Can topological summary statistics be sufficient?

Katharine Turner

University of Chicago

kate@math.uchicago.edu

February 4, 2014
Topological summary statistics
Outline

Topological summary statistics

Likelihood function
Outline

Topological summary statistics

Likelihood function

Bayesian basics
Outline

Topological summary statistics

Likelihood function

Bayesian basics

Sufficient statistics
Topological summary statistics

Likelihood function

Bayesian basics

Sufficient statistics

Euler Characteristic Transform and Persistent Homology Transform
Outline

Topological summary statistics

Likelihood function

Bayesian basics

Sufficient statistics

Euler Characteristic Transform and Persistent Homology Transform

Studying bones and silhouette pictures via the PHT
Topological summary statistics

Likelihood function

Bayesian basics

Sufficient statistics

Euler Characteristic Transform and Persistent Homology Transform

Studying bones and silhouette pictures via the PHT

Asymptotically sufficient statistics
Different stages we could use TDA

Each data entry could be a single point in a point cloud - we want to discover the distribution of the density and we use TDA mainly in an exploratory way

OR

Each data entry could be a functions/shapes/graph/point cloud - we want to compare different objects (qualitatively and quantitatively) and topological summaries give us some methods to do so.
Different stages we could use TDA

Each data entry could be a single point in a point cloud - we want to discover the distribution of the density and we use TDA mainly in an exploratory way

OR

Each data entry could be a functions/shapes/graph/point cloud - we want to compare different objects (qualitatively and quantitatively) and topological summaries give us some methods to do so.

We may even want to consider each function/shape as a location in function/shape space and look at distributions of functions/shapes. We could indeed use TDA at both levels which would be rather meta.
A summary statistic summarizes some information about the data.

Real valued data:
- mean
- median
- variance
- maximum
- minimum

Euclidean vectors:
- mean (a vector)
- covariance matrix

Sets of graphs (networks):
- number vertices/edges,
- clustering coefficient
- degree distribution

Sets of simplicial complexes:
- Homology groups
- Betti numbers
- Euler characteristic
Examples of Summary statistics

A summary statistic summarizes some information about the data.

Real valued data:
- mean
- median
- variance
- maximum
- minimum

Euclidean vectors:
- mean (a vector)
- covariance matrix
Examples of Summary statistics

A summary statistic summarizes some information about the data.

Real valued data:
- mean
- median
- variance
- maximum
- minimum

Euclidean vectors:
- mean (a vector)
- covariance matrix

Sets of graphs (networks):
- number vertices/edges,
- clustering coefficient
- degree distribution

Sets of simplicial complexes:
- Homology groups
- Betti numbers
- Euler characteristic
Examples of Summary statistics

A summary statistic summarizes some information about the data.

Real valued data:
- mean
- median
- variance
- maximum
- minimum

Euclidean vectors:
- mean (a vector)
- covariance matrix

Sets of graphs (networks):
- number vertices/edges,
- clustering coefficient
- degree distribution

Sets of simplicial complexes:
- Homology groups
- Betti numbers
- Euler characteristic
Care about how the topology changes over time (or some other filter). Examples include:

- Merge tree
- Reeb Graph
- Euler characteristic curve
- Barcode/Persistence diagram
Euler characteristic curve

Given a simplicial complex $M$ the Euler characteristic is

$$\chi(M) = \sum_{i=0}^{\infty} (-1)^i \dim H_i(M) = \sum_{i=0}^{\infty} (-1)^i c_i$$

where $c_i$ is the number of $i$-cells.
Given a simplicial complex $M$ the Euler characteristic is

$$\chi(M) = \sum_{i=0}^{\infty} (-1)^i \dim H_i(M) = \sum_{i=0}^{\infty} (-1)^i c_i$$

where $c_i$ is the number of $i$-cells.

