Outline: Three (Related) Topics

Part I: Estimating Manifolds
Collaborators: Chris Genovese, Marco Perone-Pacifico, Isa Verdinelli

Part II: Estimating Ridges
Collaborators: Yen-Chi Chen, Chris Genovese, Marco Perone-Pacifico, Isa Verdinelli

Part III: Inference For Persistent Homology
Collaborators: Sivaraman Balakrishnan, Brittany Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh
Motivating Example: the Cosmic Web

Gnedin (2005)
Our method on a data slice (Sloan Digital Sky Survey)

(thanks to Yen-Chi Chen)
Part I: Summary: Manifold Estimation

\[
\inf_{\hat{M}} \sup_Q \mathbb{E}_Q \left[ \text{dist}(\hat{M}, M) \right] \geq \frac{1}{\log n} = \text{hopeless?}
\]
Part II: Summary: Ridge Estimation

Ridge $R$ of density approximates (geometrically and topologically) manifold $M$. $R$ can be estimated well.
Part III: Summary: Persistent Homology

Barcodes

Birth
Death

\[ H_0, H_1, H_2 \]

\[ \varepsilon \]

0.0 0.5 1.0 1.5 2.0 2.5 3.0

0 7 / 76
Part III: Summary: Persistent Homology

Our contribution:

\[ \begin{array}{ccc}
\text{Birth} & | & \text{Death} \\
\text{noise} & | & \text{signal} \\
\hline
\end{array} \]
PART I
ESTIMATING MANIFOLDS
Manifolds and Manifold Complexes

Manifolds give a useful representation of low dimensional structure.

A **manifold** is a space that looks locally like a Euclidean space of some dimension (called the dimension of the manifold).
Examples: point (0-dim), filaments (1-dim), surface of the sphere or torus (2-dim), three-dimensional sphere, space-time (4-dim).

Want to allow for intersections and other complexities.
Consider a *union* of manifolds embedded in $\mathbb{R}^D$ with maximal dimensions $d < D$.
We’ll call this a $d$-dimensional **manifold complex** (stratified space).
**Challenge**: Given a point cloud sampled from a manifold complex and then perturbed by noise, **accurately estimate the manifold complex**.
Challenge: Given a point cloud sampled from a manifold complex and then perturbed by noise, **accurately estimate the manifold complex.**
A Synthetic Example

An smooth manifold with $d = 2, D = 3$
A Synthetic Example

An smooth manifold with $d = 2, D = 3$ plus 3-dimensional additive noise
A Synthetic Example

The data drawn from the additive model
Models for Manifold Estimation

Suppose $M$ belongs to a class $\mathcal{M}$ (to be defined shortly) of $d$-dimensional “smooth” manifolds embedded in $\mathbb{R}^D$ for $D > d$.

$G$ is a distribution on $M$, with density bounded away from 0 and $\infty$.

Draw $X_1, \ldots, X_n$ from $G$ and then draw $Y_1, \ldots, Y_n$ according to one of four noise models:

1. **noiseless**: $Y_i = X_i$.
2. **clutter**: $Y_i = X_i$ with probability $p$, otherwise $Y_i \sim \text{Uniform}$.
3. **perpendicular**: $Y_i = X_i + Z_i$ where $Z_i$ is normal to $M$.
   (See also Niyogi, Smale, Weinberger 2008.)
4. **additive**: $Y_i = X_i + Z_i$ and $Z_i \sim \Phi$ (e.g., spherical Normal).

Want to estimate $M$ from $Y_1, \ldots, Y_n$.

The noise model strongly affects the difficulty of this problem.
Existing Literature

**Computational geometry** (e.g., Cheng et al. 2005, Dey 2006)  
Here, “noise” does *not* have the statistical meaning of points drawn randomly from a distribution; instead, data must be close to $M$ but not too close to each other. (There are a few notable exceptions.)

