Four Lectures on Foundations of Statistics on Non-Euclidean Spaces Held at The Second Conference of Interactions of Statistics and Geometry (ISAG II, Oct 14-25, 2024) at the Institute for Mathematical Science (IMS) of the National University of Singapore (NUS)

Stephan Huckemann

December 30, 2024

Contents

1 Motivation and Generalized Fréchet Means

1.1 Intro and Brief Statistics Recap

Given data (a sample) $X_1, \dots, X_n \overset{i.i.d.}{\sim} X$ in some metric data space (Q, d) of "dimension" $m \in \mathbb{N}$, i.e.

$$X_i, X : (\Omega, \mathcal{A}, \mathbb{P}) \to (Q, \mathcal{B}(Q)) \text{ measurable }, i = 1, \dots, n\,,$$
$$\mathbb{P}\{X_1 \in B_1, \dots, X_n \in B_n\} = \Pi_{j=1}^n \mathbb{P}\{X_j \in B_j\} \text{ for all } B_1, \dots, B_n \in \mathcal{B}(Q)\,.$$

our most important tool is:

$$\mathbb{E}[f \circ X] := \int_\Omega f \circ X(\omega)\, d\mathbb{P}(\omega) = \int_Q f(x)\, d\mathbb{P}^X(x)$$

Think first of $Q = \mathbb{R}^m$ but aim at more general spaces, see further below.

1. Dimension reduction:
   - the typical data point (dimension 0), e.g. their mean
   - their main mode of variation (dimension 1),
   - represent the data in a suitable k-dimensional subspace $0 \leq k < m$ (principal component analysis (PCA), multidimensional scaling (MDS).
2. Parametric models, regression, clustering, classification, discrimination, etc.
3. Asymptotic inference from samples:
   - sample descriptor (estimator) $\hat{\mu}_n \to \mu$ population descriptor (parameter)

$$\hat{\mu}_n \overset{\text{a.s.}}{\to} \mu \quad \Leftrightarrow \quad \mathbb{P}\{\lim_{n\to\infty} d(\hat{\mu}_n, \mu) > 0\} = 0$$
$$\hat{\mu}_n \overset{\mathbb{P}}{\to} \mu \quad \Leftrightarrow \quad \lim_{n\to\infty} \mathbb{P}\{d(\hat{\mu}_n, \mu) > \epsilon\} \to 0 \quad \forall \epsilon > 0$$
$$\mu_n \overset{\mathcal{D}}{\to} Y \quad \Leftrightarrow \quad \mathbb{E}[f \circ \mu_n] \to \mathbb{E}[f \circ Y] \quad \forall f : Q \to \mathbb{R} \text{ bounded and continuous}$$

   - two-sample tests, inference on models, e.g. via MANOVA, model selection?

We build on

Theorem 1.1. *Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} X$ be random variables on $Q = \mathbb{R}^m$.*

Strong Law of Large Numbers (LLN): *if $\mathbb{E}[\|X\|] < \infty$, then*

$$\bar{X}_n := \frac{1}{n} \sum_{j=1}^n X_j \overset{\text{a.s.}}{\to} \mathbb{E}[X]\,.$$

Central Limit Theorem (CLT): *if $\mathbb{E}[\|X\|^2] < \infty$, then*

$$\sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \operatorname{cov}[X]) \text{ where } \operatorname{cov}[X] := \mathbb{E}\big[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\big]\,.$$

1.2 Examples

Canonical smooth (quasi-)metrics on spheres $\mathbb{S}^{k\times(m-1)-1} = \{X \in \mathbb{R}^{k\times m} : \|X\| = 1\}$:

- extrinsic (chordal) $d_e(X, Y) = \|X - Y\| = \sqrt{2(1 - \langle X, Y\rangle)}$,
- intrinsic (spherical) $d_i(X, Y) = \arccos\langle X, Y\rangle$,
- residual (tangent space projection) $d_r(X, Y) = \|X - \langle X, Y\rangle Y\| = \sqrt{1 - \langle X, Y\rangle^2}$,

where $\langle X, Y\rangle := \operatorname{tr}(X^T Y)$.

1.3 Principal Component Analysis (PCA)

Find an affine subspace of dimension $k \in \{0, 1 \ldots, m-1\}$ best approximating a random variable X in $\mathbb{R}^m$ with $\mathbb{E}[\|X\|^2] < \infty$.

Parametrize each such subspace

$$A_{V,\alpha} := \{x \in \mathbb{R}^m : V^T x = \alpha\}$$

which has $m-k$ o.n. vectors $v_i \in \mathbb{R}^m$ to it, each translated by a certain amount $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, m-k$. In particular, every element in the *o.n. Stiefel manifold*

$$V \in O(m, m-k) := \{V = (v_1, \ldots, v_{m-k}) \in \mathbb{R}^{m \times (m-k)} : v_i^T v_j = \delta_{ij}, 1 \leq i, j \leq m-k\}$$

represents $A_{V,0}$ and $(V, \alpha), (V', \alpha') \in O(m, m-k) \times \mathbb{R}^{m-k}$ represent the same subspace if and only if there is a $S \in O(k)$ with $S.(V, \alpha) := (VS, S^T \alpha)$.

Theorem 1.2. *The family of k-dimensional affine subspaces in $\mathbb{R}^m$ is uniquely parametrized by the smooth manifold*

$$A(m, m-k) = (O(m, m-k) \times \mathbb{R}^{m-k})/O(m-k)$$

of dimension $m(m-k) + m - (m-k)(m-k+1)/2 - (m-k)(m-k-1)/2 = m + (m-k)k$ with canonical Ziezold metric if $O(m, m-k)$ carries the extrinsic metric.

Proof. Apply Theorem 4.2 since $O(k)$ acts freely, isometrically and properly ($O(k)$ is even compact). □

Remark 1.3. *$O(m, m-k)/O(m-k) =: G(m, m-k) \cong G(m, k) := O(m, m-k)/O(m-k)$ are the* Grassmannian *manifolds.*

Theorem 1.4 (Euclidean PCA is Nested)**.** *With the distance*

$$\rho(y, A_{V,\alpha}) = \inf_{x \in A_{V,\alpha}} \|y - x\| = \|V^T y - \alpha\| \text{ for } y \in \mathbb{R}^m \tag{1.1}$$

(apply Lagrange minimization) and random X in $\mathbb{R}^m$ with $\mathbb{E}[\|X\|^2] < \infty$ and spectral decomposition

$$\mathrm{cov}[X] = (v_1, \ldots, v_m)\,\mathrm{diag}(\lambda_1, \ldots, \lambda_m)(v_1, \ldots, v_m)^T, \quad \lambda_1 \geq \ldots \lambda_m \geq 0,\ (v_1, \ldots, v_m) \in SO(m)\,,$$

then, for each $k \in \{0, \ldots, m-1\}$

$$\mathrm{argmin}_{O(m-k).(V,\alpha) \in A(m,m-k)}\, \mathbb{E}[\rho(X, A_{V,\alpha})]$$

has a solution with representative $V = (v_{k+1}, \ldots, v_m)$, $\alpha = V^T \mathbb{E}[X]$, i.e. $A_{V,\alpha}$ is spanned by $v_1, \ldots, v_k$, with base point $\mathbb{E}[X]$. The space is unique if and only if $\lambda_k > \lambda_{k+1}$. For $k = 0$ it is uniquely $\{\mathbb{E}[X]\}$.

This leads to a *flag* of subspaces (*complete*) if all eigenvalues are different carrying canonical metrics/distances, see Definition 4.3.

