Persistent homology and directional statistics

Peter Bubenik

(joint work with Peter Kim)

peter.bubenik@epfl.ch

http://igat.epfl.ch/bubenik/talks/

Ecole Polytechnique Fédérale de Lausanne

(Swiss Federal Institute of Technology at Lausanne)

1. Introduction

- Point Cloud Data from densities
- Densities on S^1

Point Cloud Data

Point Cloud Data (PCD) is a sequence of sampled points $X = (X_1, X_2, \ldots, X_n)$.

Usually the PCD is assumed to have been sampled uniformly from some manifold.

From this PCD one can use persistent homology to determine Betti barcodes.

Point Cloud Data

Point Cloud Data (PCD) is a sequence of sampled points $X = (X_1, X_2, \ldots, X_n)$.

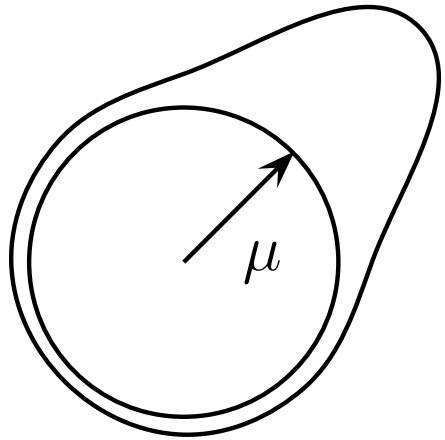
Usually the PCD is assumed to have been sampled uniformly from some manifold.

From this PCD one can use persistent homology to determine Betti barcodes.

Our point of view: We will assume that the PCD is sampled according some density on a manifold.

Densities on S^1

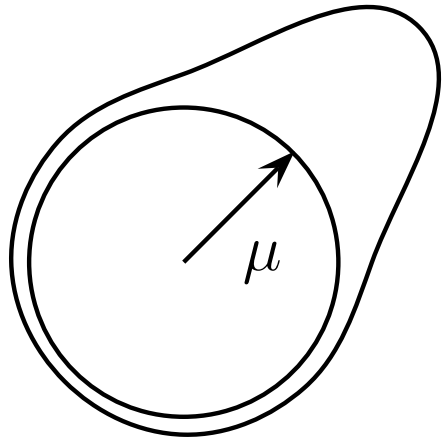
We will focus on unimodal densities on S^1, called von Mises distributions.



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

Densities on S^1

We will focus on unimodal densities on S^1, called von Mises distributions.

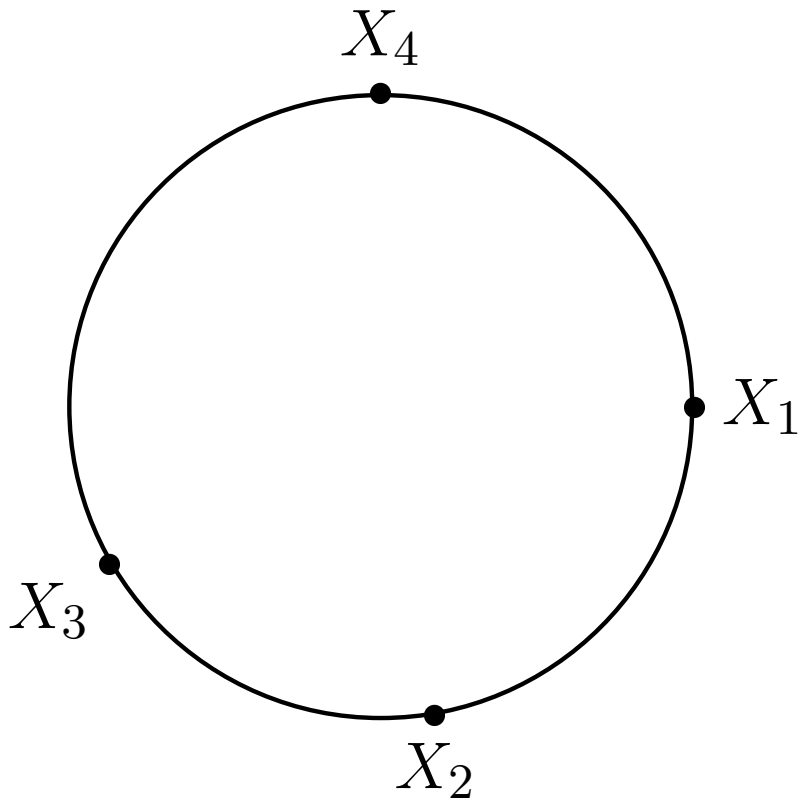


$$f_{\mu,\kappa}(x) = c(\kappa)e^{\kappa x^t \mu}.$$

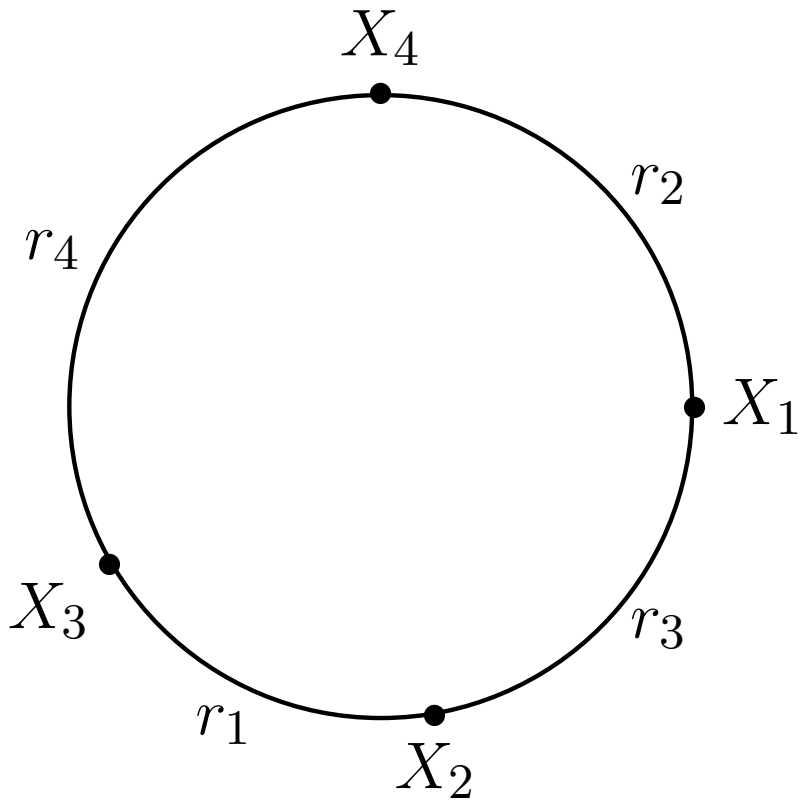
Sample $X = (X_1, \ldots, X_n) \subset S^1$ according to $f_{\mu,\kappa}$

