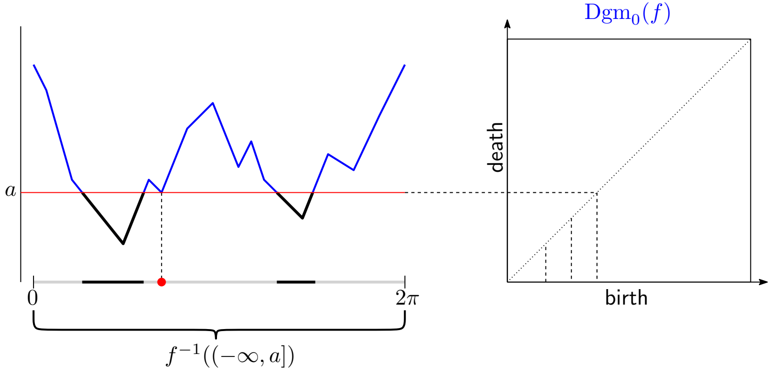
Persistence diagrams

Evolution of homology as birth-death pair.



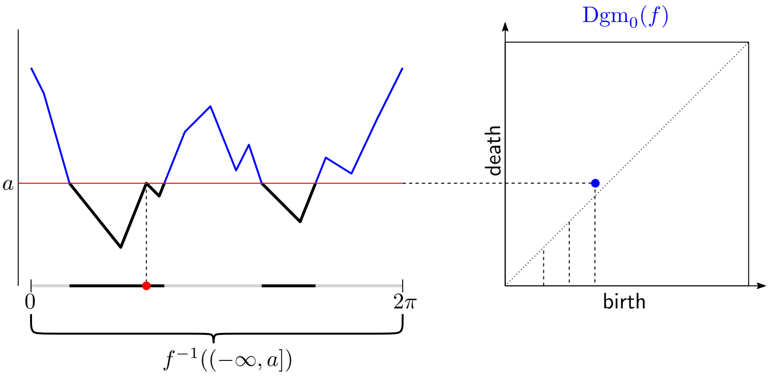
Persistence diagrams

Evolution of homology as birth-death pair.



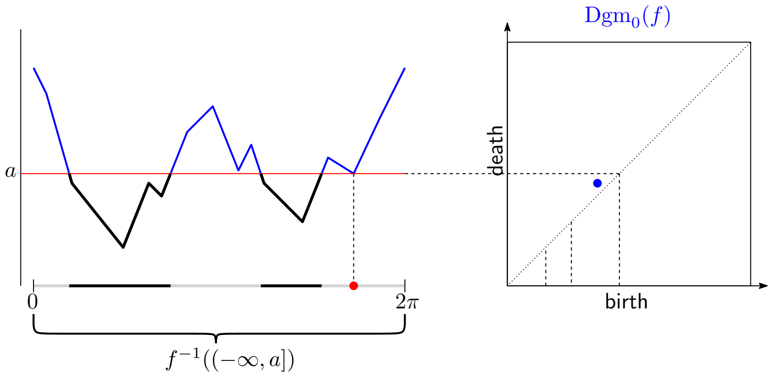
Persistence diagrams

Evolution of homology as birth-death pair.



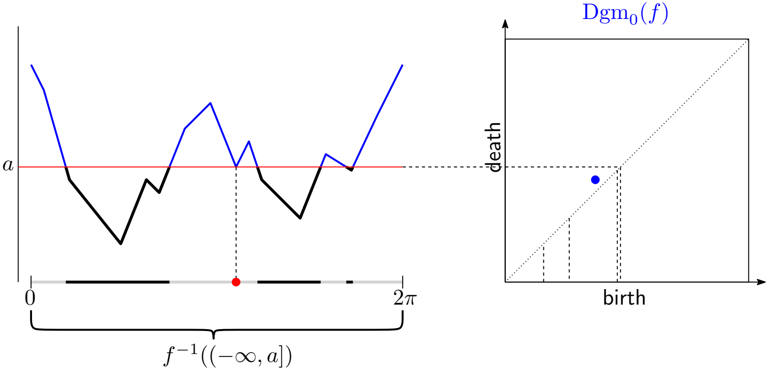
Persistence diagrams

Evolution of homology as birth-death pair.



