Extrinsic means of isotropic distributions on hyperbolic spaces

Emmanuel CHEVALLIER[1]

[1]Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel,
52 Av. Escadrille Normandie Niemen, 13013 Marseille, France
emmanuel.chevallier@univ-amu.fr

Résumé – Dans cet article, nous montrons que pour des distributions de probabilités isotropes dans un espace hyperbolique, la moyenne de Fréchet et la moyenne extrinsèque définie à partir du modèle de l'hyperboloïde coïncident. L'analyse des taux de convergences montre que l'espérance de l'erreur quadratique de l'estimation de la moyenne extrinsèque, et de la moyenne de Fréchet, décroissent à la même vitesse. La faible complexité algorithmique et la facilité d'implémentation de la moyenne extrinsèque en font une alternative intéressante à la moyenne de Fréchet.

Abstract – In this paper, we show that for isotropic probability distributions on hyperbolic spaces, the Fréchet mean and the extrinsic mean defined from the hyperboloid model coincide. The analysis of convergence rates shows that the expected squared error of the estimation of the extrinsic mean, and of the Fréchet mean, decrease at the same speed. The low computational complexity and programming efforts required for the computation of the extrinsic mean make it an interesting alternative to the Fréchet mean.

1 Introduction

Hyperbolic spaces appear in various data science problems, such as graph embedding, [15, 2], radar signals, [1, 3, 7], and color science [10, 16]. Several authors have proposed to use isotropic probability distributions to model random phenomenons on the hyperbolic space, see for instance [17, 13, 9, 12, 5]. The notion of isotropy of a distribution is always defined with respect to a point p of the space. Under an existence condition, the parameter p is the Fréchet mean of the distribution, and can be estimated by the empirical Fréchet mean. Our main contribution is to show that for isotropic distributions, the Frechet mean coincides with the notion of 'hyperbolic barycenter' proposed in [11]. Due to its geometric construction, we call this 'hyperbolic barycenter' the extrinsic mean. Our result is analogous to the Theorem 3.3 of [4], where authors proved that for certain submanifolds of a Euclidean space, the extrinsic mean and the Frechet mean of isotorpic distribution coincide. Their result apply in particular to spheres and certain projective spaces. The extrinsic mean analysed here is based on an embedding in a pseudo-Euclidean space, and the projection on the embedded manifold differs from the one of [4]. Hence it requires a separate treatment.

The paper is almost entirely self contained. In section 2, we provide the necessary background on hyperbolic geometry and on the hyperboloid model. Section 3 states the definition of isotropic probability distribution. The section 4 contains the main result. We start by basic facts about the Fréchet mean and then we show that it coincides with the extrinsic mean when the distribution is isotropic. In section 5, we show that the mean squared errors of empirical estimators of both means decrease in $\frac{1}{N}$, with N the number of samples. We conclude the paper in section 6 by indicating several possible generalizations.

2 Hyperbolic geometry

2.1 Riemannian manifolds

Hyperbolic spaces are a particular type of Riemannian manifolds. In terms of topological and differentiable manifolds, their structures are similar to $\mathbb{R}^n$. Note $\mathbb{H}_n$ a hyperbolic space of dimension n. There exists a diffeomorphism $\varphi : \mathbb{H}_n \to \mathbb{R}^n$. As a Riemannian manifold, tangent spaces at each $p \in \mathbb{H}_n$ are endowed with a local inner product

$$(T_p\mathbb{H}_n, \langle \cdot, \cdot \rangle_p),$$

which enable to define length of differentiable path $\gamma : [a, b] \to \mathbb{H}_n$ as,

$$L(\gamma) = \int_a^b \sqrt{\langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)}} dt$$

and distances between points as the length of the shortest curve joining them,

$$d(p, q) = \inf_{\gamma} L(\gamma)$$

where the infimum is taken over all differentiable path joining p and q. There exists several way to construct hyperbolic spaces of dimension n, such as the hyperboloid model, the Klein model, the Poincaré ball model, or the Poincaré half space model. All these models of hyperbolic geometry are different as sets but are all isometric Riemannian manifolds : between each model, there is a bijection which preserves distances.