1.4 Principal Nested Spheres (PNS)

PNS by Jung et al. (2012). For X on a sphere $\mathbb{S}^m \subset \mathbb{R}^{m+1}$, every k-dimensional small subsphere is some

$$A_{V,\alpha} \cap \mathbb{S}^m \text{ for some } A_{V,\alpha} \in A(m+1, m-k) \text{ with } \|\alpha\| < 1\,.$$

Thus, imposing a spherical distance (for which Pythagoras' theorem is not valid) instead of the Euclidean in (1.1), the corresponding subspheres are in general no longer nested, but one can enforce this by *principal backward nested spheres* (PNS) analysis, yielding a (random, see next section) *flag* of affine subspaces:

$$\mathbb{R}^{m+1} = A_{V^{(m+1)},\alpha^{(m+1)}} \supset A_{V^{(m)},\alpha^{(m)}} \supset \ldots \supset A_{V^{(2)},\alpha^{(2)}} \supset\in A_{V^{(1)},\alpha^{(1)}} = \{p\}$$

with $A_{V^{(k)},\alpha^{(k)}} \in A(m+1, m-k)$ and $\|\alpha^{(1)}\| = 1$.

This motivates the following section.

1.5 Generalized Fréchet Means

Definition 1.5. *Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} X$ be Borel random variables on a topological space Q (data space) linked to a metric space (P, d) (descriptor space) via a continuous function $\rho : Q \times P \to \mathbb{R}$. Then*

$$F(p) := \mathbb{E}[\rho(X, p)], \quad F_n(p) := \frac{1}{n}\sum_{j=1}^{n} \rho(X_i, p)$$

are the (random) generalized Fréchet functions *and the (random) closed sets*

$$E[X] := \operatorname{argmin}_{p\in P} F(p)\,, \quad E_n := \operatorname{argmin}_{p\in P} F_n(p)$$

is the (possibly empty) sets of generalized population *and* sample Fréchet means*, respectively.*

If $P = Q$ and $\rho = d^2$ then this definition gives classical Fréchet means *as introduced by Fréchet (1948).*

We require $F(p) < \infty$ for all $p \in E[X]$.

Above, P can be taken as a flag manifold (PCA,PNS), or one of the $A(m, m-k)$, with suitable ρ, see, e.g. Lecture III and with more notational effort, this can be generalized to more general nested descriptor sequences, cf. Huckemann and Eltzner (2018) with application to stem cell differentiation.

Remark 1.6 (FMs are Extension of Euclidean Means)**.** *Indeed, for a classical random vector X in $\mathbb{R}^m$ with metric $d(p, p') = \|p - p'\|$, with $E[\|\|X\|^2] < \infty$, $p \mapsto \mathbb{E}[\|X - p\|^2], \mathbb{R}^m \to [0, \infty)$ is a well defined smooth function and*

$$\mathbb{E}[\|X - p\|^2] = \mathbb{E}[\|X\|^2] - 2p^T\mathbb{E}[X] + \|p\|^2$$

is uniquely minimized by $p = \mathbb{E}[X]$.

Does $E[\|\|X\|] < \infty$ suffice to define Fréchet means? Yes, see below.

Generalized Fréchet means are

- closed sets (as empty sets or as preimages of closed sets under a continuous function),
- nonempty sets, if ρ is bounded from below (e.g. $\rho \geq 0$) the descriptor space (P, d) is complete (Cauchy sequences converge), and the Fréchet function F is finite for at least one point $p \in P$ (in a metric space a set is closed if and only if it contains all of its cluster points),
- in particular, sample Fréchet mean sets are nonempty, if (P, d) is complete.

2 "Classical" Fréchet Means

Throughout this section consider a random variable X on a complete metric space $(Q, d) = (P, d)$ with $\rho = d^2$.

Theorem 2.1. *[Folklore, Mattner, and Sturm (2003b)] Fix $p_0 \in Q$ and consider*

$$F^{p_0}(p) := \mathbb{E}\big[d(X,p)^2 - d(X,p_0)^2\big]\,, \quad p \in Q\,.$$

Then

(i) $F^{p_0}(p)$ exists for all $p \in Q$ if $\mathbb{E}[d(X,p_0)] < \infty$,

(ii) if $\mathbb{E}\big[d(X,p_0)^2\big] < \infty$ then

$$E[X] = \operatorname{argmin}_{p\in Q} F^{p_0}(p)\,.$$

Proof. (i): Using twice the triangle inequality, observe,

$$\begin{aligned}
\mathbb{E}\big[\big|d(X,p)^2 - d(X,p_0)^2\big|\big] &= \mathbb{E}\Big[\big(d(X,p) + d(X,p_0)\big)\,\big|d(X,p) - d(X,p_0)\big|\Big] \\
&\le \Big(\mathbb{E}[d(X,p)] + \mathbb{E}[d(X,p_0)]\Big)\, d(p,p_0) \\
&\le \Big(d(p,p_0) + 2\mathbb{E}[d(X,p_0)]\Big)\, d(p,p_0)\,.
\end{aligned}$$

(ii): Using again the triangle inequality, observe,

$$\begin{aligned}
\mathbb{E}\big[d(X,p)^2\big] &\le \mathbb{E}\big[\,(d(X,p_0) + d(p_0,p))^2\,\big] \\
&= \mathbb{E}\big[d(X,p_0)^2\big] + 2d(p,p_0)\mathbb{E}\big[d(X,p_0)\big] + d(p,p_0)^2\,.
\end{aligned}$$

□

This gives the reason for the positive answer in Remark 1.6.

2.1 Extrinsic Means for Embeddings

Every $Q \subset \mathbb{R}^m$ and carries the *extrinsic metric* $d(p,q) = \|p-q\|$ for all $p,q \in Q$ (verify at once that is it a metric). Then for every $x \in \mathbb{R}^m$ its *orthogonal projection*

$$\Phi(x) := \operatorname{argmin}_{p\in Q} \|x-p\| = \operatorname{argmin}_{p\in Q} \|x-p\|^2$$

is a nonempty set if Q is closed.

The points $x \in \mathbb{R}^m$ with nonunique projection are called *focal points*. For instance, if Q is a proper ellipse then all points between its foci are focal points; for a sphere its center is a focal point.

Fréchet means $E[X]$ with respect to an extrinsic metric are called *extrinsic means*.

Theorem 2.2. *If $\mathbb{E}[X] \in \mathbb{R}^m$ exists then*

$$E[X] = \Phi(\mathbb{E}[X])\,.$$

If $\mathbb{E}[X]$ is nonfocal then $E[X]$ is unique.

Proof. First note that every minizer $p \in Q$ of

$$\mathbb{E}[\|X-p\|^2] = \mathbb{E}[\|X\|^2] - 2\mathbb{E}[X]^T(p-p_0) + \|p\|^2$$

is also a minimizer of

$$\|\mathbb{E}[X]-p\|^2 = \|\mathbb{E}[x]\|^2 - 2\mathbb{E}[X]^T(p-p_0) + \|p\|^2\,,$$

and vice versa. Hence $E[X] = \Phi(\mathbb{E}[X])$ in case of $\mathbb{E}[\|X\|^2] < \infty$. Next, utilize Theorem 2.1 to see that $\mathbb{E}[\|X\|] < \infty$ suffices. □

In consequence the spherical extrinsic Fréchet mean is either unique or the entire sphere. More generally, in the spirit of Sard's theorem we have:

Lemma 2.3 (Bhattacharya and Patrangenaru (2003))**.** *If Q is an embedded smooth manifold in $\mathbb{R}^m$ then its focal points form a closed set of Lebegue measure zero.*

2.2 Residual Means for Embedded Manifolds

We now assume that $Q \subset \mathbb{R}^m$ is an embedded k-dimensional smooth manifold, $1 \leq k < m$. Hence for every $p \in Q$ there is an open $U \subseteq \mathbb{R}^m$ with $p \in U$ and smooth $u : U \to \mathbb{R}^m$ such that

$$U \cap Q = \{x \in U : u^{k+1}(x) = \ldots = u^m(x) = 0\}\,.$$

Then, with $v_i := (\partial_i)_p$, $1 \leq i \leq m$[1] and $V = (v_1, \ldots, v_k)$

$$Q \to T_pQ, \quad q \mapsto \frac{VV^T}{\|V\|^2}q$$

is the *orthogonal projection* of Q to the *tangent space* $T_pM \subset \mathbb{R}^m$ at p. Further,

$$\frac{VV^T}{\|V\|^2}q \text{ and } d_r(p,q) := \left\|\frac{VV^T}{\|V\|^2}q\right\|$$

are called the *residual tangent space coordinate* of $q \in Q$ at p and the residual distance of q from p. While on spheres, the residual distance agrees with the definition in Section **??**, which is symmetric but does not allow for the triangle inequality, on general spaces it is not even symmetric. Still, residual Fréchet means can be defined on any embedded manifold.