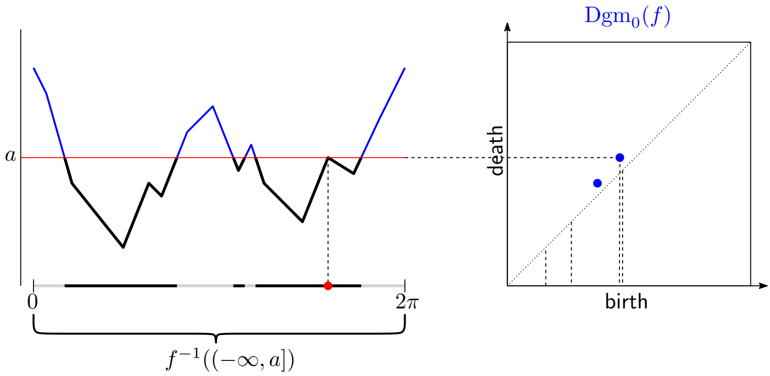
Persistence diagrams

Evolution of homology as birth-death pair.



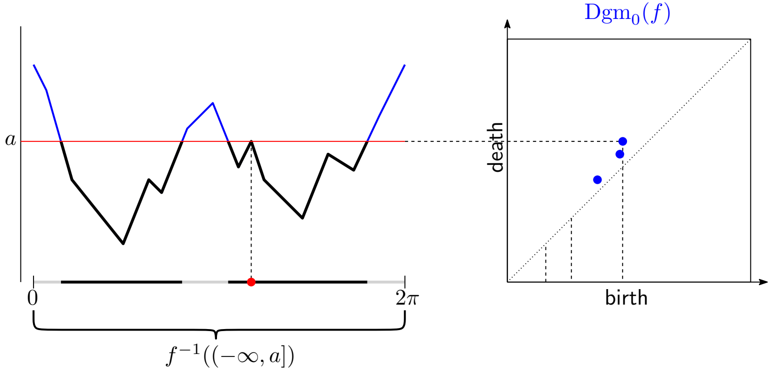
Persistence diagrams

Evolution of homology as birth-death pair.



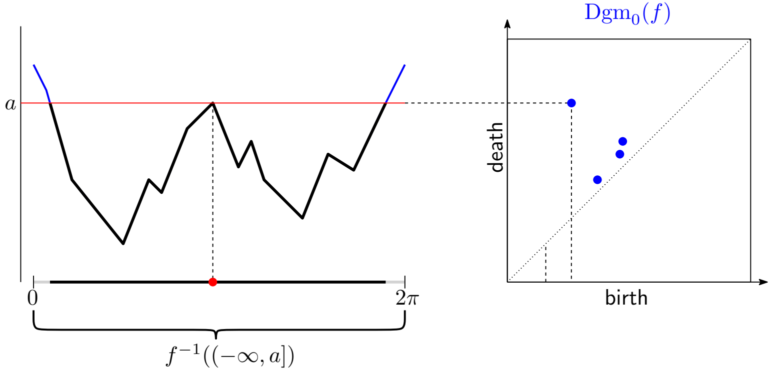
Persistence diagrams

Evolution of homology as birth-death pair.



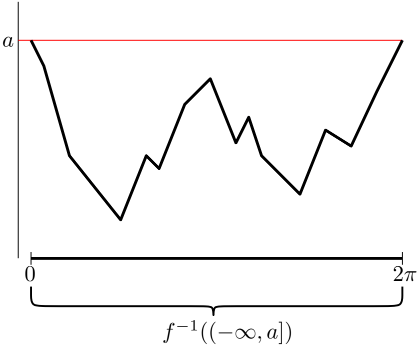
Persistence diagrams

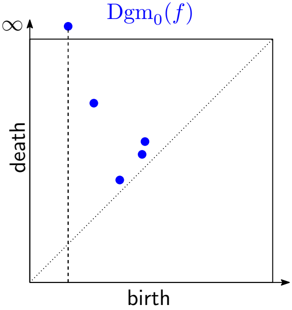
Evolution of homology as birth-death pair.



