A Statistical Approach to Persistent Homology

Peter Bubenik[1] Peter T. Kim[2]

[1]Cleveland State University

[2]University of Guelph

May 31, 2006 - SSC 2006 @ UWO

Aim

To combine statistical and topological ideas in the study of data.

Our present focus is on data sampled from parametric families of directional densities.

Outline

What is topology?

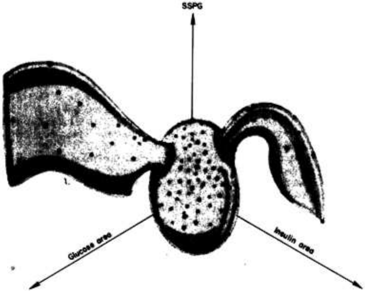
Definition

Topology is the branch of mathematics that studies spatial features which do not change under continuous deformations (can stretch, twist and compress, but not tear or glue together).

0-dimensional features

Example

Connected components are 0-dimensional topological features.



Source: R.G. Miller, Jr. Annals of Statistics. **13** (2) (1985) p.511.

1-dimensional features

Example

Holes and tunnels are 1-dimensional topological features.

1-dimensional features

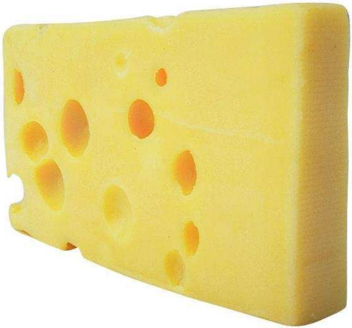
Example

Periodic motion can be revealed by 1-dimensional topological features.

2-dimensional features

Example

Voids are 2-dimensional topological features.



Betti numbers and homology

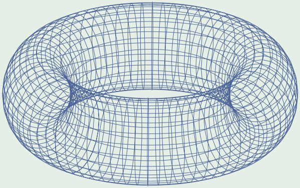
- For each dimension, homology gives a vector space.
- The Betti numbers give the dimension of this vector space.

Betti numbers and homology

- For each dimension, homology gives a vector space.
- The Betti numbers give the dimension of this vector space.

Example

the torus



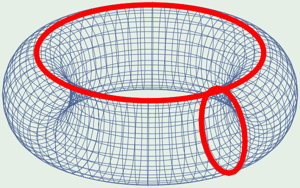
$\beta_0 = 1,$

Betti numbers and homology

- For each dimension, homology gives a vector space.
- The Betti numbers give the dimension of this vector space.

Example

the torus



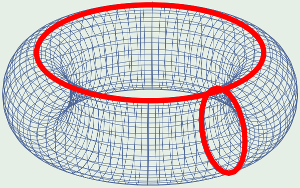
$$\beta_0 = 1, \quad \beta_1 = 2,$$

Betti numbers and homology

- For each dimension, homology gives a vector space.
- The Betti numbers give the dimension of this vector space.

Example

the torus



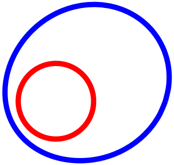
$$\beta_0 = 1, \quad \beta_1 = 2, \quad \beta_2 = 1, \quad \beta_k = 0 \text{ for } k \geq 3$$

From topology to statistical topology

- Instead of considering the topology of geometric objects, we will consider the topology of densities.
- Our calculational tool will be persistent homology.

The von Mises distribution

Our basic examples are unimodal densities on the unit circle S^1, called von Mises distributions:



$$f_{\mu,\kappa}(x) = c(\kappa)\exp\{\kappa x^t \mu\}.$$

where $\mu \in S^1$and $\kappa \in [0, \infty)$.

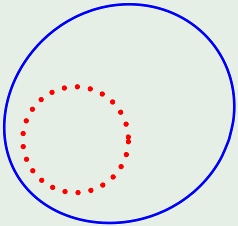
Persistent homology

- Persistent homology describes the (homological) features which persist as a single parameter changes.
- In many previous works, this parameter was taken to be the scale at which neighboring data points are connected.
- We take this parameter to be a threshold on the density of the space from which we are sampling.

Two filtrations on the circle

Example

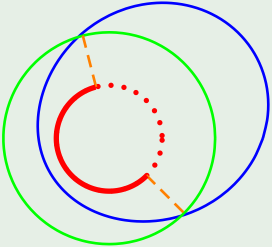
- We calculate the persistent homology of S^1 with density $f_{\mu,\kappa}$ by filtering S^1 according to the density.



