

Statistics with the boundary at infinity on Hadamard spaces

Ha-Young Shin

Department of Statistics and Actuarial Science, Soongsil University

Table of Contents

1 Introduction

- Geometric statistics
- Hadamard spaces

2 Quantiles on Hadamard spaces

- Euclidean quantiles
- Quantiles on Hadamard spaces
- Application to permutation testing

3 Causal inference on Hadamard spaces

- Causal inference basics
- Treatment effects on Hadamard spaces

Table of Contents

1 Introduction

- Geometric statistics
- Hadamard spaces

2 Quantiles on Hadamard spaces

- Euclidean quantiles
- Quantiles on Hadamard spaces
- Application to permutation testing

3 Causal inference on Hadamard spaces

- Causal inference basics
- Treatment effects on Hadamard spaces

Mean

- Let us begin with the mean.
- Given a sample $X_1, \dots, X_N \in \mathbb{R}$, the mean is defined as

$$\bar{X}_N := \frac{X_1 + \dots + X_N}{N}$$

- Let's look at this from a different perspective: What is the value of p that minimizes

$$\frac{1}{N} \sum_{i=1}^N (X_i - p)^2?$$

- Taking the derivative d/dp , we find that the solution satisfies

$$-\frac{2}{N} \sum_{i=1}^N (X_i - p) = 0 \implies p = \bar{X}_N.$$

- That is, the mean of a sample is equivalent to the point that minimizes the mean squared distance to the data.
- In fact, since they are equivalent, the sample mean can be defined in this way.

Mean

- More generally, the mean of a random variable X can be defined as

$$\arg \min_{p \in \mathbb{R}} E[(X - p)^2].$$

- Similarly, the mean of a multivariate sample $X_1, \dots, X_N \in \mathbb{R}^n$ can be defined as

$$\arg \min_{p \in \mathbb{R}^n} \frac{1}{N} \sum_{i=1}^N \|X_i - p\|^2$$

and the mean of a multivariate random vector X as

$$\arg \min_{p \in \mathbb{R}^n} E[\|X - p\|^2].$$

- Thus we have reinterpreted a statistical concept (the mean) from a geometric perspective (since distance is a geometric concept).

Median

- Can we do something similar for the median?
- The median of a sample $X_1, \dots, X_N \in \mathbb{R}$ is defined to be the middle value in terms of size.
- Then the median is equivalent to the value of p that satisfies

$$\sum_{i=1}^N I(p > X_i) = \sum_{i=1}^N I(p < X_i) \implies \sum_{i=1}^N [I(p > X_i) - I(p < X_i)] = 0.$$

- Since

$$\frac{d}{dp} |X_i - p| = I(p > X_i) - I(p < X_i)$$

the median can be shown to be

$$\arg \min_{p \in \mathbb{R}} \frac{1}{N} \sum_{i=1}^N |X_i - p|$$

(non-differentiability at X_i needs to be accounted for; we skip these details.)

Median

- The median of a random variable can be defined as $\arg \min_{p \in \mathbb{R}} E[|X - p|]$.
- Then how do we define a multivariate median?
- The basic definition of median involves ordering the sample, but multivariate samples have no order.
- There are several ideas you might have (e.g. just take the median of each coordinate; this works for the mean). This has problems.
- We can instead generalize the above idea to define the geometric median: For a sample $X_1, \dots, X_N \in \mathbb{R}^n$, the median is

$$\arg \min_{p \in \mathbb{R}^n} \frac{1}{N} \sum_{i=1}^N \|X_i - p\|$$

and for a random vector X , its median is $\arg \min_{p \in \mathbb{R}^n} E[\|X - p\|]$

Mean and median

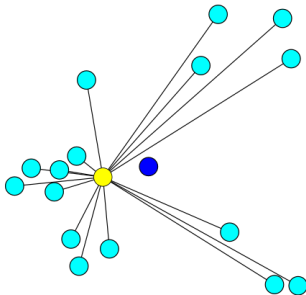


Figure 1: Geometric median (yellow) and mean (dark blue) of a data set (teal).

