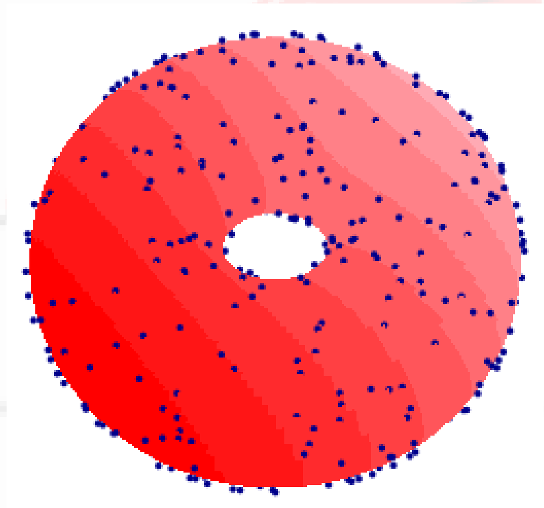
Simulation of PCD* on parametrized manifolds
Simulation methods

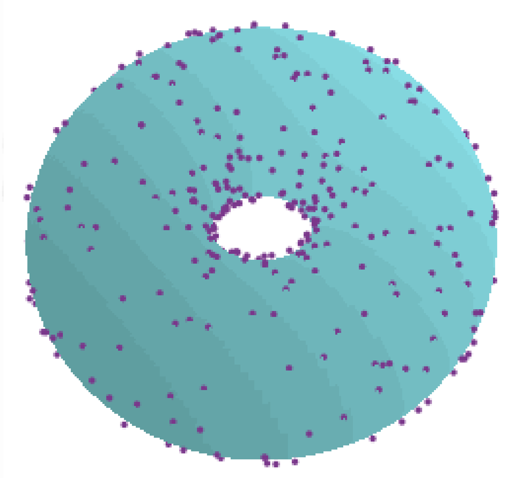
47th Annual John H. Barrett Memorial Lectures
University of Tennessee, Knoxville, May 1-3, 2017
Yair Hernández, Gilberto Flores
Universidad de Guanajuato-CIMAT

*Point Cloud Data

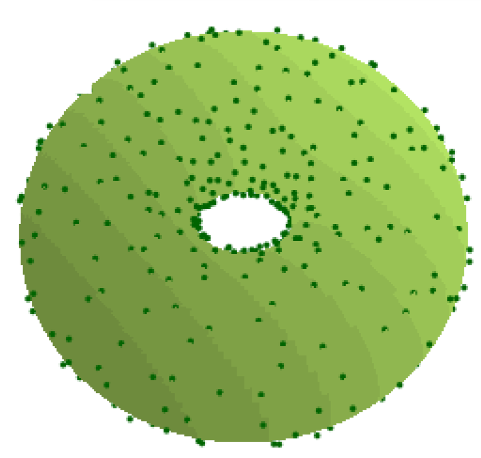
How to sprinkle a donut?



Useful for testing TDA algorithms



How to sample "uniformly" from a (parametrized) manifold $\mathbb{M}$?



How does the distribution matter?

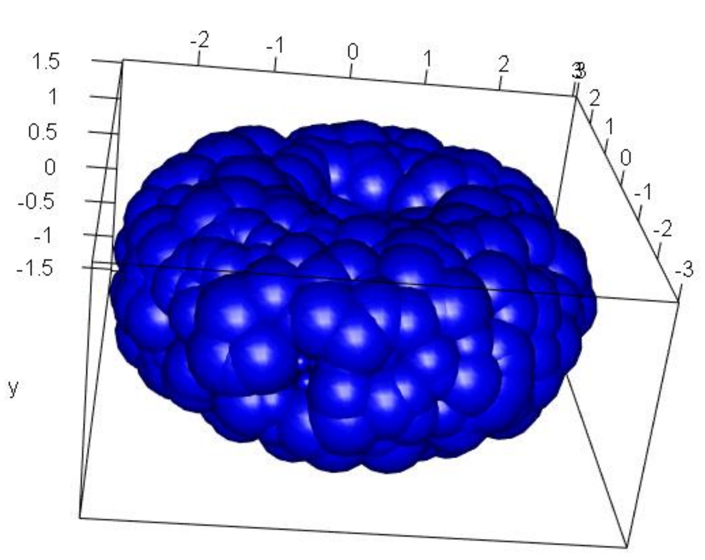
What, exactly, does "uniform" mean?

- Uniform with respect to a measure (sets with same measure have same probability)
- Length, Area, Volume, …

$A \subset \mathbb{R}^n$ — Volume of unitary (m-1)-sphere

$$\mathcal{H}^m(A) = \lim_{\delta \to \infty} \inf_{\substack{A \subseteq \bigcup B_i, \\ \operatorname{diam}(B_i) \le \delta}} \sum \alpha_m \left(\frac{\operatorname{diam}(B_i)}{2} \right)^m$$

