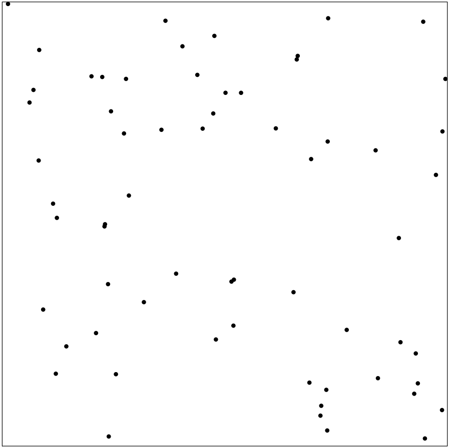
The accumulated persistence function, a new useful functional summary statistic for topological information, with an application to spatial point processes

Christophe A. N. Biscio and Jesper Møller, Aalborg University

May 26, 2016

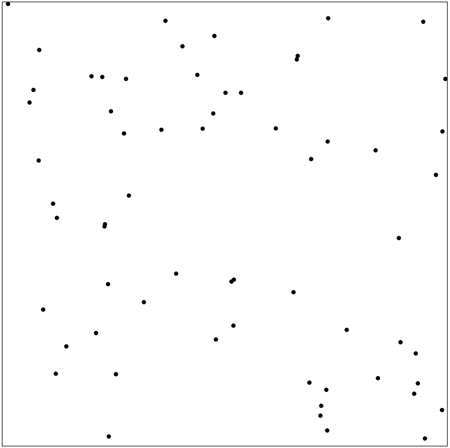
Persistent homology - Delaunay complex filtration

- Persistent homology recovers the tolopological information from a spatial point process **X**.



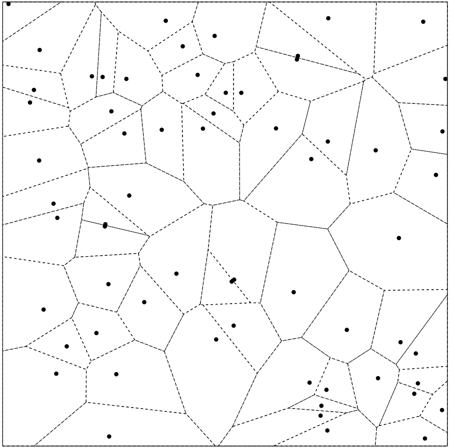
Persistent homology - Delaunay complex filtration

- Persistent homology recovers the tolopological information from a spatial point process $\mathbf{X}$.
- We construct a so called Delaunay complex filtration.



Persistent homology - Delaunay complex filtration

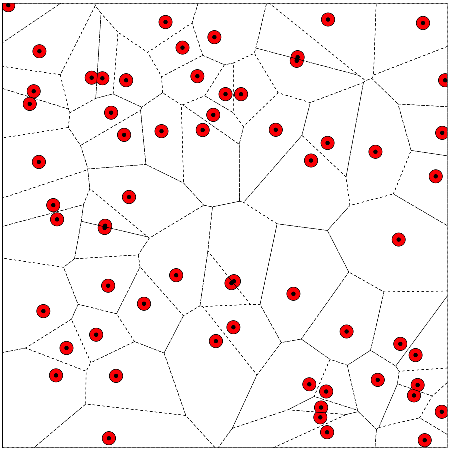
- Persistent homology recovers the tolopological information from a spatial point process **X**.
- We construct a so called Delaunay complex filtration.
- The union of balls of radius r intersected with the Voronoi tessellation.



- $r = 0$, a point $\sim$ a connected component.