For Procrustes analysis, the following is important

Theorem 2.4. *Let X be a random variable on a sphere $\mathbb{S}^{m-1} \subset \mathbb{R}^m$ with $2 \leq m \in \mathbb{N}$ with $\mathbb{E}[\|X\|^2] < \infty$. Then, with a spectral decomposition*

$$\mathbb{E}[XX^T] = V\Lambda V^T, \quad V = (v_1, \ldots, v_m) = V \in SO(m), \quad \Lambda = (\lambda_1, \ldots, \lambda_m), \quad \lambda_1 \geq \ldots \geq \lambda_m\,,$$

the residual mean set of X satisfies

$$E(X) = \mathbb{S}^{m-1} \cap \operatorname{span}\{v_1, \ldots, v_k\}$$

if $\lambda_1 = \ldots = \lambda_k$ and $k = m$ or if $\lambda_k > \lambda_{k+1}$. In particular, if $\lambda_1 > \lambda_2$ then

$$\mathbb{E}[X] = \{-v_1, v_1\}\,.$$

Proof. This follows rather directly from the fact that the o.g. projection to $T_p\mathbb{S}^{m-1}$ is given by $q \mapsto (I_m - pp^T)q$. □

2.3 Intrinsic Means

Let X be a random variable on a complete connected Riemannian manifold Q with the induced geodesic distance d and we consider *intrinsic Fréchet means*

$$E(X) = \operatorname{argmin}_{p \in Q} F(p) \text{ with } F(p) = \mathbb{E}[d(X,p)^2]\,.$$

Recall CAT(κ) spaces from Ezra's talk. For instance, a sphere of radius $\sqrt{\kappa}$, $\kappa > 0$, is a CAT(κ) manifold.

Theorem 2.5 (Afsari's Theorem (2011))**.** *If Q is additionally a CAT(κ) space, $\kappa > 0$ and if X has support*[2] *in a* geodesic half ball

$$B_{R_\kappa}(p_0) = \{p \in Q : d(p,p_0) < R_\kappa\} \textit{ with } R_\kappa = \frac{\pi}{2\sqrt{\kappa}}$$

about some $p_0 \in Q$, then X has a unique Fréchet mean and this is contained in $B_{R_\kappa}(p_0)$.

[1] for $f \in \mathcal{C}^\infty(M \to \mathbb{R})$, $(\partial_i)_q f := \frac{d}{dt} f \circ u^{-1}(q + te_i)$ with the i-th standard unit vector e_i, $1 \leq i \leq m$

[2] $\operatorname{supp}(X) = \bigcap_{\mathbb{P}\{X \in A\}=1, A \text{ closed}} A$

Recall that a connected Riemannian manifold Q is a *symmetric space* if for every $p \in Q$ there is a smooth isometric mapping $j : Q \to Q$ such that $j(\exp_p v) = \exp_p(-v)$ for all $v \in T_pM$. In particular, j is *involutive*, i.e. $j^2 = id_Q$.

Theorem 2.6 (Le's Theorem (1998)). *If Q is additionally a symmetric space and if X has a density $f : Q \to [0, \infty), p \mapsto f(p)$ with respect to the Riemannian volume with the property $f(p) = g(d(p, p^*))$ for some $p^* \in Q$ with a nonincreasing function function $g : [0, \infty) \to [0, \infty)$ that is strictly decreasing in (r_1, r_2) with some $0 \le r_1 < r_2 \le R_\kappa$, then p^* is the unique Fréchet mean of X.*

Proof. Let $p' \in Q$ be arbitrary. Since Q is connected and complete, it can be reached by a geodesic from p^*. Take an involutive isometry $j : Q \to Q$ with $j(p'') = p''$ for the midpoint p'' on that geodesic from p^* to p'. Then $j(p^*) = p'$ and let

$$A = \{z \in Q : d(p^*, z) < d(p^*, j(z))\}.$$

By hypothesis $A \cup j(A) \cup N = Q$ with a set N that Riemannian volume zero. Then the assertion follows from

$$\begin{aligned} F(p') - F(p^*) &= \int_Q \left(d(p', z)^2 - d(p^*, z)^2\right) g(d(z, p^*))\, d\,\mathrm{vol}(z) \\ &= \int_A \left(d(p', z)^2 - d(p^*, z)^2\right) g(d(z, p^*))\, d\,\mathrm{vol}(z) \\ &\quad + \int_A \left(d(p', j(z))^2 - d(p^*, j(z))^2\right) g(d(j(z), p^*))\, d\,\mathrm{vol}(z) \\ &= \int_A \left(d(p', z)^2 - d(p^*, z)^2\right) \left(g(d(z, p^*)) - g(d(j(z), p^*)\right)\, d\,\mathrm{vol}(z) \\ &> 0 \end{aligned}$$

since, by hypothesis $d(p', z) = d(p^*, j(z)) > d(p^*, z)$ and hence $g(d(z, p^*)) \ge g(d(j(z), p^*)$ for all $z \in A$, where, the latter inequality is strict for a set of positive volume. □

Remark 2.7. *Unaware of Le's result – she actually proved it for the special symmetric spaces of Σ_2^k, see Definition 4.5 below, the proof translates one-to-one to general symmetric spaces – Aveni and Mukherjee (2024) give the above proof in a more general setting.*

Sample means tend to be more unique than population means.

Theorem 2.8 (Arnaudon and Miclo (2014)). *If $X_1, \dots, X_n \overset{i.i.d.}{\sim} X$ is absolutely continuous with respect to Riemannian volume, then E_n is a.s. a unique point.*

A role of the cut points $C(p) := \{q \in Q :$ there are at least two length minimizing geodesics from p to $q\}$ for $p \in Q$. It is well known that the Riemannian volume of $C(p)$ is zero. Also:

Theorem 2.9 (Le and Barden (2014)). *Suppose that $\mu \in Q$ is an intrinsic Fréchet mean of X. Then $\mathbb{P}\{X \in C(\mu)\} = 0$.*

Notably, the *cut locus* is $\{q \in Q :$ every length minimizing geodesic from p to q is no longer length minimizing beyond $q\}$ and it can also include *conjugate points* which are not cut points.

Intrinsic means on the circle. Now, X is a random variable on $Q = \mathbb{S}^1 = \{e^{it} : -\pi \le t < \pi\} \subset \mathbb{C}$.

Theorem 2.10 (Hotz and Huckemann (2015)). *Consider the distribution of X, decomposed into the part which is absolutely continuous with respect to arc-length measure, with density f, and η, the part singular to arc-length measure. If*

1. $S_1, \dots, S_k$ are the distinct open arcs on which $f < (2\pi)^{-1}$,

2. they are all disjoint from $\mathrm{supp}(\eta)$,

3. and $\{x \in \mathbb{S}^1 : f(x) = (2\pi)^{-1}\}$ is a circular null-set,

then X has at most k intrinsic means and every antipodal arc of S_j contains at most one candidate, $1 \le j \le k$.