**Manifold learning** (e.g., Isomap, LLE)  
Often, the primary focus has been on dimension reduction.

**Computational Topology** (e.g., Niyogi, Smale, and Weinberger 2009)  
Focus on topological rather than geometric information

Filaments, principle curves, support estimation, ...  
e.g., Hastie and Stuetzle (1989); Tibshirani (1992); Arias-Castro et al. (2006);  
Cheng, Hall, and Hartigan (2004); Hall et al. (1992); Wegman and Luo (2002)
Minimax Manifold Estimation

Assume $M \in \mathcal{M}$, where

\[ \mathcal{M} = \{ M : \text{reach}(M) \geq \kappa \} . \]

Model

\[ Q = \bigcup_{M \in \mathcal{M}} \{ Q_{M,G} : G \in \mathcal{G}(M) \} \]

where

\[ Q_{M,G}(A) = \int_{M} \Phi(Y \in A \mid X = x) \, dG(x) \]

Draw a random sample $Y_1, Y_2, \ldots, Y_n$ from $Q_M$.

**Goal**: find the minimax risk:

\[ R_n = \inf_{\hat{M}_n} \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q \text{Haus}(\hat{M}_n, M). \]
The Reach of a Manifold

Define the reach of a manifold $M$ as follows:

reach$(M)$ is the largest (sup) $r$ such that $d(x, M) \leq r$ implies that $x$ has a unique projection onto $M$.

This is also called the thickness or condition number of the manifold; see Niyoki, Smale, and Weinberger (2009).

Intuitively, a manifold $M$ with reach$(M) = \kappa$ has two constraints:

1. **Curvature.** A ball or radius $r \leq \kappa$ can roll freely and smoothly over $M$, but a ball or radius $r > \kappa$ cannot.

2. **Separation.** $M$ is at least $2\kappa$ from self-intersecting.
Reach in One Dimension

circles have radius $r$

$\kappa > r$

$\kappa > 2r$

$\kappa < r$

$\kappa < 2r$
Reach

Normals of size $< \text{reach}(\mathcal{M})$ do not cross.
Hausdorff Distance

Given two subsets of $\mathbb{R}^D$, $A$ and $B$:

$$\text{Haus}(A, B) = \inf \{ \epsilon : A \subset B \oplus \epsilon \text{ and } B \subset A \oplus \epsilon \}$$

where $A \oplus \epsilon = \bigcup_{x \in A} B(x, \epsilon)$ and $B(x, \epsilon) = \{ y : \|x - y\| \leq \epsilon \}$.

Example:

$$\text{Haus}(A, B) = \max \{ 2.5, 1.5 \} = 2.5$$
## Minimax Rates

<table>
<thead>
<tr>
<th>Noise Model</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clutter/Noiseless</td>
<td>$n^{-\frac{2}{d}}$</td>
</tr>
<tr>
<td>Perpendicular Compact</td>
<td>$n^{-\frac{2}{2+d}}$</td>
</tr>
<tr>
<td>Additive sub-Gaussian c.f.</td>
<td>$(\log n)^{-1}$</td>
</tr>
<tr>
<td>Additive Compact/Polynomial c.f.</td>
<td>??</td>
</tr>
</tbody>
</table>

Ambient $D$ and reach $\kappa$ are in the constants.
Proof Sketch

The lower bound is established with Le Cam’s Lemma.
Then for any pair $Q_0, Q_1 \in \mathcal{Q}$

$$ \text{Risk} \geq d(M_0, M_1) \left[ 1 - \text{TV}(Q_0, Q_1) \right]^{2n}, $$

where

$$ \text{TV}(Q_0(A), Q_1(A)) = \sup_A |Q_0(A) - Q_1(A)|. $$

We want to choose a least favorable pair of manifolds whose distributions are as hard to distinguish as possible.
Perpendicular Noise: Sketch of Lower Bound

Construct $M_0$ and $M_1$ such that:

- $M_i \in \mathcal{M}_\kappa$
- $\text{Haus}(M_1, M_0) = \gamma$
- $\text{TV} \equiv \int |q_1 - q_0| = O(\gamma^{(d+2)/2})$, which is minimum possible.

Apply Le Cam’s Lemma: For any $\hat{\mathcal{M}}$:

$$\sup_{Q \in \mathcal{Q}} \mathbb{E} Q^n \text{Haus}(M, \hat{\mathcal{M}}) \geq \text{Haus}(M_1, M_0) \times (1 - \text{TV})^{2n}$$

$$= \gamma (1 - c \gamma^{(d+2)/2})^{2n}.$$ 

Setting $\gamma = n^{-2/(d+2)}$ yields the result.

