On the f-divergences between hyperboloid and Poincaré distributions

Frank Nielsen (Sony CSL), Kazuki Okamura (Shizuoka Univ.)

GSI '23

Introduction

Embedding: from **discrete** graph to **continuous** space
e.g. Sarker (2012): embedding of trees in **hyperbolic** plane with low distortion (not Euclidian plane)
$\longrightarrow$ probability distribution on hyperbolic space.

Review:

- hyperboloid distributions on the Minkowski space by Jensen (1981)
  —— analogy to the von-Mises Fisher distributions on the sphere
- Souriau-Gibbs distributions on by Barbaresco (2019)
  —— in the Poincaré disk with its Fisher information metric = the Poincaré hyperbolic Riemannian metric

Poincaré distributions

Tojo and Yoshino (2020); hyperboloid distribution realized on the upper half-plane $\mathbb{H}$

- **Parameter space**:
  $\Theta := \{(a, b, c) \in \mathbb{R}^3 \ : a > 0, c > 0, \ ac - b^2 > 0\} \simeq \mathrm{Sym}^+(2, \mathbb{R})$ by $(a, b, c) \simeq \begin{bmatrix} a & b \\ b & c \end{bmatrix}$.
- $|\theta| := ac - b^2 > 0$ and $\mathrm{tr}(\theta) := a + c$ for $\theta = (a, b, c)$.
- **pdf**:

$$p_\theta(x, y) := \frac{\sqrt{|\theta|}\exp(2\sqrt{|\theta|})}{\pi} \exp\left(-\frac{a(x^2 + y^2) + 2bx + c}{y}\right) \frac{1}{y^2}, (x, y) \in \mathbb{H}$$

$\exists$ q-**deformed** Poincaré distributions. (Tojo and Yoshino)

f-divergence

The **f-divergence** induced by a convex generator $f : (0, \infty) \to \mathbb{R}$ between two pdfs $p(x, y)$ and $q(x, y)$ on $\mathbb{H}$:

$$D_f[p : q] := \int_{\mathbb{H}} p(x, y)\, f\left(\frac{q(x, y)}{p(x, y)}\right) \mathrm{d}x\, \mathrm{d}y.$$

This measures dissimilarity between two distributions.

Theorem 1

Every f-divergence between two Poincaré distributions p_θ and $p_{\theta'}$ is a function of $\left(|\theta|, |\theta'|, \mathrm{tr}\left(\theta'\theta^{-1}\right)\right)$

Proof components:
(1) $D_f\left[p_\theta : p_{\theta'}\right]$ is invariant wrt $\mathrm{SL}(2, \mathbb{R})$-action:

$$D_f\left[p_\theta : p_{\theta'}\right] = D_f\left[p_{g^{-\top}\theta g^{-1}} : p_{g^{-\top}\theta' g^{-1}}\right],\ \ g \in \mathrm{SL}(2, \mathbb{R})$$

(2) Every action-invariant function $g(\theta, \theta')$ on $\mathbb{H}^2$ is a function of $\left(|\theta|, |\theta'|, \mathrm{tr}\left(\theta'\theta^{-1}\right)\right)$ — **maximal invariant** of the action.

Importance of the concept of maximal invariant

- Assume that one has a problem for which a function f which is **invariant** wrt some group action $f(gx) = f(x)$ but difficult to solve explicitly $f()$ from scratch
- For the group action, one finds a **maximal invariant** $m()$: It is an invariant and maximal, i.e.

$$m(x) = m(y) \Longrightarrow \exists g \text{ s.t. } y = gx$$

- Then, $\exists h$ s.t. $f(x) = h(m(x))$. Solving/finding $h()$ *may be simpler* than solving/finding the original $f()$

See the book by Eaton(1989)

Proposition 1 (explicit formulae)