Given a filtration $M_t$ of $M$ the corresponding Euler characteristic curve is the function $f : \mathbb{R} \to \mathbb{Z}$ with $f(t) = \chi(M_t)$. This is very fast to compute.
Given a simplicial complex $M$ the Euler characteristic is

$$\chi(M) = \sum_{i=0}^{\infty} (-1)^i \dim H_i(M) = \sum_{i=0}^{\infty} (-1)^i c_i$$

where $c_i$ is the number of $i$-cells.

Given a filtration $M_t$ of $M$ the corresponding Euler characteristic curve is the function $f : \mathbb{R} \to \mathbb{Z}$ with $f(t) = \chi(M_t)$. This is very fast to compute.

In particular, if we pick our filtration to be the height in the direction of $v$ we have

$$f(t) = \chi(\{x \in M : x \cdot v \leq t\}).$$
Likelihood function

In statistics, a likelihood function is a function of the parameters $\theta$ of a statistical model given the data $x$.

$$\mathcal{L}(\theta|x) = f_{\theta}(x)$$

where:

If $X$ is a discrete valued random variable then $f_{\theta}(x) = P(X = x|\theta)$

If $X$ is a continuous valued random variable then $f_{\theta}$ is the (joint) pdf of $X$ given the parameter $\theta$.

If $X_1, \ldots, X_n$ are iid random variables then $f_{\theta}(\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}) = \prod_{i=1}^{n} f_{\theta}(x_i)$.
In statistics, a likelihood function is a function of the parameters $\theta$ of a statistical model given the data $x$.

$$ \mathcal{L}(\theta|x) = f_\theta(x) $$

where:

- If $X$ is a discrete valued random variable then $f_\theta(x) = \mathbb{P}(X = x|\theta)$
Likelihood function

In statistics, a likelihood function is a function of the parameters $\theta$ of a statistical model given the data $x$.

$$\mathcal{L}(\theta | x) = f_{\theta}(x)$$

where:

- If $X$ is a discrete valued random variable then $f_{\theta}(x) = P(X = x | \theta)$
- If $X$ is a continuous valued random variable then $f_{\theta}$ is the (joint) pdf of $X$ given the parameter $\theta$.

If $X_1, \ldots X_n$ are iid random variables then

$$f_{\theta}({X_1 = x_1, X_2 = x_2, \ldots X_n = x_n}) = \prod_{i=1}^{n} f_{\theta}(x_i).$$
Likelihood function

In statistics, a likelihood function is a function of the parameters $\theta$ of a statistical model given the data $x$.

$$L(\theta|x) = f_\theta(x)$$

where:

- If $X$ is a discrete valued random variable then $f_\theta(x) = P(X = x|\theta)$
- If $X$ is a continuous valued random variable then $f_\theta$ is the (joint) pdf of $X$ given the parameter $\theta$.

If $X_1, \ldots X_n$ are iid random variables then

$$f_\theta(\{X_1 = x_1, X_2 = x_2, \ldots X_n = x_n\}) = \prod_{i=1}^{n} f_\theta(x_i).$$

To infer $\theta$ from $X$ often we just use the $\theta$ with the maximal likelihood.
The is to use Bayes rule:

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \]

This can be written in terms of conditional densities for continuous valued variables.
The is to use Bayes rule:

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \]

This can be written in terms of conditional densities for continuous valued variables.

Bayes rule follows from \( P(B|A)P(A) = P(A \text{ and } B) = P(A|B)P(B). \)
Say we want to find the probability that some parameter $\Theta$ in a model is $\theta$ using the information from data $x$. 

Prior distribution of $\theta$, denoted $p(\theta) := P(\Theta = \theta)$. This is like our belief system of what $\theta$ should be before we see any data.

Posterior distribution of $\theta$, denoted $\pi(\theta | x) := P(\Theta = \theta | X = x)$. This is like our belief system but updated using the evidence of the data.
Say we want to find the probability that some parameter $\Theta$ in a model is $\theta$ using the information from data $x$.

