PCD Simulation with Uniform Distribution on a Klein Bottle

Part 2: Other Distributions and Hausdorff Distance

Gilberto Flores⋆, Yair Hernández⋆, Fermín Reveles∗

⋆DEMAT-UG, ∗CIMAT

In this part we compare the samples taken from different distibutions on the parameters.

A naive approach for simulating on a parametrized manifold is to take each parameter with uniform distribution. It is worth noting that in many cases this will yield point clouds with properties different to those presented in the first part. We illustrate this in the Klein bottle:

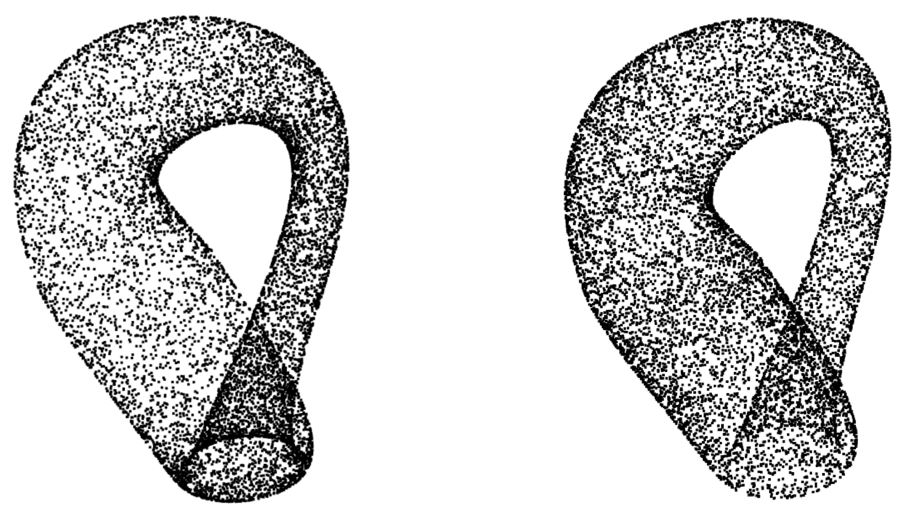


Figure 1: *Left*: point cloud simulated by taking a sample from the uniform distribution on the parameters. *Right*: point cloud simulated from the distribution described in part 1.

One difference that arises from the simulations above is shown in their corresponding persistence diagrams, using the Morse-Smale filtration. This difference is expected, since the kernel density estimator should vary if the distribution in the domain is changed.

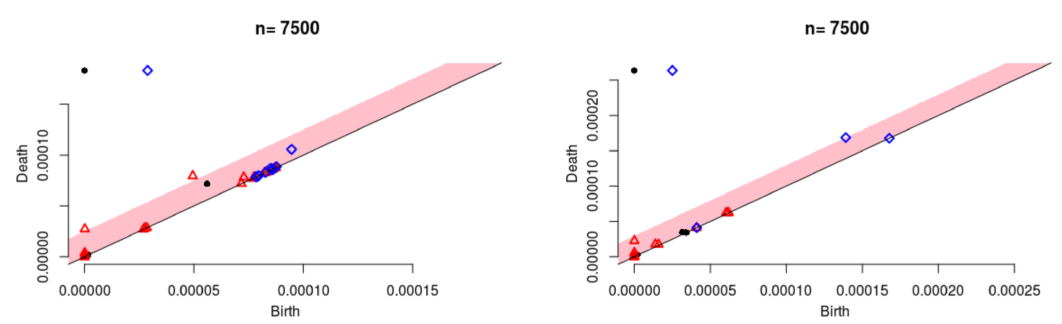


Figure 2: *Left*: Persistence diagram for the sample taken from the uniform distribution on the bottle. *Right*: Persistence diagram for the sample taken from the uniform distribution on the parameters.

There are other properties that remain similar if we change the distribution on the parameters. The Hausdorff distance will be useful showing this. This will be the main focus in this part.

In each of the following examples we will compute the Hausdorff distance between two samples. One of the samples will be simulated uniformly over the Klein bottle (distribution described in part 1). We take this distribution because it covers the manifold "evenly": as it assigns equal probability to parts with equal area, there are no parts which concentrate more points. For the other sample we will be using different distributions.

For each example we will show an image with colors assigned as follows: we take contour lines from the density on the parameters, and assign a color to each one, so each colored region on the bottle is the image of a region in the parametrization's domain, $[0, 2\pi] \times [0, 2\pi]$.

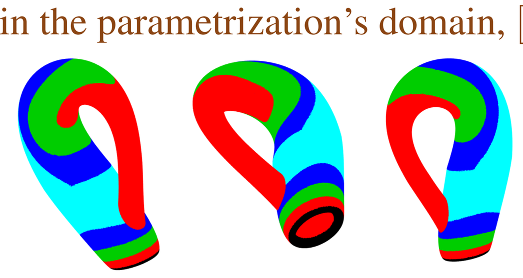


Figure 3: Density for the uniform distribution on the Klein bottle.