Countable cover of A



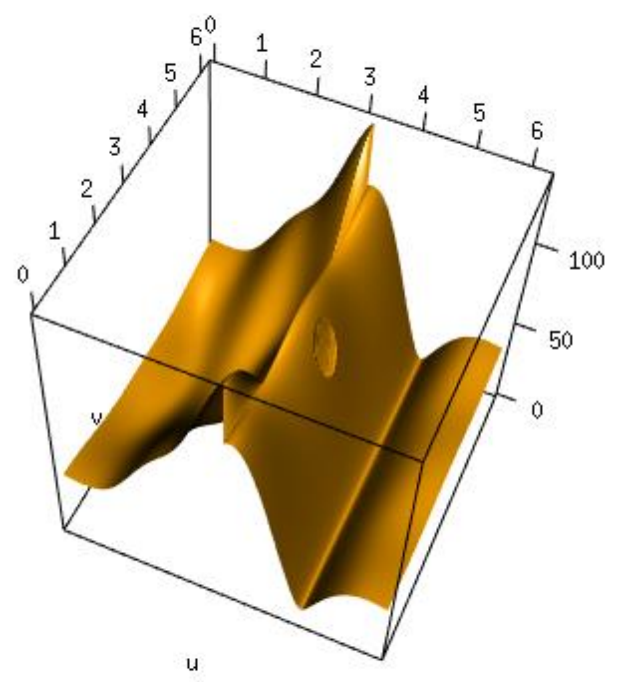
Torus covered by 2-spheres

Method by Diaconis et al. [2]

Sprinkling uniformly (1)

Based on Area Formula:

$\mathbb{M}$ parametrized by f, $x = f(\theta)$



m-dimensional Jacobian of f; Lebesgue measure

$$\int_A g(f(x)) J_m f(x) \lambda^m(dx)$$

$$= \int_{\mathbb{R}^n} g(y) N(f|_A, y) \mathcal{H}^m(dy)$$

Cardinality of the fiber of y; Hausdorff measure

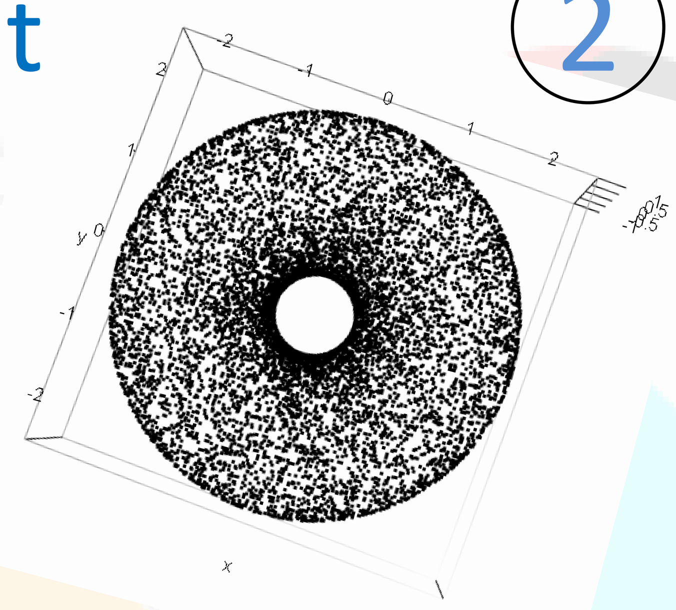
If $g = \frac{1}{\mathcal{H}^m(\mathbb{M})}$ we can sample $\mathcal{H}^m$-uniformly, sampling from $\frac{J_m f}{\mathcal{H}^m(\mathbb{M})}$ on the domain of f.

Torus is only example in [2], already implemented in TDA R library.

A simple and natural method

Sprinkling simply, but non-uniformly (2)

Sampling uniformly from the domain of the parametrization



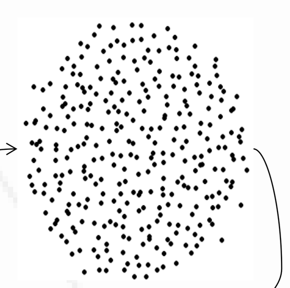
- Tempting to use
- Eventually covers the manifold
- Linear complexity on sample size
- If $J_m f$ is constant we are sampling $\mathcal{H}^m$-uniformly

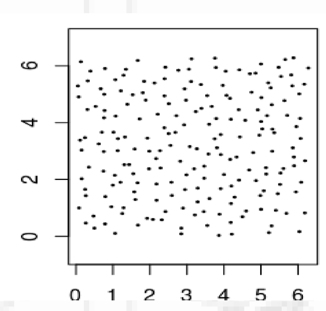
Torus: Regions with highest concentration have highest Gaussian curvature.

A useful method for TDA

Sprinkling with repulsion (3)

Our proposal: Use eigenvalues of a random matrix with i.i.d. entries





- Slow
- Based on Circular Law for random matrices [1]

(Distance between points matters)

Asymptotic behavior

Application of concentration inequality in Fasy et al. [3]

Bottleneck distance

d_A, function distance to closed subset A

Hausdorff distance

n points from a distribution on $\mathbb{M}$

$$\rho(x,t) = \frac{\mathbb{P}(B_x(t/2))}{t^d} \qquad \rho(t) = \inf_{x \in \mathbb{M}} \rho(x,t)$$

$$\mathbb{P}\left(W_\infty\left(\mathrm{dgm}(d_{S_n}), \mathrm{dgm}(d_{\mathbb{M}})\right) > t\right) \le \mathbb{P}(H(S_n, \mathbb{M}) > t) \le \frac{2^d}{\rho(t/2)t^d} e^{-n\rho(t)t^d}$$

