Geometrical sensitivity indices using homological constructions on manifolds

Maikol Solís

@maikol_solis

maikol-solis

joint work with:

Alberto Hernández & Ronald Zúñiga

UNIVERSIDAD DE COSTA RICA

@UniversidadCR

CIMPA Centro de Investigación en Matemática Pura y Aplicada

@CIMPAUCR

ESCUELA DE MATEMÁTICA
UNIVERSIDAD DE COSTA RICA

@EscuelaMateUCR

ConectaR
Del 24 al 26 de enero de 2019
San José, Costa Rica

San José, Costa Rica

January 24th, 2019

Outline

Motivation

Assume that $\mathbf{X} = (X_1, \ldots, X_p) \in \mathbb{R}^p$ produces the output $Y \in \mathbb{R}$ linked by the model

$$Y = \psi(X_1, \ldots, X_p). \tag{1}$$

The function ψ could be known or unknown.

Generally is a complex function.

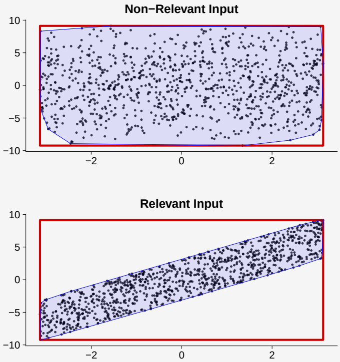
Questions

- ¿How sensitive is each input with respect to Y?
- ¿How much changes the output Y if there exist a perturbation in the inputs?

The geometric correlation

If X_i is geometrically uncorrelated with respect to Y then blue area is similar to the red box

Otherwise, the blue area does not cover all the domain.



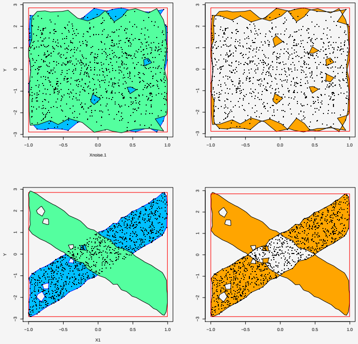
The geometric correlation is

$$\rho_i^{\text{Geom}} = 1 - \frac{\text{Object Area for variable i}}{\text{Box Area for variable i}}.$$

The geometric sensitivity

If X_i is geometrical irrelevant or insensible with respect to Y then a symmetric pattern will appear on its reflection.

Otherwise, the manifolds and its reflection does not overlap completely.

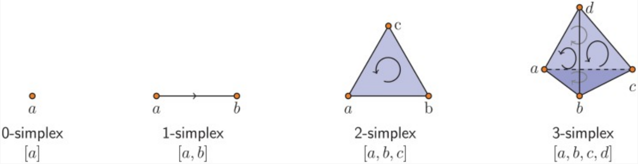


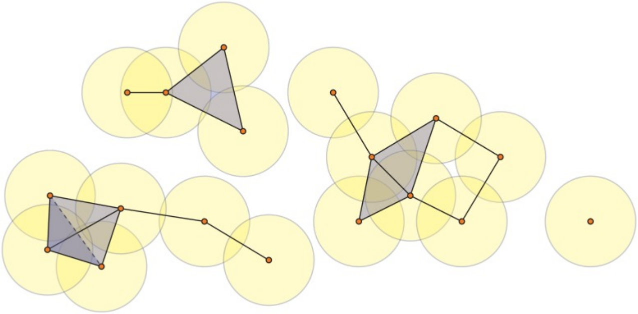
The sensitivity index is

$$S_i^{\text{Geom}} = \frac{\text{Symmetric Difference Area for variable i}}{2 \times \text{Manifold Area for variable i}}.$$

Outline

Simplicials and the Vietoris-Rip Complex



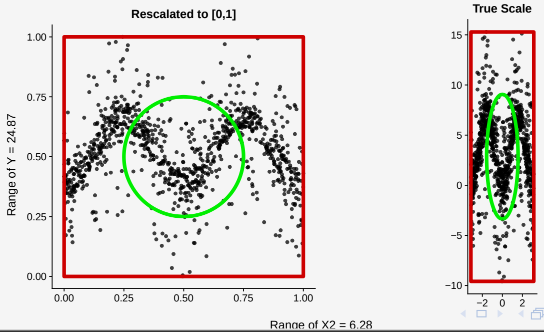


C. M. Topaz, L. Ziegelmeier, and T. Halverson. 2015. Topological Data Analysis of Biological Aggregation Models:1–26

The 1-simplex

To preserve the dimensions of the x and y axis, we rescaled all the data-points to the square $[0,1] \times [0,1]$.

1-simplex: With the rescaled variables ($[0,1] \times [0,1]$), construct pairwise the edges of the data points with distance less to r with `igraph`.