We first take the uniform distribution on the parameters. As this distribution is a constant, we don't show the colorized image of the bottle. We can see that the Hausdorff distance between the two point clouds approach to 0 as the number of points increase.

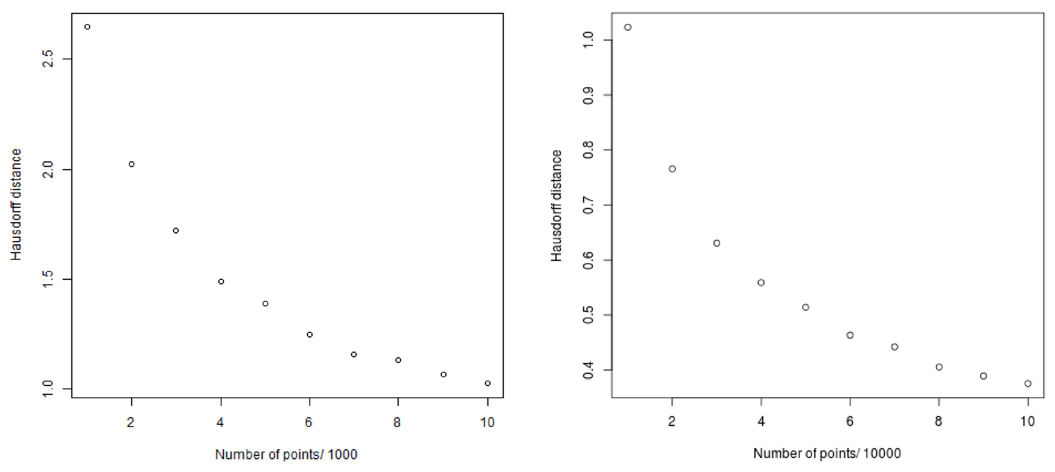


Figure 4: Hausdorff distance between noiseless point clouds

This doesn't occur, at least in an obvious manner, if we add noise to both samples.

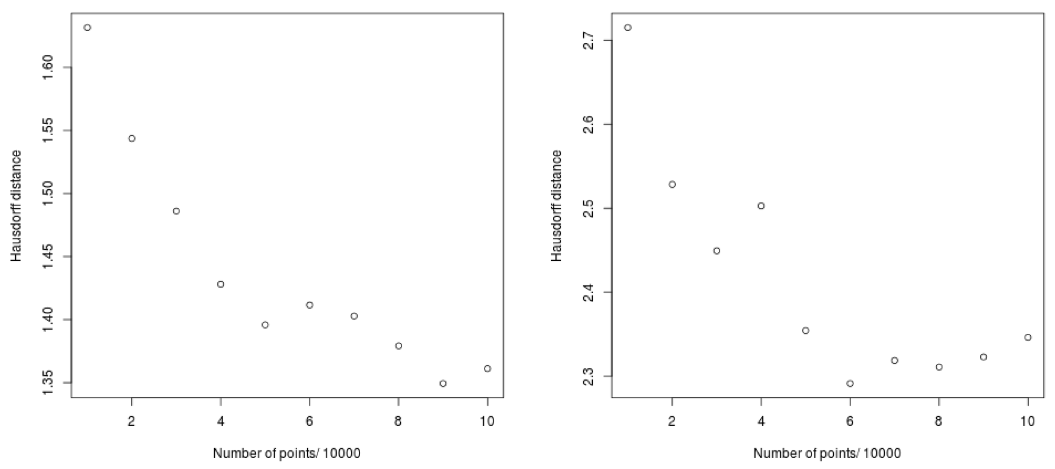
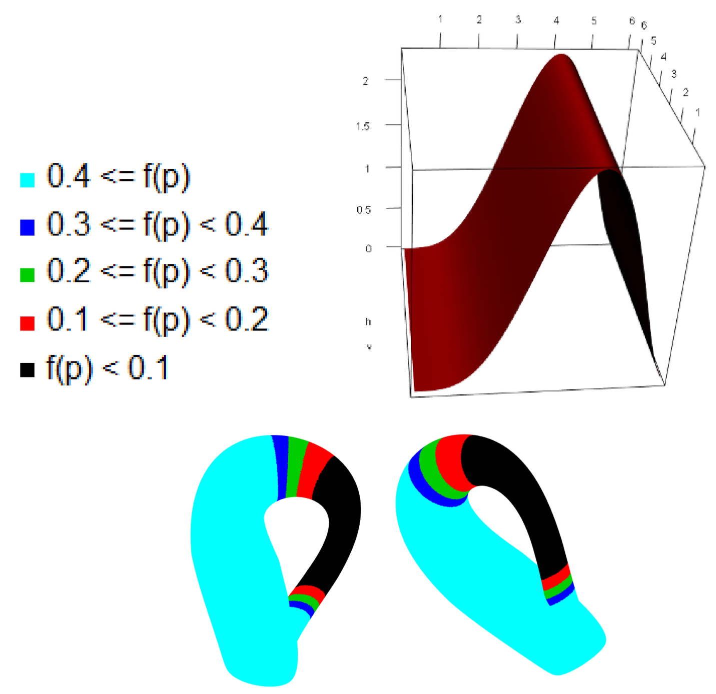
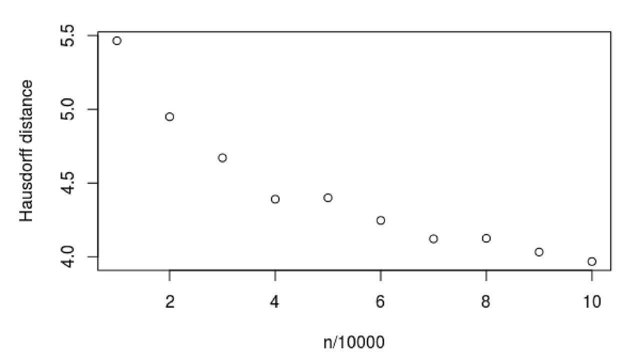