In particular, $f > (2\pi)^{-1}$ at the antipode of an intrinsic mean is not possible

This extends at once to the *torus* $\mathbb{T}^m := (\mathbb{S}^1)^m$.

Intrinsic means on Hadamard spaces. Euclidean spaces are CAT(0) spaces, so are hyperbolic spaces in information geometry, and so are BHV trees spaces. The latter are no longer manifolds. *Hadamard* spaces are complete CAT(0) spaces. Leaving the manifold world, returning to just metric spaces, we have:

Theorem 2.11 (Sturm (2003b))**.** *A X be a random variable on a Hadamard space (Q, d) with finite first moment $\mathbb{E}[d(X, p)] < \infty$ for some $p \in Q$ has a unique Fréchet mean.*

2.4 Computing Intrinsic Fréchet Sample Means

N.B.: The proof of Theorem 2.9 relies on showing that

$$\log_\mu = (\exp_\mu)^{-1} \qquad \text{is a.s. well defined} \tag{2.1}$$

Thus, for data on a Riemannian manifold, the most straightforward method is to find the root $p \in Q$ of

$$0 = \mathrm{grad}_p \, \mathbb{E}[d(X,p)^2] = \mathrm{grad}_p \, \mathbb{E}[\| \log_p X \|^2] = \mathrm{grad}_p \, \mathbb{E}[\| \log_X p \|^2] = -2\mathbb{E}[\log_p X] \,. \tag{2.2}$$

On Hadamard spaces there is an algorithm converging in probability and if $\mathrm{supp}(X)$ is bounded even a.s. (Sturm, 2003a):

Algorithm 2.12 (Sturm (2003a))**.** *Computing the Fréchet mean of a random variable X in a separable Hadamard space (Q, d) based on $X_1, X_2, \dots \overset{i.i.d.}{\sim} X$*

Iterate until $d(Z^{(k)}, Z^{(k-1)})$ is below a given threshold and return $Z^{(k)}$.

Step 1 *: Set $Z^{(1)} = X_1$*

Step $k \ge 2$ *: Set*

$$Z^{(k)} := \gamma\left(\frac{1}{k} d(Z^{(k-1)}, X_k)\right)$$

with the geodesic γ from $Z^{(k-1)}$ to X_k, parametrized by arc length

N.B.: the deterministic version also converges (Bacák, 2014).

On the circle the following gives an $O(n)$ algorithm to compute intrinsic sample means.

Corollary 2.13 (Hotz and Huckemann (2015))**.** *Let $X_1 = e^{iT_1}, \dots, X_n = e^{iT_n}$ be a sample on $\mathbb{S}^1$, $-\pi \le T_j < \pi$, $1 \le j \le n$ with $\bar{T}_n = \frac{1}{n}\sum_{j=1}^n T_j$. Then the candidates for their intrinsic sample mean are*

$$e^{i\left(\bar{T}_n + \frac{2\pi k}{n}\right)} \quad 1 \le k \le n$$

Proof. If $\mu = e^{i\nu}$, $-\pi \le \nu < \pi$ is an intrinsic mean then $\log_\mu X_j = T_j + 2\pi k_j$ with suitable $k_j \in \{-1, 0, 1\}$ such that $-\pi + \nu < T_j + 2\pi k_j < \pi + k_j$, $1 \le j \le m$ and by (2.2), $\nu = \bar{X}_n + \frac{1}{n}\sum_{j=1}^n 2\pi k_j$. □

This extends at once to the torus $\mathbb{T}^m := (\mathbb{S}^1)^m$.

3 Some Asymptotics for Generalized Fréchet Means

In this section we return to Definition 1.5, $X_1, \dots, X_n \overset{i.i.d.}{\sim} X$ are Borel random variables on a topological space Q (data space) linked to a separable metric space (P, d) (descriptor space) via a continuous $\rho : Q \times P \to \mathbb{R}$ giving rise to

$$F(p) := \mathbb{E}[\rho(X,p)], \quad F_n(p) := \frac{1}{n}\sum_{j=1}^{n} \rho(X_i, p)$$

are the (random) *generalized Fréchet functions* and the (random) closed sets

$$E[X] := \operatorname{argmin}_{p\in P} F(p)\,, \quad E_n := \operatorname{argmin}_{p\in P} F_n(p)$$

is the (possibly empty) sets of *generalized population* and *sample Fréchet means*, respectively.

3.1 Strong Laws of Large Numbers

We require additionally:

Assumption 3.1. *Assume that there exist*

1. $\dot\rho : Q \times P \to [0,\infty)$ *continuous in* P *and measurable in* Q *with* $\mathbb{E}[\dot\rho(X,p)] < \infty$ *for all* $p \in P$,
2. $h : [0,\infty) \to [0,\infty)$ *continuous with* $h(0) = 0$*, and*
3. $\delta > 0$*, such that for every* $p, p' \in P$ *with* $d(p,p') < \delta$ *and all* $q \in Q$,
$$|\rho(q,p) - \rho(q,p')| \le \dot\rho(q,p)\ h\big(d(p,p')\big)\,.$$

Assumption 3.2. $\exists C \le \infty$ *such that* $\forall$ *random discrete (no cluster points) sequences* $p_n \in P$,

$$\liminf_{n\to\infty} \rho(X, p_n) \ge C \ a.s.$$

Theorem 3.3 (Version by Wiechers et al. (2023))**.** *Under Assumption 3.1,* Ziezold strong consistency *(ZSC, Ziezold (1977)) holds*

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_k} \subseteq E(X), \quad a.s. \tag{3.1}$$

If, additionally, $E(X) \neq \emptyset$*, then, under Assumption 3.2,* Bhattacharya-Patrangenaru strong consistency *(PBSC, Bhattacharya and Patrangenaru (2003)) holds:*

$$\forall \epsilon > 0\ \exists\ a.s. N_\epsilon \in \mathbb{N}\,\text{such that}\ \bigcup_{k=n}^{\infty} E_k \subseteq \{p \in P : d(E(X), p) \le \epsilon\}\ \forall n \ge N_\epsilon\,.$$

Remark 3.4. *Assumption 3.1 covers*

1. *quasi-metrics for* $P = Q$, $\rho = d^2$ *(original proof for ZSC by Ziezold (1977)), there* $\dot\rho = 1$ *and* $h(t) = t$
2. *metrics for* $P = Q$, $\rho = d^2$ *(original proof for BPSC by Bhattacharya and Patrangenaru (2003), for* (Q,d) *is Heine-Borel), again there* $\dot\rho = 1$ *and* $h(t) = t$,
3. *in fact,* (P,d) *is Heine-Borel implies Assumption 3.2, cf. Wiechers et al. (2023),*
4. *(flags for) PCA and PNS (Huckemann and Eltzner, 2018), geodesic PCA (Huckemann et al., 2010).*
5. *unlcear whether it covers barycentric subspaces (Pennec, 2018).*
6. *minimizing negative log-likelihoods, for this reason* ρ *may be negative as well (estimating drift models for ENDOR in Wiechers et al. (2023), say), the version with nonnegative* ρ *is from Huckemann (2011b).*
7. *Schötz (2022) and Evans and Jaffe (2024) have further extended SLLNs.*

3.2 The Central Limit Theorem

Now, P is a m-dimensional Riemannian manifold with induced geodesic distance d and we require additionally:

Assumption 3.5. *Assume that*

1. $\mu \in P$ *is the unique generalized Fréchet mean of a random variable* X *on* Q
2. μ_n *is a measurable selection of sample means* $\stackrel{\mathbb{P}}{\to} \mu$
3. $\phi : U \to \mathbb{R}^m$ *is local chart at* $p \in U$ *with* $\phi(\mu) = 0$
4. $x \mapsto \tau(X, x) := \rho(X, \phi^{-1}(x))$ *is a.s. smooth*
5. $\Sigma := \operatorname{cov}[\operatorname{grad}_2 \tau(X, 0)]$ *and* $H := \mathbb{E}(\operatorname{Hess}_2 \tau(X, x))$ *exist for* $x = 0$,
6. $\operatorname{Hess} F_n(\tilde{x}_n) \stackrel{\mathbb{P}}{\to} H$ *for every random sequence* $\tilde{x}_n \to 0$ *with* $\tilde{x}_n$ *between* x_n *and* 0.