We have some **prior** distribution of $\theta$, denoted

$$p(\theta) := \mathbb{P}(\Theta = \theta).$$

This is like our belief system of what $\theta$ should be before we see any data.
Prior and posterior probabilities

Say we want to find the probability that some parameter $\Theta$ in a model is $\theta$ using the information from data $x$.

We have some prior distribution of $\theta$, denoted

$$p(\theta) := \mathbb{P}(\Theta = \theta).$$

This is like our belief system of what $\theta$ should be before we see any data.

We then have a posterior distribution of $\theta$, denoted

$$\pi(\theta|x) := \mathbb{P}(\Theta = \theta|X = x)$$

given data $x$. This is like our belief system but updated using the evidence of the data.
By Bayes rule:

\[ \pi(\theta|x) = \mathbb{P}(\Theta = \theta|X = x) = \frac{\mathbb{P}(X = x|\Theta = \theta)\mathbb{P}(\Theta = \theta)}{\mathbb{P}(X = x)} = \frac{\mathcal{L}(\theta|x)p(\theta)}{\mathbb{P}(X = x)} \]
Prior and posterior probabilities

By Bayes rule:

\[ \pi(\theta|x) = \mathbb{P}(\Theta = \theta|X = x) \]
\[ = \frac{\mathbb{P}(X = x|\Theta = \theta)\mathbb{P}(\Theta = \theta)}{\mathbb{P}(X = x)} \]
\[ = \frac{\mathcal{L}(\theta|x)p(\theta)}{\mathbb{P}(X = x)} \]

Since \( \mathbb{P}(X = x) \) is common for all \( \theta \) we have

Posterior \( \propto \) Likelihood \( \times \) Prior
Example: posterior for pants

Suppose there is a mixed school having 60% boys and 40% girls as students. The girls wear pants or skirts in equal numbers; the boys all wear trousers. The parameter is gender and the data will be pants vs skirt.
Suppose there is a mixed school having 60% boys and 40% girls as students. The girls wear pants or skirts in equal numbers; the boys all wear trousers. The parameter is gender and the data will be pants vs skirt.

**Prior:**
Before we have any information to update our estimates we have the prior distribution of $p(\text{boy}) = 0.6$ and $p(\text{girl}) = 0.4$. 
Suppose there is a mixed school having 60% boys and 40% girls as students. The girls wear pants or skirts in equal numbers; the boys all wear trousers. The parameter is gender and the data will be pants vs skirt.

**Prior:**
Before we have any information to update our estimates we have the prior distribution of $p(\text{boy}) = 0.6$ and $p(\text{girl}) = 0.4$.

An observer sees a (random) student from a distance; all the observer can see is that this student is wearing pants. We want the posterior distribution given this extra information.
Example: posterior for pants

We have the prior $p(\text{boy}) = 0.6$ and $p(\text{girl}) = 0.4$, the likelihoods $\mathcal{L}(\text{girl}|\text{pants}) = 0.5$ and $\mathcal{L}(\text{boy}|\text{pants}) = 1$, and

$$P(\text{pants}) = P(\text{pants and girl}) + P(\text{pants and boy}) = 0.5 \times 0.4 + 0.6 = 0.8.$$ 

$$\pi(\text{girl}|\text{pants}) = \frac{\mathcal{L}(\text{girl}|\text{pants})p(\text{girl})}{P(\text{pants})} = \frac{0.5 \times 0.4}{0.8} = 0.25$$

$$\pi(\text{boy}|\text{pants}) = \frac{\mathcal{L}(\text{boy}|\text{pants})p(\text{boy})}{P(\text{pants})} = \frac{1 \times 0.6}{0.8} = 0.75$$
A statistic $T$ is sufficient for $\theta$ if
$$\mathbb{P}(X = x | T(X) = T(x)) = \mathbb{P}(X = x | T(X) = T(x), \theta) \text{ for all } X = \{x_1, \ldots, x_n\}.$$
A statistic $T$ is sufficient for $\theta$ if

$$P(X = x | T(X) = T(x)) = P(X = x | T(X) = T(x), \theta) \text{ for all } X = \{x_1, \ldots x_n\}.$$ 

Equivalently $T$ is sufficient if $L(\theta | T(x)) = L(\theta | x)$. 
A statistic $T$ is sufficient for $\theta$ if
$$P(X = x | T(X) = T(x)) = P(X = x | T(X) = T(x), \theta)$$
for all $X = \{x_1, \ldots, x_n\}$.