The 2-simplex

Using the work of Zomorodian[1], we estimate the 2-simplex as the cliques of dimension 3 of a graph using the package igraph,

```
clq <- igraph::cliques(graphBase, min = 3, max = 3)
TwoSimplex <- do.call("rbind", clq)
```

1. A. Zomorodian. 2010. Fast construction of the Vietoris-Rips complex. *Computers & Graphics* 34 (3): 263–271.

Transforming the 2-complex to Polygons

Each 2-simplex is a Triangle which is stored into a `POLYGON` of the package `sf` (simple features):

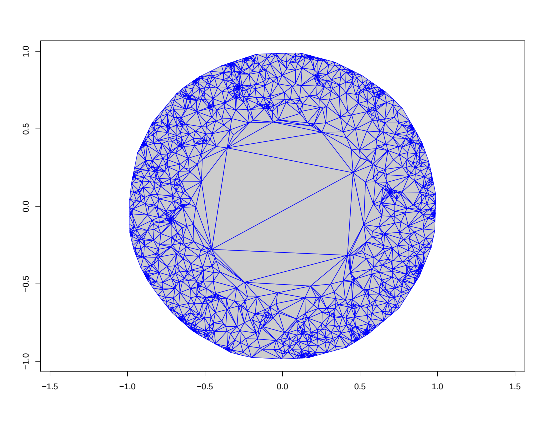
```
p <- sf::st_polygon(list(Triangle))
```

With all the Polygons we form a list `l` and then create a `multipolygon` object:

```
mp <- sf::st_multipolygon(l)
mp_union <- sf::st_union(mp)
```

Outline

Delaunay triangulation



J. R. Shewchuk. 2002. Delaunay refinement algorithms for triangular mesh generation. *Computational Geometry* 22 (1-3): 21–74

Implementation (Ongoing work)

The package `sfdct` has ported all the features of `RTriangle` to `sf`.

The `q` parameter means that each angle is no less than 30º.

```
triangulation <-
    sfdct::ct_triangulate(sf::st_union(datapoints),
                          q = 30)
```

We remove all the triangles with large areas compared to others and call this object `mp_union`.

Outline

Index estimation

We estimate the Area of the all Triangles and the box contained all the points.

```
bb <- sf::st_make_grid(x = mp_union, n = 1)
Manifold.Area <- sf::st_area(mp_union)
Box.Area <- sf::st_area(bb)
```

Finally, the geometric correlation is:

```
rho_i = 1 -Manifold.Area / Box.Area
```

Outline

We estimate the symmetric reflection over the x-axis using an affine transformation. Then we add 2 times the centroid over the y-axis to locate the reflected polygon over the original one.

```
reflectionx <-  matrix(c(1, 0, 0, -1), 2, 2)
mp_reflectionx <- mp_union * reflectionx +
                  c(0, 2 * sf::st_centroid(mp)[2])
mp_sym_difference <-
  sf::st_sym_difference(mp_union, mp_reflectiony)
Symmetric.Diff.Area <- sf::st_area(mp_sym_difference)
```

Finally, the geometric correlation is:

```
S_i = Symmetric.Diff.Area/(2 * Manifold.Area)
```

Outline

Ilustrative models

Linear: $Y = 2X_1 + X_2$ and X_3, X_4, X_5 variables of uniform pure noise.

Circle with hole:

$$\begin{cases} X_1 = r\cos(\theta) \\ Y = r\sin(\theta) \end{cases}$$

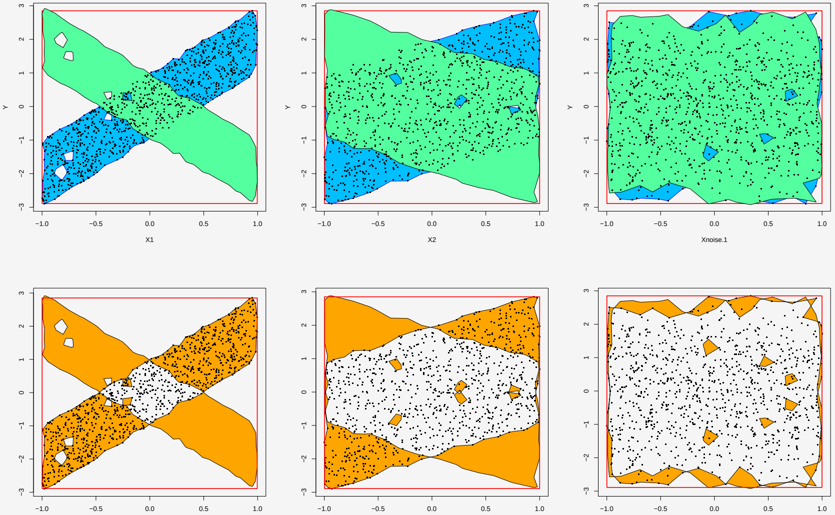
with $r \sim \mathrm{Unif}(0.5, 1)$ and $\theta \sim \mathrm{Unif}(0, 2\pi)$ random. The variable X_2 is pure noise.