Theorem 3.6 (Bhattacharya and Patrangenaru (2005))**.** *Under Assumption 3.5 we have*

$$\sqrt{n} H \phi(\mu_n) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, \Sigma)$$

and, if H *is invertible,*

$$\sqrt{n}\phi(\mu_n) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0, H^{-1}\Sigma H^{-1})\,. \tag{3.2}$$

Proof. Let $x_n = \phi(\mu_n)$ and in abuse of notation write $F(x), F_n(x)$ for $F \circ \phi^{-1}(x)$ and $F_n \circ \phi^{-1}(x) as$, respectively. Then

$$\begin{aligned} 0 &= \sqrt{n}\, \operatorname{grad} F_n(x_n) \\ &= \sqrt{n}\, (\operatorname{grad} F_n(0) + \operatorname{Hess} F_n(\tilde{x}) x_n) \\ &= \sqrt{n}\frac{1}{n}\sum_{j=1}^{n} \operatorname{grad}_2 \tau(X_j, 0) + \operatorname{Hess} F_n(\tilde{x})\sqrt{n} x_n \end{aligned}$$

with some $\tilde{x}$ between x_n and 0. Then, the assumptions in combination with Slutky's Lemma for real valued random variables

$$Y_n \stackrel{\mathcal{D}}{\to} Y, Z_n \stackrel{\mathbb{P}}{\to} z \text{ (deterministic) } \Rightarrow Z_n Y_n \stackrel{\mathcal{D}}{\to} zY\,,$$

yield the assertion.

□

Remark 3.7. *Bhattacharya and Patrangenaru (2005) were the first to prove along the above lines the case of intrinsic means on manifolds, see also Bhattacharya and Lin (2017). The general version for generalized Fréchet means is from Huckemann (2011a).*

A more refined CLT (Benjamin's tutorial) will deal making $\operatorname{Hess} F_n(\tilde{x}_n) \stackrel{\mathbb{P}}{\to} H$ *more accessible as well as dealing with* H *not of full rank. The latter leads to* smeariness*.*

A special case is a new and quick proof of Anderson (1963) and its extensions, e.g. Davis (1977); Tyler (1981).

Theorem 3.8. *Let* $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} X$ *be a Gaussian r.v. on* $\mathbb{R}^m$ *with* $\mathbb{E}[X] = 0$, *with population PCs* $\gamma_1, \ldots, \gamma_m$ *and corresponding eigenvalues* $\lambda_1 \geq \ldots \geq \lambda_m$ *as in Lecture 1. If* λ_k *is simple and* $\hat{\gamma}_k$ *the* k*-th sample PC then*

$$\sqrt{n}(\hat{\gamma}_k - \gamma_k) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(0, \sum_{k \neq j=1}^{m} \frac{\lambda_k \lambda_j}{(\lambda_k - \lambda_j)^2} \gamma_j \gamma_j^T\right) \stackrel{\mathcal{D}}{\leftarrow} \sqrt{n}(\hat{\gamma}_k - \hat{\gamma}_k \gamma_k^T \gamma_k)$$

Proof. Using the residual distance and coordinates, H can be explicitly computed and Hess $F_n(\tilde{x}_n) \xrightarrow{\mathbb{P}} H > 0$ can be verified, see Huckemann and Eltzner (2019) for $k = 1$. $\square$

Remark 3.9. *1. For asymptotics of higher dim. subspaces such as repeated eigenvalues, with the Grassmannian-like $A(m, m-k)$ as in Theorem 1.4 define*

$$\rho(X, A_{W,0}) = \|X - WW^T X\|^2 \,,$$

see Huckemann and Eltzner (2020). There, the CLT of Eltzner and Huckemann (2019), see Benjamin's talk, can be employed.

2. Rabenoro and Pennec (2022) obtain similar CLTs, also with explicit covariances and Hessians, also for entire flags, using a geodesic, not residual distances and coordinates.

3. The CLT is holds also for the asymptotics of suitable general nested subspaces, e.g. for PNS flags, see Huckemann and Eltzner (2018).

3.3 Two-Sample Tests and the Bootstrap

If H is not available and thus ignored, quantile based tests have wrong level. Hence, resort to bootstrapping.

The Euclidean two-sample Hotelling T^2 test. For two independent samples $X_1, \ldots, X_{n_1} \overset{i.i.d.}{\sim} \mathcal{N}(\mu_1, \Sigma_1)$, $Y_1, \ldots, Y_{n_2} \overset{i.i.d.}{\sim} \mathcal{N}(\mu_2, \Sigma_2)$, $\mu_1, \mu_2 \in \mathbb{R}^m$, $\Sigma_1, \Sigma_2 \in SPD(m)$ with $X = (X_1, \ldots, X_{n_1}), Y = (Y_1, \ldots, Y_{n_2})$ and the *samples' Mahalanobis distance*

$$d(X, Y) := \sqrt{(\bar{X} - \bar{Y})' \Big((n_1 + n_2 - 2)^{-1} \big(n_1 \operatorname{Scov}(X) + n_2 \operatorname{Scov}(Y) \big) \Big)^{-1} (\bar{X} - \bar{Y})} \,,$$

under $H_0 : \mu_1 = \mu_2, \Sigma_1 = \Sigma_2$ we have

$$\hat{T}^2 := \frac{n_1 n_2}{n_1 + n_2} d(X, Y)^2 \sim T^2(m, n_1 + n_2 - 2) \sim \frac{m(n_1 + n_2 - 2)}{n_1 + n_2 - 1 - m} F_{m, n_1 + n_2 - 1 - m} \,.$$

Here $\bar{X} := \frac{1}{n_1} \sum_{j=1}^{n_1} X_j$, $\bar{Y} := \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j$ and

$$\operatorname{Scov}(X) = \frac{1}{n_1} \sum_{j=1}^{n_1} (X_j - \bar{X})(X_j - \bar{X})^T, \quad \operatorname{Scov}(Y) = \frac{1}{n_2} \sum_{j=1}^{n_2} (Y_j - \bar{Y})(Y_j - \bar{Y})^T \,.$$

Reject H_0 if $\hat{T}^2 > F_{m, n_1 + n_2 - 1 - m, 1 - \alpha}$ ($1 - \alpha$ quantile) at level $\alpha \in (0, 1)$; typically $= 0.05$ for *significance* and 0.01 for *high significance*.

This two-sample test for $\mathbb{E}[X] = \mathbb{E}[Y]$ is asymptotically ($n_1, n_2 \to \infty$) robust under nonnormality if $n_1/n_2 \to 1$ or $\operatorname{cov}[X] = \operatorname{cov}[Y]$ (Romano and Lehmann, 2005, Scn. 11.3).

"Naïve" applications to manifold data in a local chart ϕ at pooled sample Fréchet mean $\mu_{n_1+n_2}$, replace $\bar{X}$ by $\overline{\phi(X)}_{n_1}$, $\bar{Y}$ by $\overline{\phi(Y)}_{n_2}$, $\operatorname{Scov}(X)$ by $\operatorname{Scov}(\phi(X))$, $\operatorname{Scov}(Y)$ by $\operatorname{Scov}(\phi(Y))$, but miss H and thus the asymptotic level α, see Benjamin's tutorial.