Geometric statistics

- Data sets and distributions possess geometric structure and shape, and so can be studied through a geometric lens, using geometric tools.
- This is what we mean by *geometric statistics*.
- Traditional data, as in the above examples, lie in vector spaces (Euclidean spaces).
- My research focuses on non-Euclidean statistics, analyzing data that lie in non-linear spaces.
- This is important because many modern data sets have complex structures that cannot be fully exploited in Euclidean space and are best analyzed as points in some non-linear space.
- Geodesic regression, which generalizes linear regression to Riemannian manifolds, is an example of some of the fruitful work done in this field.

Metric spaces

- A *metric space* is a space in which distances can be defined (these distances must satisfy certain properties).
- e.g. think of Euclidean space, where the distance between two points $p, q \in \mathbb{R}^n$ can be defined as $\|p - q\|$.
- Riemannian manifolds (e.g. a sphere), which can be thought of as smooth (hyper)surfaces on which tangent vectors and inner products are defined, are the classic examples.

Hadamard spaces

- My recent research has focused on statistics with a special class of useful metric spaces called Hadamard spaces (also called global non-positive curvature spaces and complete CAT(0) spaces).
- These show up in diverse fields such as phylogenetics, image processing, computational linguistics, and developmental biology.
- Hadamard spaces have recently received great interest among statisticians and machine learning researchers.
- We'll skip the technical definition of a Hadamard space, and just say that it is a space in which triangles are *at least as thin* as their counterparts in Euclidean space.
- (Compare this to e.g. a sphere, in which triangles are fatter than their Euclidean counterparts.)

Hadamard spaces

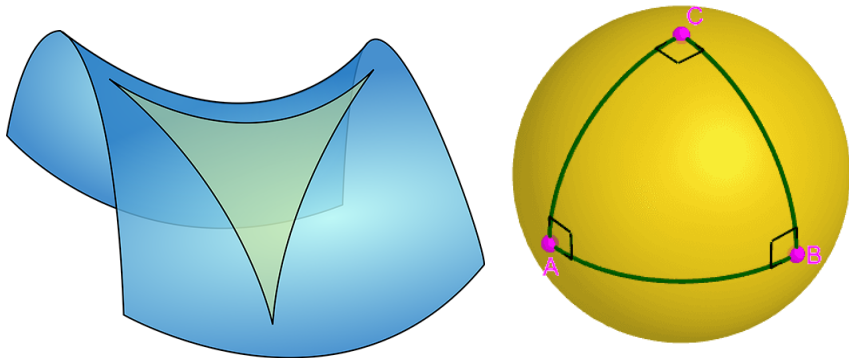


Figure 2: Thin (left) vs fat (right) triangles.

Examples of Hadamard spaces

- The most common examples of Hadamard spaces are Hadamard manifolds: Riemannian manifolds that are Hadamard spaces (equivalently, complete, simply connected Riemannian manifolds of non-positive sectional curvature).
- Examples include:
 - Euclidean space (constant zero sectional curvature).
 - Hyperbolic space (constant negative sectional curvature).
 - The space $\mathcal{P}_m$ of $m \times m$ symmetric positive definite (SPD) matrices is an $m(m+1)/2$ -dimensional Hadamard manifold.

Hyperbolic space

- Hyperbolic spaces in particular are becoming increasingly popular among machine learning researchers, being incorporated into neural nets and other machine learning algorithms (e.g. hyperbolic representation learning, hyperbolic classification, hyperbolic deep learning), as they are natural homes for *hierarchical data*.
- By hierarchical data, we mean data that can be organized in tree-like structures, with parent-child relationships.

