Hilbert's simplex distance:
A non-separable information monotone distance

Frank Nielsen
Frank.Nielsen@acm.org

October 6, 2021

Abstract

This note shows that the Hilbert's metric distance in the probability simplex is a non-separable distance which satisfies the information monotonicity.

Consider the open cone $\mathbb{R}^d_{++}$ of positive measures (i.e., histograms with d positive bins) with its open probability simplex subset $\Delta_d = \{(x_1, \dots, x_d) \in \mathbb{R}^d_+ \ : \ \sum_{i=1}^d x_i = 1\}$.

The f-divergence [1] between $p, q \in \Delta_d$ is defined for a convex function $f(u)$ such that $f(1) = 0$ and $f(u)$ strictly convex at 1 by:

$$I_f[p : q] := \sum_{i=1}^{d} p[i] f(q[i]/p[i]) \geq 0.$$

For example, the Kullback-Leibler divergence is a f-divergence for $f(u) = -\log u$.

All f-divergences are separable by construction: That is, they can be expressed as sum of coordinate-wise scalar divergences: Here, $I_f[p : q] := \sum_{i=1}^d i_f(p[i] : q[i])$, where i_f is a scalar f-divergence. Moreover, f-divergences are information monotone: That is, let $\mathcal{X} = \{X_1, \dots, X_m\}$ be a partition of $\{1, \dots, n\}$ into $m \leq n$ pairwise disjoint subsets X_i's. For $p \in \Delta_n$, let $p_{|\mathcal{X}} \in \Delta_m$ denote the induced probability mass function with $p_{|\mathcal{X}}(i) = \sum_{j \in X_i} p(i)$. Then we have

$$I_f[p_{|\mathcal{X}} : q_{|\mathcal{X}}] \leq I_f[p : q], \quad \forall \mathcal{X}$$

Moreoever, it can be shown that the only separable divergences satisfying this partition inequality are f-divergences [1] when $n > 2$. The special curious binary case $n = 2$ is dealt in [5].

Now, consider the non-separable Hilbert distance in the probability simplex [6]:

$$D_{\mathrm{Hilbert}}[p, q] = \log \frac{\max_i \frac{p_i}{q_i}}{\min_i \frac{p_i}{q_i}}.$$

This dissimilarity measure is a projective distance on $\mathbb{R}^d_{++}$ (Hilbert's projective distance) because we have $D_{\mathrm{Hilbert}}[\lambda p, \lambda' q] = D_{\mathrm{Hilbert}}[p, q]$ for any $\lambda, \lambda' > 0$. However, the Hilbert distance is a metric distance on Δ_d.

We state the main theorem:

Theorem 1 *The Hilbert distance on the probability simplex is an information monotone non-separable distance.*

Proof: We can represent the coarse-graining mapping $p \mapsto p_{|\mathcal{X}}$ by a linear application with a $m \times n$ matrix A with columns summing up to one (i.e., positive column-stochastic matrix):

$$p_{|\mathcal{X}} = A \times p.$$

For example, the partition $\mathcal{X} = \{X_1 = \{1,2\}, X_2 = \{3,4\}\}$ (with $n = 4$ and $m = 2$) is represented by the matrix

$$A = \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right].$$

Now, a key property of Hilbert distance is Birkhoff's contraction mapping theorem [2, 3]:

$$D_{\mathrm{Hilbert}}[Ap, Aq] \leq \tanh\left(\frac{1}{4}\Delta(A)\right) D_{\mathrm{Hilbert}}[p, q],$$

where $\Delta(A)$ is called the projective diameter of the positive mapping A:

$$\Delta(A) = \sup\{D_{\mathrm{Hilbert}}[Ap, Aq] \ : \ p, q \in \mathbb{R}^d_{++}\}.$$

Since $0 \leq \tanh(x) \leq 1$ for $x \geq 0$, we get the property that Hilbert distance on the probability simplex is an information monotone non-separable distance:

$$D_{\mathrm{Hilbert}}[p_{|\mathcal{X}}, q_{|\mathcal{X}}] \leq D_{\mathrm{Hilbert}}[p, q]$$

$\square$

Another example of non-separable information monotone distance is Aitchison's distance on the probability simplex [4] (using for compositional data analysis).

References

[1] Shun-ichi Amari. *Information geometry and its applications*, volume 194. Springer, 2016.

[2] Garrett Birkhoff. Extensions of Jentzsch's theorem. *Transactions of the American Mathematical Society*, 85(1):219–227, 1957.

[3] PJ Bushell. On the projective contraction ratio for positive linear mappings. *Journal of the London Mathematical Society*, 2(2):256–258, 1973.

[4] Ionas Erb and Nihat Ay. The information-geometric perspective of compositional data analysis. In *Advances in Compositional Data Analysis*, pages 21–43. Springer, 2021.

[5] Jiantao Jiao, Thomas A Courtade, Albert No, Kartik Venkat, and Tsachy Weissman. Information measures: the curious case of the binary alphabet. *IEEE Transactions on Information Theory*, 60(12):7616–7626, 2014.

[6] Frank Nielsen and Ke Sun. Clustering in Hilbert's projective geometry: The case studies of the probability simplex and the elliptope of correlation matrices. In *Geometric Structures of Information*, pages 297–331. Springer, 2019.