Persistence diagrams

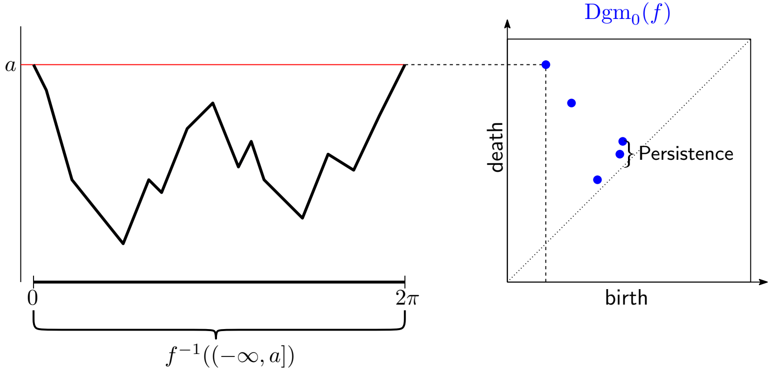
Evolution of homology as birth-death pair.



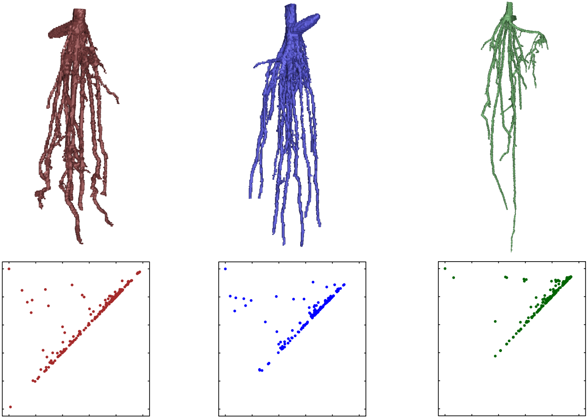


Persistence diagrams

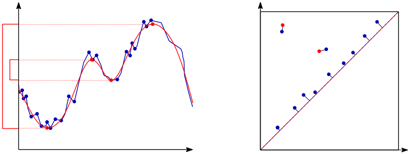
Evolution of homology as birth-death pair.



Roots



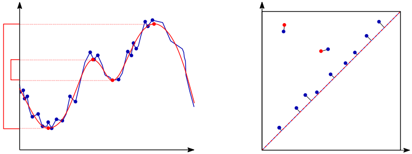
Metrics on diagrams



p-Wasserstein distance

$$d_{W_p}(X,Y)^2 = \left(\inf_{\phi:X\to Y} \sum_{x\in X} \|x-\phi(x)\|_\infty^p \right)^{1/p}$$

Metrics on diagrams



L^2-Wasserstein distance

$$d_{L^2}(X, Y)^2 = \inf_{\phi: X \to Y} \sum_{x \in X} \|x - \phi(x)\|^2$$

Picking a metric

1. We will use L^2-Wasserstein distance.

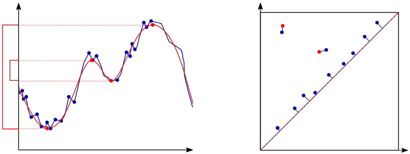
Picking a metric

1. We will use L^2-Wasserstein distance.
2. Some of our results hold for both metrics.

Picking a metric

1. We will use L^2-Wasserstein distance.
2. Some of our results hold for both metrics.
3. For practical applications use more recent robust metrics.