Persistent homology - Delaunay complex filtration

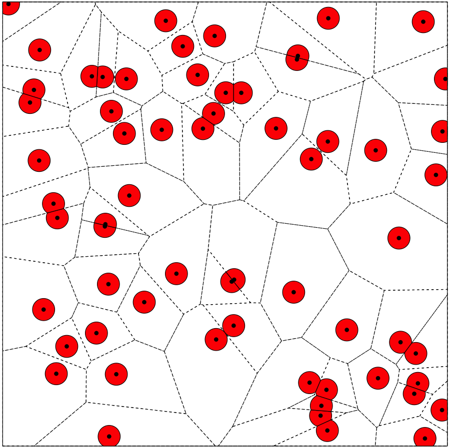
- Persistent homology recovers the tolopological information from a spatial point process $\mathbf{X}$.
- We construct a so called Delaunay complex filtration.
- The union of balls of radius r intersected with the Voronoi tessellation.



- $r = 0$, a point $\sim$ a connected component.

Persistent homology - Delaunay complex filtration

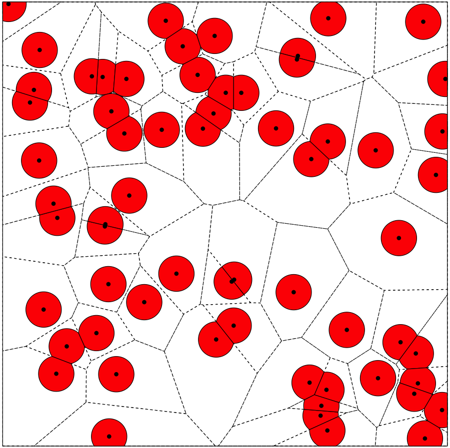
- Persistent homology recovers the tolopological information from a spatial point process $\mathbf{X}$.
- We construct a so called Delaunay complex filtration.
- The union of balls of radius r intersected with the Voronoi tessellation.



- $r = 0$, a point $\sim$ a connected component.

Persistent homology - Delaunay complex filtration

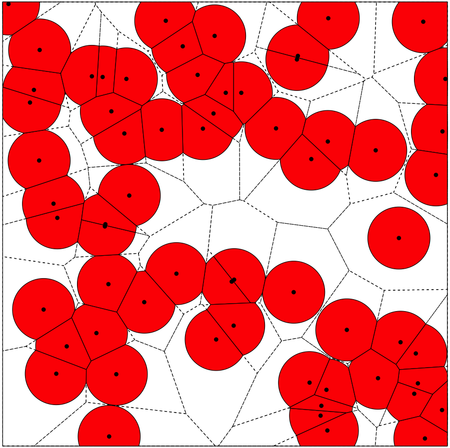
- Persistent homology recovers the tolopological information from a spatial point process $\mathbf{X}$.
- We construct a so called Delaunay complex filtration.
- The union of balls of radius r intersected with the Voronoi tessellation.



- $r = 0$, a point $\sim$ a connected component.
- Balls intersect $\sim$ connected components die.

Persistent homology - Delaunay complex filtration

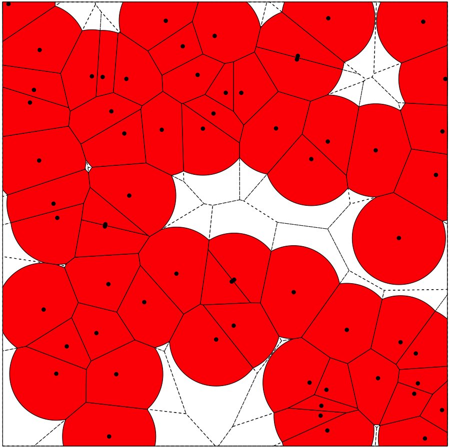
- Persistent homology recovers the tolopological information from a spatial point process $\mathbf{X}$.
- We construct a so called Delaunay complex filtration.
- The union of balls of radius r intersected with the Voronoi tessellation.



- $r = 0$, a point $\sim$ a connected component.
- Balls intersect $\sim$ connected components die.
- Holes may appear (birth).