2.2 Hyperboloid models

Hyperbolic spaces can be seen as duals of spheres. A sphere S_n is obtained as the set of unit norm vectors x of $\mathbb{R}^{n+1}$. As a Riemannian manifold, the local metric of S_n is given by the canonical inner product of $\mathbb{R}^{n+1}$ restricted to tangent spaces T_xS_n. The hyperboloid model of hyperbolic geometry is obtained in a similar way but with a modified inner product. In the rest of the paper, the inner product $\langle ., . \rangle$ on $\mathbb{R}^{n+1}$ is defined as the following pseudo-Euclidean inner product,

$$\forall x, y \in \mathbb{R}^{n+1},\ \langle x, y\rangle = x_0 y_0 - \sum_{i=1}^{n} x_i y_i.$$

The set of vectors $x \in \mathbb{R}^{n+1}$ with $\langle x, x\rangle = 1$ is a hyperboloid with 2 sheets. Note H_n the sheet defined by

$$H_n = \{x \in \mathbb{R}^{n+1} : \langle x, x\rangle = 1 \text{ and } x_0 > 0\}.$$

H_n is a hypersurface of $\mathbb{R}^{n+1}$ parametrized freely by the n coordinates $x_1, ..., x_n$: it is diffeomorphic to $\mathbb{R}^n$. Let $p \in H_n$. The tangent space of H_n at p can be characterized by

$$T_pH_n = \{u \in \mathbb{R}^{n+1} | \langle p, u\rangle = 0\}.$$

For arbitrary $u, v \in T_pH_n$, define

$$\langle u, v\rangle_p = -\langle u, v\rangle.$$

The bilinear form $\langle ., .\rangle_p$ is positive definite on T_pH_n because an orthogonal basis $u_1, \ldots, u_n$ of T_pH_n completed by p is an orthogonal basis of $\mathbb{R}^{n+1}$ and the signature of the bilinear form $\langle ., .\rangle$ is $(1, n)$ so that $\langle u_1, u_1\rangle, \ldots, \langle u_n, u_n\rangle$ must be negative.

H_n endowed with the local inner products $\langle ., .\rangle_p$ is a Riemannian manifold and a model of hyperbolic geometry. As for spheres, it can be shown that the distance d between points has an explicit form, for $p, q \in H_n$,

$$d(p, q) = \operatorname{arcosh}(\langle p, q\rangle).$$

The trigonometric functions of the spherical geometry are simply replaced by hyperbolic functions.

2.3 The isometries of the hyperboloid model

Note $SO(1, n)$ the group of $(n + 1) \times (n + 1)$ real matrices with determinant 1 which preserve the pseudo-inner product $\langle ., .\rangle$. The action of $SO(1, n)$ preserves the hyperboloid of two sheets of unit norm. Call $SO^+(1, n) \subset SO(1, n)$ the group of matrices that also preserve each sheet. By construction of the metric on H_n, matrices of $SO^+(1, n)$ preserve the hyperbolic distance on H_n,

$$\forall x, y \in H_n, \forall M \in SO^+(1, n),\ d(x, y) = d(Mx, My).$$

Homogeneity : If $x \in H_n$ then there exists an orthogonal basis $e_1, \ldots, e_n$ of the orthogonal of x with respect to $\langle ., .\rangle$. Since the signature of $\langle ., .\rangle$ is $(1, n)$, we can suppose that $\langle e_i, e_i\rangle = -1$. It follows that the linear map that maps the canonical basis of $\mathbb{R}^{n+1}$ to $x, e_1, \ldots, e_n$, is in $SO^+(1, n)$. It follows that the group $SO^+(1, n)$ acts transitively on H_n,

$$\forall x, y \in H_n, \exists M \in SO^+(1, n), Mx = y, \tag{1}$$

hence H_n is a homogeneous Riemannian manifold.

Isotropy : Since any Euclidean rotation that fixes the vector $\mathbf{1} = (1, 0, ..., 0)^T \in H_n$ is in $SO^+(1, n)$, for each pair of unit norm tangent vectors u, v in the tangent space $T_\mathbf{1}H_n$, $\exists M \in SO^+(1, n), M\mathbf{1} = \mathbf{1}$ and $Mu = v$. Since H_n is homogeneous, it follows that H_n is an isotropic Riemannian manifold : $\forall p \in H_n, \forall u, v \in T_pH_n, \langle u, u\rangle = \langle v, v\rangle$,

$$\exists M \in SO^+(1, n), Mp = p \text{ and } Mu = v. \tag{2}$$

3 Isotropic probability distribution

Let P be a probability distribution on H_n. P is said isotropic with respect to $\overline{x} \in H_n$ when all the isometries which fix $\overline{x}$ preserve P. In particular, for all $M \in SO^+(1, n)$ with $M\overline{x} = \overline{x}$, and for all measurable $A \subset H_n$,

$$P(A) = P(M^{-1}A),$$

where $M^{-1}A = \{x \in H_n | Mx \in A\}$.