$\xrightarrow{\text{persistent homology}}$ Betti barcodes

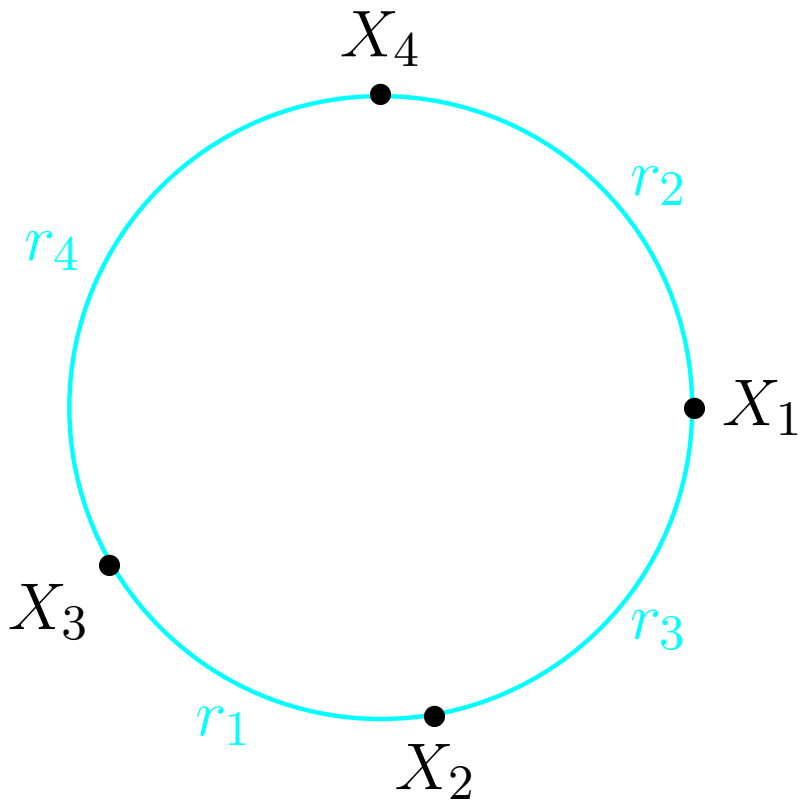
2. Persistent Homology



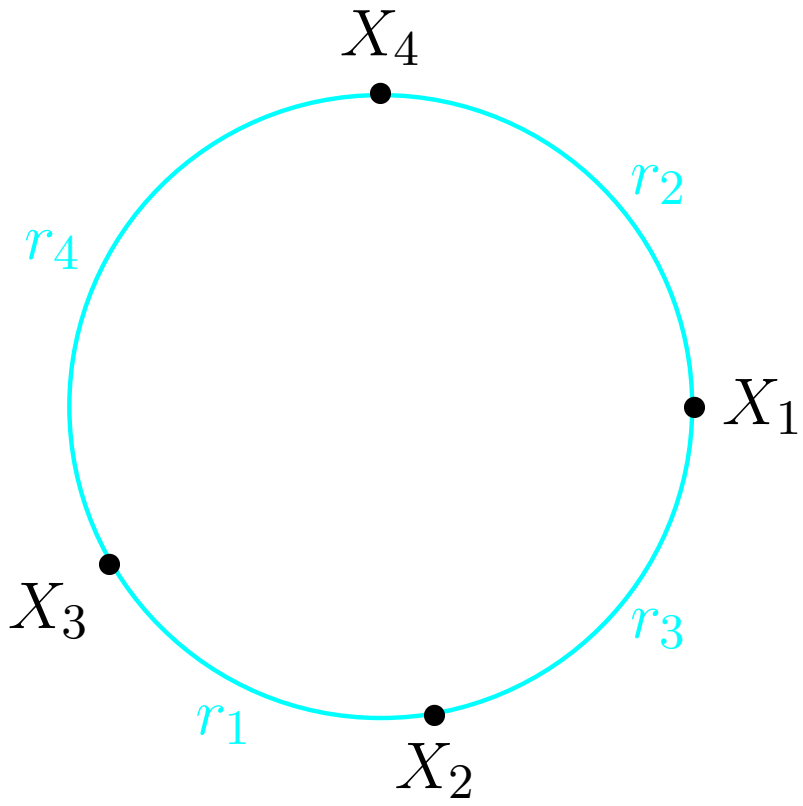
2. Persistent Homology

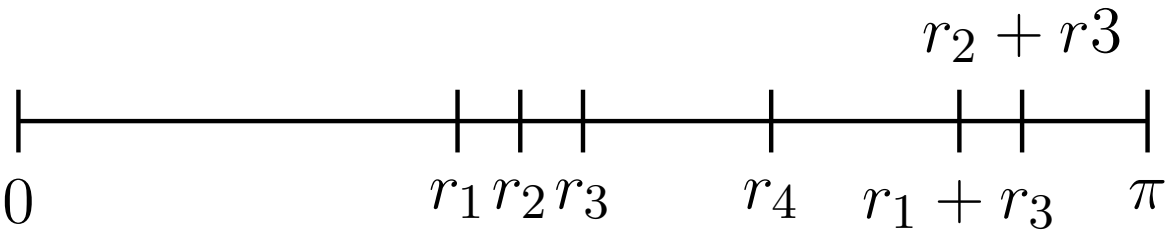


The Rips complex

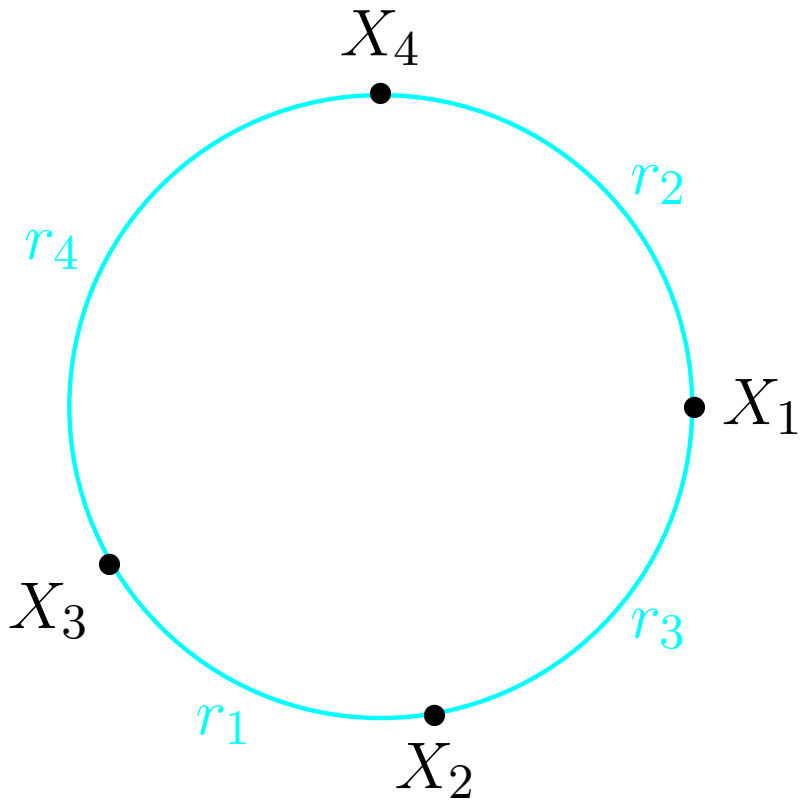


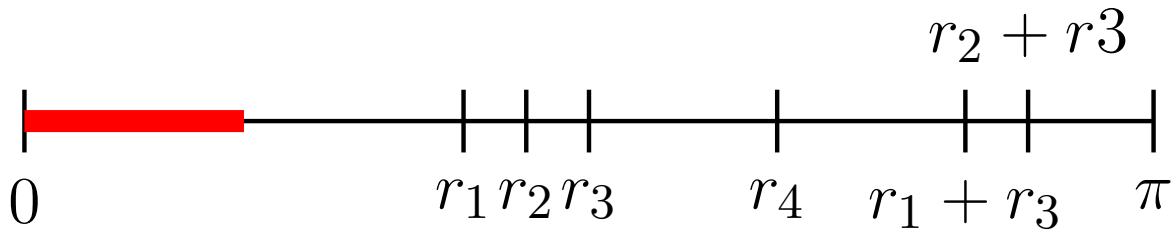
The Rips complex



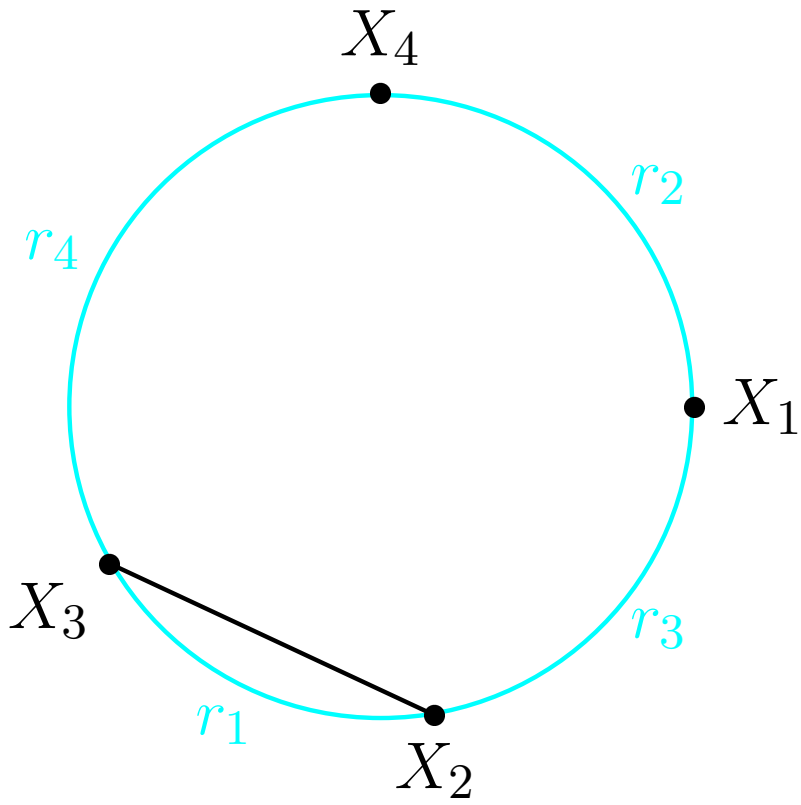


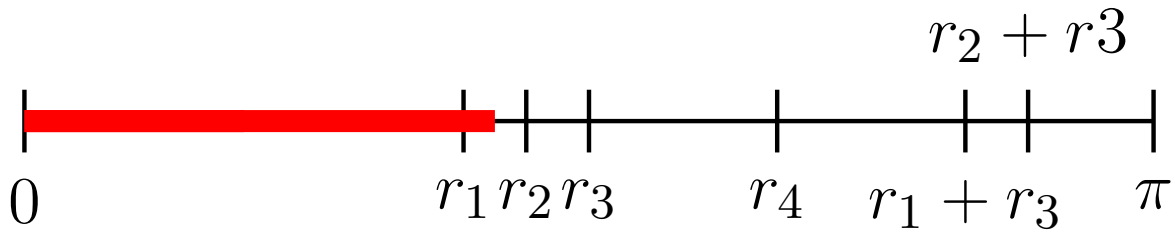