Least Favorable Pair $M_0$ and $M_1$: $M_0 = \text{plane}$ and $M_1 = \text{“flying saucer”}$. 
Constructing $M_1$

Start with $M_0 \ldots$
Constructing $M_1$

Push up $\kappa$-ball,
Constructing $M_1$

Push up $\kappa$-ball, through the plane to height $\gamma$. But reach still 0 ...
Constructing $M_1$

But reach still 0, so smooth the corners.
Constructing $M_1$

Smooth the corners . . .
Constructing $M_1$

Flying Saucer $M_1$
Perpendicular Noise: Sketch of Upper Bound

Construct an “estimator” that achieves the bound:

1. Split the data into two halves.

2. Using the first half, construct a pilot estimator. This is a (sieve) maximum likelihood estimator.

3. Cover the pilot estimator with thin, long, slabs.

4. Using the second half of the data, fit local linear estimators $\hat{M}_j$ in slab $j$.

5. $\hat{M} = \bigcup_j \hat{M}_j$.

The details are messy and the estimator is not practical, but it suffices for establishing the bound.
Clutter Model

Suppose

\[ Y_1, \ldots, Y_n \sim Q \equiv (1 - p)U + pG \]

where \( 0 < p \leq 1 \), \( U \) is uniform on the compact set \( K \subset \mathbb{R}^D \), and \( G \) supported on \( M \) as before.

Then,

\[
\inf_{\hat{M}} \sup_{Q \in Q} \mathbb{E}_{Q^n} \operatorname{Haus}(\hat{M}, M) \asymp^* C \left( \frac{1}{np} \right)^{\frac{2}{d}}.
\]

(The \( \asymp^* \) means I am hiding log factors.)

Lower bound uses the same least favorable pair.
Clutter Model: Upper Bound

Let

- $\epsilon_n = (\log n / n)^{2/d}$.
- $\hat{Q}_n$ be the empirical measure.
- $S_M(y)$ denotes a $\epsilon^{d/2} \times \epsilon^{D-d}$ slab:

$$\begin{align*}
&\text{y} \\
&b_1 \epsilon_n \\
&b_2 \epsilon_n
\end{align*}$$

Define

$$s(M) = \inf_{y \in M} \hat{Q}_n[S_M(y)] \quad \text{and} \quad \hat{M}_n = \arg\max_M s(M).$$

This works (i.e. achieves the lower bound). But it is not computable. Finding a computable estimator that achieves the bound is an open question.
Additive Error

\[ X_1, X_2, \ldots, X_n \sim G \text{ where } \text{support}(G) = M, \text{ and} \]

\[ Y_i = X_i + Z_i, \quad i = 1, \ldots, n, \]

where \( Z_i \sim \Phi = \text{Gaussian} \).

This is analogous to an errors-in-variables problem, except:

1. We want to estimate the support of \( G \) not \( G \) itself.
2. \( G \) is singular.
3. The underlying object is a manifold not a function.
Additive Error

For technical reasons, we allow the manifolds to be noncompact. Define a truncated loss function,

\[ L(M, \hat{M}) = H(M \cap \mathcal{K}, \hat{M} \cap \mathcal{K}). \]

Then,

\[ \inf_{\hat{M}} \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[L(M, \hat{M})] \geq \frac{C}{\log n}. \]

Rate is similar to deconvolution but the proof is somewhat different (since \( Q_0 \) and \( Q_1 \) have different supports). Least favorable pair:
Additive Error: Upper Bound

Let $\hat{g}$ be a deconvolution density estimator (though $G$ has no density), and let $\hat{M} = \{\hat{g} > \lambda\}$. Fix any $0 < \delta < 1/2$.

$$\inf_{\hat{M}} \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[L(\hat{M}, \hat{M})] \leq C \left( \frac{1}{\log n} \right)^{\frac{1-\delta}{2}}.$$ 

In some special cases, we can achieve $\frac{1}{\log n}$ but, in general, not.
Summary So Far

In the most realistic case (additive Gaussian error) the rate is logarithmic.

Does this mean it is hopeless?
PART II
RIDGES
Surrogates

Problems with rates like $1/\log n$ seem to offer little practical hope for good performance. But it is sometimes possible to define a surrogate for the true object. We focus on estimating the surrogate accurately instead of the object itself. A surrogate will be \textit{good} to the degree that it
\begin{itemize}
  \item captures essential features of the true object, and
  \item can be estimated with a good rate of convergence.
\end{itemize}
Example: Uniform confidence bands (Genovese and Wasserman 2008).
Surrogates for Manifold Estimation

Strategy: Define a surrogate $R$, called the ridge, for the manifold complex $M$.