Persistence diagrams

Stability theorem

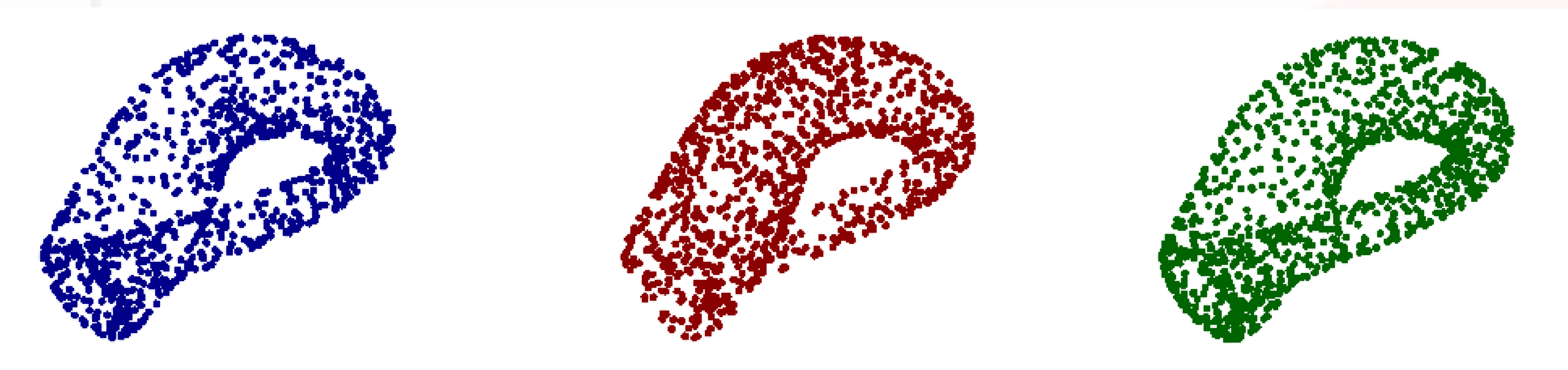
Asumptions about how small it can be (bounded away from zero)

Asymptotic covering with $\mathcal{H}^m$-uniform distribution

Examples

Klein Bottle

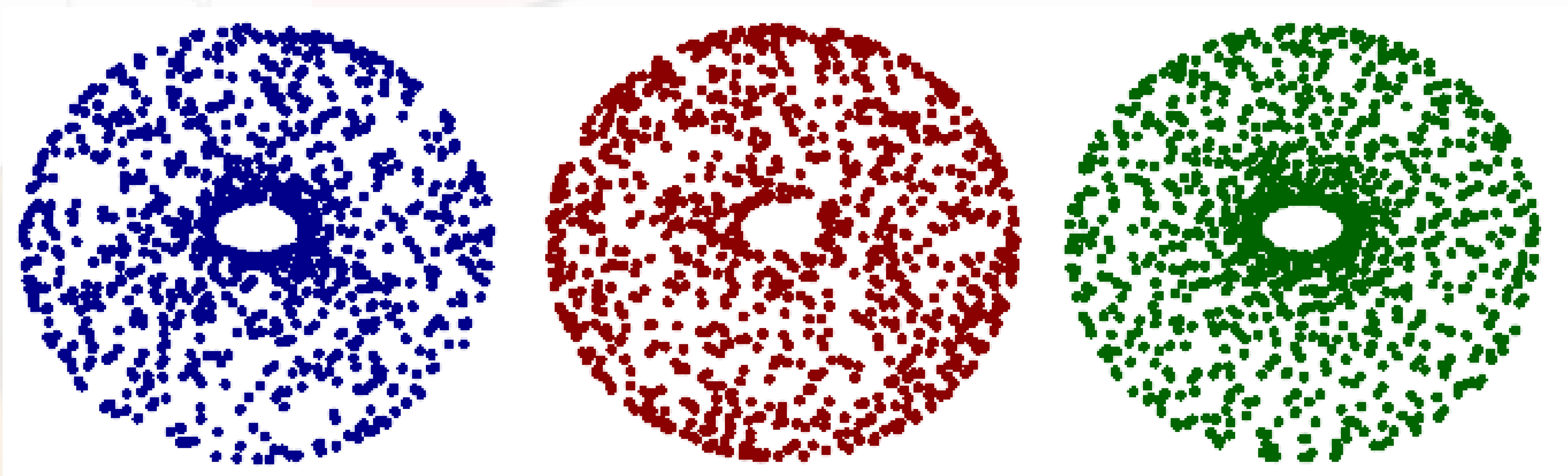
(Using a parametrization for the Klein Bottle in [4])