Persistent homology - Delaunay complex filtration

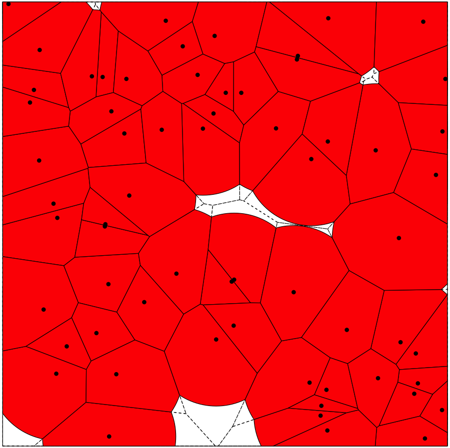
- Persistent homology recovers the tolopological information from a spatial point process $\mathbf{X}$.
- We construct a so called Delaunay complex filtration.
- The union of balls of radius r intersected with the Voronoi tessellation.



- $r = 0$, a point $\sim$ a connected component.
- Balls intersect $\sim$ connected components die.
- Holes may appear (birth).
- Holes may disappear (death).

Persistent homology - Delaunay complex filtration

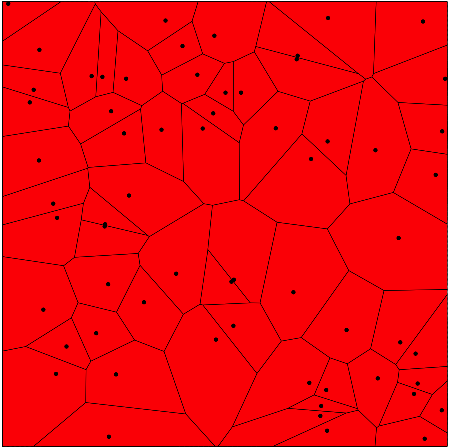
- Persistent homology recovers the tolopological information from a spatial point process $\mathbf{X}$.
- We construct a so called Delaunay complex filtration.
- The union of balls of radius r intersected with the Voronoi tessellation.



- $r = 0$, a point $\sim$ a connected component.
- Balls intersect $\sim$ connected components die.
- Holes may appear (birth).
- Holes may disappear (death).

Persistent homology - Delaunay complex filtration

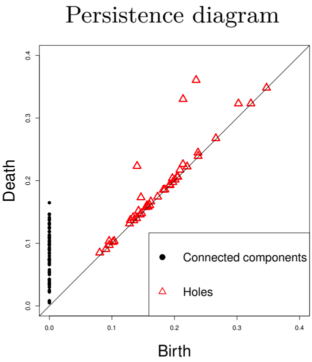
- Persistent homology recovers the tolopological information from a spatial point process $\mathbf{X}$.
- We construct a so called Delaunay complex filtration.
- The union of balls of radius r intersected with the Voronoi tessellation.



- $r = 0$, a point $\sim$ a connected component.
- Balls intersect $\sim$ connected components die.
- Holes may appear (birth).
- Holes may disappear (death).

Persistent homology - Persistence diagram

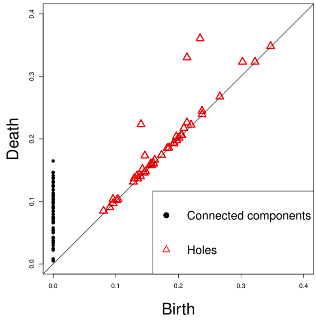
- A persistent diagram consists of points (b_i, d_i) representing as r varies a connected components (holes) appearing at $r = b_i$ (birth) and disappearing at $r = d_i$ (death),
- possibly with multiplicity c_i.