Therefore bootstrap: With the sample Fréchet means μ_{n_1} and μ_{n_2} of the $X_1, \ldots, X_{n_1}$ and the $Y_1, \ldots, Y_{n_2}$, respectively,

(i) for $1 \leq b \leq B$ (large, e.g. 1000), resample $X_1^{*,b}, \ldots, X_{n_1}^{*,b} \overset{i.i.d.}{\sim} \frac{1}{n_1} \sum_{j=1}^{n_1} \delta_{X_j}$ and $Y_1^{*,b}, \ldots, Y_{n_2}^{*,b} \overset{i.i.d.}{\sim} \frac{1}{n_2} \sum_{j=1}^{n_2} \delta_{Y_j}$ to obtain sample Fréchet means $\mu_{n_1}^{*,b}, \mu_{n_2}^{*,b}$, respectively, to obtain with the same chart ϕ as before,

$$d_1^{*,b} := \phi(\mu_{n_1}^{*,b}) - \phi(\mu_{n_1}), \quad d_2^{*,b} := \phi(\mu_{n_2}^{*,b}) - \phi(\mu_{n_2})$$

and

$$\Sigma_1^* := \frac{1}{B}\sum_{b=1}^{B} d_1^{*,b}(d_1^{*,b})^T, \quad \Sigma_2^* := \frac{1}{B}\sum_{b=1}^{B} d_2^{*,b}(d_2^{*,b})^T, \quad A := \Sigma_1^* + \Sigma_2^*;$$

(ii) Bootstrap independently again, as above $d_1^{*,b}$ and $d_2^{*,b}$, to obtain

$$(T^{*,b})^2 := (d_1^{*,b} - d_2^{*,b})^T A^{-1}(d_1^{*,b} - d_2^{*,b}), \quad 1 \le b \le B$$

to obtain an empirical $1-\alpha$ quantile $(T^*)^2_{1-\alpha}$.

Then, rejecting H_0 if

$$\big(\phi(\mu_{n_1}) - \phi(\mu_{n_1})\big)^T A^{-1}\big(\phi(\mu_{n_1}) - \phi(\mu_{n_1})\big) > (T^*)^2_{1-\alpha}$$

is a test, under no smeariness (see Benjamin's talk), of asymptotic level $\alpha \in (0,1)$ (Eltzner and Huckemann, 2017). Notably, for increased power, Step (ii) mimics H_0 even if H_1 is true.

4 Some Statistics on Stratified Spaces: Manifold Stability, Optimal Lifting, and More

4.1 Shape Spaces

Example of micromolecules. Given locations of k atom nuclei $x_1, \dots, x_k \in \mathbb{R}^m$ (here, $m = 3$), describe their orientation preserving Euclidean (modulo translation and rotation) shape.

- $X = (x_1, \dots, x_m) \in \mathbb{R}^{m\times k}$ is a *landmark configuration matrix*
- $X \sim Y$ if exists $R \in SO(m)$, $a \in \mathbb{R}^m$ such that $RX + a \cdot 1_k^T = Y$
- here, $O(m) = \{A \in \mathbb{R}^{m\times m} : A^{-1} = A^T\}$ and $SO(m) = \{A \in O(m) : \det(A) = 1\}$,
- $[X] := \{Y \in \mathbb{R}^{m\times k} : Y \sim X\}$ and
- $S\Sigma_m^k := \{[X] : X \in \mathbb{R}^{m\times k}\}$ is the *size-and-shape space*
- $d([X],[Y]) := \inf_{R\in SO(m), a\in\mathbb{R}^m} \|RX + a\cdot 1_k^T - Y\|$ where

$$\|X - Y\|^2 := \operatorname{tr}\big((X-Y)^T(X-Y)\big).$$

Note (isometric action):

$$\inf_{\substack{R_i \in SO(m)\\ a_i \in \mathbb{R}^m\\ i=1,2}} \|R_1X + a_1\cdot 1_k^T - R_2Y - a_2\cdot 1_k^T Y\| = \inf_{R\in SO(m), a\in\mathbb{R}^m} \|RX + a\cdot 1_k^T - Y\|.$$

Hence, consider only centered configurations X, i.e. $X1_k = 0$, or, more subtlely with any o.g. complement H of $\frac{1}{\sqrt{k}}1_k$, i.e. $\left(H\middle|\frac{1}{\sqrt{k}}1_k\right) \in SO(k)$

$$XH \in \mathbb{R}^{m\times(k-1)},$$

e.g. via Helmertizing (Dryden and Mardia, 2016).

Definition 4.1. *A Lie group G (a group that is an analytic manifold with analytic group operations) acts on a smooth ($\mathcal{C}^\infty$) manifold that is a metric space (M,d) with smooth d if*

- *$g : M \to M, p \mapsto g.p$ is well defined and smooth or all $g \in G$,*
- *$(gh).p = g.(h.p)$ for all $g,h \in G$ and $p \in M$,*

- *$e.p = p$ for all $p \in M$ where e is the unit element in G.*

The action is

- *free if $I_p = \{e\}$ (unit element) for all $p \in M$, where $I_p = \{g \in G : g.p = p\}$ (isotropy group at p)*
- *isometric if $d(g.p, g.p') = d(p, p')$ for all $g \in G$ and $p, p' \in M$*
- *proper if preimages of compact sets are compact again, i.e. $\{(g,p) \in G \times M : (p, g.p) \in K\}$ is compact whenever $K \subset M \times M$ is compact.*

Set $M^* := \{p \in M : I_p = \{e\}\}$.

Theorem 4.2. *If a Lie group G acts isometrically and properly on a smooth manifold M with smooth metric d then*

$$Q := M/G = \{G.p : p \in M\} \text{ with } d_Q(G.p, G.p') := \min_{g \in G} d(g.p, p'), \text{ for all } p, p' \in M$$

is a metric Hausdorff space and M^* is open in M. If $M^* \neq \emptyset$, then moreover $Q^* := M^*/G$ carries a canonical smooth structure of a manifold of dimension $\dim(M) - \dim(G)$ with

$$T_pM \cong T_p[p] \oplus T_{[p]}Q^* \text{ for all } p \in M^*$$

Proof. The first assertions is rather straightforward. More technically, due to the free action, the implicit function theorem can be invoked, then providing for local charts □

Definition 4.3. *If, with the above assumptions, $M \hookrightarrow \mathbb{R}^m$ is embedded then with extrinsic metric d_e, residual distance d_r and intrinsic metric d_i, the quotient distance*

$$d_Q(G.p, G.p') := \min_{g \in G} d(p, g.p'), \quad p, p' \in M$$

is the

Ziezold metric *for $d = d_e$,*

intrinsic metric *for $d = d_i$, and*

Procrustean distance *for $d = d_r$ (Gower, 1975).*

Note that for $m \geq 3$, if all landmark columns of $X \in \mathbb{R}^{m \times k}$ lie on a line through the origin, then rotation about this axis leaves X invariant, i.e. $[X] \notin (S\Sigma_m^k)^*$. In order to apply the CLT, Theorem 3.6, we would want to have Fréchet means on the manifold part, though. Fortunately this is not an issue.

Theorem 4.4 (Manifold Stability Theorem)**.** *Suppose that X is a random variable on Q with existing intrinsic mean $E(X)$. If $\mathbb{P}\{X \in Q^*\} > 0$ then $E(X) \subset Q^*$.*

A slightly less general version is in Huckemann (2012), Susovan will prove the general version in his talk.

The reason is that proper and isometric actions, or more generally well behaved submersions, never decrease sectional curvatures (O'Neill, 1966). So the singularities $Q \setminus Q^*$ are like cones with an opening angle $< 2\pi$.

For the bootstrap two-sample test of Section 3.3, how do we map data on Q to the orthogonal complement of $T_p[p]$? This can be done by horizontal lifts through sample means. Schould we use separate lifts, each one for each sample? Do, under H_0, such lifts converge to one-another? This will be answered in Susovan's talk.