Figure 5: Hausdorff distances between noisy pount clouds

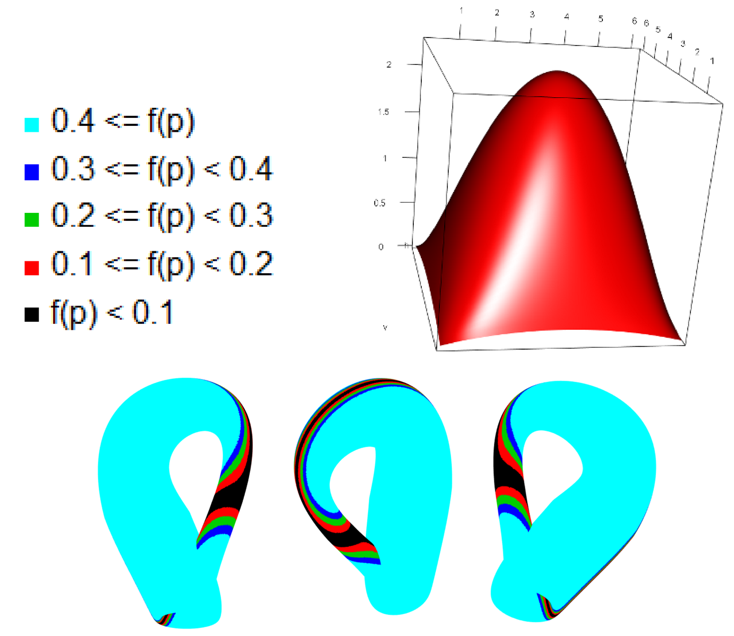
We now sample each parameter for the beta distribution (we multiply each value by 2π). For the next simulation we sample u from a $B(\alpha = 5, \beta = 3)$ (beta) distribution (each value multiplied by 2π and v from the uniform distribution on $[0, 2\pi]$.



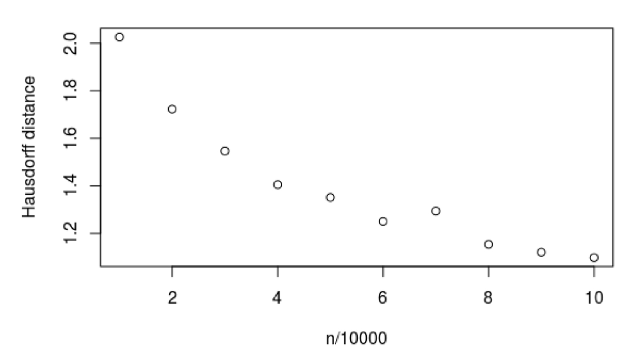
It is not clear if the Hausdorff distance converges to some value; if it converges to 0, it would be slower compared to the previous example.