Stability



Theorem (Turner-Milyeko-M-Harer)

f, g are tame Lipschitz functions $f, g : \mathbb{X} \to \mathbb{R}$

$$d_{L^2}(Diag(f), Diag(g)) \leq 2^{\frac{k+2}{2}} \left[C \|f - g\|_\infty^{2-k} \right]^{1/2},$$

$k \in [1, 2)$ and C depends on Lipschitz properties of f, g.

Persistence diagram

Definition

A generalized persistence diagram is a countable multiset of points in $\mathbb{R}^2$ along with the diagonal $\Delta = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$, where each point on the diagonal has infinite multiplicity.

Space of persistence diagrams

Definition

The space of persistence diagrams is defined as

$$\mathcal{D} = \{x \mid d(x, \emptyset) < \infty\}.$$

where $\emptyset$ is a diagram with only a diagonal.

Complete and separable

Theorem (Milyeko-M-Harer)

The space $\mathcal{D}$ is complete and separable.

Fréchet expectation

Given a probability space $(\mathcal{D}, \mathcal{B}(\mathcal{D}), \mathcal{P})$ the quantity

$$\mathsf{Var}_{\mathcal{P}} = \inf_{Y \in \mathcal{D}} \left[F := \int_{\mathcal{D}} d(X, Y)^2 \mathrm{d}\mathcal{P}(X) < \infty \right],$$

is the Fréchet variance of $\mathcal{P}$ and the set at which the value is obtained

$$\mathbb{E}_{\mathcal{P}} = \{ Y \in \mathcal{D} \mid F(Y) = \mathsf{Var}_{\mathcal{P}} \},$$

is the Fréchet expectation.

Fréchet expectation

Given a probability space $(\mathcal{D}, \mathcal{B}(\mathcal{D}), \mathcal{P})$ the quantity

$$\text{Var}_{\mathcal{P}} = \inf_{Y \in \mathcal{D}} \left[F := \int_{\mathcal{D}} d(X, Y)^2 \mathrm{d}\mathcal{P}(X) < \infty \right],$$

is the Fréchet variance of $\mathcal{P}$ and the set at which the value is obtained

$$\mathbb{E}_{\mathcal{P}} = \{ Y \in \mathcal{D} \mid F(Y) = \text{Var}_{\mathcal{P}} \},$$

is the Fréchet expectation.
Does it exist ?

Fréchet expectation

Theorem (Milyeko-M-Harer)

Let $\mathcal{P}$ be a probability measure on $(\mathcal{D}, \mathcal{B}(\mathcal{D}))$ with finite second moment. If $\mathcal{P}$ has compact support then $\mathbb{E}_{\mathcal{P}} \neq \emptyset$.

Tail conditions

Definition (Tight)

A measure $\mathcal{P}$ on a measurable metric space $(\mathbb{X}, \mathcal{P})$ is tight if $\forall \varepsilon > 0$ there exists a compact set $\mathbb{S} \subset \mathbb{X}$ such that $\mathcal{P}(\mathbb{X} - \mathbb{S}) < \varepsilon$.

Tail conditions

Definition (Tight)

A measure $\mathcal{P}$ on a measurable metric space $(\mathbb{X}, \mathcal{P})$ is tight if $\forall \varepsilon > 0$ there exists a compact set $\mathbb{S} \subset \mathbb{X}$ such that $\mathcal{P}(\mathbb{X} - \mathbb{S}) < \varepsilon$.

Definition (Decay of measure)

A measure $\mathcal{P}$ on $(\mathbb{X}, \mathcal{P})$ has a rate of decay q if for $x_0 \in \mathbb{X}$ there exist $K, R > 0$ such that for $r \geq R$ we have $\mathcal{P}(B^r(x_0)) \leq Kr^{-q}$, with $B^r(x_0) = \{x \in \mathbb{X} \mid \mathcal{P}(x, x_0) \geq r\}$.

Fréchet expectation

Theorem (Milyeko-M-Harer)

Let $\mathcal{P}$ be a tight probability measure on $(\mathcal{D}, \mathcal{B}(\mathcal{D}))$ with rate of decay $q > 2$. Then $\mathbb{E}_{\mathcal{P}} \neq \emptyset$.