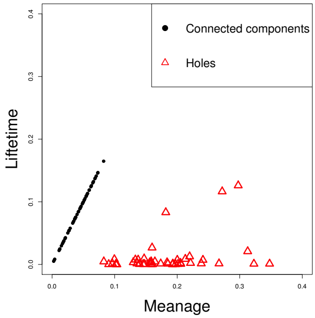
Persistent homology - Persistence diagram

- A persistent diagram consists of points (b_i, d_i) representing as r varies a connected components (holes) appearing at $r = b_i$ (birth) and disappearing at $r = d_i$ (death),
- possibly with multiplicity c_i.
- $(b_i, d_i) \leftrightarrow (m_i, l_i)$, where $m_i = \frac{b_i + d_i}{2}$ and $l_i = d_i - b_i$.

Persistence diagram



Rotated and rescaled persistence diagram



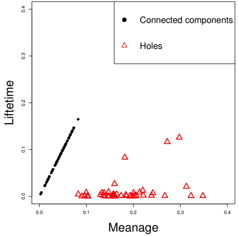
A new functional summary statistic

The **accumulated persistence function** (where $k = 0$ if connected components are considered, $k = 1$ if holes):

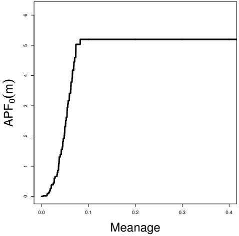
$$\mathrm{APF}_k(m) = \sum_i c_i l_i 1(m_i \leq m), \quad m \geq 0.$$

Example:

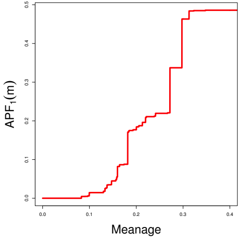
Rotated and rescaled persistent diagram



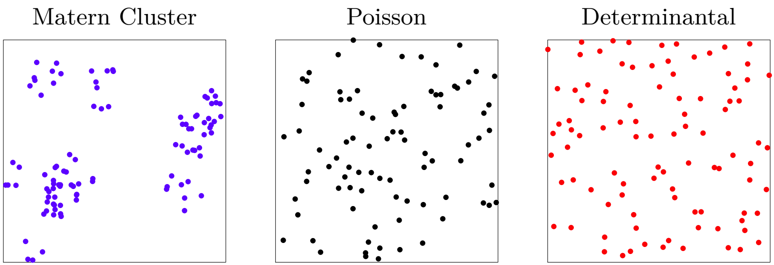
Connected components

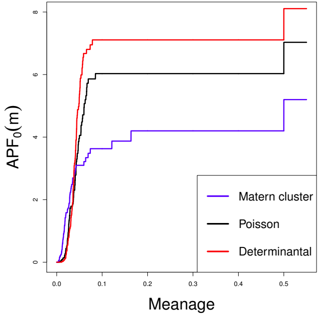


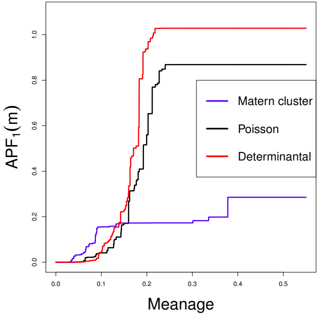
Holes



Inhibitive - Aggregative behaviour







Applications

- A single APF
  - Transfer confidence region from persistence diagram Fasy *et al.*, 2014 and Chazal *et al.*, 2014
  - Extreme rank envelope
- A sample of APFs
  - Functional boxplot
  - Confidence region for the mean of APFs
- Two or more samples of APFs
  - Two sample problem
  - Clustering
  - Supervised classification