Equivalently $T$ is sufficient if $L(\theta | T(x)) = L(\theta | x)$.

We don’t lose information by throwing away $x$ and just keeping $T(x)$!
**Theorem (Fisher-Neyman)**

A statistic $T = T(X)$ is sufficient for $\theta$ if and only if the likelihood function factorizes into the form

$$\mathcal{L}(\theta|x) = g(\theta, T(x))h(x)$$

for some functions $g$ and $h$. 
Examples of sufficient statistics

Trivially $T(X) = X$ must be a sufficient statistic of $X$. 
Examples of sufficient statistics

Trivially $T(X) = X$ must be a sufficient statistic of $X$.

For the Gaussian distribution $N(\mu, \sigma^2)$
- the pair of (sample mean, sample variance) is a sufficient statistic
- if the variance is known then the sample mean is sufficient
- the variance is not sufficient even with known mean
Euler Characteristic of a height function

Let $M$ be a subset of $\mathbb{R}^d$. For each unit vector $v \in S^{d-1}$ we can create an Euler characteristic curve $\chi(M, v)$ by the filtration by the height function in the direction of $v$. That is

$$\chi(M, v)(t) = \chi(\{x \in M : x \cdot v \leq t\})$$

picture from Mao Li.
We can consider the characteristic functions for height functions of different directions.
Euler Characteristic of a height function

**Definition**

The Euler characteristic transform of $M \subset \mathbb{R}^d$ is the function

$$ECT(M) : S^{d-1} \rightarrow L^2(\mathbb{R})$$

$$\nu \mapsto \chi(M, \nu).$$
Let $\mathcal{D}$ denote the space of persistence diagrams. Similar to the ECT we can define the persistence homology transform (PHT).

**Definition**

The Persistent Homology transform (PHT) of $M \subset \mathbb{R}^d$ is the function

$$PHT(M) : S^{d-1} \rightarrow \mathcal{D}$$

where $PHT(M)(v)$ is the persistence diagram corresponding to the filtration by the height function in the direction of $v$. 
Euler transform

Theorem (T-Mukherjee-Boyer)

Let $\mathcal{M}_d$ be the space of finite simplicial complexes in $\mathbb{R}^d$. The ECT and the PHT are both injective for $d = 2, 3$.

We conjecture that they are also injective for higher dimensions.
Distances on $M_d$

We can define various distance functions between simplicial complexes $M_1, M_2$ in $M_d$. In particular:

$$d_{ECT}(M_1, M_2) := \int_{S}^{d-1} d(\chi(M_1, v), \chi(M_2, v)) dv.$$  

for whatever distance function $d(,)$ you want on the space of functions

or

$$d_{PHT}(M_1, M_2) := \int_{S^{d-1}} d(PHT(M_1)(v), PHT(M_2)(v)) dv$$

for whatever distance function $d(,)$ you want on the space of persistence diagrams.
Analyzing bones
We used the a distance function for (already aligned) bones of

\[ d_{PHT}(M_1, M_2) := \left( \int_{S^{d-1}} d_{L^2}(PHT(M_1)(v), PHT(M_2)(v)) \, dv \right)^{1/2} \]

which is the $L^2$ version of the distances between the sets of persistence diagrams.
Displaying the clusters after PCA

Phenetic clustering of phylogenetic groups of primate calcanei (n = 106)
"In at least one way the method matched shapes with family groups better than any of the other previous methods... it linked a Hylobates specimen with the the other ape specimens (pan, gorilla, pongo, and oreopithecus). Previous both hylobatids (which ARE apes) always ended up closest to some Alouatta specimens."
Modifying the distance algorithm we can make it scale, rotation and translation invariant and applied to a subset of the silhouette images in MPEG-7 CE Shape-1 Part-B:
Shapes in a plane
Multidimensional Persistence transform

ECT and PHT not great when the data is not isomorphic.