The Rips complex



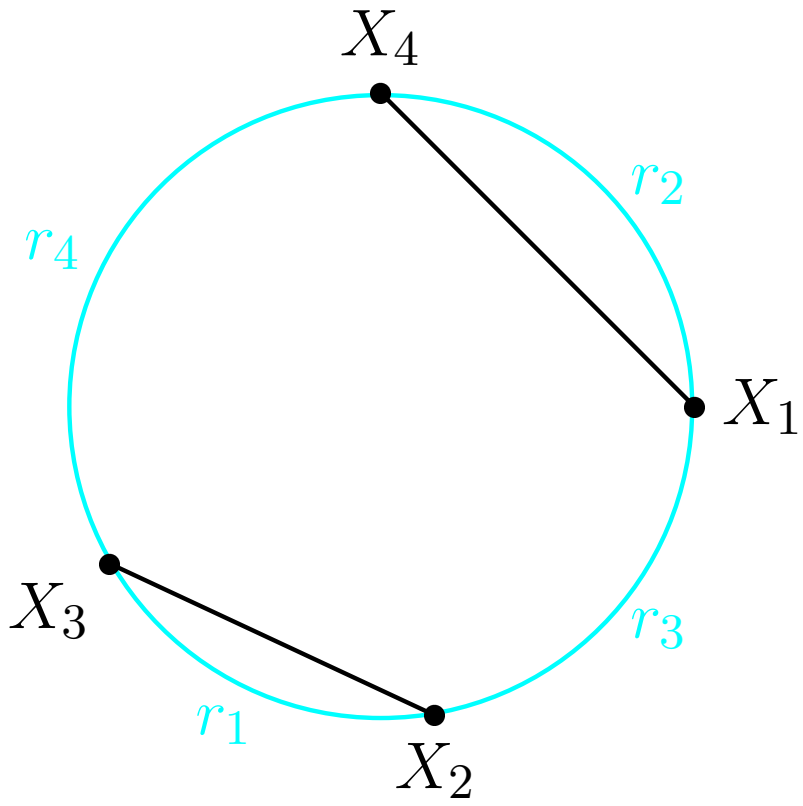


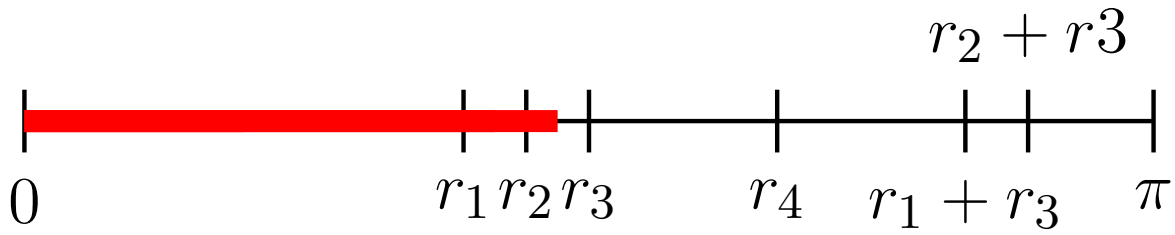
The Rips complex



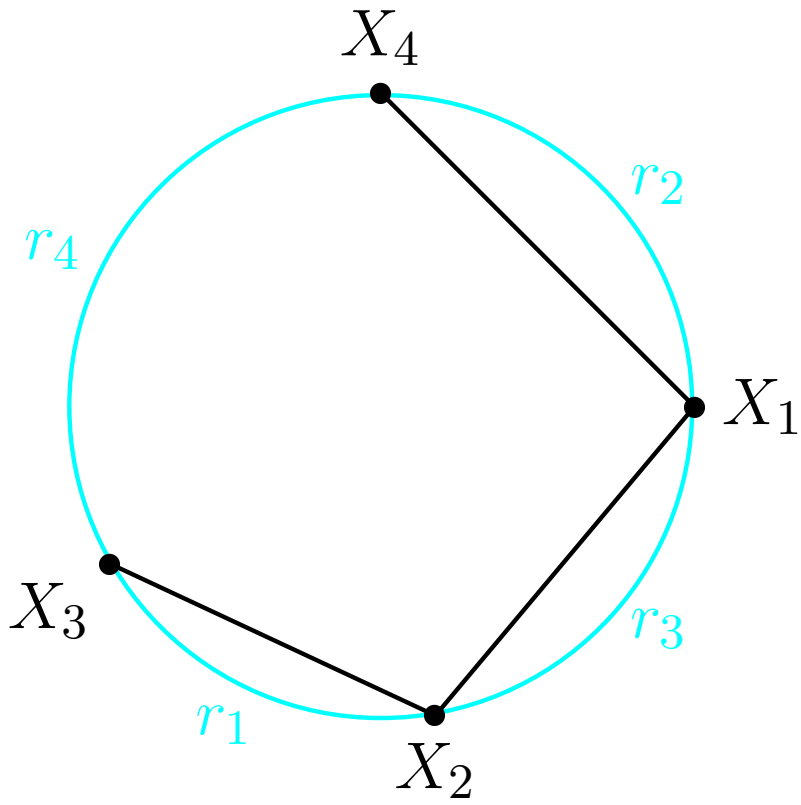


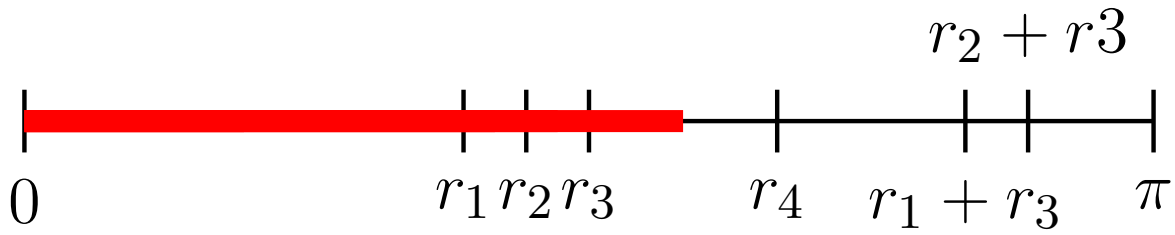
The Rips complex



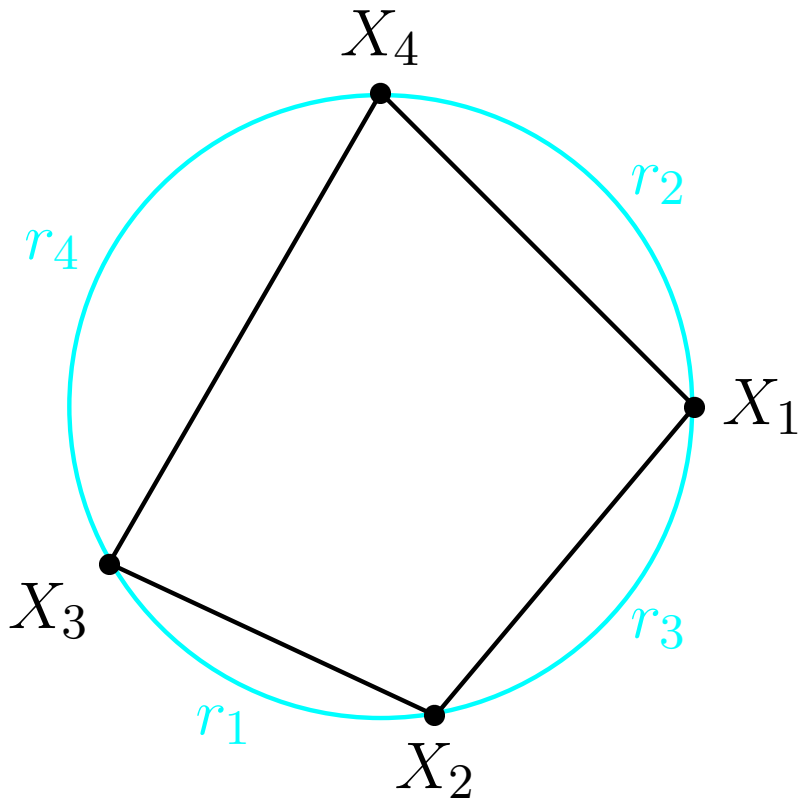


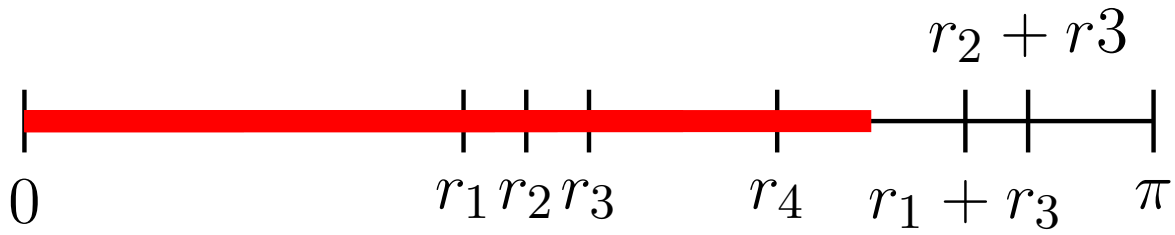
The Rips complex



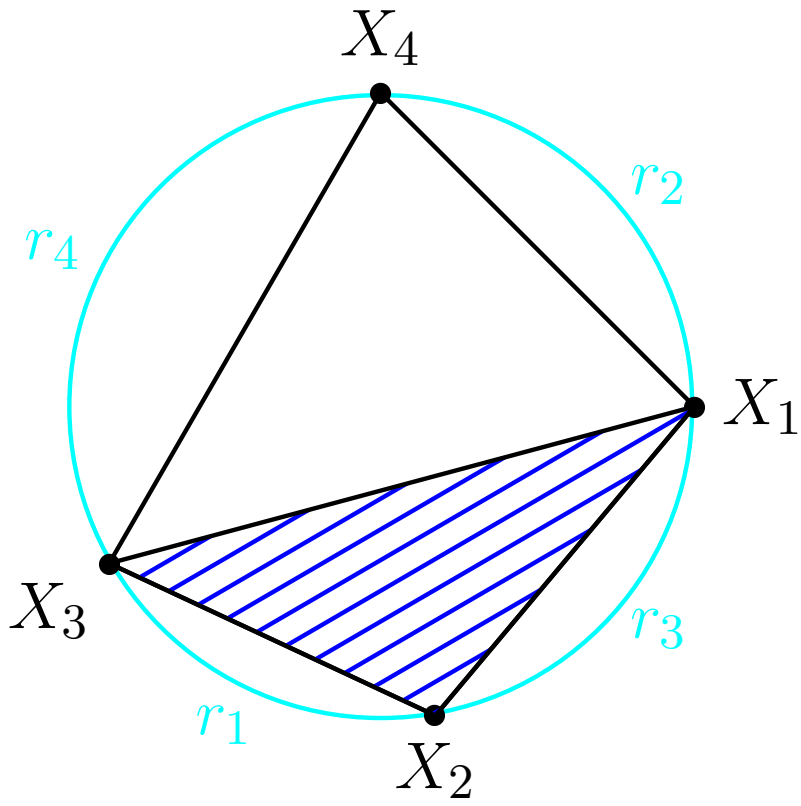


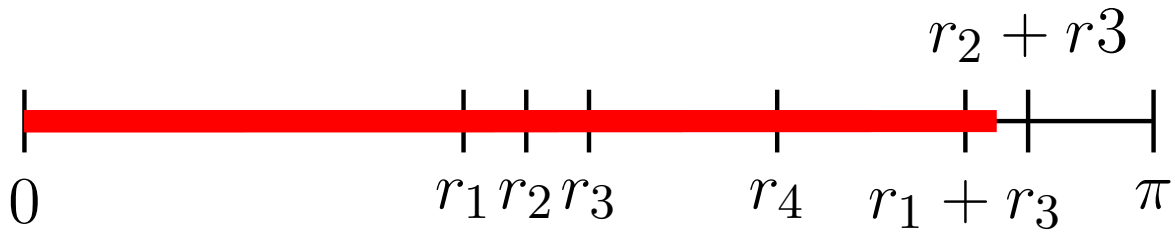
The Rips complex