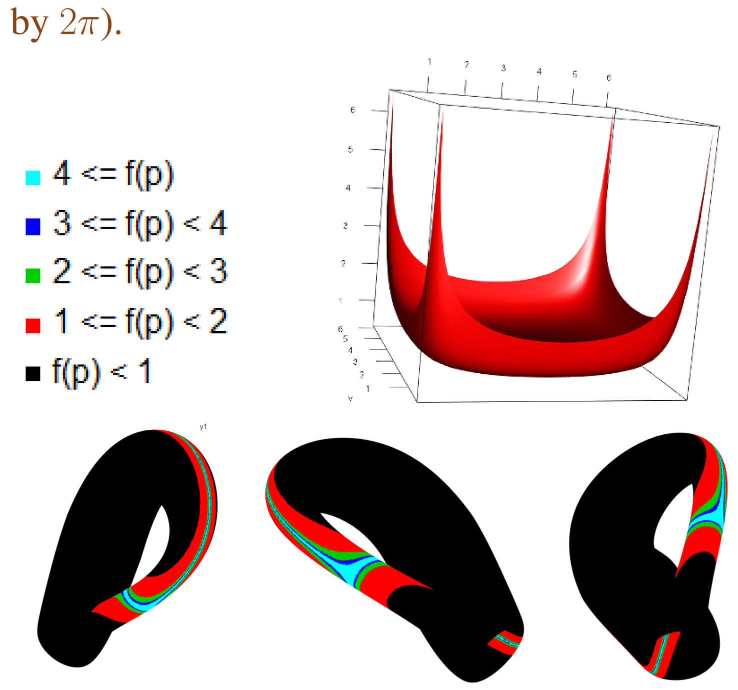
For the next simulation we sample both parameters from a $B(\alpha = 2, \beta = 2)$ (multiplying each value by 2π).



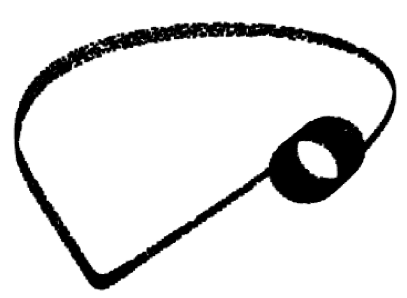
This example behaves similarly to the last one: it is not clear if the Hausdorff distance will converge to 0, or if it does slowly.



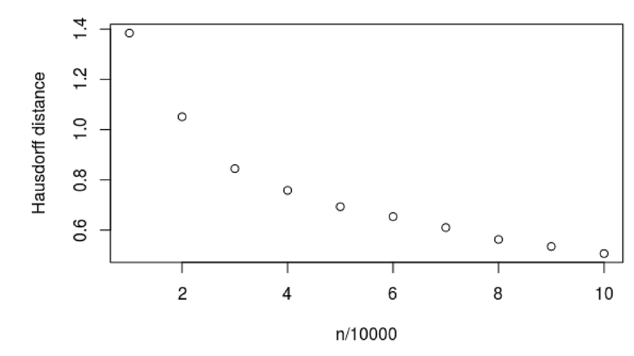
For the last simulation we sample both parameters from a $B(\alpha = 0.5, \beta = 0.5)$ distribution (arcsine distribution; as always, multiplying each value by 2π).



In this case we make the following special remark. In data sets with the shape of a Klein bottle that could arise from real data, we cannot expect them to be distributed uniformly on the Klein bottle. Instead, the points will be concentrated on certain regions. Therefore, we should look for topological features of the parts were the points are concentrated. For the example above, the region with a **higher probabilty** has the following shape, which is a punctured Klein bottle:



As for the convergence of the Hausdorff distance for the last example, we see it (possibly) converges to 0:



The convergence of the Hausdorff distance under certain assumptions is studied in Fasy et al. [1]. The following theorem establishes the relationship between the Hausdorff distance and the distribution of a sample: it is a concentration inequality. In Fasy et al.[1] it is used to estimate confidence sets for persistence diagrams. It links the stability theorem and the distribution of the sample:

Theorem 0.1. *For all $t > 0$*

$$\mathbb{P}\left(W_\infty(\hat{\mathcal{P}}, \mathcal{P}) > t\right) \le \mathbb{P}\left(d_H(S_n, \mathbb{M}) > t\right) \le \frac{2^d}{\rho(t/2)\, t^d} \exp\left(-n\rho(t)t^d\right),$$

where $\rho(t)$ is defined by

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

under some hypothesis on $\rho(x,t)$.

A desirable goal is to use the theorem in order to calculate the size of a sample, given a distribution, such that with high probability it covers the manifold.

Forthcoming Work

The concentration inequality involves persistence diagrams from the Vietoris-Rips filtration, so future work will be oriented towards obtaining a similar relation for the Morse-Smale filtration. Another application of this inequality is to analize the eficiency of algorithms in this context.

It should also be studied the behavior of de-noising algorithms in this context.

Recalling that the classification theorem states that any connected closed surface is homeomorphic to the sphere, connected sum of g-tori or connected sum of k real projective planes, there is potential in studying the results of applying these methods to those families of surfaces.

References

[1] B.T. Fasy, 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.