$\mathcal{D}$ as a metric space

Alexandrov space bounded from below: Given a geodesic space $\mathbb{X}$ with metric d' for any geodesic $\gamma : [0,1] \to \mathbb{X}$ from X to Y and any $Z \in \mathbb{X}$

$$d'(Z, \gamma(t))^2 \geq td'(Z, Y)^2 + (1-t)d'(Z, X)^2 - t(1-t)d'(X, Y)^2.$$

$\mathcal{D}$ as a metric space

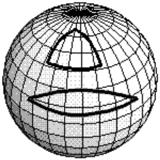
Alexandrov space bounded from below: Given a geodesic space $\mathbb{X}$ with metric d' for any geodesic $\gamma : [0, 1] \to \mathbb{X}$ from X to Y and any $Z \in \mathbb{X}$

$$d'(Z, \gamma(t))^2 \geq td'(Z, Y)^2 + (1 - t)d'(Z, X)^2 - t(1 - t)d'(X, Y)^2.$$

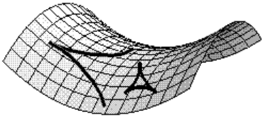
Theorem (Turner-Milyeko-M-Harer)

$(\mathcal{D}, d_{L^2})$ is a geodesic space and is a non-negatively curved Alexandrov space.

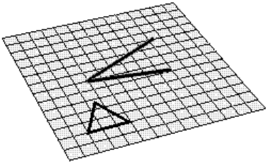
Comparison triangles



Universe with *positive* curvature. Diverging line converge at great distances. Triangle angles add to more than 180°.



Universe with *negative* curvature. Lines diverge at ever increasing angles. Triangle angles add to less than 180°.



Universe with no curvature. Lines diverge at constant angle. Triangle angles add to 180°.

Barycenter of diagrams

Given diagrams $\{X_i\}_{i=1}^{n}$ a Fréchet mean Y satisfies

$$\min_{Y \in \mathcal{D}} \left[F_n := \int_{\mathcal{D}} d(X, Y)^2 \mathrm{d}\mathcal{P}_n(X) \right],$$

where $\mathcal{P}_n := n^{-1} \sum_{i=1}^{n} \delta_{X_i}$.

Gradient decent algorithm

Input: diagrams $\{X_1, \ldots, X_n\}$
Output: Fréchet mean $\{Y\}$

1. $i \sim \text{Uniform}(1, ..., n)$;
2. Initialize $Y \leftarrow X_i$;
3. stop $=$ false;
4. Repeat until stop $=$ false
    1. $K = |Y|$;
    2. for $i = 1, \ldots n$, $\quad (y^j, x_i^j) \leftarrow \text{Hungarian}(Y, X_i)$
    3. for $j = 1, \ldots K$, $\quad y^j \leftarrow \text{mean}_{i=1,..,n}(x_i^j)$
    4. If $\text{Hungarian}(Y, X_i) = (y_j, x_i^j)$ then stop $=$ true

Law of large numbers

Fréchet function and empirical Fréchet function:

$$\begin{aligned} \mathbf{Y} &= \left\{ \min_{Z\in\mathcal{D}} \left[F := \int_{\mathcal{D}} d(X,Z)^2 \mathrm{d}\mathcal{P}(X) \right] \right\}, \\ \mathbf{Y}_n &= \left\{ \min_{Z\in\mathcal{D}} \left[F_n := \int_{\mathcal{D}} d(X,Z)^2 \mathrm{d}\mathcal{P}_n(X) \right] \right\}. \end{aligned}$$

What can we say about concentration of a Y_n computed by our algorithm to a Y ?