Framework

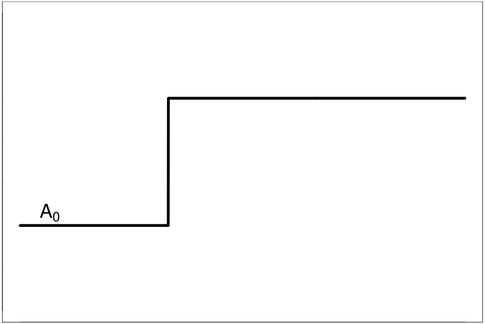
- Given a realization of $\mathbf{X}_0$ ("the data") and a model for $\mathbf{X}_0$: assess goodness of fit?
- Simulate r IID copies of $\mathbf{X}_0$: $\mathbf{X}_1, \ldots, \mathbf{X}_r$.
- NB: $(\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_r)$ is **exchangeable**.
- $A_0, A_1, \ldots A_r$ the corresponding APFs.
- **Aim:** construct confidence regions for $A_0(m)$ with $0 \leq m \leq T$ (a given number) under $\mathcal{H}_0$: $(A_0, A_1, \ldots, A_r)$ is exchangeable.

The k-th bounding curves, $k = 1, 2 \dots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\dots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\dots,r}{}^k A_i(m), \quad 0 \le m \le T.$$

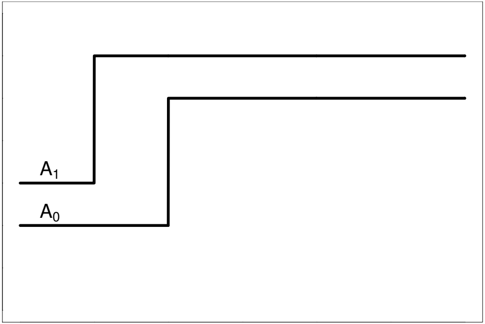
The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \le m \le T.$$



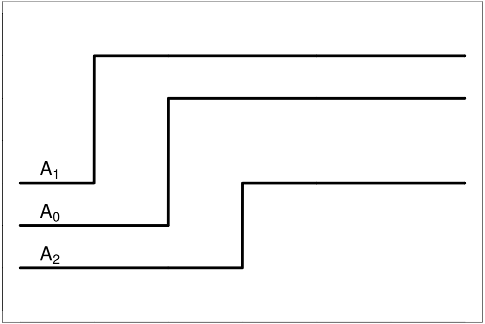
The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \le m \le T.$$



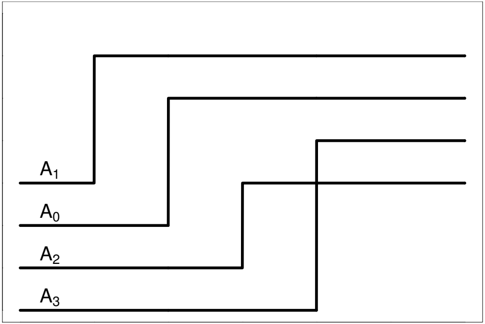
The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\mathrm{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\mathrm{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \le m \le T.$$



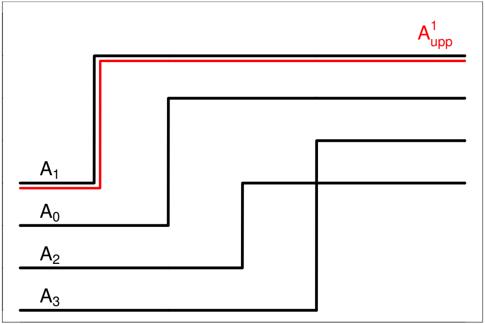
The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \leq m \leq T.$$



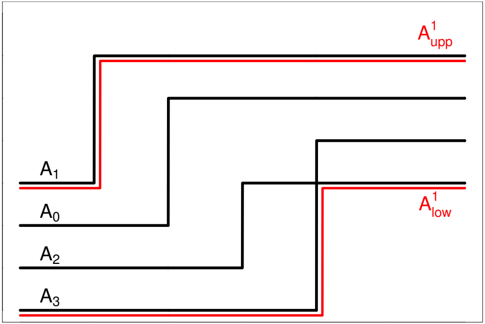
The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \le m \le T.$$