(i) (Kullback-Leibler) Let $f(u) = -\log u$. Then,

$$D_f\left[p_\theta : p_{\theta'}\right] = \frac{1}{2}\log\frac{|\theta|}{|\theta'|} + 2\left(\sqrt{|\theta|} - \sqrt{|\theta'|}\right) + \left(\frac{1}{2} + \sqrt{|\theta|}\right)(\mathrm{tr}(\theta'\theta^{-1}) - 2).$$

(ii) (squared Hellinger) Let $f(u) = (\sqrt{u} - 1)^2/2$. Then,

$$D_f[p_\theta : p_{\theta'}] = 1 - \frac{2|\theta|^{1/4}|\theta'|^{1/4}\exp\left(|\theta|^{1/2} + |\theta'|^{1/2}\right)}{|\theta + \theta'|^{1/2}\exp\left(|\theta + \theta'|^{1/2}\right)}.$$

(iii) (Neyman χ^2) Let $f(u) := (u - 1)^2$. Assume that $2\theta' - \theta \in \Theta$. Then,

$$D_f[p_\theta : p_{\theta'}] = \frac{|\theta'|\exp(4|\theta'|^{1/2})}{|\theta|^{1/2}|2\theta' - \theta|^{1/2}\exp\left(2(|\theta|^{1/2} + |2\theta' - \theta|^{1/2})\right)} - 1.$$

$|\theta + \theta'|$ and $|2\theta' - \theta|$ can be expressed by $|\theta|, |\theta'|$, and tr$(\theta'\theta^{-1})$.

2D hyperboloid distributions

Barndorff-Nielsen (1978); Jensen (1981)
Lobachevskii space:

$$\mathbb{L}^2 := \left\{ (x_0, x_1, x_2) \in \mathbb{R}^3 : x_0 = \sqrt{1 + x_1^2 + x_2^2} \right\} \simeq \mathbb{R}^2$$

Minkowski inner product:

$$[(x_0, x_1, x_2), (y_0, y_1, y_2)] := x_0 y_0 - x_1 y_1 - x_2 y_2.$$

- **Parameter space**:

$$\Theta_{\mathbb{L}^2} := \left\{ (\theta_0, \theta_1, \theta_2) \in \mathbb{R}^3 : \theta_0 > \sqrt{\theta_1^2 + \theta_2^2} \right\}.$$

- **pdf**: For $\theta \in \Theta_{\mathbb{L}^2}$,

$$p_\theta(x_1, x_2) := \frac{|\theta| \exp(|\theta|)}{2(2\pi)^{1/2}} \frac{\exp(-[\theta, \widetilde{x}])}{\sqrt{1 + x_1^2 + x_2^2}}, \ (x_1, x_2) \in \mathbb{R}^2$$

where we let $\widetilde{x} := \left(\sqrt{1 + x_1^2 + x_2^2}, x_1, x_2 \right) \in \mathbb{L}^2$ and $|\theta| := [\theta, \theta]^{1/2}$.

Theorem 2

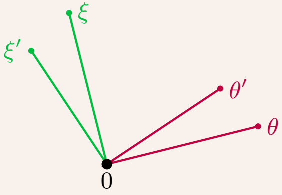
Every f-divergence between p_θ and $p_{\theta'}$ is a function of the triplet $([\theta,\theta],[\theta',\theta'],[\theta,\theta'])$, i.e., the pairwise Minkowski inner products of θ and θ'.

Geometric interpretation: $D_f\left[p_\theta : p_{\theta'}\right] \longleftrightarrow \triangle 0\theta\theta'$
Thm 2 $\longleftrightarrow$ **side-angle-side theorem** in Euclidean and *hyperbolic* geometry

$$\left([\theta,\theta],[\theta',\theta'],[\theta,\theta']\right) = \left([\xi,\xi],[\xi',\xi'],[\xi,\xi']\right)$$

$$\Longrightarrow \triangle 0\theta\theta' \equiv \triangle 0\xi\xi' \Longrightarrow D_f\left[p_\theta : p_{\theta'}\right] = D_f\left[p_\xi : p_{\xi'}\right]$$

Proof strategy is similar to the one of Theorem 1.