Two filtrations on the circle

Example

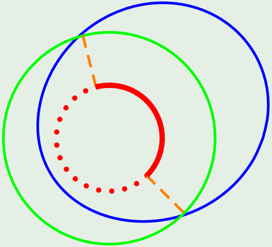
- We calculate the persistent homology of S^1 with density $f_{\mu,\kappa}$ by filtering S^1 according to the density.
- Either we take the portion of S^1 which has density
  - less than some threshold (the Morse filtration), or



Two filtrations on the circle

Example

- We calculate the persistent homology of S^1 with density $f_{\mu,\kappa}$ by filtering S^1 according to the density.
- Either we take the portion of S^1 which has density
  - less than some threshold (the Morse filtration), or
  - greater than some threshold (the Čech filtration).



Betti barcodes for the circle

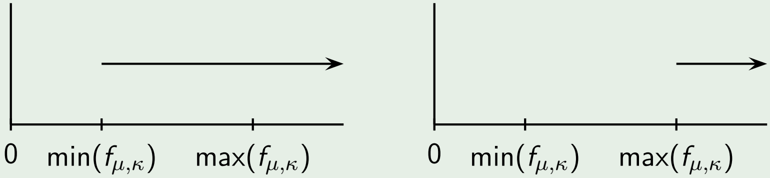
In each dimension, a Betti barcode describes the persistent homology.

Example

using the Morse filtration on S^1 with $f_{\mu,\kappa}$

Betti-0 barcode

Betti-1 barcode



Betti barcodes for the circle

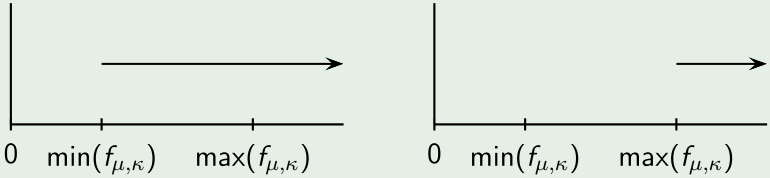
In each dimension, a Betti barcode describes the persistent homology.

Example

using the Morse filtration on S^1 with $f_{\mu,\kappa}$

Betti-0 barcode

Betti-1 barcode



$(\min(f_{\mu,\kappa}) = c(\kappa)e^{-\kappa}$ and $\max(f_{\mu,\kappa}) = c(\kappa)e^{\kappa})$

Estimates from samples on the circle

Let $X = (X_1, X_2, \dots X_n)$ be a sample on S^1 according the the von Mises density $f_{\mu,\kappa}$. Let $\bar{X}$ be the sample mean.

$$\bar{X} = ||\bar{X}|| \frac{\bar{X}}{||\bar{X}||}$$

The maximum likelihood estimators are given by:

$$\hat{\mu} = \frac{\bar{X}}{||\bar{X}||} \text{ and } \hat{\kappa} = A^{-1}(||\bar{X}||), \text{ where } A(\lambda) = \frac{I_1(\lambda)}{I_0(\lambda)}.$$

Errors of the estimates

The large sample mean-squared error:

$$E(\hat{\kappa} - \kappa)^2 \sim \frac{1}{A'(\kappa)} \frac{1}{n} \quad \text{as} \quad n \to \infty.$$

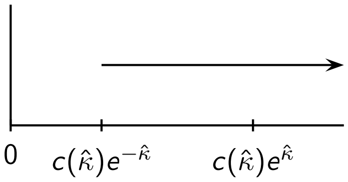
Errors of the estimates

The large sample mean-squared error:

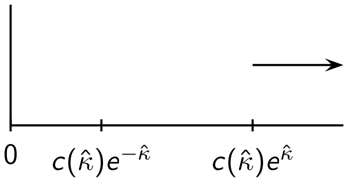
$$E(\hat{\kappa} - \kappa)^2 \sim \frac{1}{A'(\kappa)} \frac{1}{n} \quad \text{as} \quad n \to \infty.$$

Using the estimator $\hat{\kappa}$ we obtain estimates for the Betti barcodes.

Betti-0 barcode



Betti-1 barcode



The matrix von Mises distribution

Next we consider the consider the space of all orientations of an arbitrary 3-dimensional object. This is the rotation group $S0(3)$.