How to compute sample means? Iterate: lift data in Q into optimal position of a representative of a candidate mean, compute their mean in M (easiest if extrinsic/Ziezold, still easy residual/Procrustean, else see Section 2.4), lift data in Q into optimal position of this, ... (Groisser, 2005; Dryden and Mardia, 2016).

Minutiae in fingerprints, taken by scanners, are again k landmark configurations in $\mathbb{R}^m$ (now $m = 2$). They are, again, considered modulo Euclidean motions, but, due to different scanner types and their pixel sizes, also modulo scaling:

- $X \sim Y$ if exists $R \in SO(m)$, $a \in \mathbb{R}^m$ and $\lambda > 0$ such that $\lambda RX + a \cdot 1_k^T = Y$
- $[X] := \{Y \in \mathbb{R}^{m\times k} : Y \sim X\}$ and
- $Q := \{[X] : X \in \mathbb{R}^{m\times k}\}$ is the (naive) *shape space*
- $d([X], [Y]) := \inf_{R\in SO(m),a\in\mathbb{R}^m,\lambda_1,\lambda_2>0} \|\lambda_1 RX + a \cdot 1_k^T - \lambda_2 Y\|$

Why naive? $d \equiv 0$ is not a metric (scaling acts neither properly nor isometrically).

- Workaround: consider only configurations $XH \neq 0$ and their representatives $\frac{XH}{\|XH\|} \in \mathbb{S}^{m\times(k-1)-1}$ modulo $SO(m)$.

Definition 4.5. *On* Kendall's shape spaces $\Sigma_m^k := \mathbb{S}^{m\times(k-1)-1}/SO(m)$ *the* Ziezold metric *is the quotient of the extrinsic, the* intrinsic metric *of the spherical and the* Procrustes metric *(only for even m, else quasi-) of the residual, cf. Theorem 4.2.*

The following theorem is a straightforward (technical) consequence:

Theorem 4.6 (Kendall (1984))**.** $\Sigma_2^k = \mathbb{C}^{k-1}/\mathbb{S}^1 = \mathbb{C}P^{k-2}$ *(complex projective space of complex dimension $k-2$) is the* Hopf fibration*, allowing for the* Veronese-Whitney embedding

$$\mathbb{C}P^{k-2} \hookrightarrow \mathbb{C}^{(k-2)\times(k-2)} \cong \mathbb{R}^{2(k-2)\times 2(k-2)}, \quad [Z] \mapsto ZZ^*$$

Z^* is the Hermitian conjugate and $[Z] = \{e^{it}Z : -\pi \leq t < \pi\}$, $Z \in \mathbb{C}^{k-1}$.

Further, the extrinsic metric w.r.t. the Veronese-Whitney metric agrees with the Procrustes metric.

4.2 Other Stratified Spaces and Stickiness

Phylogenetic trees with N labelled taxa (including a root) = vertices with degree 1 and internal nonlabeled vertics of degrees ≥ 3. Every edge is then a *split* of the label set

$$A|B = B|A, \quad A, B \text{ a partition of } \{1, \ldots, N\}, A \neq \emptyset \neq B$$

There are $\left(\left|\{0,1\}^{\{1,\ldots,N\}}\right| - 2\right)/2 = 2^{N-1} - 1$ possible splits. Of these are N *pendant*, the other *interior*. In a tree there are at most (induction) $N - 3$ interior edges. Modeling the length of each split in $\mathbb{R}_+$

$$\mathcal{T}_N = \mathbb{R}_+^N \times BHV_{N-1} \subset \mathbb{R}_+^{2^{N-1}-(N-1)-1}$$

where BHV_{N-1} is a manifold stratified set containing flat orthants of dimension 0 up to $N - 3$.

BHV spaces are Hadamard spaces (Billera et al., 2001) and thus have unique Fréchet means, see Theorem 2.11.

Stickiness. Consider $N = 4$ where $BHV_3 \cong \bigcup_{i=1,2,3}\{te_i : t \geq 0\} \hookrightarrow \mathbb{R}^3$ with the standard unit vectors e_1, e_2, e_3. Then $X \sim \frac{1}{3}\sum_{i=1}^{3} \delta_{e_i}$ has unique Fréchet mean 0. What about the mean of a perturbed random variable

$$X_t \sim \frac{1}{3}(\delta_{(1+t)e_1} + \delta_{e_1} + \delta_{e_1}$$

for some $t > 0$. By symmetry the mean has form se_1 for some $s \geq 0$. Solving

$$\operatorname{argmin}_{s\geq 0} F(se_1) = \operatorname{argmin}_{s\geq 0} \frac{1}{3}\left((1+t-s)^2 + 2(1+s)^2\right)$$

we see that the function to be minimized attains its for minimum at $3s = t-1$ which is nonnegative only if $t \geq 1$. For small $0 < t < 1$, the mean of X_t *sticks* to the origin, and, one can show (Hotz et al., 2013) that beyond a finite random sample size, the sample Fréchet mean also sticks to the origin, i.e. not featuring asymptotic fluctuation. A dead end for statistics? Note quite as Lammers et al. (2024) teaches.

Alternatively, more biologically motivated, model the correllation of taxa x, y (including the root) via

$$\rho(x, y) = e^{-d(x,y)}$$

with the tree graph distance $d(x, y)$. Including forests and degree 2 labels, obtain the *wald space*

$$\mathcal{W}_N \hookrightarrow SDP(N) = \{\Sigma \in \mathbb{R}^{m\times m} : \Sigma = \Sigma^T > 0\},$$

a Riemann stratified space, if SPD is equipped with the information geometry of $\mathcal{N}(0, \Sigma)$ (Cartan-Killing metric).

The information geometry for the parameter space $\Theta \subset \mathbb{R}^k$ of a statistical model

$$\mathcal{P} = \{\mathbb{P}_\theta^X = f_\theta(x)\, d\mu(x) : \theta \in \Theta\}$$

for vectorvalued X (dominating measure μ) equipps $T_\theta\Theta$ with the Riemannian metric tensor

$$I(\theta) = \operatorname{cov}_\theta[\operatorname{grad}_\theta \log f_\theta(X)] = -\mathbb{E}[\operatorname{Hess}_\theta \log f_\theta(X)]\,.$$

For instance, for $\mathcal{N}(\mu, \sigma^2)$, $\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\} = \mathbb{H}$, we have

$$d\mathbb{P}_{\mu,\theta}^X(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

yielding

$$I(\mu, \sigma) = \frac{1}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

which yields upon reparametrization $\sigma \to \sigma/\sqrt{2}$ twice the metric tensor for the hyperbolic geometry of $\mathbb{H}$.

The scaling rotation geometry for $SPD(N)$ by Jung et al. (2015) is motivated by tensor diffusion imaging and gives a (not fully understood?) stratified (?) space, see Armin Schwartzman's talk.