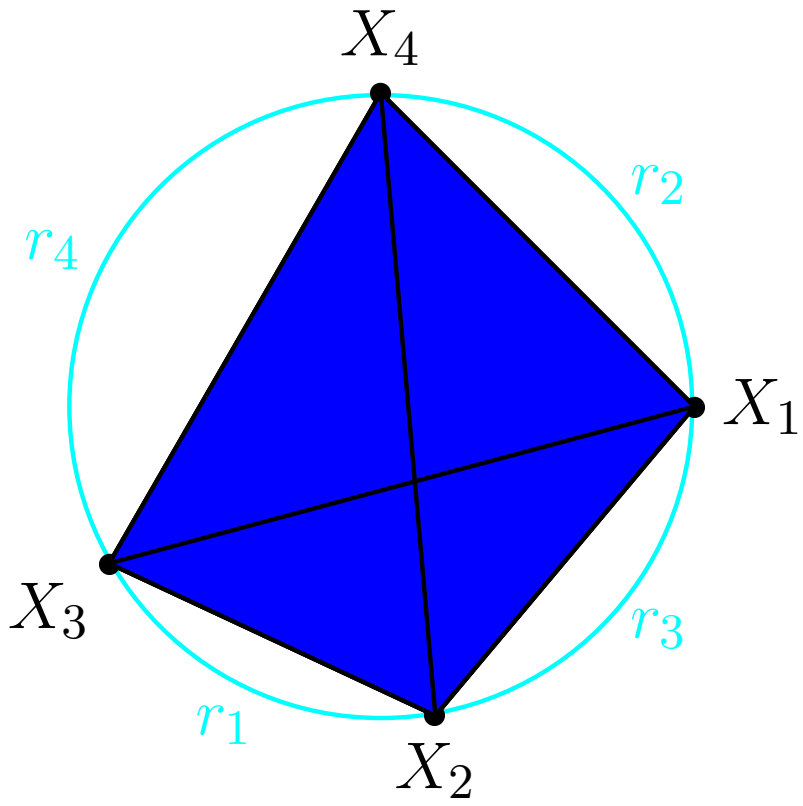


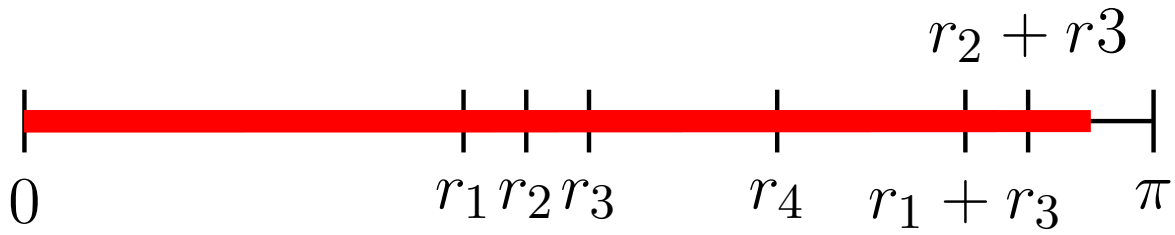
The Rips complex



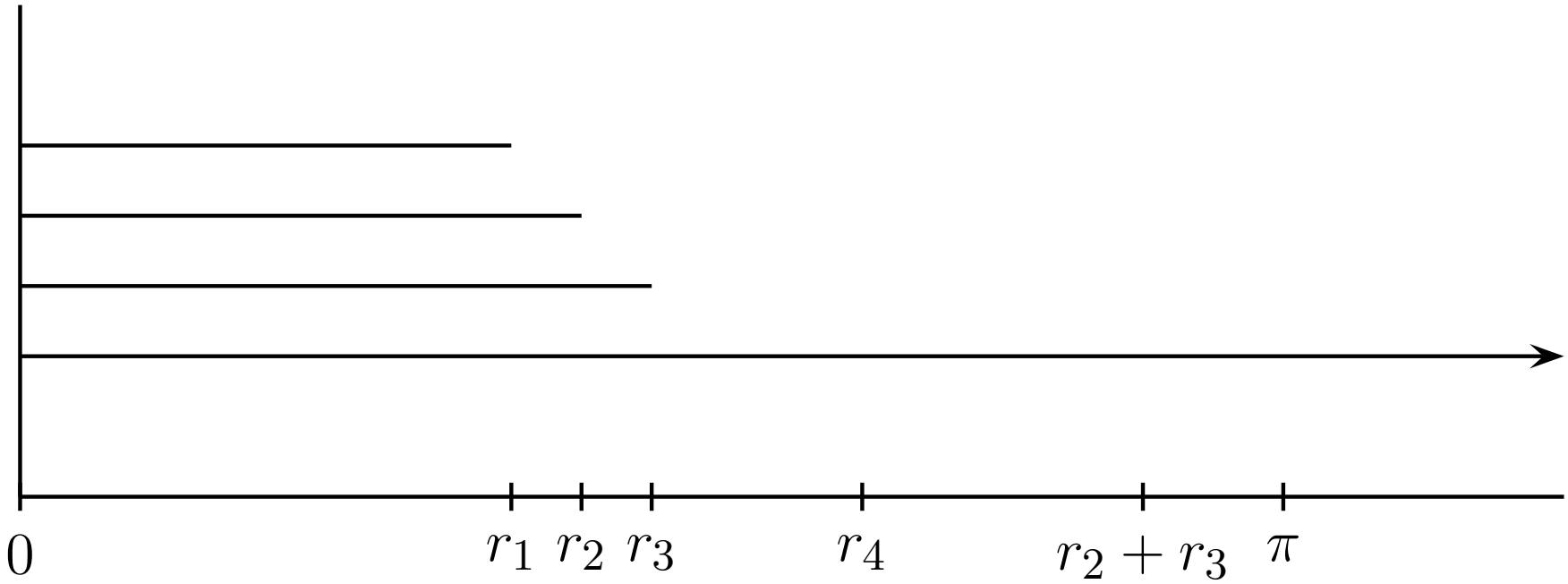


The Rips complex

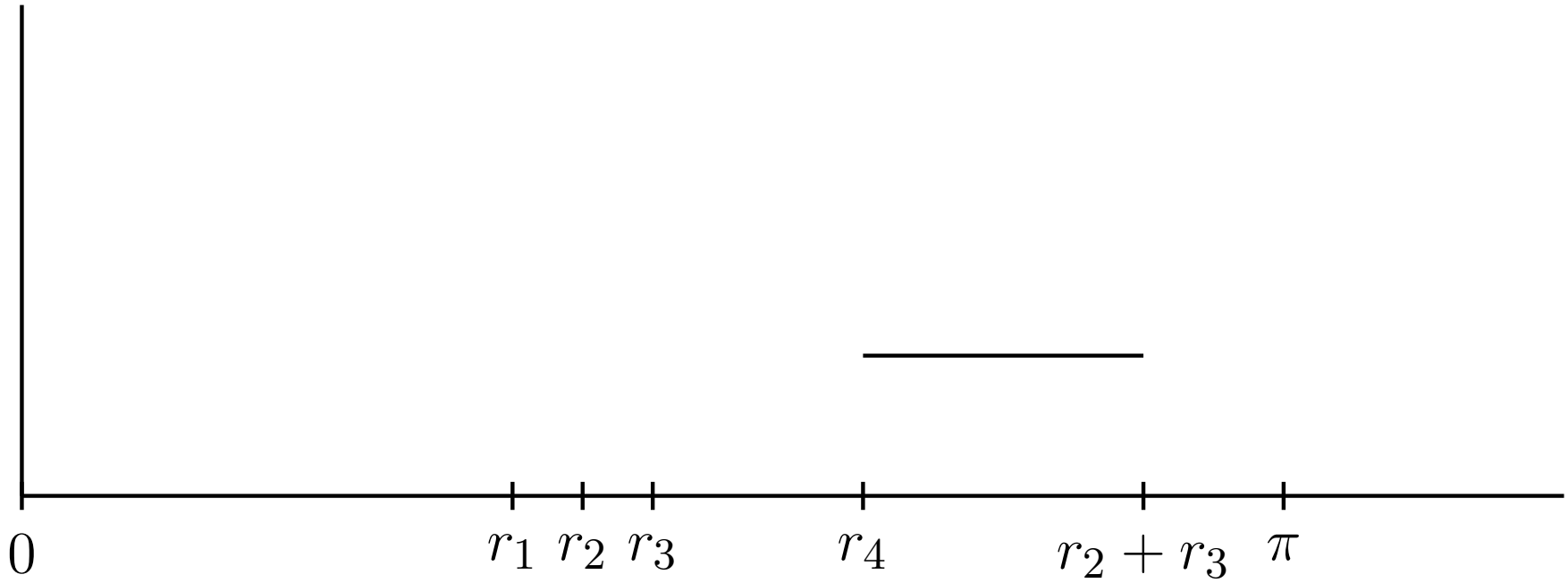




Betti 0-barcode

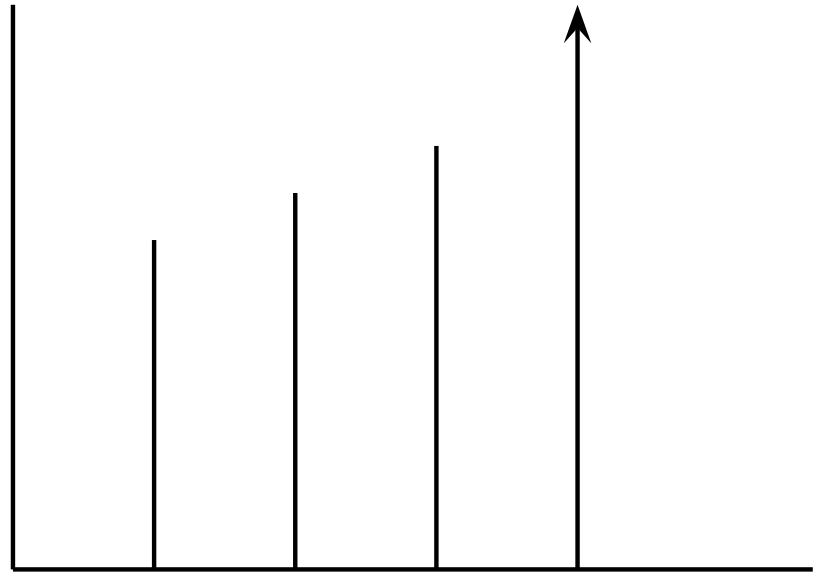


Betti 1-barcode



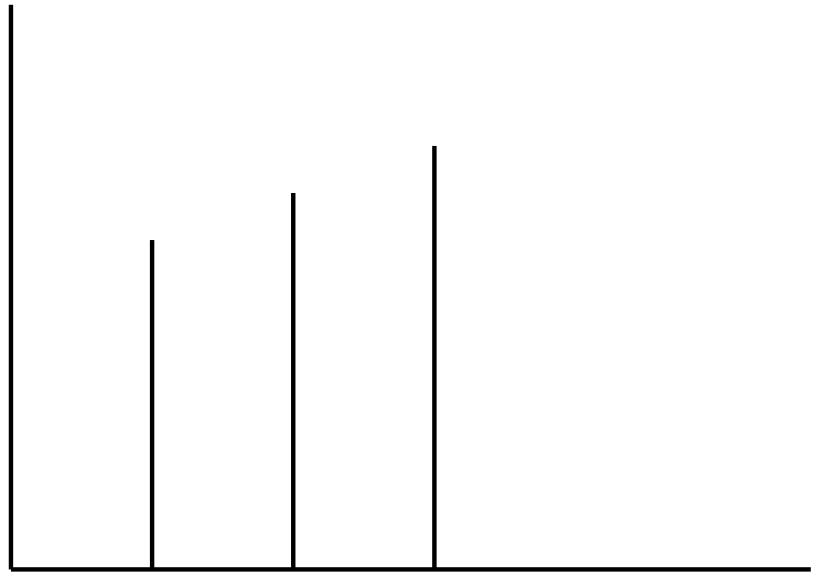
A different view

Consider the following view of the Betti-0 barcode.



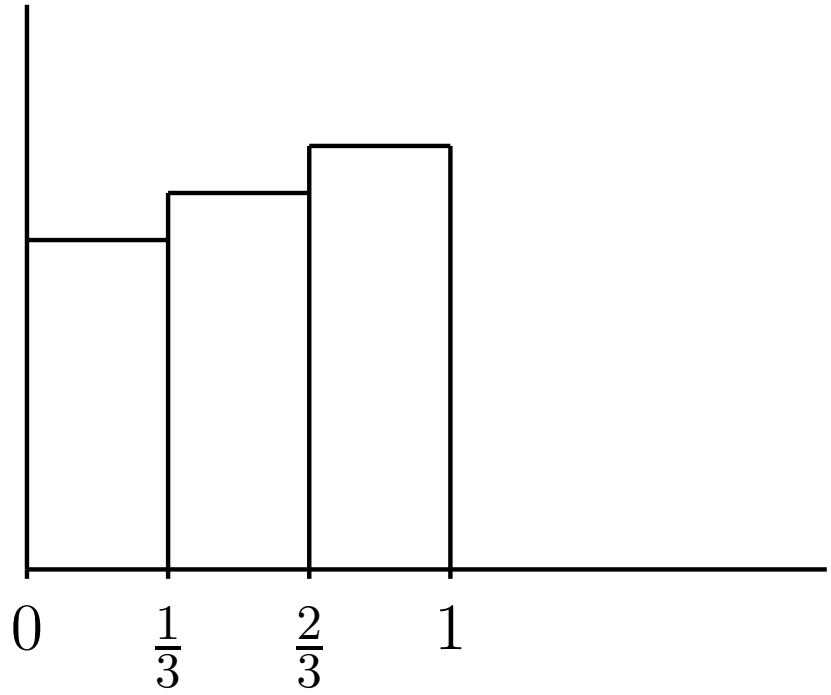
A different view