Law of large numbers

Fréchet function and empirical Fréchet function:

$$\begin{aligned} \mathbf{Y} &= \left\{ \min_{Z \in \mathcal{D}} \left[F := \int_{\mathcal{D}} d(X,Z)^2 \mathrm{d}\mathcal{P}(X) \right] \right\}, \\ \mathbf{Y}_n &= \left\{ \min_{Z \in \mathcal{D}} \left[F_n := \int_{\mathcal{D}} d(X,Z)^2 \mathrm{d}\mathcal{P}_n(X) \right] \right\}. \end{aligned}$$

What can we say about concentration of a Y_n computed by our algorithm to a Y ?
If $\mathcal{P} = \frac{1}{m} \sum_{i=1}^{m} \delta_{Z_i}$.

Concentration

Theorem (Turner-Milyeko-M-Harer)

Let F be the Fréchet function corresponding to $\mathcal{P}$ and Y be a local minimum of F. Let F_n be the Fréchet function corresponding to empirical measure $\mathcal{P}_n$.

Concentration

Theorem (Turner-Milyeko-M-Harer)

Let F be the Fréchet function corresponding to $\mathcal{P}$ and Y be a local minimum of F. Let F_n be the Fréchet function corresponding to empirical measure $\mathcal{P}_n$.
There exists a local minimum Y_n of F_n such that with probability greater than $1-\delta$

$$d(Y, Y_n)^2 \leq \frac{m^2 F(Y)}{n} \ln\left(\frac{m}{\delta}\right),$$

for $n \geq 8m \ln \frac{m}{\delta}$ and $\frac{m^2 F(Y)}{n} \ln\left(\frac{m}{\delta}\right) < r^2$ where r characterizes the separation between local minima of F.

Local minima

Theorem (Turner-Milyeko-M-Harer)

The number of local minima of F_n is bounded by

$$\prod_{i=1}^{m}(k_i+1)^{k_1+k_2+\ldots+k_m}.$$

Local minima

Theorem (Turner-Milyeko-M-Harer)

The number of local minima of F_n is bounded by

$$\prod_{i=1}^{m}(k_i+1)^{k_1+k_2+\ldots+k_m}.$$

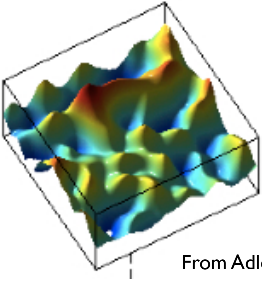
We strongly suspect under some general position conditions there is one local minima.

F has a unique minima

Theorem (Turner-Milyeko-M-Harer)

If F has a unique minima Y then with probability 1 the Hausdorff distance between Y and $\mathbf{Y}_n$ goes to 0 as n goes to infinity.