Uniform on domain | $\mathcal{H}^m$-uniform | Repulsion

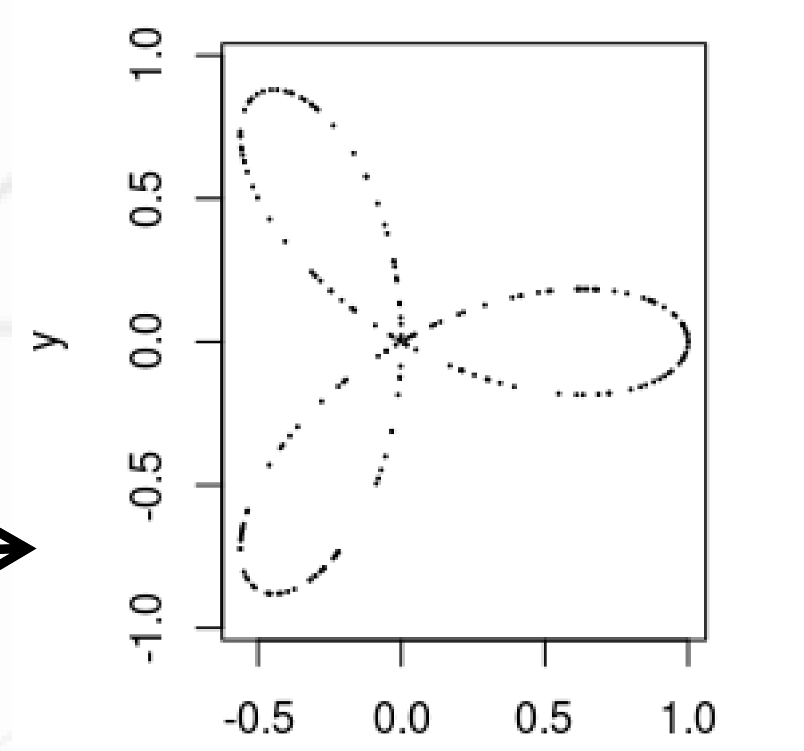
2-Torus

Uniform on domain | $\mathcal{H}^m$-uniform | Repulsion

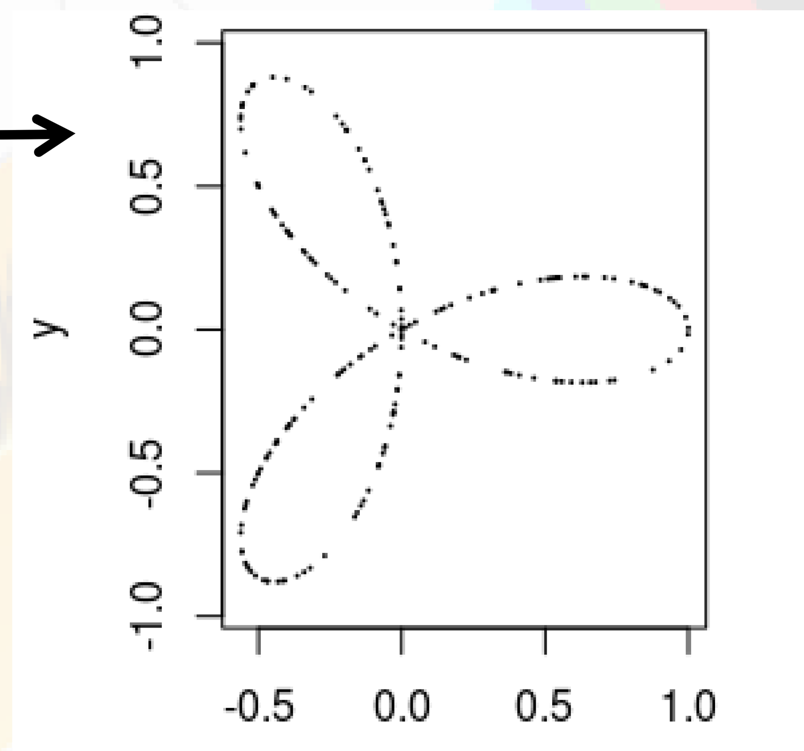


Rose

Uniform on domain



$\mathcal{H}^m$-uniform

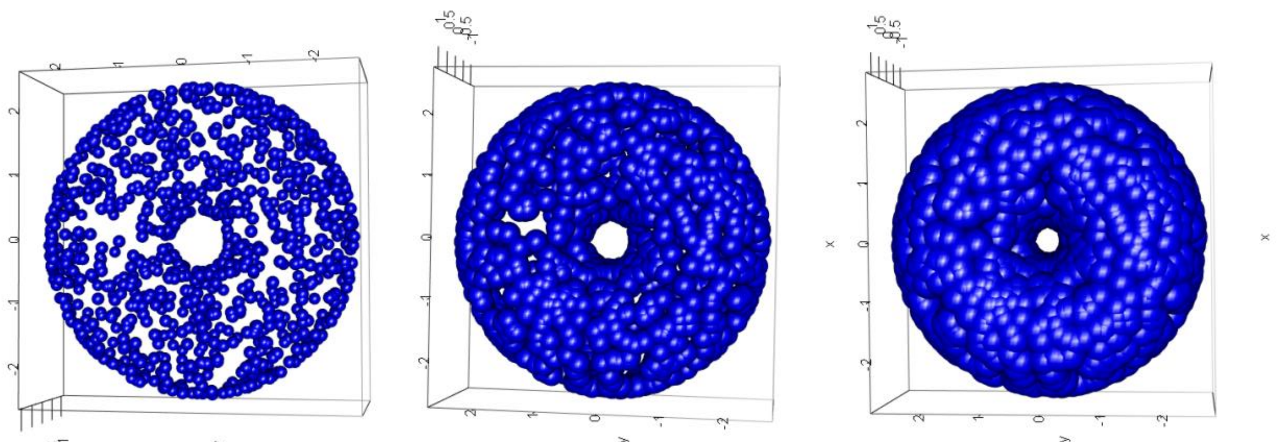


Simulation of PCD* on parametrized manifolds

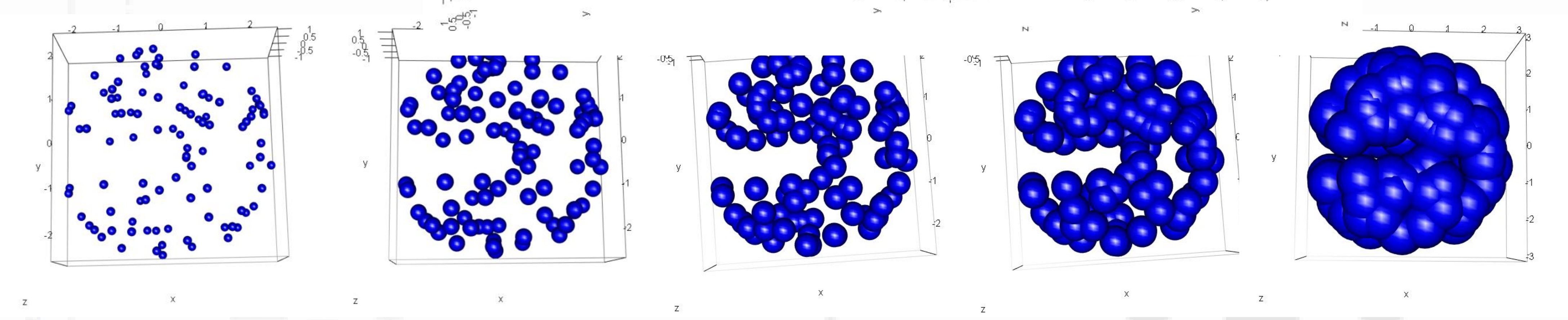
Topological Data Analysis

Some TDA techniques

- Real function defined on a manifold or simplicial complex (g)
- Computation of topological features in each upper level or sublevel sets (Betti numbers)
- Summarizes critical points (dgm(g))