Consider the following view of the Betti-0 barcode.



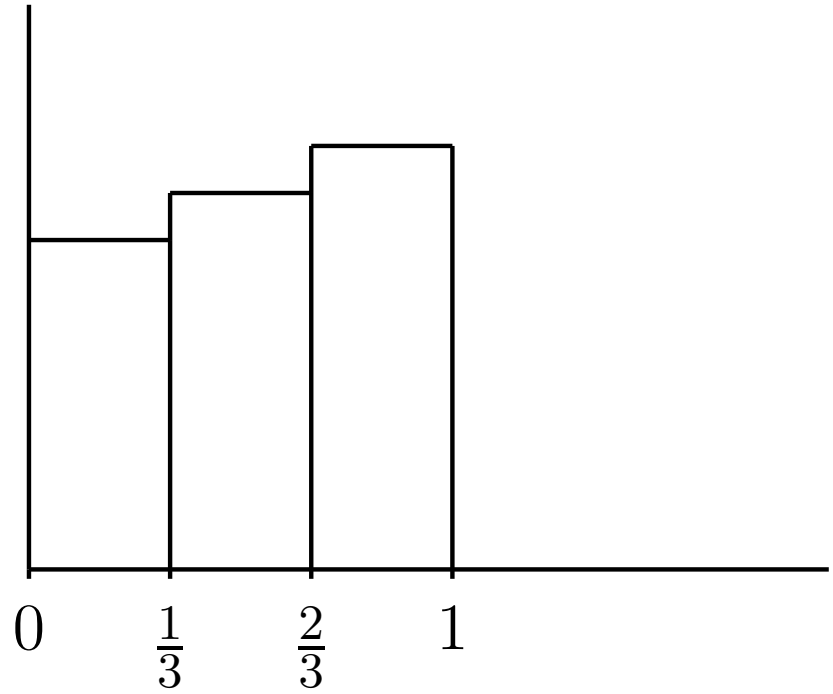
A different view

Consider the following view of the Betti-0 barcode.



A different view

Consider the following view of the Betti-0 barcode.



This gives us a function ${}_nB_0 : [0,1] \to \mathbf{R}_{\geq 0}$,
where ${}_nB_0(x)$ is the length on the k-th longest 0-homology interval, where $k = \lceil (n-1)x \rceil$.

Normalization