Proposition 2 (explicit formulae)

(i) (Kullback-Leibler) Let $f(u) = -\log u$. Then,

$$D_f[p_\theta : p_{\theta'}] = \log\left(\frac{|\theta|}{|\theta'|}\right) - |\theta'| + \frac{[\theta,\theta']}{[\theta,\theta]} + \frac{[\theta,\theta']}{|\theta|} - 1.$$

(ii) (squared Hellinger) Let $f(u) = (\sqrt{u} - 1)^2/2$. Then,

$$D_f[p_\theta : p_{\theta'}] = 1 - \frac{2|\theta|^{1/2}|\theta'|^{1/2}\exp\left(|\theta|/2 + |\theta'|/2\right)}{|\theta+\theta'|\exp\left(|\theta+\theta'|/2\right)}.$$

(iii) (Neyman χ^2) Let $f(u) := (u-1)^2$. Assume that $2\theta' - \theta \in \Theta_{\mathbb{L}^2}$. Then,

$$D_f[p_\theta : p_{\theta'}] = \frac{|\theta'|^2\exp(2|\theta'|)}{|\theta||2\theta'-\theta|\exp(|\theta| + |2\theta'-\theta|)} - 1.$$

This corresponds to Proposition 1 for Poincaré distributions.

Correspondence

Proposition 3 (Correspondence between the parameter spaces)

A bijection:

$$\Theta \longrightarrow \Theta_{\mathbb{L}}$$
$$\theta := (a, b, c) \mapsto \theta_{\mathbb{L}} := (a + c, a - c, 2b)$$

By this map,
(i) For $\theta, \theta' \in \Theta_{\mathbb{H}}$,

$$|\theta_{\mathbb{L}}|^2 = [\theta_{\mathbb{L}}, \theta_{\mathbb{L}}] = 4|\theta|, \ |\theta'_{\mathbb{L}}|^2 = \left[\theta'_{\mathbb{L}}, \theta'_{\mathbb{L}}\right] = 4|\theta'|, \ \left[\theta_{\mathbb{L}}, \theta'_{\mathbb{L}}\right] = 2|\theta|\mathrm{tr}(\theta'\theta^{-1}).$$

(ii) For every f and $\theta, \theta' \in \mathbb{H}$,

$$D_f^{\mathbb{L}}\left[p_{\theta_{\mathbb{L}}} : p_{\theta'_{\mathbb{L}}}\right] = D_f^{\mathbb{H}}\left[p_{\theta} : p_{\theta'}\right].$$

There is also a correspondence between the sample spaces, which is compatible with the above one.

Some references

Barbaresco, F.: *Lie group machine learning and Gibbs density on Poincaré unit disk from Souriau Lie groups thermodynamics and SU(1,1) coadjoint orbits.* In: International Conference on GSI. pp. 157-170. Springer (2019)

Barndorff-Nielsen, O.: *Hyperbolic distributions and distributions on hyperbolae*. Scandinavian Journal of Statistics 5, 151-157 (1978)

Eaton, M.L.: *Group invariance applications in statistics*. Hayward, CA: Institute of Mathematical Statistics; Alexandria, VA: American Statistical Association (1989)

Jensen, J.L.: *On the hyperboloid distribution*. Scandinavian Journal of Statistics 8, pp. 193-206 (1981)

Sarkar, R.: *Low distortion Delaunay embedding of trees in hyperbolic plane*, Graph drawing, LNCS 7034, 355-366, (2012)

Tojo, K., Yoshino, T.: *An exponential family on the upper half plane and its conjugate prior*. In: Workshop on Joint Structures and Common Foundations of Statistical Physics, Information Geometry and Inference for Learning. pp. 84-95. Springer (2020)