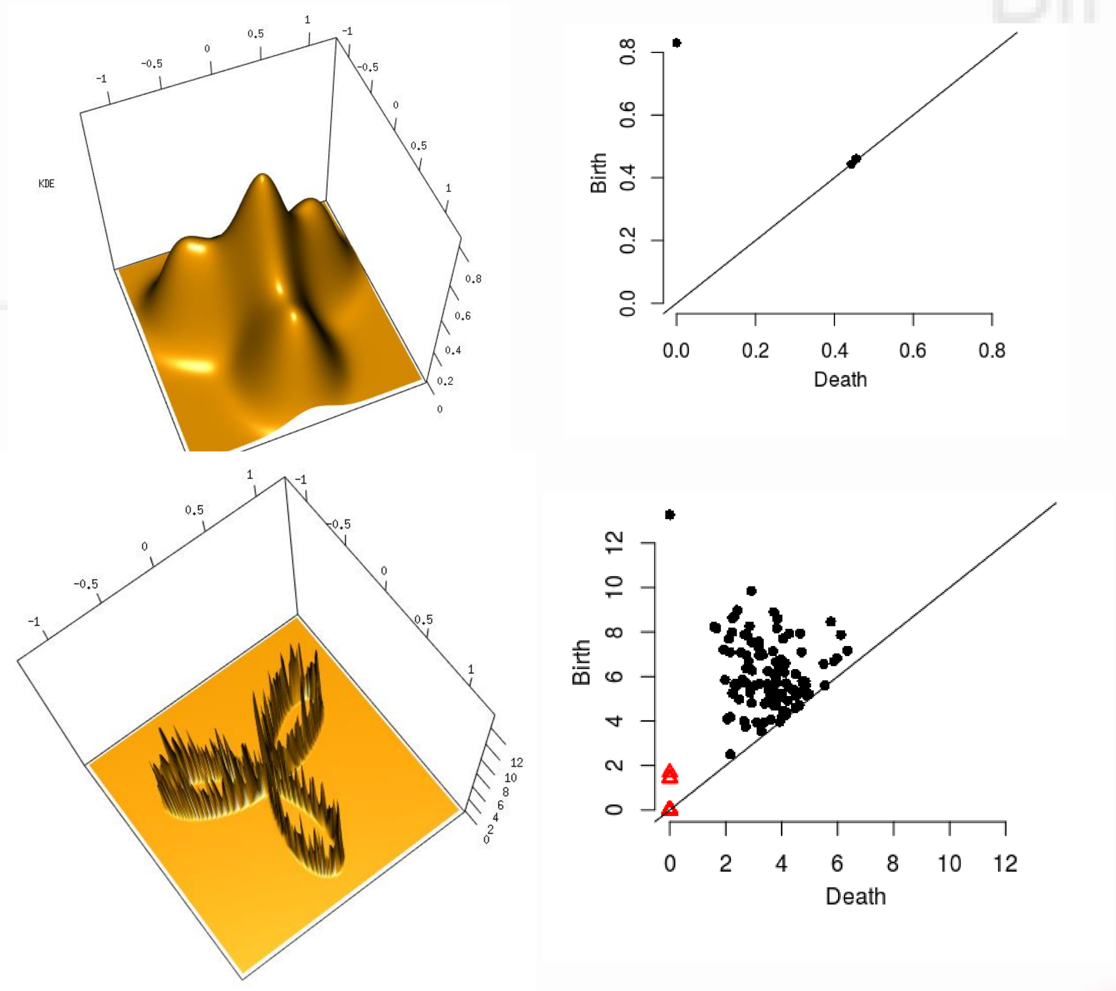


Distances (Vietoris-Rips filtration)