4 Means

4.1 Frechet mean

For a probability distribution P on H_n, let $F : H_n \to \mathbb{R}$ be the function defined by

$$F(x) = \int_{y\in H_n} d(x, y)^2 \mathrm{d}P. \tag{3}$$

A global minimum of F is called a Fréchet mean. We will check that if F takes a finite value for some $x \in H_n$, F has a unique Fréchet mean. Since $F(x)$ goes to infinity in every direction in the hyperbolic space, there exists a Fréchet mean. By the median inequality, for all $x, x', y \in H_n$,

$$d(m, y)^2 \leq \frac{1}{2}(d(y, x)^2 + d(y, x')^2) - \frac{1}{4}d(x, x')^2$$

where m is the midpoint of the geodesic segment joining x to x' (see [6]). Integrating both side of the inequality with respect to P, we obtain $F(m) \leq \frac{1}{2}(F(x) + F(x')) - \frac{1}{4}d(x, x')^2$ which in turn implies the uniqueness of the minimum. Note

$$\mathcal{F}(P) = \operatorname{argmin}_{x\in H_n}(F(x)),$$

the Fréchet mean.

Theorem 1 *Assume that P has a Fréchet mean. If P is isotropic with respect to $\overline{x}$, then*

$$\mathcal{F}(P) = \overline{x},$$

$\overline{x}$ is the Fréchet mean of P.

Proof 1 *Let q_1 and let $q_2 = M(q_1)$ where M is a isometry that fixes $\overline{x}$. Since P is M-invariant,*

$$F(q_1) = \int d(q_1, y)^2 dP(y) = \int d(M(q_1), M(y))^2 dP(y)$$
$$= \int d(q_2, M(y))^2 dP(y) = \int d(q_2, y) dP(y) = F(q_2).$$

Therefore, if $\gamma : [-a, a] \to H_n$ is any geodesic such that $\gamma(0) = \overline{x}$, then $F(\gamma(a)) = F(\gamma(-a))$. Now, $F(\overline{x}) \leq \frac{1}{2}(F(\gamma(a)) + \gamma(-a)) - \frac{1}{4}d(\gamma(a), \gamma(-a))^2$, hence $F(\overline{x}) \leq F(\gamma(a))$.

4.2 Extrinsic mean

We will now provide an alternative definition of the mean, and see that for isotropic distributions it coincides with the Fréchet mean, when they exist.

We define the extrinsic mean of a probability distribution P on H_n as the traditional vector mean of P, rescaled by the appropriate factor to lie on H_n. When it exists, let $V(P)$ be the vector mean,

$$V(P) = \int_{x \in H_n} x dP.$$

Vectors $x \in \mathbb{R}^{n+1}$ with $x_0 > 0$ and $\langle x, x \rangle > 0$ form a convex cone C_+. Hence $\langle V(P), V(P) \rangle > 0$. Let $\pi : C_+ \to H_n$ be the projection defined by

$$\pi(x) = \frac{x}{\sqrt{\langle x, x \rangle}},$$

and define the extrinsic mean as

$$E(P) = \pi(V(P)) = \frac{V(P)}{\sqrt{\langle V(P), V(P) \rangle}}.$$

As for the Fréchet means, we will now prove that if the vector mean of a distribution isotropic with respect to a point $\overline{x}$ exists, then $\overline{x}$ is the extrinsic mean.

Let $M \in SO^+(1,n)$. For an arbitrary distribution P, note MP the probability defined by $MP(A) = P(M^{-1}A)$, for $A \subset H_n$. We have

$$\begin{aligned} V(MP) &= \int_{x \in H_n} x \mathrm{d}MP \\ &= \int_{x \in H_n} Mx \mathrm{d}P \\ &= M \int_{x \in H_n} x \mathrm{d}P \\ &= MV(P), \end{aligned}$$

without surprises, the vector mean commutes with M.