Hyperbolic space

- What's the problem with embedding hierarchical data in Euclidean space?
- The number of nodes increases exponentially with radius, but volumes increase only polynomially in Euclidean space, so the nodes at the bottom are forced to be too close together.
- However, in hyperbolic space, volumes increase exponentially!

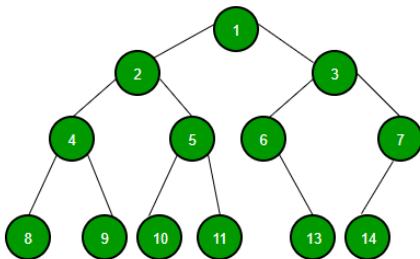


Figure 3: Tree-like structure

Hyperbolic space

- Hierarchical data is of great interest to machine learning researchers.
- For example, much NLP data is obviously hierarchical, and it has been argued by computer vision researchers that image data is inherently hierarchical and thus finds a natural home in hyperbolic space.
- In my own research I have worked with mRNA sequencing data and blood cell developmental data from single-cell gene expression embedded in the hyperbolic plane.

Poincaré disk

- What is hyperbolic space? How do we visualize it?
- Unlike e.g. the sphere, it is impossible to visualize hyperbolic space in Euclidean space without distortions.
- There are many ways to embed hyperbolic space in $\mathbb{R}^n$ *with distortions*.
- One of the most famous is the so-called Poincaré ball (Poincaré disk for 2D cases).
- The entire space is conceptualized as being contained in the interior of a ball, while distances approach infinity as you approach the boundary.

Hyperbolic space

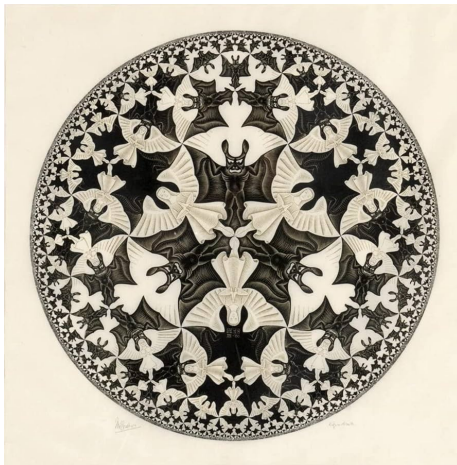


Figure 4: M. C. Escher's *Angels and Devils*, drawn in the Poincaré disk.

Hierarchical image data embedded in hyperbolic space

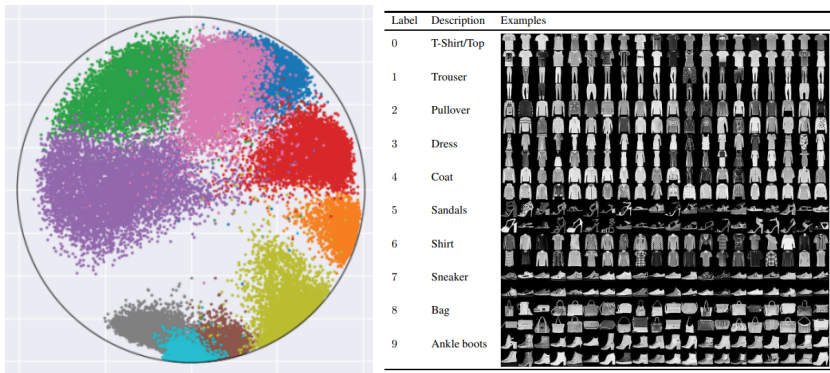


Figure 5: Example of image data embedded in hyperbolic space (Poincaré disk). Key: 0=blue=t-shirt, 1=orange=trouser, 2=green=pullover, 3=red=dress, 4=purple=coat, 5=brown=sandals, 6=pink=shirt, 7=gray=sneaker, 8=olive=bag, 9=cyan=ankle boots.