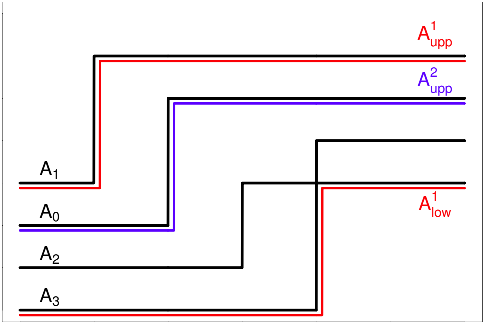
The k-th bounding curves, $k = 1, 2 \dots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\dots,r}{}^{k} A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\dots,r}{}^{k} A_i(m), \quad 0 \le m \le T.$$



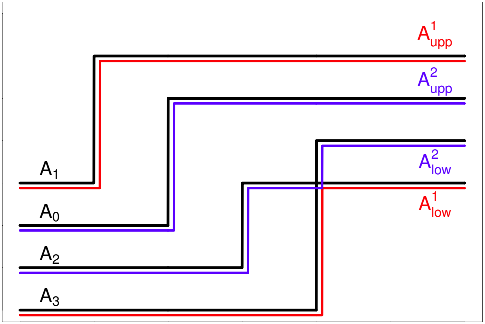
The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \leq m \leq T.$$



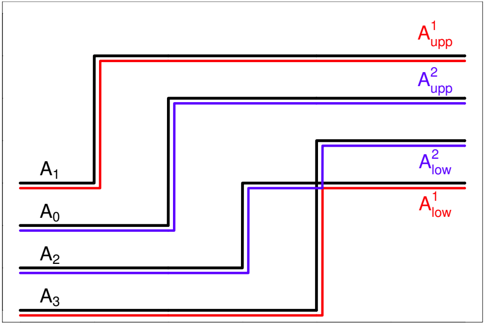
The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \leq m \leq T.$$



The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\mathrm{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\mathrm{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \leq m \leq T.$$

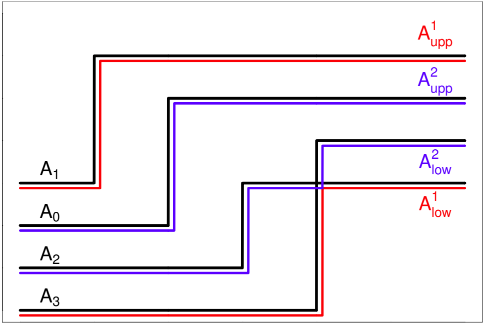


The i-th extreme rank, $i = 0, \ldots, r$, of A_i

$$R_i = \max\left\{k : \; A^k_{\mathrm{low}}(m) \leq A_i(m) \leq A^k_{\mathrm{upp}}(m), \quad \text{for all } m \in [0, T]\right\}.$$

The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \le m \le T.$$



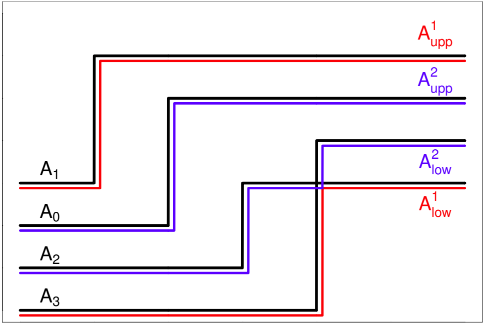
- $R_0 = 2$

The i-th extreme rank, $i = 0, \ldots, r$, of A_i

$$R_i = \max \left\{ k : \; A^k_{\text{low}}(m) \le A_i(m) \le A^k_{\text{upp}}(m), \quad \text{for all } m \in [0, T] \right\}.$$

The k-th bounding curves, $k = 1, 2\ldots,$

$$A^k_{\rm low}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\rm upp}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \le m \le T.$$