It has the unimodal densities ,

$$f_{A,\kappa}(X) = d(\kappa)\exp\{\kappa\,\mathrm{tr}(X^tA)\},$$

called matrix von Mises densities, which are parametrized by $A \in SO(3)$ and $\kappa \geq 0$.

Parametrizing the rotations

There is the Cayley-Klein map, $\rho : S^3 \longrightarrow SO(3)$,

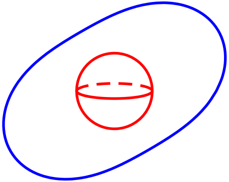
$$\rho \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} a^2+b^2-c^2-d^2 & 2bc-2ad & 2ac+2bd \\ 2ad+2bc & a^2-b^2+c^2-d^2 & 2cd-2ab \\ 2bd-2ac & 2ab+2cd & a^2-b^2-c^2+d^2 \end{pmatrix}$$

from the unit quaternions which parametrizes the rotations.

The inverse image of the matrix von Mises density

The preimage of the matrix von Mises distribution under this map is the following bimodal distribution on S^3:

$$f_{a,\lambda}(x) = K(\lambda)\exp\{\lambda(x^t a)^2\},$$



called the Watson distribution.

Samples of rotations

The Watson density $f_{a,\lambda}(x) = K(\lambda)\exp\{\lambda(x^t a)^2\}$ can be analyzed analogously to our analysis of the von Mises density on S^1.

Given a sample $X_1, \dots, X_n$ on $SO(3)$ we take their preimages $x_1, \dots, x_n$ on S^3, and we can calculate the maximum likelihood estimators $\hat{a}$ and $\hat{\lambda}$ for the Watson density.

From these we obtain maximum likelihood estimators $\hat{A} = \rho(\hat{a})$ and $\hat{\kappa}$ for the matrix von Mises density.

Estimating the Watson parameters

Watson density: $f(x|a,\lambda) = K(\lambda)\exp\{\lambda(x^t a)^2\}$

log-likelihood function: $\ell(a,\lambda) = n\log K(\lambda) + \lambda\sum_{j=1}^n (x_j^t a)^2$

The estimators $\hat{a}, \hat{\lambda}$, are the solutions of $\nabla_{a,\lambda}\ell(a,\lambda) = 0$.

large sample asymptotics:

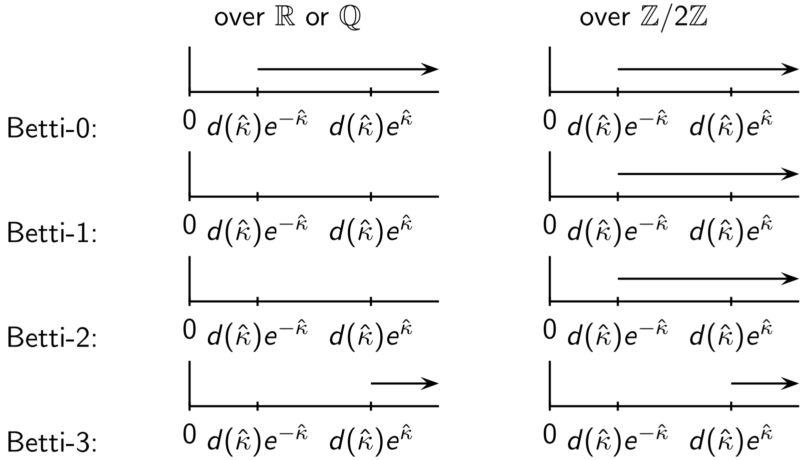
$$\sqrt{n}\left[\begin{pmatrix}\hat{a}\\ \hat{\lambda}\end{pmatrix} - \begin{pmatrix}a\\ \lambda\end{pmatrix}\right] \to_d N_4\left(0,\ I(a,\lambda)^{-1}\right) \text{ as } n\to\infty$$

$$\sqrt{n}\,(\hat{\lambda}-\lambda) \to_d N\left(0,\ \bar{K}(\lambda)\right) \text{ as } n\to\infty$$

The interesting topology of the rotations

- S0(3) is more interesting topologically than S^3.
- There is a nontrivial loop γ in $S0(3)$ such that 2γ is trivial.
- So we get different barcodes depending on the choice of field of coefficients.

Betti barcodes of the matrix von Mises density



Summary

- Topological features of densities can be detected using persistent homology.
- For several parametric problems, statistical estimators can be used to recover the topology of the distribution from which the data has been sampled.