$R$ is, roughly speaking, a smoother, slightly biased version of $M$. Under weak conditions, we can estimate the ridge accurately, and the ridge is a good surrogate in the sense that

– It is close to the manifold (complex), *relative to the noise level*.
– It has essentially the same topology as the manifold (complex).
What’s a Ridge?

A 0-dimensional ridge is a mode. \( \nabla p(x) = 0 \) and \( p''(x) < 0 \).

A mode can also be thought of as the destination of a gradient ascent path, \( \pi \):

\[
\pi'(t) = \nabla p(\pi(t)).
\]

The modes of \( p \) can be found by the mean-shift algorithm.
A Ridge
Ridge versus Underlying Manifold
Ridge $R(p)$

$Y_1, \ldots, Y_n \sim P = (1 - \pi)U + \pi \cdot (G \ast \Phi_\sigma)$.

Let

- Density $p$.
- Gradient $g(x)$, Hessian $H(x) = U(x)\Lambda(x)U^T(x)$.
- $\lambda_1(x) \geq \lambda_2(x) \geq \cdots \geq \lambda_d(x) \geq \lambda_{d+1}(x) \geq \cdots \geq \lambda_D(x)$.
- $U(x) = [W(x) : V(x)]$.
- $L(x) = V(x)V^T(x)$ projector onto (local) tangent space.
- Projected gradient $G(x) = L(x)g(x)$.

$$x \in R(p) \text{ iff } \lambda_{d+1}(x) < 0 \text{ and } G(x) = 0.$$ 

Note: $\text{dimension}(R) = d$
Ridges (cont’d)

Ridges can also be thought of as the destinations of projected gradient ascent paths:

\[ \pi'(t) = G(\pi(t)) \]

where \( G(x) \) is the projected gradient.

\( x \) is on the ridge if

\[ \lim_{t \to \infty} \pi(t) = x. \]

This definition corresponds exactly to the ridge finding algorithm of Ozertem and Erdogmus (2011). (Mean shift with projected gradient replacing the gradient.)
Several definitions of ridge have been proposed in the literature. There is not one *right* definition. (And there is controversy about what a good definition is.)

But ours satisfies four useful properties

1. $\hat{p}$ close to $p$ implies $R(\hat{p})$ close to $R(p)$
2. $\sigma$ small implies $R$ close to $M$.
3. $R$ is topologically similar to $M$.
4. It corresponds to an algorithm: the projected gradient ascent path (Ozertem and Erdogmus).
Modified Mean-Shift Methods

Genovese, Perone-Pacifico, Verdinelli and Wasserman (2009) use the mean-shift trajectories to trace out ridges of the density and find filaments.

Ozertem and Erdogmus (2011) take this further, projecting each mean-shift point onto the space spanned by the smallest (most-negative) $D - d$ eigenvectors of Hessian($\hat{q}$).

The latter is called the subspace-constrained mean-shift algorithm (SCMS).

Our ridge estimator uses SCMS.
A Ridge Estimator (SuRF)

Subspace Ridge Finding (SuRF)

Steps:

1. **Estimation**: estimate the density, its gradient, and its Hessian.
2. **Denoising**: remove background clutter and low-probability regions, restricting attention to a set where the density is not too small;
3. **Mean-Shift**: apply the SCMS algorithm within the restriction set.

The denoising step is essential for good performance.
Assumptions

Given a suitably smooth density (or log density) $q$ on a compact set, assume that:

A0. $g(x), H(x),$ and $H'(x)$ exist for all $x$

A1. For $x \in \mathbb{R}$, $\lambda_{d+1} < -\beta$ and $\lambda_d(x) - \lambda_{d+1}(x) > \beta$.

A2. $\|H'(x)\|_{\max} \leq \beta^2$.

From A0-A2 can show that $R(q)$ exists and is stable: if $\tilde{q}$ is close to $q$ the the ridge of $\tilde{q}$ is close to the ridge of $q$.

\[
\begin{array}{ccc}
\lambda_{d+1} & -\beta & 0 \\
\end{array}
\]
Performance of the Estimator

1. \( \hat{R} \) is a consistent estimator of \( R \) and:
   \[
   \text{Haus}(R, \hat{R}) = O_P \left( n^{-\frac{2}{8+D}} \right) \equiv O_P(\psi_n)
   \]

2. For fixed bandwidth \( h \) (which still captures the shape),
   \[
   \text{Haus}(R, \hat{R}) = O_P \left( \sqrt{\log n/n} \right).
   \]

3. Small dilations of \( \hat{R} \) are topologically equivalent to (negligible dilations of) \( R \).
   \[
   \hat{R} \oplus O(\psi_n) \text{ homotopic to } R \oplus \eta \quad \text{for all } \eta > 0.
   \]

4. \( H(R, M) = O(\sigma^2) \).