4.3 The Rabbit's Ears and Its Scalp

If $\{x \in \mathbb{S}^1 : f(x) = (2\pi)^{-1}\}$ is not a circular null-set, strange things can happen. Choose $0 < b < \pi$ and uniformly distribute a total mass of b/π on $\{e^{it} : \pi - b \leq t \leq \pi + b\}$ the rest of the corresponding probability mass, namely, $1 - b/\pi$ place on $1 = e^{i0}$. Consider X thus distributed. Then, for $0 \leq p < \pi$ we have

$$\begin{aligned}
F(p) - F(0) &= \int_{-\pi+p}^{\pi} (p-x)^2 \, d\mathbb{P}^X(x) + \int_{-\pi}^{-\pi+p} \big(p - (x+2\pi)\big)^2 \, d\mathbb{P}^X(x) - \int_{-\pi}^{\pi} x^2 \, d\mathbb{P}^X(x) \\
&= \int_{-\pi}^{\pi} \left(\left(p-x\right)^2 - x^2\right) d\mathbb{P}^X(x) - 4\pi \int_{-\pi}^{-\pi+p} (p - x - \pi) \, d\mathbb{P}^X(x) \\
&= p^2 - 2p\mathbb{E}[X] - 4\pi \int_{-\pi}^{-\pi+p} (p - x - \pi) \, 1_{-\pi \leq x \leq b} \frac{dx}{2\pi} \\
&= \begin{cases} p^2(1-1) = 0 & \text{if} \quad p \leq b \\ p^2 - b(2p-b) = (p-b)^2 & \text{if} \quad p > b \end{cases}
\end{aligned}$$

since $\big(p - (x+2\pi)\big)^2 = (p-x)^2 - 4\pi(p-x) + 4\pi^2$, by symmetry $\mathbb{E}[X] = 0$ and

$$\int_{-\pi}^{-\pi+p} (p - x - \pi) \, 1_{-\pi \leq x \leq b} \, dx = p(p-\pi) - \frac{x^2}{2}\bigg|_{x=-\pi}^{x=-\pi+p} = \frac{p^2}{2} \quad \text{if} \quad p \leq b$$

$$\int_{-\pi}^{-\pi+p} (p - x - \pi) \, 1_{-\pi \leq x \leq b} \, dx = b(p-\pi) - \frac{x^2}{2}\bigg|_{x=-\pi}^{x=-\pi+b} = \frac{b(2p-b)}{2} \quad \text{if} \quad p > b$$

since

$$x^2\bigg|_{x=-\pi}^{x=-\pi+b} = (\pi - b)^2 - \pi^2 = b(b - 2\pi)\,.$$

In consequence $E[X] = \{e^{it} : -b \leq t \leq b\}$. Sample Fréchet means, however, seem to accumulate near $\pm b$ (ears) and the open interval $\{e^{it} : -b < t < b\}$ (scalp) seems to be visited less often. This hints to

a) smeariness, discussed in Benjamin's tutorial

b) the question whether in (3.1) strict inequality holds (Evans and Jaffe (2024) has given such an example) $\rightsquigarrow$ sampling measure on $E(X)$?

References

Anderson, T. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist. 34*(1), 122–148.

Arnaudon, M. and L. Miclo (2014). Means in complete manifolds: uniqueness and approximation. *ESAIM: Probability and Statistics 18*, 185–206.

Aveni, A. and S. Mukherjee (2024). Uniform consistency of generalized Fréchet means. *arXiv preprint arXiv:2408.07534*.

Bacák, M. (2014). Computing medians and means in hadamard spaces. *SIAM journal on optimization 24*(3), 1542–1566.

Bhattacharya, R. and L. Lin (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proceedings of the American Mathematical Society 145*(1), 413–428.

Bhattacharya, R. N. and V. Patrangenaru (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds I. *The Annals of Statistics 31*(1), 1–29.

Bhattacharya, R. N. and V. Patrangenaru (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds II. *The Annals of Statistics 33*(3), 1225–1259.

Billera, L., S. Holmes, and K. Vogtmann (2001). Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics 27*(4), 733–767.

Davis, A. W. (1977). Asymptotic theory for principal component analysis: non-normal case. *Australian Journal of Statistics 19*, 206–212.

Dryden, I. L. and K. V. Mardia (2016). *Statistical Shape Analysis* (2nd ed.). Chichester: Wiley.

Eltzner, B. and S. Huckemann (2017). Bootstrapping descriptors for non-Euclidean data. In *Geometric Science of Information 2017 proceedings*, pp. 12–19. Springer.

Eltzner, B. and S. F. Huckemann (2019). A smeary central limit theorem for manifolds with application to high-dimensional spheres. *Ann. Statist. 47*(6), 3360–3381.

Evans, S. N. and A. Q. Jaffe (2024). Limit theorems for Fréchet mean sets. *Bernoulli 30*(1), 419–447.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut de Henri Poincaré 10*(4), 215–310.

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika 40*, 33–51.

Groisser, D. (2005). On the convergence of some Procrustean averaging algorithms. *Stochastics: Internatl. J. Probab. Stochstic. Processes 77*(1), 51–60.

Hotz, T. and S. Huckemann (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics 67*(1), 177–193.

Hotz, T., S. Huckemann, H. Le, J. S. Marron, J. Mattingly, E. Miller, J. Nolen, M. Owen, V. Patrangenaru, and S. Skwerer (2013). Sticky central limit theorems on open books. *Annals of Applied Probability 23*(6), 2238–2258.

Huckemann, S. (2011a). Inference on 3D Procrustes means: Tree boles growth, rank-deficient diffusion tensors and perturbation models. *Scandinavian Journal of Statistics 38*(3), 424–446.

Huckemann, S. (2011b). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *The Annals of Statistics 39*(2), 1098–1124.

Huckemann, S. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Annals of the Institute of Statistical Mathematics 64*(6), 1227–1259.

Huckemann, S., T. Hotz, and A. Munk (2010). Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). *Statistica Sinica 20*(1), 1–100.

Huckemann, S. F. and B. Eltzner (2018). Backward nested descriptors asymptotics with inference on stem cell differentiation. *The Annals of Statistics* (5), 1994 – 2019.

Huckemann, S. F. and B. Eltzner (2019). On central limit theorems on manifolds and stratified spaces. *Oberwolfach report 2019/28*, 1715–1717.

Huckemann, S. F. and B. Eltzner (2020). Statistical methods generalizing principal component analysis to non-Euclidean spaces. In *Handbook of Variational Methods for Nonlinear Geometric Data*, pp. 317–388. Springer.

Jung, S., I. L. Dryden, and J. S. Marron (2012). Analysis of principal nested spheres. *Biometrika 99*(3), 551–568.

Jung, S., A. Schwartzman, and D. Groisser (2015). Scaling-rotation distance and interpolation of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications 36*(3), 1180–1201.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc. 16*(2), 81–121.

Lammers, L., T. M. Nye, and S. F. Huckemann (2024). Statistics for phylogenetic trees in the presence of stickiness. arXiv 2407.03977.

Le, H. and D. Barden (2014). On the measure of the cut locus of a Fréchet mean. *Bulletin of the London Mathematical Society 46*(4), 698–708.

O'Neill, B. (1966). The fundamental equations of a submersion. *Michigan Math. J. 13*(4), 459–469.

Pennec, X. (2018). Barycentric subspace analysis on manifolds. *The Annals of Statistics 46*(6A), 2711–2746.

Rabenoro, D. and X. Pennec (2022). A geometric framework for asymptotic inference of principal subspaces in PCA. *arXiv preprint arXiv:2209.02025*.

Romano, J. P. and E. L. Lehmann (2005). *Testing statistical hypotheses*. Springer, Berlin.

Schötz, C. (2022). Strong laws of large numbers for generalizations of Fréchet mean sets. *Statistics 56*(1), 34–52.

Sturm, K. (2003a). Probability measures on metric spaces of nonpositive curvature. *Contemporary mathematics 338*, 357–390.

Sturm, K.-T. (2003b). Probability measures on metric spaces of nonpositive curvature, heat kernels and analysis on manifolds, graphs, and metric spaces. *Contemporary Mathematics 338*, 357–390.

Tyler, D. (1981). Asymptotic inference for eigenvectors. *Ann. Statist. 9*(4), 725–736.

Wiechers, H., M. Zobel, M. Bennati, I. Tkach, B. Eltzner, S. Huckemann, and Y. Pokern (2023). Drift models on complex projective space for electron-nuclear double resonance. *arXiv preprint arXiv:2307.12414*.

Ziezold, H. (1977). Expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transaction of the 7th Prague Conference on Information Theory, Statistical Decision Function and Random Processes A*, 591–602.