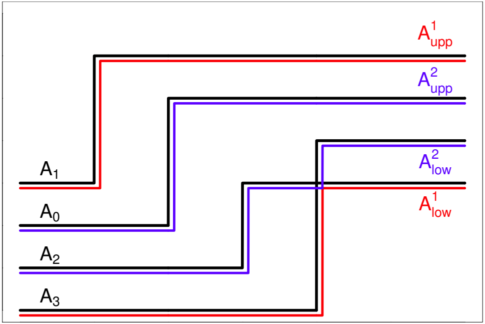
- $R_0 = 2$
- $R_1 = 1$

The i-th extreme rank, $i = 0, \ldots, r$, of A_i

$$R_i = \max\left\{k : \; A^k_{\rm low}(m) \le A_i(m) \le A^k_{\rm upp}(m), \quad \text{for all } m \in [0, T]\right\}.$$

The k-th bounding curves, $k = 1, 2 \dots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\dots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\dots,r}{}^k A_i(m), \quad 0 \le m \le T.$$



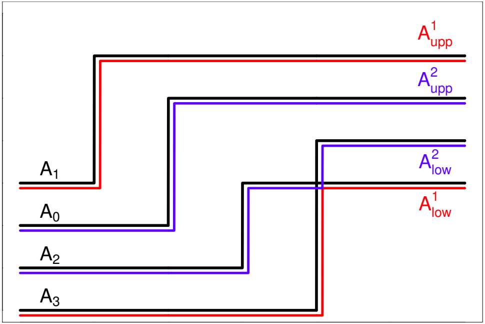
- $R_0 = 2$
- $R_1 = 1$
- $R_2 = 1$

The i-th extreme rank, $i = 0, \dots, r$, of A_i

$$R_i = \max\left\{k : \; A^k_{\text{low}}(m) \le A_i(m) \le A^k_{\text{upp}}(m), \quad \text{for all } m \in [0, T]\right\}.$$

The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \le m \le T.$$



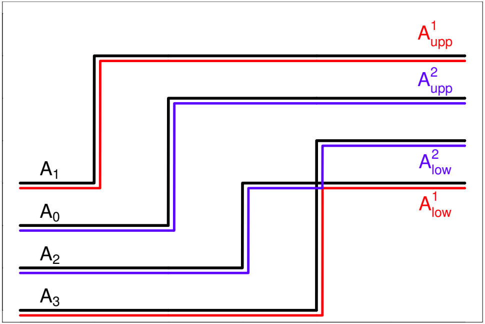
- $R_0 = 2$
- $R_1 = 1$
- $R_2 = 1$
- $R_3 = 1$

The i-th extreme rank, $i = 0, \ldots, r$, of A_i

$$R_i = \max \left\{ k : \; A^k_{\text{low}}(m) \le A_i(m) \le A^k_{\text{upp}}(m), \quad \text{for all } m \in [0, T] \right\}.$$

The k-th bounding curves, $k = 1, 2 \ldots,$

$$A^k_{\text{low}}(m) = \min_{i=0,\ldots,r}{}^k A_i(m) \quad \text{and} \quad A^k_{\text{upp}}(m) = \max_{i=0,\ldots,r}{}^k A_i(m), \quad 0 \le m \le T.$$



- $R_0 = 2$
- $R_1 = 1$
- $R_2 = 1$
- $R_3 = 1$

The i-th extreme rank, $i = 0, \ldots, r$, of A_i

$$R_i = \max\left\{k : \; A^k_{\text{low}}(m) \le A_i(m) \le A^k_{\text{upp}}(m), \quad \text{for all } m \in [0, T]\right\}.$$

The larger R_i is, the "deeper" or "more central" A_i is among $A_0, \ldots, A_r$.