Normal fly wings [photos from David Houle’s lab]:

Topologically abnormal veins:
Mulitdimensional Persistence transform

One approach is to consider the generalized persistence given pairs of filtrations

\[(\text{distance to the set, height in direction } v)\]

or an Euler Characteristic surface with these two parameters.
One approach is to consider the generalized persistence given pairs of filtrations

\[(\text{distance to the set, height in direction } v)\]

or an Euler Characteristic surface with these two parameters.

Alternatively we may be given a function over the plane rather than a compact set. For example, grey scale images. We could do multidimensional persistence with the pairs of filtrations

\[(\text{function of interest, height in direction } v)\]

or an Euler Characteristic surface with these two parameters.
A shape is transformed into collection of the curves \( \{ \chi(M, v) \} \). We may want an exponential family of sets of curves.

A natural exponential family model for the collection of these curves is a multivariate Gaussian process

\[
X = [\chi(M, v_1), \chi(M, v_2), \chi(M, v_3), \ldots \chi(M, v_L)] \sim \mathcal{GP}_L(\mu, K).
\]

Perhaps we could finite dimensional subspaces which approximates the subspace of the image in function space well and project onto that. We could then have finite dimensional multivariate Gaussian distributions, etc.
Asymptotically Sufficient Statistics

Definition

Let $k \geq 1$ be an integer, $\pi(\theta|\mathbf{x}^{(n)})$ a posterior density distribution function of $\theta$ with respect to some prior and $\psi(\theta|T_n)$ a probability density function determined by some statistic $T = T_n(\mathbf{x}^{(n)})$. For every $\theta_0$ and $\epsilon > 0$ if

$$\lim_{n \to \infty} \mathbb{P}_{\theta_0}^N \{ \mathbf{x} \in X^N : d(\pi(\theta|\mathbf{x}), \psi(\theta|T_n)) \geq \epsilon n^{-(k-1)/2} \} = 0$$

then $T_n$ is a $k$th order asymptotically sufficient statistic.

There are a variety of choices for type of convergence.
The idea is that our best guess of the distribution using just the summary statistic \( T(X) \) is close to the posterior distribution given \( X \) most of the time and that as \( n \) increases our notion of “close” and “most” become more strict.
Asymptotically Sufficient Statistics

The idea is that our best guess of the distribution using just the summary statistic $T(X)$ is close to the posterior distribution given $X$ most of the time and that as $n$ increases our notion of “close” and “most” become more strict.

When $T(X)$ is a topological summary statistic we expect that our best guess using $T(X)$ will only work with high probability (rather than for all $X$) because of sampling theory - we normally will need a suitably nice set of points.
Suppose that the model is that points are being drawn from a circle of radius $r$ without noise and we wish to find the parameter $r$. 
Suppose that the model is that points are being drawn from a circle of radius \( r \) without noise and we wish to find the parameter \( r \).

It is possible to do so with three data points but can we also do so with persistence diagrams?
Suppose that the model is that points are being drawn from a circle of radius $r$ without noise and we wish to find the parameter $r$.

It is possible to do so with three data points but can we also do so with persistence diagrams?

Let $T(X)$ be the $H_1$ persistence diagram for the Čech complex. Generically we would think that $T(X)$ will have exactly one point off the diagonal and it will have $y$ coordinate of $r_0$ if the radius of the circle was $r_0$. 
Toy example - sufficient?

Is $T(X)$ a sufficient statistic?
Is $T(X)$ a sufficient statistic?