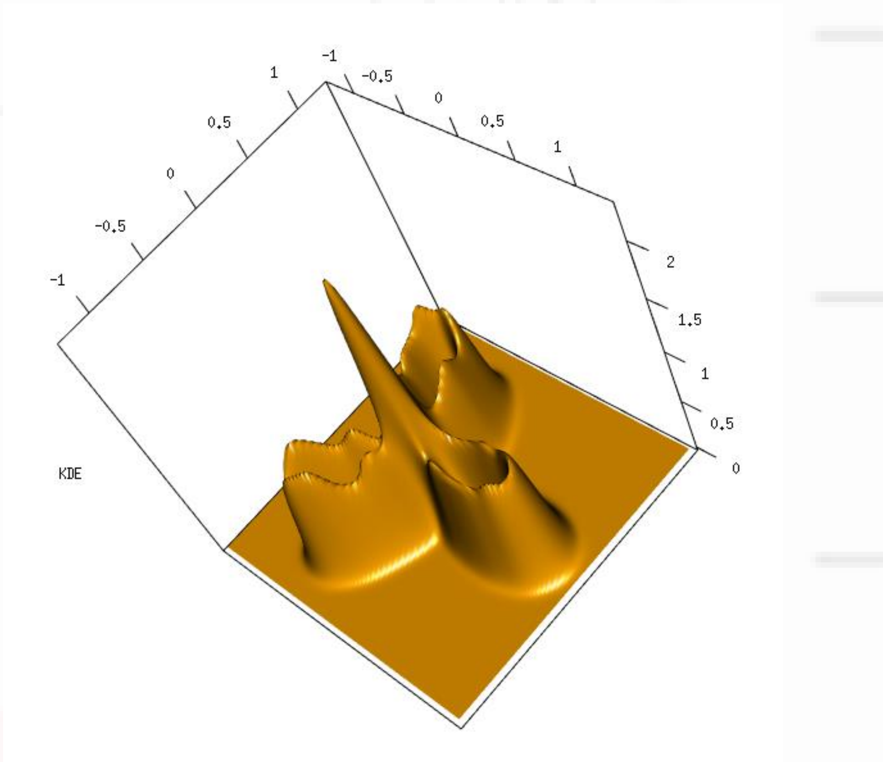


Kernel density estimator (KDE)

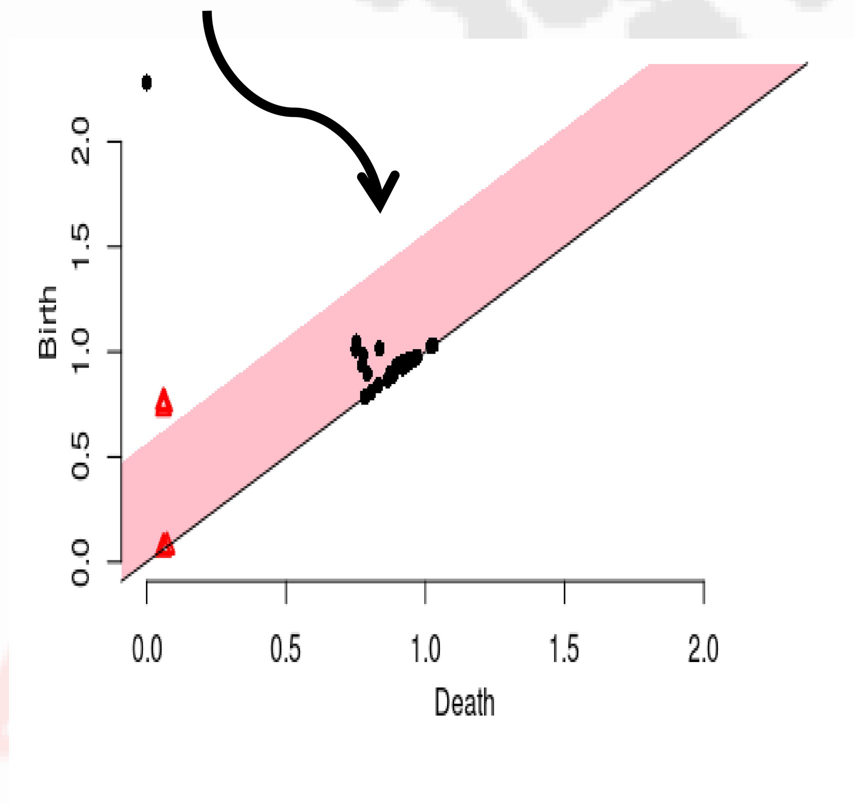
Choose carefully smoothing parameter



An Illustrative example



Confidence band

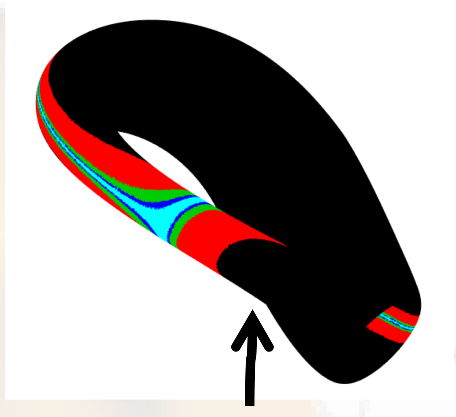


Examples on a Klein Bottle

Sampling from other distributions on domain.

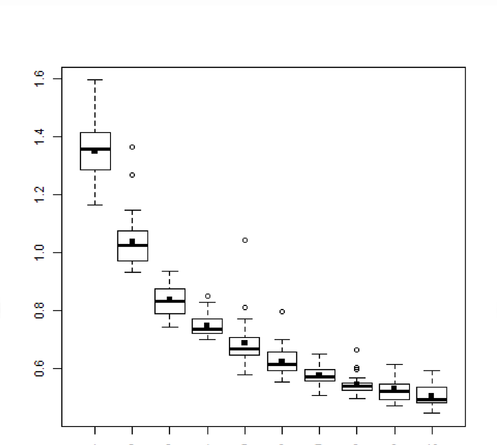
Points distribution

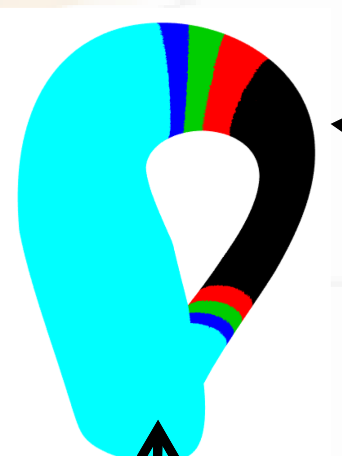
Hausdorff distance ($\mathcal{H}^m$-uniform vs distribution on the left)