5. \( R \oplus \sigma^2 \) homotopic to \( M \).
The Ridge of a “Hidden” Manifold

Suppose \( Q = (1 - p)U + p \cdot (G \ast \Phi_\sigma) \) where \( G \) is supported by a manifold \( M \) with reach \( \kappa \). Let \( R_\sigma \) be the ridge of \( q \).

**Surrogate Theorem.** Let \( K_\sigma^2 = \sigma^2 \log(1/\sigma) \). For all small \( \sigma > 0 \),

1. \( R_\sigma \) satisfies A1 and A2 with \( \beta \equiv \beta_\sigma = c\sigma^{-(D-d+2)} \)
2. \( \text{Haus}(M, R_\sigma) = O(K_\sigma^2) \)
3. \( R_\sigma \oplus O(K_\sigma^2) \) is homotopic to \( M \oplus \eta \) for every \( \eta > 0 \).
Example 1
Example 1 “Phase transition” in bandwidth
Example 2
Example 2
Example 2
Example 2 (cont’d)

But we need to denoise first or else ...
Bootstrap

(with Yen-Chi Chen)
Bootstrap resampling from the data, generating perturbed ridge sets.

Assess stability and variation for the estimated ridges. (This is very heuristic so far.)

\[
\mathbb{P}(R \subset \hat{R} \oplus t_\alpha) = 1 - \alpha
\]

approximate \( t_\alpha \) from:

\[
\mathbb{P}(\hat{R} \subset \hat{R}^* \oplus T_\alpha|X_1, \ldots, X_n) \approx \frac{1}{B} \sum_j I(\hat{R} \subset \hat{R}_j^* \oplus T_\alpha|X_1, \ldots, X_n) = 1 - \alpha.
\]
Dimensional Peeling

Want to distinguish manifold pieces of different dimensions in the same data set.

Peel off components (and associated data) as we move from low to high dimensions.

Run SuRF for \( d = 0, 1, 2, \ldots \).
Criterion: when next higher dimensional structure does not explain sufficient variation on associated data.
In Progress: Intersections

(with Yen-Chi Chen, Fabrizio Lecci, Ale Rinaldo)

PART III
PERSISTENT HOMOLOGY
Homology In One Slide

\[ \beta_0 = 1, \beta_1 = 1 \]

\[ \beta_0 = 1, \beta_1 = 2 \]

\[ \beta_0 = 2, \beta_1 = 1 \]
Persistent Homology
Barcode

Barcodes

H_2

H_1

H_0

\( \varepsilon \)
Persistence Diagram

Close diagonal = “noise”
Our Contribution: Confidence Band
Bottleneck Distance

Bottleneck distance between two persistence diagrams $\mathcal{P}$ and $\mathcal{Q}$: $W_\infty(\mathcal{P}, \mathcal{Q}) = \text{cost of best matching}$:
$\mathcal{P} = \text{diagram of } M$.

$\hat{\mathcal{P}} = \text{diagram of data.}$

Find confidence bound $c_n$ such that

$$\limsup_{n \to \infty} \mathbb{P}(W_\infty(\hat{\mathcal{P}}, \mathcal{P}) > c_n) \leq \alpha.$$
CONFIDENCE INTERVALS

\( X_1, \ldots, X_n \sim P. \)
Support of \( P \) is \( \mathbb{M} \).
\( S_n = \{X_1, \ldots, X_n\} \).
Four methods:

1. Subsampling
2. Concentration of measure
3. Shells
4. Density Estimation + Bootstrap
Method I: SUBSAMPLING

Choose $b = b_n$ where $b_n \to \infty$ but $b_n = o(n)$.