Ishigami:

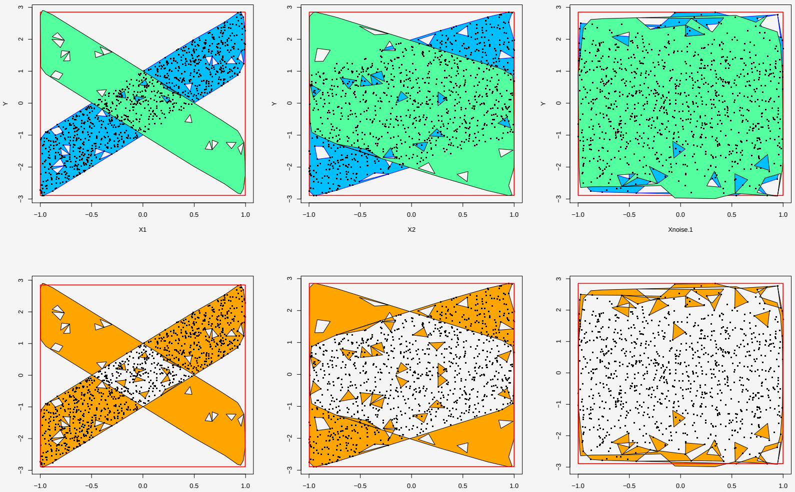
$$Y = \sin X_1 + 7\ \sin^2 X_2 + 0.1\ X_3^4 \sin X_1$$

where $X_i \sim \mathrm{Uniform}(-\pi, \pi)$ for $i = 1, 2, 3$.

Linear model with VR construction



Linear model with Delaunay construction



Linear model results

VR construction:

	Max. Radius	Manifold	Box	Correlation	Sensitivity
X1	0.08	3.68	11.46	0.68	0.77
X2	0.11	7.54	11.47	0.34	0.28
X3	0.12	10.08	11.47	0.12	0.06

Delaunay construction:

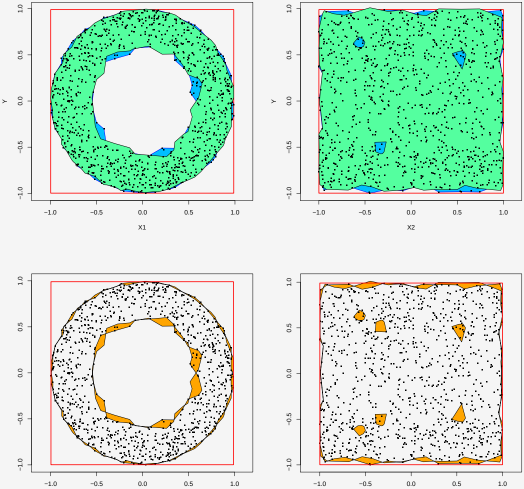
	Min. Triangle*	Manifold	Box	Correlation	Sensitivity
X1	0.00	3.73	11.46	0.67	0.76
X2	0.00	7.47	11.47	0.35	0.28
X3	0.00	10.07	11.47	0.12	0.07

* The areas are smaller than 2-digits precision.

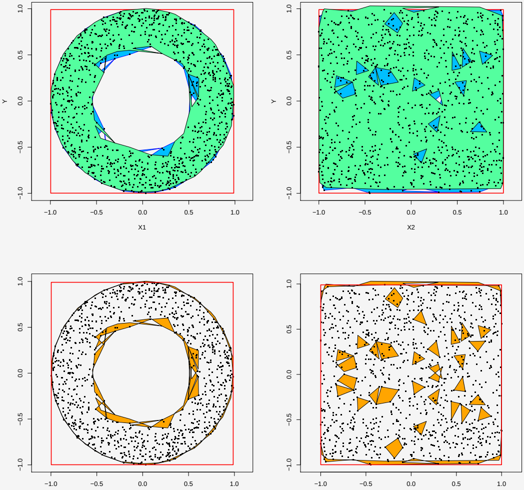
True Sobol indices:

$$X_1 = 0.80 \quad X_2 = 0.20 \quad X_3 = 0.00 \quad X_4 = 0.00 \quad X_5 = 0.00.$$

Circle with hole with VR construction



Circle with hole with Delaunay construction



Circle with hole results

VR construction:

	Max. Radius	Manifold	Box	Correlation	Sensitivity
X1	0.10	2.01	3.94	0.49	0.06
X2	0.13	3.75	3.98	0.06	0.03