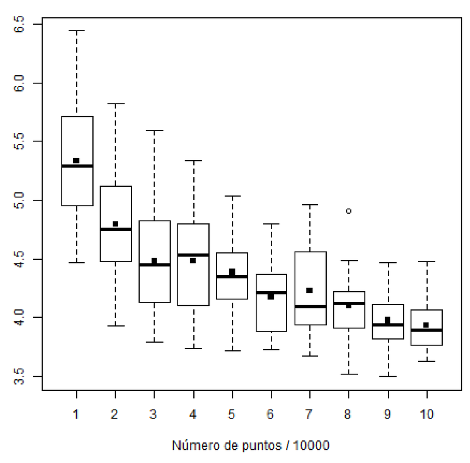
High concentration

Arcsine in both parameters

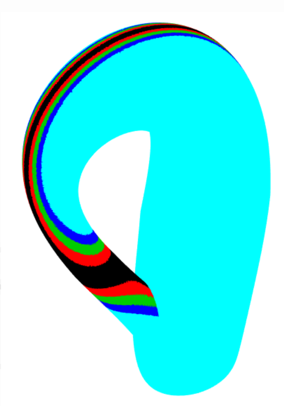




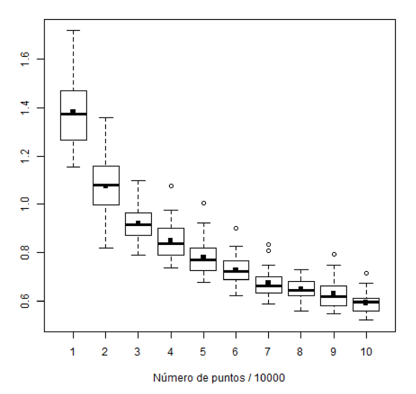
Beta in one parameter and uniform in other



Low concentration

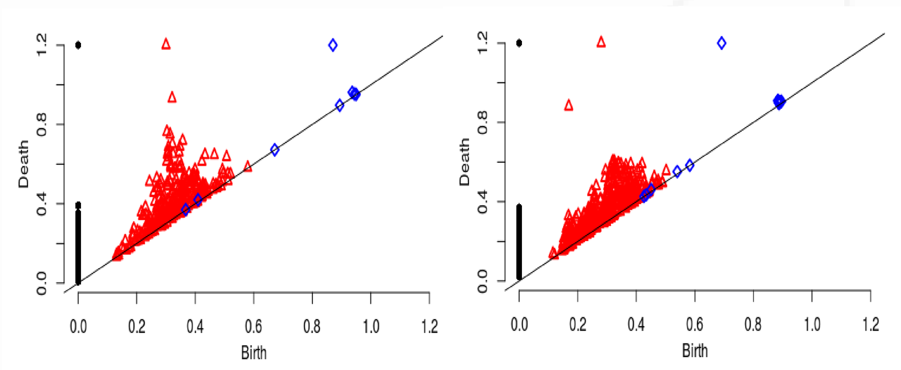


Semicircle in both parameters

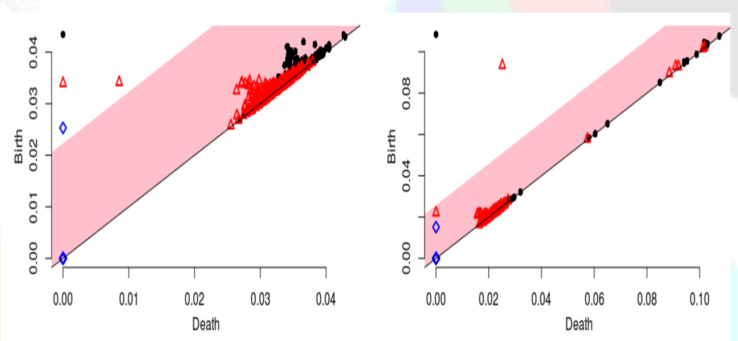


TDA applied to simulated PCD

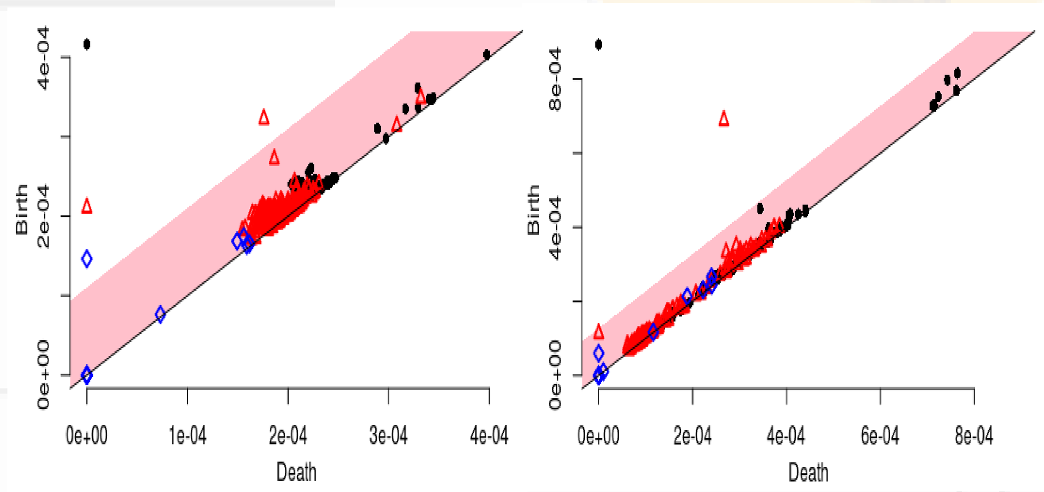
Torus, Vietoris-Rips ($\mathcal{H}^m$-uniform vs repulsion)