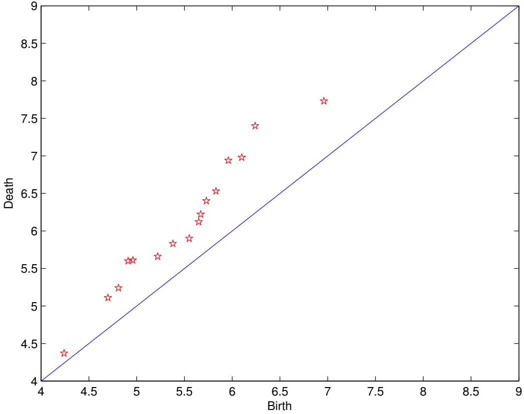
Gaussian random fields



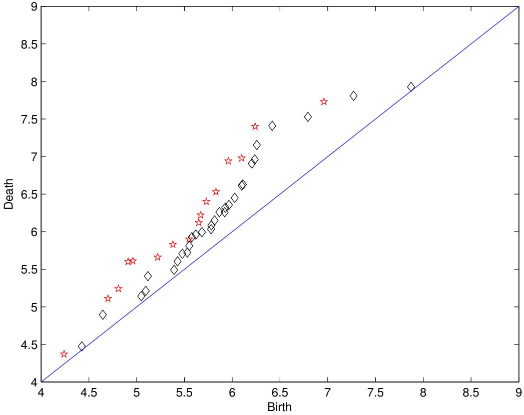
From Adler et al

Unit square with $c(p) = \exp(-\alpha\|p\|^2)$, $\alpha = 100$

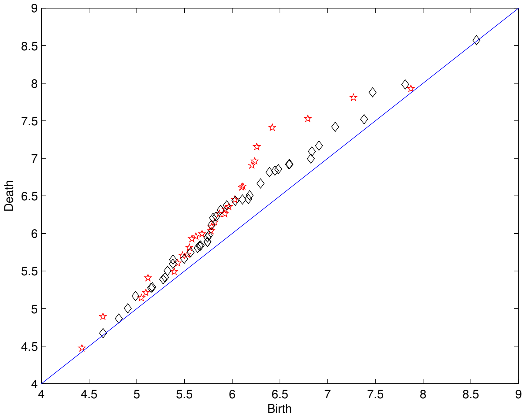
Diagram of a random field



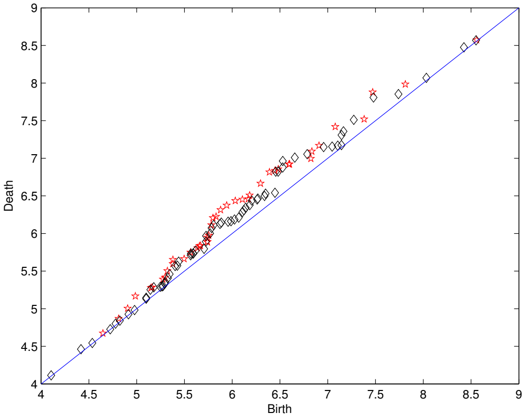
1 vs. 5



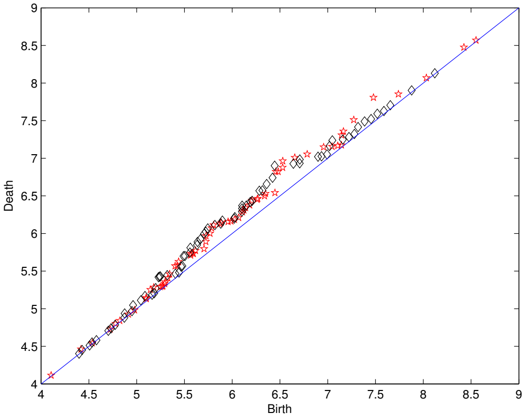
5 vs. 10



10 vs. 20



20 vs. 40



Likelihood of data

Assume our point cloud data $Z \equiv \{X_1, ..., X_n\}$ are n points drawn iid from F_θ

$$\text{Lik}(Z;\theta) \equiv f_\theta(X).$$

Likelihood of data

Assume our point cloud data $Z \equiv \{X_1, ..., X_n\}$ are n points drawn iid from F_θ

$$\text{Lik}(Z;\theta) \equiv f_\theta(X).$$

Z is an element of the product space $(\mathbb{X}^n, \Sigma^n, \mathcal{P}_\theta^n)$.

Induced measure $\mathcal{P}$

A persistence diagram is a map $g : \mathbb{X}^n \to \mathcal{D}$.

Induced measure $\mathcal{P}$

A persistence diagram is a map $g : \mathbb{X}^n \to \mathcal{D}$.

The map g (under weak conditions) is measurable for every $A \in \mathcal{B}(\mathcal{D})$

$$g^{-1}(A) = \{\omega : g(\omega) \in A\} \in \Sigma^n.$$

Induced measure $\mathcal{P}$

A persistence diagram is a map $g : \mathbb{X}^n \to \mathcal{D}$.