To see the effect of κ as $n \to \infty$, we normalize the above function.

$$ {}_n\bar{B}_0(x) = \frac{{}_nB_0(x)}{\int_0^1 {}_nB_0(t)dt} $$

We call this the Betti-0 function.

3. Spacings

Let $X = (X_1, \ldots X_n) \subset S^1$ be a sample.

3. Spacings

Let $X = (X_1, \ldots X_n) \subset S^1$ be a sample.
Choose α such that $X_1 = e^{2\pi i \alpha}$.

3. Spacings

Let $X = (X_1, \ldots X_n) \subset S^1$ be a sample.
Choose α such that $X_1 = e^{2\pi i \alpha}$.
For $2 \leq k \leq n$, choose $U_k \in [0, 1]$ such that

$$X_k = e^{2\pi i(\alpha + U_k)}.$$

3. Spacings

Let $X = (X_1, \ldots X_n) \subset S^1$ be a sample.
Choose α such that $X_1 = e^{2\pi i \alpha}$.
For $2 \leq k \leq n$, choose $U_k \in [0, 1]$ such that

$$X_k = e^{2\pi i(\alpha + U_k)}.$$

Order $\{U_2, \ldots U_n\}$ to obtain

$$0 < U_{n:1} < U_{n:2} < \ldots < U_{n:n-1} < 1.$$

Let $U_{n:0} = 0$ and $U_{n:n} = 1$.