No.
We could be very unlucky and draw all the data points in one half of the circle; $T(X)$ will have no off diagonal points and we could not discover the radius. In contrast, we can always determine the radius from the raw data.
Fix the parameter of the model as $r_0$. The data determines the radius so $\pi(r_0|x) = \delta_{r_0}$.
Fix the parameter of the model as $r_0$. The data determines the radius so $\pi(r_0|x) = \delta_{r_0}$. $T(X)$ contains at most one point. Set

$$\psi(\theta|T(X)) := \begin{cases} \delta_b & (= \delta_{r_0}) \text{ if } (a, b) \text{ is the off diagonal point in } T(X) \\ \delta_0 & \text{if there are no off diagonal points in } T(X) \end{cases}$$
Fix the parameter of the model as $r_0$. The data determines the radius so $\pi(r_0|x) = \delta_{r_0}$. $T(X)$ contains at most one point. Set

$$
\psi(\theta|T(X)) := \begin{cases} 
\delta_b & \text{if } (a, b) \text{ is the off diagonal point in } T(X) \\
\delta_0 & \text{if there are no off diagonal points in } T(X)
\end{cases}
$$

Sampling theory shows $\mathbb{P}(X^{(n)}: \pi(r|X) \neq \psi(r|T(X))) \to 0$ as $n \to \infty$ so $T(X)$ is asymptotically sufficient.
Suppose we draw points from a uniform measure between different latitudes of a sphere lying in the southern hemisphere. Use $a, b$ for the parameters of the 90 minus the latitudes that bound this area. Let us have a uniform prior except $a < b$. All distances will be the distances in the sphere.
Toy example - annular region in a sphere

Suppose we draw points from a uniform measure between different latitudes of a sphere lying in the southern hemisphere. Use $a, b$ for the parameters of the 90 minus the latitudes that bound this area. Let us have a uniform prior except $a < b$. All distances will be the distances in the sphere.

For now on let $A(a, b)$ denote the area between latitudes $a$ and $b$. 

\[
L(a, b | x, y) = \mathbf{1}\{a < x, b > y\} \frac{n_{A(x, y)}}{n - 1} A(a, b) \frac{1}{n} \int_{\{a' < x, b' > y\}} \frac{n_{A(a', b')}}{n} \]

and hence the posterior is

\[
\pi(a, b | x, y) = \mathbf{1}\{a < x, b > y\} \frac{1}{A(a, b)} \frac{1}{n} \int_{\{a' < x, b' > y\}} \frac{n_{A(a', b')}}{n} \]
Suppose we draw points from a uniform measure between different latitudes of a sphere lying in the southern hemisphere. Use $a, b$ for the parameters of the 90 minus the latitudes that bound this area. Let us have a uniform prior except $a < b$. All distances will be the distances in the sphere.

For now on let $A(a, b)$ denote the area between latitudes $a$ and $b$.

A sufficient statistic for this the minimum $x$ and maximum $y$ latitude seen in the data.

$$\mathcal{L}(a, b|x, y) = 1_{\{a<x,b>y\}} \frac{nA(x, y)^{n-1}}{A(a, b)^n}$$

and hence the posterior is

$$\pi(a, b|x, y) = 1_{\{a<x,b>y\}} \frac{1}{A(a, b)^n} \int_{\{a'<x,b'>y\}} \frac{1}{A(a', b')^n}$$
Fix parameters $a_0, b_0$. Split the sphere by longitudes into 6 regions. Say we have a “nice” sample with respect to epsilon if we have points $x_1, x_2, \ldots x_6$ in each of the 6 regions with latitudes in $[a + 0, a + 0 + \epsilon]$ and points $y_1, y_2, \ldots y_6$ in each of the 6 regions with latitudes in $[b_0 - \epsilon, b_0]$ and no huge gaps (relative to $a_0$) inside the bar of latitudes.
Fix parameters $a_0, b_0$. Split the sphere by longitudes into 6 regions. Say we have a “nice” sample with respect to epsilon if we have points $x_1, x_2, \ldots x_6$ in each of the 6 regions with latitudes in $[a + 0, a + 0 + \epsilon]$ and points $y_1, y_2, \ldots y_6$ in each of the 6 regions with latitudes in $[b_0 - \epsilon, b_0]$ and no huge gaps (relative to $a_0$) inside the bar of latitudes.