Extreme rank envelope

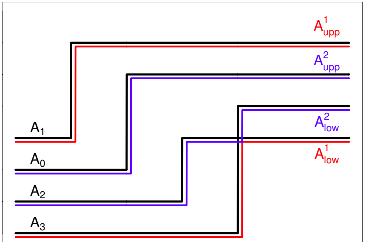
Proposition (Myllymäki *et al., 2015*)

Given $\alpha \in (0,1)$, define

$$k_\alpha = \max \left\{ k \ : \ \frac{1}{r+1} \sum_{i=0}^{r} 1(R_i < k) \leq \alpha \right\}.$$

Under $\mathcal{H}_0$, with at least probability greater or equal than $1-\alpha$,

$$A_{\text{low}}^{k_\alpha}(m) \leq A_0(m) \leq A_{\text{upp}}^{k_\alpha}(m) \quad \text{for all } m \in [0, T].$$

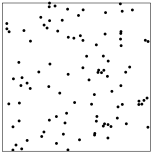


- $\sim$ conservative extreme rank envelope test at level α.

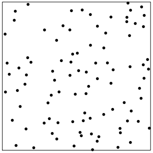
Test for CSR

- Suppose $\mathbf{X}_0$ is modelled as a Poisson point process on $[0,1]^2$ with **known** intensity ρ: $\mathcal{P}(\rho, [0,1]^2)$.
- Simulate $r = 2500$ IID copies of $\mathcal{P}(\rho, [0,1]^2)$.
- Compute extreme rank envelope test at level 5% when the true model for $\mathbf{X}_0$ is either:

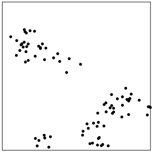
CSR - Poisson process



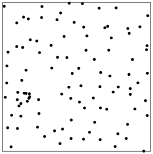
Inhibitive - DPP



Aggregation Matérn cluster



Baddeley-Silverman cell process



Test for CSR

Percentage of simulated point patterns rejected by the 95%-extreme rank envelope test.

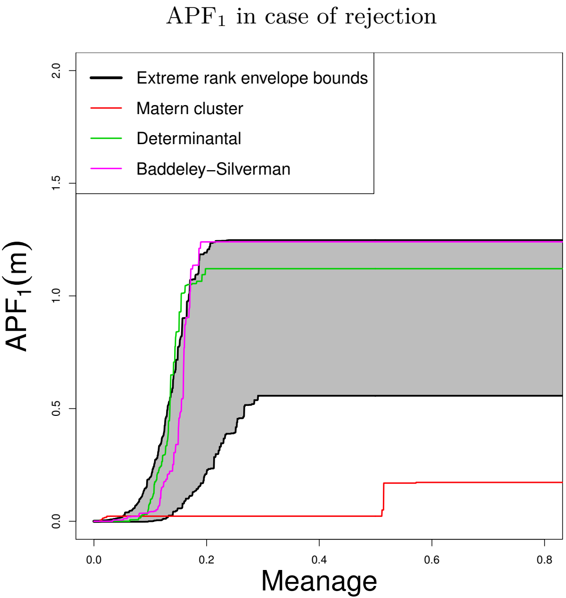
	Poisson		Determinantal		Matérn cluster		Baddeley-Silverman	
	$\rho = 100$	$\rho = 300$	$\rho = 100$	$\rho = 300$	$\rho = 100$	$\rho = 300$	$\rho = 100$	$\rho = 300$
APF_0	5	3.5	83.5	100	100	100	50.5	87.5
APF_1	5	4	14.5	46.5	100	100	52	94.5
K	3.2	2.8	98.8	100	100	100	51	48
F	2.2	1.8	31	57	99.8	100	58	100

- Decent detection for Baddeley-Silverman cell process.
- Good detection for inhibitive model but only when considered the connected components.
- Excellent detection for cluster model.
- The power increases with the number of points.

Extreme rank envelope, $k = 0$

Short lifetime features matters.

Extreme rank envelope, $k = 1$



Conclusions:

- We introduced a new functional summary statistics.
- We successfully use it in different situations.
- We focused on an extreme rank envelope test and the corresponding plot.
- This plot can be very informative for suggesting alternative models.

Perspectives:

- Define new spatial point process models based on their persistence diagrams.

Thank you for your attention.