3. Spacings

Let $X = (X_1, \ldots X_n) \subset S^1$ be a sample.
Choose α such that $X_1 = e^{2\pi i \alpha}$.
For $2 \leq k \leq n$, choose $U_k \in [0, 1]$ such that

$$X_k = e^{2\pi i(\alpha + U_k)}.$$

Order $\{U_2, \ldots U_n\}$ to obtain

$$0 < U_{n:1} < U_{n:2} < \ldots < U_{n:n-1} < 1.$$

Let $U_{n:0} = 0$ and $U_{n:n} = 1$.
For $1 \leq k \leq n$, let

$$S_k = U_{n:k} - U_{n:k-1}.$$

These are called the spacings.

Ordered Spacings

Order S:

$$0 < S_{n:1} < S_{n:2} < \ldots < S_{n:n} < 1$$

Ordered Spacings

Order S:

$$0 < S_{n:1} < S_{n:2} < \ldots < S_{n:n} < 1$$

Recall: ${}_nB_0(x)$ equals the length of the k-th longest homology interval, where $k = \lceil (n-1)x \rceil$.

Ordered Spacings

Order S:

$$0 < S_{n:1} < S_{n:2} < \ldots < S_{n:n} < 1$$

Recall: ${}_nB_0(x)$ equals the length of the k-th longest homology interval, where $k = \lceil (n-1)x \rceil$.

Therefore,

$${}_nB_0(x) \propto S_{n:\lceil (n-1)x \rceil}.$$

Uniform spacings

If X is sampled uniformly from S^1 then each U_k is uniform on $[0, 1]$ and $S = (S_1, \ldots, S_n)$ are called uniform spacings.

Assume X is sampled uniformly.

Uniform spacings

If X is sampled uniformly from S^1 then each U_k is uniform on $[0,1]$ and $S = (S_1, \ldots, S_n)$ are called uniform spacings.

Assume X is sampled uniformly.

Lemma: $S = (S_1, \ldots, S_n)$ has a constant density on the standard $(n-1)$-simplex Δ^{n-1}

$$S_k \geq 0, \quad \sum_{k=1}^{n} S_i = 1.$$

Expected spacings

For $1 \leq k \leq n$, $ES_{n:k}$ can be found be integrating over Δ^{n-1}.

Expected spacings

For $1 \leq k \leq n$, $ES_{n:k}$ can be found be integrating over Δ^{n-1}.

Proposition:

$$ES_{n:k} = \frac{1}{n} \sum_{j=n+1-k}^{n} \frac{1}{j}$$

Expected Betti-0 function

Recall that ${}_nB_0(x) \propto S_{n:\lceil (n-1)x \rceil}$.

Thus,

$$\begin{aligned} E({}_nB_0)(x) &\propto ES_{n:\lceil (n-1)x \rceil} \\ &= \frac{1}{n} \left(\sum_{j=n+1-\lceil (n-1)x \rceil}^{n} \frac{1}{j} \right) . \end{aligned}$$

Hence,

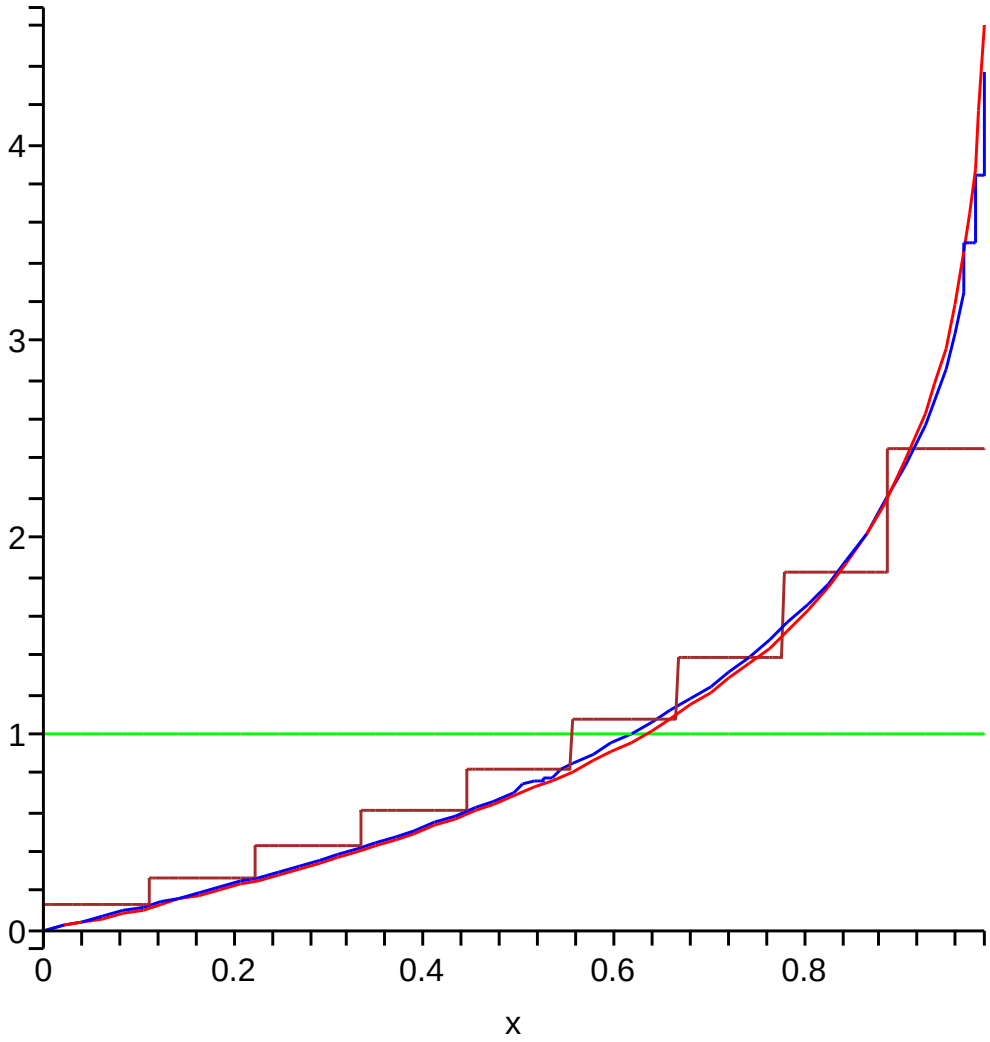
$$E({}_n\bar{B}_0)(x) = \frac{c_n}{n} \left(\sum_{j=n+1-\lceil (n-1)x \rceil}^{n} \frac{1}{j} \right)$$

Limiting Betti-0 function

Proposition[B-K]:

$$E({}_n\bar{B}_0)(x) \to -\log(1-x) \text{ as } n \to \infty.$$

Betti-0 Graphs



Summary

- We would like to be able to understand the persistent homology of PCD sampled from densities.
- To understand the limiting behavior of the Betti-0 barcode, it helps to consider the Betti-0 function.
- For densities on S^1 the persistent homology is determined by the spacings.
- For the uniform density we can explicitly calculate the expected persistent homology.
- The limiting Betti-0 function for the uniform density is $-\log(1-x)$.