Sampling theory tells us that for certain functions of $n$, such as $\epsilon = 1/n^\alpha$ with $\alpha \in (0, 1)$, we have a “nice” sample with probability going to 1 as $n$ goes to infinity.
Fix parameters $a_0, b_0$. Split the sphere by longitudes into 6 regions.
Say we have a “nice” sample with respect to epsilon if we have points $x_1, x_2, \ldots x_6$ in each of the 6 regions with latitudes in $[a + 0, a + 0 + \epsilon]$ and points $y_1, y_2, \ldots y_6$ in each of the 6 regions with latitudes in $[b_0 - \epsilon, b_0]$ and no huge gaps (relative to $a_0$) inside the bar of latitudes.

Sampling theory tells us that for certain functions of $n$, such as $\epsilon = 1/n^\alpha$ with $\alpha \in (0, 1)$, we have a “nice” sample with probability going to 1 as $n$ goes to infinity.

Under these sampling conditions the $H_1$ persistence diagram will have two significant features - $\tilde{x} \in [a + 0, a + 0 + \epsilon]$ and $\tilde{w} \in [\pi - b_0, \pi - b_0 + \epsilon]$. 
Fix parameters $a_0, b_0$. Split the sphere by longitudes into 6 regions.
Say we have a “nice” sample with respect to epsilon if we have points $x_1, x_2, \ldots x_6$ in each of the 6 regions with latitudes in $[a + 0, a + 0 + \epsilon]$ and points $y_1, y_2, \ldots y_6$ in each of the 6 regions with latitudes in $[b_0 - \epsilon, b_0]$ and no huge gaps (relative to $a_0$) inside the bar of latitudes.

Sampling theory tells us that for certain functions of $n$, such as $\epsilon = 1/n^\alpha$ with $\alpha \in (0, 1)$, we have a “nice” sample with probability going to 1 as $n$ goes to infinity.

Under these sampling conditions the $H_1$ persistence diagram will have two significant features - $\tilde{x} \in [a + 0, a + 0 + \epsilon]$ and $\tilde{w} \in [\pi - b_0, \pi - b_0 + \epsilon]$.

Set $\psi(a, b|\tilde{x}, \tilde{w}) = \pi(a, b|\tilde{x}, \pi - \tilde{w})$. 

Fix parameters $a_0, b_0$. Split the sphere by longitudes into 6 regions. Say we have a “nice” sample with respect to epsilon if we have points $x_1, x_2, \ldots x_6$ in each of the 6 regions with latitudes in $[a + 0, a + 0 + \epsilon]$ and points $y_1, y_2, \ldots y_6$ in each of the 6 regions with latitudes in $[b_0 - \epsilon, b_0]$ and no huge gaps (relative to $a_0$) inside the bar of latitudes.

Sampling theory tells us that for certain functions of $n$, such as $\epsilon = 1/n^\alpha$ with $\alpha \in (0, 1)$, we have a “nice” sample with probability going to 1 as $n$ goes to infinity.

Under these sampling conditions the $H_1$ persistence diagram will have two significant features - $\tilde{x} \in [a + 0, a + 0 + \epsilon]$ and $\tilde{w} \in [\pi - b_0, \pi - b_0 + \epsilon]$.

Set $\psi(a, b|\tilde{x}, \tilde{w}) = \pi(a, b|\tilde{x}, \pi - \tilde{w})$.

We can show that $\psi(a, b|\tilde{x}, \tilde{w})$ is suitably close to $\pi(a, b|x, y)$ if we pick $\epsilon$ well.