Boundary at infinity

- Most important for my research is a property called the boundary at infinity.
- This boundary at infinity can be conceptualized as consisting of points at infinity, and crucially, *there is exactly one ray connecting any given point in a Hadamard space to any given point at infinity*; thus, we can define a point at infinity to be a direction, and the aforementioned rays to all be pointing in this direction.
- In Euclidean space, for any given point at infinity, those rays are parallel to each other, which is intuitively satisfying.
- Alternative notions of directions involve arbitrary decisions and thus cannot be canonical.

Boundary at infinity

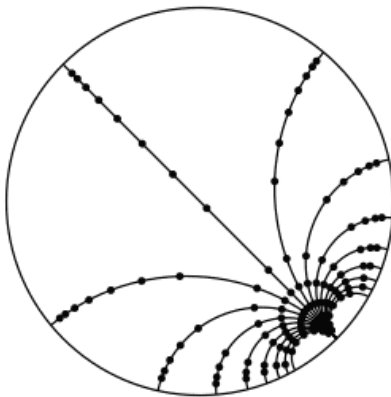


Figure 6: The boundary at infinity, illustrated in the Poincaré disk representation of hyperbolic space.

Hadamard manifolds

- On a Hadamard manifold M , there are two important functions at every point $x \in M$: the exponential map $\exp_x : T_x M \rightarrow M$ and the inverse exponential map $\log_x : M \rightarrow T_x M$, where $T_x M$ is the tangent space at x .
- $\exp_x$ is defined with geodesics (these are generalizations of straight lines in Euclidean space and show the unique shortest path between two points): $\exp_x(v) := \gamma_v(1)$, where γ_v is the geodesic issuing from x with initial velocity v .
- $\log_x$ is simply the inverse of $\exp_x$.

Hadamard manifolds

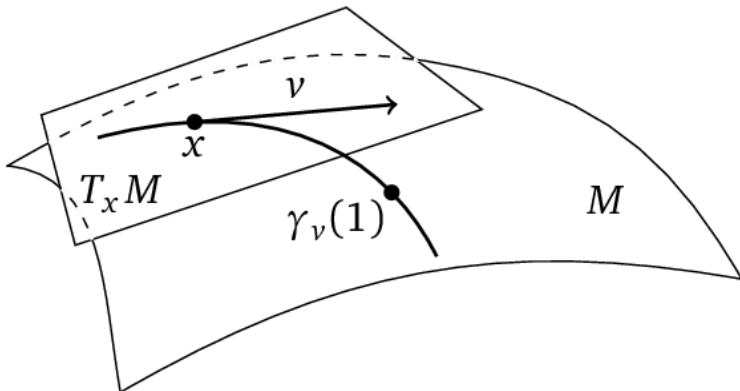


Figure 7: Illustration of the exponential and inverse exponential maps.

Radial fields

- On Hadamard manifolds, for any $x \in M$ and $\xi \in \partial M$, there exists a unique unit vector $\xi_x \in T_x M$ that points in the direction of ξ from x (that is, $t \mapsto \exp_x(t\xi_x)$ defined on $[0, \infty)$ gives the aforementioned ray connecting x to ξ).
- The vector field on M defined by $x \mapsto \xi_x$ is called a *radial field*.
- So there is a direct one-to-one connection between radial fields and points at infinity.

SPD matrix space

- SPD matrix-valued data is common.
- For example:
 - Covariance matrices can be random $\mathcal{P}_m$ -valued objects in their own right e.g. as sample covariance matrices or as parameters in a Bayesian framework.
 - Diffusion tensor imaging (DTI) expresses the diffusion of water molecules at a location in the brain as a 3×3 SPD matrix called a diffusion tensor (DT). The space $\mathcal{P}_3$ is a six-dimensional Hadamard manifold.