Subsamples of size $b$: $\Omega_1, \ldots, \Omega_N$

$$T_j = H(S_n, \Omega_j).$$

$$L(t) = \frac{1}{N} \sum_{j=1}^{N} I(T_j > t)$$

$$c = 2L^{-1}(\alpha).$$

Then

$$\mathbb{P}(W_\infty(\hat{P}, P) > c_n) \leq \alpha + O\left(\sqrt{\frac{b \log n}{n}}\right) + \frac{2^d}{n \log n}.$$
Method II: CONCENTRATION OF MEASURE

\[
\mathbb{P}(W_\infty(\hat{\mathcal{P}}, \mathcal{P}) > c_n) \leq \frac{2^{d+1}}{t^d \rho} \exp \left( -\frac{n \rho t^d}{2} \right)
\]

where \( \rho = \inf_{x \in M} \rho(x) \), \( \rho(x) = \lim_{t \to 0} \frac{P(B(x, t/2))}{t^d} \).

Need to estimate \( \rho \), set RHS equal to \( \alpha \) and find \( t \).

\[
\hat{\rho} = \min_i \frac{P_n(B(X_i, r_n/2))}{r_n^d}
\]

where \( r_n = (\log n / n)^{\frac{1}{2+d}} \). We get

\[
\mathbb{P}(W_\infty(\hat{\mathcal{P}}, \mathcal{P}) > c_n) \leq \alpha + O \left( \sqrt{\frac{\log n}{n}} \right)^{\frac{1}{2+d}}.
\]
Method III: SHELLS

Apply previous inequality over shells \( \{ x : \gamma_j < \rho(x) < \gamma_{j+1} \} \). Let shells shrink and we get:

\[
\mathbb{P}(W_\infty(\hat{\mathcal{P}}, \mathcal{P}) > t) \leq \frac{2^{d+1}}{t^d} \int g(v) \frac{e^{-nvt^d/2}}{v} dv
\]

where

\[
g(v) = G'(v)
\]

\[
G(v) = \mathbb{P}(\rho(X) \leq v).
\]

Note: \( g \) is the “density of the density.”
We estimate $g$ by

$$
\hat{g}(\nu) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{b} K \left( \frac{||\nu - V_i||}{b} \right)
$$

and

$$
V_i = \frac{P_n(B(X_i, r_n/2))}{r_n^d}.
$$

Set RHS equal to $\alpha$ and solve for $t$:

$$
P(W_\infty(\hat{P}, P) > t) \leq \alpha + O \left( \frac{\log n}{n} \right)^{\frac{1}{2+d}}.
$$
Method IV: Via DENSITY ESTIMATION

Recall that $X_1, \ldots, X_n \sim P$. $\mathbb{M} \subset \text{support}(P)$. Let

$$\hat{p}_h(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^D} K \left( \frac{||x - X_i||}{h} \right)$$

which has mean

$$p_h(x) = \mathbb{E}(\hat{p}_h(x)) = P \ast K_h.$$ 

Let $\mathcal{P}_h$ be the persistence diagram of the upper level sets

$$\{x : p_h(x) > t\}.$$ 

We estimate $\mathcal{P}_h$ with $\hat{\mathcal{P}}_h$, the persistence diagram of the upper level sets

$$\{x : \hat{p}_h(x) > t\}.$$
We have that
\[
W_{\infty}(\hat{P}_h, P_h) \leq \|\hat{p}_h - p_h\|_{\infty}.
\]

Define \(Z_\alpha\) such that:
\[
P\left( \sqrt{nh^D} \|\hat{p}_h^* - \hat{p}_h\|_{\infty} > Z_\alpha \mid X_1, \ldots, X_n \right) = \alpha.
\]

Then:
\[
P\left( W_{\infty}(\hat{P}_h, P_h) > \frac{Z_\alpha}{\sqrt{nh^D}} \right) \leq \alpha + O\left( \sqrt{\frac{1}{n}} \right).
\]
Why Do This?

1. Probe cosmological features. (See, for example, Sousbie (2013) arxiv.org/abs/1302.6221).

2. Compare sky surveys to cosmological simulations.

3. Currently: applying persistent homology to cosmological data (with Jessi Cisewski).
Conclusion

1. Topological (and geometric) statistical inference is about finding “topological signal” in point clouds.

2. Computational topologists have developed very sophisticated theory and methods.

3. Statisticians can contribute by developing ways to measure uncertainty (confidence sets etc.)

4. Lots of things to do:
   - intersections (stratified spaces)
   - improvements
   - optimality and minimaxity
   - choosing tuning parameters
   - spatial adaptivity.
THE END