Let $\mathbf{1} = (1, 0, ..., 0)^T \in H_n$. Show now the desired result for a distribution $P_{\mathbf{1}}$ isotropic with respect to $\mathbf{1}$. Call $G_{\mathbf{1}} \subset SO^+(1,n)$ the set of $(n+1) \times (n+1)$ matrices of the form

$$M = \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix},$$

with $R \in SO(n)$. Elements of $G_{\mathbf{1}}$ are hyperbolic isometries fixing $\mathbf{1}$. Since $P_{\mathbf{1}}$ is isotropic with respect to $\mathbf{1}$, for all $M \in G_{\mathbf{1}}$ we have $MP_{\mathbf{1}} = P_{\mathbf{1}}$. Hence

$$V(MP_{\mathbf{1}}) = V(P_{\mathbf{1}}) = MV(P_{\mathbf{1}}),$$

$V(P_{\mathbf{1}})$ is an eigenvector of M. Since there are no eigenvector $u \in \mathbb{R}^n$ common to all the elements of $SO(n)$, $\mathbb{R}.\mathbf{1}$ are the only eigenvectors common to all the elements of $G_{\mathbf{1}}$. We have then that

$$\exists \alpha \in \mathbb{R}, V(P_{\mathbf{1}}) = \alpha \mathbf{1}.$$

which enable to conclude that $\mathbf{1}$ is the extrinsic mean,

$$E(P_{\mathbf{1}}) = \pi(V(P_{\mathbf{1}})) = \frac{\alpha \mathbf{1}}{\sqrt{\langle \alpha \mathbf{1}, \alpha \mathbf{1} \rangle}} = \mathbf{1}.$$

Consider now an arbitrary $\overline{x} \in H_n$, and a distribution $P_{\overline{x}}$ isotropic with respect to $\overline{x}$. As seen in section 2.3, there is a matrix $M \in SO^+(1,n)$ such that $M\overline{x} = \mathbf{1}$. It can be checked that $MP_{\overline{x}}$ is isotorpic with respect to $\mathbf{1}$. Hence

$$V(MP_{\overline{x}}) = MV(P_{\overline{x}}) \in \mathbb{R}.\mathbf{1}$$

and

$$V(P_{\overline{x}}) \in \mathbb{R}.M^{-1}\mathbf{1} = \mathbb{R}.\overline{x}.$$

We can then conclude that

$$E(P_{\overline{x}}) = \pi(V(P_{\overline{x}})) = \overline{x},$$

and state the following theorem.

Theorem 2 *Let P be a probability distribution on H_n isotropic with respect to $\overline{x}$ and such that the vector mean exists. Then*

$$E(P) = \overline{x}.$$

5 Convergence

Under some finiteness assumptions of second order quantities detailed in Theorem 2.1 of [4], we get from the same reference that the expected square error of the empirical estimation of the Fréchet mean decreases in $\frac{1}{N}$.

In order to show the practical interest of the extrinsic mean, we also establish that, unsurprisingly, the expected square error of the empirical estimation is bounded above by an equivalent of $\frac{1}{N}$. The error is actually equivalent to $\frac{1}{N}$ but the proof is slightly more technical.

Let (X_i) be independent random variables $\Omega \to H_n$ of distribution P. Note V_N the vector mean

$$V_N = \frac{X_1 + \ldots + X_N}{N}$$

and E_N the extrinsic mean of the empirical distribution $\frac{1}{N}\sum_i \delta_{X_i}$.

Assume that the P have a finite second order Euclidean moment. We known that

$$\mathbb{E}(\|V_N - V(P)\|^2) \sim \frac{1}{N},$$

where $\|.\|$ is the Euclidean norm of $\mathbb{R}^{n+1}$. Assume now that P has a compact support. This assumption can be relaxed, but it enables to obtain the desired result in a straightforward way. Choose $m \in \mathbb{R}$ such that the support of P is contained in

$$U = \{x = (x_0, \ldots, x_n) : 1 \leq x_0 \leq m, \langle x, x \rangle \geq 1\}.$$

By convexity, U contains the vector means $V(P)$ and V_N. The projection π is continuously differentiable in the cone C_+, hence its restriction to the compact set U is a Lipshitz map, i.e., there exists a positive constant L such that for all $x, x' \in E$,