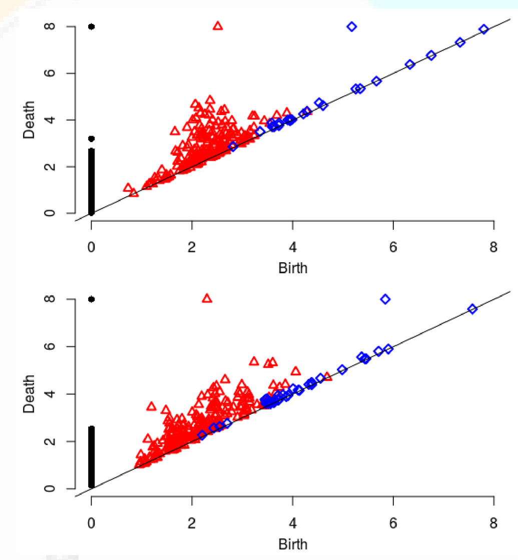
Torus, KDE ($\mathcal{H}^m$-uniform vs uniform on domain)



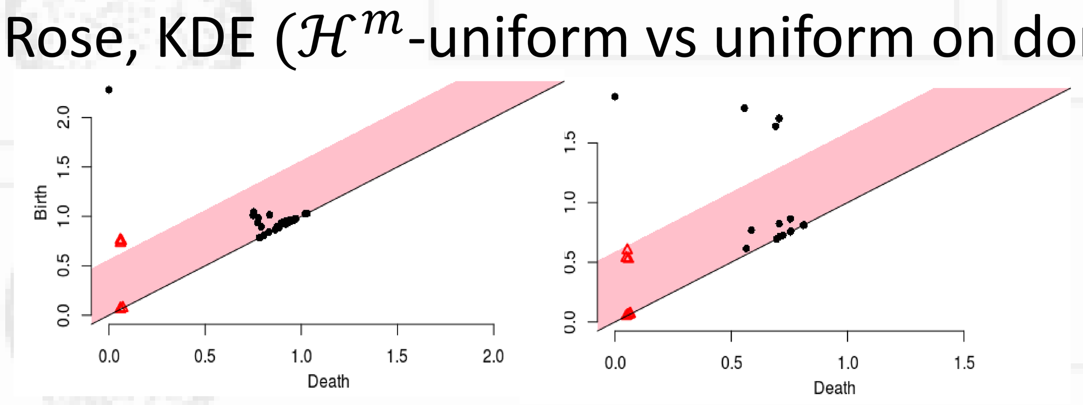
Klein Bottle KDE ($\mathcal{H}^m$-uniform vs uniform on domain)



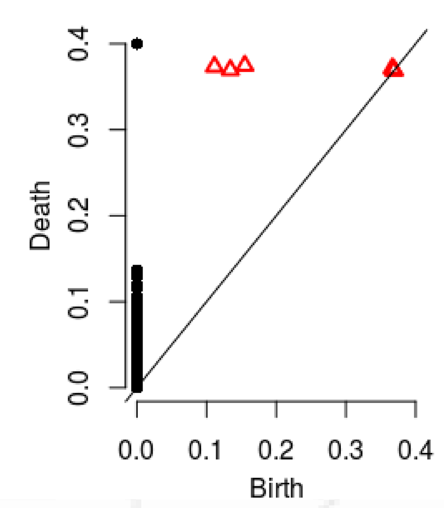
Klein Bottle Vietoris-Rips ($\mathcal{H}^m$-uniform vs repulsion)



Rose, KDE ($\mathcal{H}^m$-uniform vs uniform on domain)



Rose, Vietoris-Rips



Torus and Klein Bottle's Betti numbers: $\beta_0 = 1, \beta_1 = 2, \beta_2 = 1$

Rose's Betti numbers: $\beta_0 = 1, \beta_1 = 3$

Conclusions and forthcoming work

- Uniform on the domain is NOT equivalent to uniform on the manifold (exception: constant Jacobian).
- Choose method depending on objectives and other factors like computing cost.
- For uniform distribution on the manifold use parametrizations with a "simple" Jacobian.
- For examples in torus, simulating with repulsion yields better persistence diagrams.
- When using filtration over KDE: best result obtained with uniform distribution on manifold.
- Sampling from other distributions gives empirical insight in the identification of regions with highest and lowest probability, and on the rate of convergence in the concentration inequality since the computation of $\rho(t)$ is not in general easy.

References

1. Bordenave, C. and D. Chafaï: Lecture notes on the circular law. Proceedings of Symposis in Applied Mathematics, 72:1–34, January 2014. http://dx.doi.org/10.1090/psapm/072/00617.
2. Diaconis, P., S. Holmes, and M. Shahshahani: *Sampling from a Manifold*. Advances in Modern Statistical Theory and Applications: a Festschrift in honor of Morris L. Eaton, 10:102–125, 2013.
3. Fasy, B.T., F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh: *Confidence Sets for Persistence Diagrams*. The Annals of Statistics, 42(6):2301–2339, 2014.
4. Franzoni, G. *The Klein bottle: Variations on a theme*. Notices of the AMS, 59(8):1094–1099, 2012.

Acknowledgments

To Victor Pérez-Abreu for his constant support and guidance; Carlos Vargas for his suggestion for considering repulsion; and to Miguel Nakamura for useful comments.

Contact info

yair.hernandez@cimat.mx
gilberto.flores@cimat.mx

Partially supported by Conacyt-SNI 4337.