Delaunay construction:

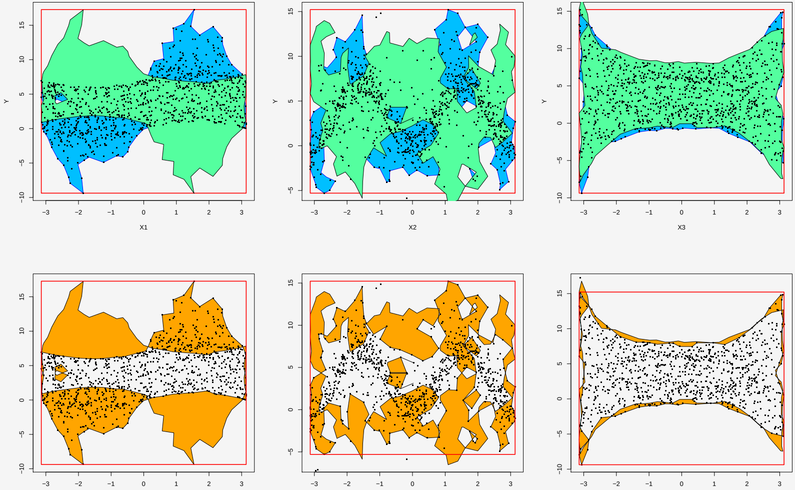
	Min. Triangle*	Manifold	Box	Correlation	Sensitivity
X1	0.00	2.07	3.94	0.47	0.05
X2	0.00	3.71	3.98	0.07	0.07

* The areas are smaller than 2-digits precision.

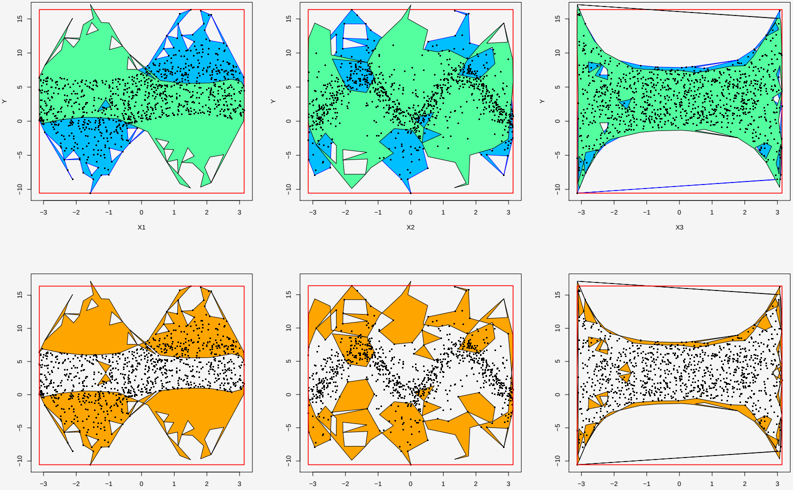
True Sobol indices:

$$X_1 = 0.00 \qquad X_2 = 0.00$$

Ishigami with VR construction



Ishigami with Delaunay construction



Ishigami results

VR construction:

	Max. Radius	Manifold	Box	Correlation	Sensitivity
X1	0.09	67.77	167.25	0.59	0.49
X2	0.08	64.73	128.51	0.50	0.43
X3	0.09	70.72	154.22	0.54	0.06

Delaunay construction:

	Min. Triangle*	Manifold	Box	Correlation	Sensitivity
X1	0.00	68.09	168.80	0.60	0.48
X2	0.00	91.17	169.07	0.46	0.33
X3	0.00	73.17	168.79	0.57	0.09

* The areas are smaller than 2-digits precision.

True Sobol indices:

$$X_1 = 0.31 \qquad X_2 = 0.44 \qquad X_3 = 0.00$$

Summary and future work

- We could capture geometric structures of the projection of each variable.
- Our method match the classic methods.
- We conjecture the method could recognize structured and non-structured noisy variables.

Future work:

- Submit this package to CRAN.
- Estimate efficiently the neighborhood graph (less than $\mathcal{O}(n^2)$).
- Study further the optimal delaunay construction to create smoother manifolds.
- Try to link these results with the classic statistic theory.
- Estimate sensitivity indices in higher dimensions (hard).

Maikol Solís

@maikol_solis

maikol-solis

joint work with:

Alberto Hernández & Ronald Zúñiga

Preprints

Hernández, Alberto, Maikol Solís, and Ronald Zúñiga. "Geometrical correlation indices using homological constructions on manifolds" Submitted to be published. (2018).

Hernández, Alberto, Maikol Solís, and Ronald Zúñiga. "Geometrical sensitivity indices through symmetric reflections" On preparation. (2019).