$d(\pi(x), \pi(x')) \leq L\|x - x'\|$ where $d(.,.)$ is the hyperbolic distance. It follows that

$$d(E_N, E(P)) = d(\pi(V_N), \pi(V(P))) \leq L\|V_N - V(P)\|$$

which implies,

$$\mathbb{E}(d(E_N, E(P))^2) \leq L^2 \mathbb{E}(\|E_N - E(P)\|^2) \sim \frac{1}{N}.$$

6 Conclusion

In this paper, we have shown that for isotropic distributions on hyperbolic spaces, the extrinsic mean is a legitimate alternative to the the Fréchet mean. The mean points coincide, and the analysis of empirical errors shows similar rates. However, the extrinsic mean has an explicit expression while the computation of the Fréchet mean requires to minimize the function $F(x)$ of Eq.3. The result presented in this paper can be generalized in two direction. On the one hand, it can be generalized to manifolds with fewer symmetries than the hyperbolic space such as symmetric positive definite matrices. On the other hand on hyperbolic spaces, the isotropy condition can be relaxed to a weaker symmetry condition, allowing the anisotropic distributions proposed in [8, 14, 13].

Références

[1] Arnaudon, M., Barbaresco, F., Yang, L. (2013). Riemannian medians and means with applications to radar signal processing. IEEE Journal of Selected Topics in Signal Processing, 7(4), 595-604.

[2] Asta, D. and Shalizi, C. (2015). "Geometric Network Comparisons," Proceedings of the 31st Annual Conference on Uncertainty in AI (UAI) .

[3] Barbaresco, F. (2009). "New foundation of radar Doppler signal processing based on advanced differential geometry of symmetric spaces : Doppler matrix CFAR and radar application." International Radar Conference. Vol. 82. France : Bordeaux

[4] Bhattacharya, R., Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. The Annals of Statistics, 31(1), 1-29.

[5] Bose, J., Smofsky, A., Liao, R., Panangaden, P., Hamilton, W. (2020, November). Latent variable modelling with hyperbolic normalizing flows. In International Conference on Machine Learning (pp. 1045-1055). PMLR.

[6] Bruhat, F., Tits, J. (1972). Groupes reductifs sur un corps local, Publ. Math. IHES 41 5–252.

[7] Chevallier, E., Barbaresco, F., Angulo, J. (2015, October). Probability density estimation on the hyperbolic space applied to radar processing. In International Conference on Geometric Science of Information (pp. 753-761). Springer, Cham.

[8] Chevallier, E. (2017, November). A family of anisotropic distributions on the hyperbolic plane. In International Conference on Geometric Science of Information (pp. 717-724). Springer, Cham.

[9] Ding, J., Regev, A. (2021). Deep generative model embedding of single-cell RNA-Seq profiles on hyperspheres and hyperbolic spaces. Nature communications, 12(1), 1-17

[10] Farup, Ivar. (2014). "Hyperbolic geometry for colour metrics." Optics Express 22.10 : 12369-12378.

[11] Galperin, G. A. (1993). A concept of the mass center of a system of material points in the constant curvature spaces. Communications in Mathematical Physics, 154(1), 63-84.

[12] Grattarola, D., Zambon, D., Livi, L., Alippi, C. (2019). Change detection in graph streams by learning graph embeddings on constant-curvature manifolds. IEEE Transactions on neural networks and learning systems, 31(6), 1856-1869.

[13] E. Mathieu, C. Le Lan, C.J. Maddison, R. Tomioka, Y.W. Teh, Y. W. (2019). Continuous hierarchical representations with poincaré variational auto-encoders. Advances in neural information processing systems, 32.

[14] Nagano, Y., Yamaguchi, S., Fujita, Y., Koyama, M. (2019, May). A wrapped normal distribution on hyperbolic space for gradient-based learning. In International Conference on Machine Learning (pp. 4693-4702). PMLR.

[15] Nickel, M., Kiela, D. (2017). Poincaré embeddings for learning hierarchical representations. Advances in neural information processing systems, 30.

[16] Provenzi, Edoardo. (2020). "Geometry of color perception. Part 1 : structures and metrics of a homogeneous color space." The Journal of Mathematical Neuroscience 10.1 : 1-19.

[17] S. Said, L. Bombrun, Y. Berthoumieu, (2014). New Riemannian priors on the univariate normal model. Entropy, 16(7), 4015-4031.