The map g (under weak conditions) is measurable for every $A \in \mathcal{B}(\mathcal{D})$

$$g^{-1}(A) = \{\omega : g(\omega) \in A\} \in \Sigma^n.$$

The induced measure $\mathcal{P}$ on $(\mathcal{D}, \mathcal{B}(\mathcal{D}))$ is

$$\mathcal{P}(A) = \mathcal{P}_\theta^n(g^{-1}(A)), \quad \text{for } A \in \mathcal{B}(\mathcal{D}).$$

Joints, conditionals, and sufficiency

We can define joint and conditional measures

$$\mathcal{P}(\mathcal{D}, \theta), \quad \mathcal{P}(\theta \mid \mathcal{D}).$$

Joints, conditionals, and sufficiency

We can define joint and conditional measures

$$\mathcal{P}(\mathcal{D}, \theta), \quad \mathcal{P}(\theta \mid \mathcal{D}).$$

Think of $\mathcal{P}(\mathcal{D})$ as a substitution likelihood in the sense of Jeffreys.

Joints, conditionals, and sufficiency

We can define joint and conditional measures

$$\mathcal{P}(\mathcal{D}, \theta), \quad \mathcal{P}(\theta \mid \mathcal{D}).$$

Think of $\mathcal{P}(\mathcal{D})$ as a substitution likelihood in the sense of Jeffreys. For what invariances of $\mathbb{X}^n$ is the persistence diagram a sufficient statistic ?

Kernel machine learning

Positive (semi)-definite kernel function $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$.
Given data $x_1, \dots x_n \in \mathbb{X}$

$$\hat{f}(x) = \sum_i c_i k(x, x_i), \quad x \in \mathbb{X}.$$

Kernel machine learning

Positive (semi)-definite kernel function $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$.
Given data $x_1, ...x_n \in \mathbb{X}$

$$\hat{f}(x) = \sum_i c_i k(x, x_i), \quad x \in \mathbb{X}.$$

Examples: $\exp(-\kappa\|u - v\|^2)$, $\langle u, v\rangle$, $(\langle u, v\rangle + 1)^\tau$.

Kernel machine learning

Positive (semi)-definite kernel function $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$.
Given data $x_1, ...x_n \in \mathbb{X}$

$$\hat{f}(x) = \sum_i c_i k(x, x_i), \quad x \in \mathbb{X}.$$

Examples: $\exp(-\kappa\|u - v\|^2)$, $\langle u, v\rangle$, $(\langle u, v\rangle + 1)^\tau$.

Set $\mathbb{X} = \mathcal{D}$, define for $u, v \in \mathcal{D}$

$$k(u, v) = h(d_{\text{metric}}(u, v)), \quad \text{where } h \text{ is monotonically decreasing.}$$

Given diagrams $d_1, ...d_n \in \mathbb{X}$

$$\hat{f}(d) = \sum_i c_i k(d, d_i), \quad d \in \mathcal{D}.$$

Open problems

- Modeling in persistence space.

Open problems

- Modeling in persistence space.
- Multivariate analysis (covariances).

Open problems

- Modeling in persistence space.
- Multivariate analysis (covariances).
- Function approximation in persistence space.

Open problems

- Modeling in persistence space.
- Multivariate analysis (covariances).
- Function approximation in persistence space.
- Limit theorems in persistence space.

Open problems

- Modeling in persistence space.
- Multivariate analysis (covariances).
- Function approximation in persistence space.
- Limit theorems in persistence space.
- Understanding persistence space as a summary statistics.

Open problems

- Modeling in persistence space.
- Multivariate analysis (covariances).
- Function approximation in persistence space.
- Limit theorems in persistence space.
- Understanding persistence space as a summary statistics.
- Extensions to other summaries such as Reeb graphs.

Open problems

- Modeling in persistence space.
- Multivariate analysis (covariances).
- Function approximation in persistence space.
- Limit theorems in persistence space.
- Understanding persistence space as a summary statistics.
- Extensions to other summaries such as Reeb graphs.
- Can we use these ideas to generalize shape space models.

Acknowledgements

Shmuel Weinberger and Eliran Subag

Acknowledgements

Shmuel Weinberger and Eliran Subag

Funding:

- Center for Systems Biology at Duke
- NSF DMS and CCF
- AFOSR
- NIH