Illustration of DTI

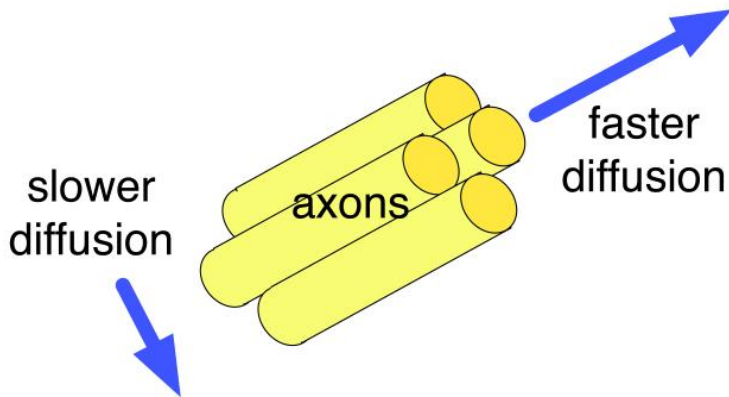


Figure 8: DTs can be used to probe the cellular structure of the brain e.g. the major eigenvector of the DT defines the principal diffusion direction (i.e. direction of fastest diffusion) at that location, and in fibrous tissues this also defines the fiber tract axis.

Main result

- Because radial fields naturally arise when studying the geometry of Hadamard manifolds, and because the spaces of SPD matrices are some of the most important examples of Hadamard manifolds, I was surprised to be unable to find an expression for the radial fields on these manifolds in the literature.
- Using Weyl's inequality, the Davis-Kahan $\sin(\Theta)$ theorem, and other results from linear algebra, I was able to prove the following, which gives an explicit expression for the radial fields on $\mathcal{P}_m$.

Theorem

For any $p \in \mathcal{P}_m$ and unit vector $z \in T_p\mathcal{P}_m$, let ξ be the unique point in $\partial\mathcal{P}_m$ satisfying $\xi_p = z$. Take any eigendecomposition VDV^T of $p^{-1/2}\xi_p p^{-1/2}$ satisfying $d_1 \geq \dots \geq d_m$, where $D := \text{diag}(d_1, \dots, d_m)$.

Take any $x \in \mathcal{P}_m$. Then, denoting the columns of the matrix $W := x^{-1/2}p^{1/2}V$ by $w_1, \dots, w_m$ so that $W = [w_1, \dots, w_m]$, let $u_1, \dots, u_m$ be an orthonormal basis of $\mathbb{R}^m$ that results from applying the Gram-Schmidt orthonormalization process to $w_1, \dots, w_m$. Then $\xi_x = x^{1/2}UDU^T x^{1/2}$, where $U := [u_1, \dots, u_m]$.

Main result

- Basically, the proof came down to the following lemma.
- Once you think about it, the lemma is very intuitive.

Lemma

For real $d_1 \geq \dots \geq d_m$ and a basis $w_1, \dots, w_m$ of $\mathbb{R}^m$, define the matrix

$$H(t) := \left[\sum_{i=1}^m e^{td_i} w_i w_i^T \right]^{1/t}.$$

Let $u_1, \dots, u_m$ be an orthonormal basis of $\mathbb{R}^m$ that results from applying the Gram-Schmidt orthonormalization process to $w_1, \dots, w_m$.

$\lim_{t \rightarrow \infty} H(t)$ exists and has eigenvalues $\exp(d_1), \dots, \exp(d_m)$ with corresponding eigenvectors $u_1, \dots, u_m$.

Smoothness of radial fields

- It turns out that radial fields are guaranteed to be C^1 , but not necessarily C^2 .
- I was able to prove that on $\mathcal{P}_m$, they are C^∞ :

Corollary

The radial fields on $\mathcal{P}_m$ are smooth.

- This has important implications for my research into using radial fields to do statistics on $\mathcal{P}_m$, as some of my theoretical results depend on the radial fields being C^2 or smoother.
- See the following paper for more details:
 - Shin, H.-Y. (2026). Radial fields on the manifolds of symmetric positive definite matrices. *SIAM Journal on Applied Algebra and Geometry*, 10:1–13.

Concrete interpretation of ξ

- In applications, a point at infinity $\xi \in \partial M$ can often be given a concrete interpretation by observing the rays in ξ . Let's do so for DTI.
- Because an SPD matrix is diagonalizable, its eigenvectors can be used to define the axes of an ellipsoid and its eigenvalues their lengths; thus DTs can be visualized as ellipsoids.
- In the figure on the following page, DTs are visualized in this way and colored based on so-called fractional anisotropy, in which brightness encodes eccentricity: ellipsoids in darker hues are closer to spheres.
- The ellipsoids have also been resized to all be on the same scale.

Concrete interpretation of ξ

- $\xi_1, \xi_2, \xi_3 \in \partial\mathcal{P}_3$ and $x_1, \dots, x_5 \in \mathcal{P}_3$ were chosen, and the i th row of the j th subfigure represents values along the ray in ξ_j issuing from x_i at times $t = 0, 1, 2, 3, 4$ i.e. $\exp_{x_i}(t(\xi_j)_{x_i})$.
- ξ_1 represents relative elongation of the ellipsoids, which can be interpreted as relatively greater diffusion of water molecules in the direction of elongation.
- ξ_2 represents a flattening in the direction roughly perpendicular to the page, interpreted as relatively greater diffusion in the plane of the page.
- ξ_3 represents greater diffusion in all directions by the same factor; because the ellipsoids have been scaled, the elements in any row look identical.
- Thus we have actual physical/biological interpretations for the points at infinity.

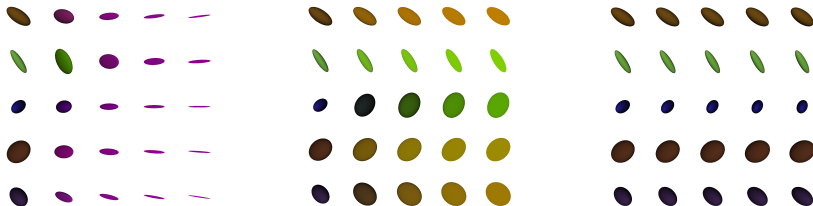


Table of Contents

1 Introduction

- Geometric statistics
- Hadamard spaces

2 Quantiles on Hadamard spaces

- Euclidean quantiles
- Quantiles on Hadamard spaces
- Application to permutation testing

3 Causal inference on Hadamard spaces

- Causal inference basics
- Treatment effects on Hadamard spaces

Quantiles

- Some Euclidean statistical methods require a notion of direction, and so with the boundary at infinity in hand, we can translate these to Hadamard spaces.
- One example on which I have worked extensively is causal inference.
- Here we will briefly explore another such example: quantiles.
- Quantiles (e.g. quartiles, quintiles, deciles, percentiles) are generalizations of the median.
- Like the median, they are defined by ordering the data.
- However, like the median (and mean), they can be defined as minimizers of a certain loss function.

Quantiles

- Fixing some $\alpha \in (0, 1)$, define a loss function by

$$|x - p| \{ (1 - \alpha) I(x \leq p) + \alpha I(x > p) \}.$$

- Then the α -quantile of a sample $X_1, \dots, X_N \in \mathbb{R}$ is

$$\arg \min_{p \in \mathbb{R}} \frac{1}{N} \sum_{i=1}^N |X_i - p| \{ (1 - \alpha) I(X_i \leq p) + \alpha I(X_i > p) \}$$

and for a random variable X is

$$\arg \min_{p \in \mathbb{R}} E[|X - p| \{ (1 - \alpha) I(X \leq p) + \alpha I(X > p) \}].$$

- ($\alpha = 0.25$ gives the 1st quartile, $\alpha = 0.5$ the median, $\alpha = 0.9$ the 9th decile etc.)

Quantiles

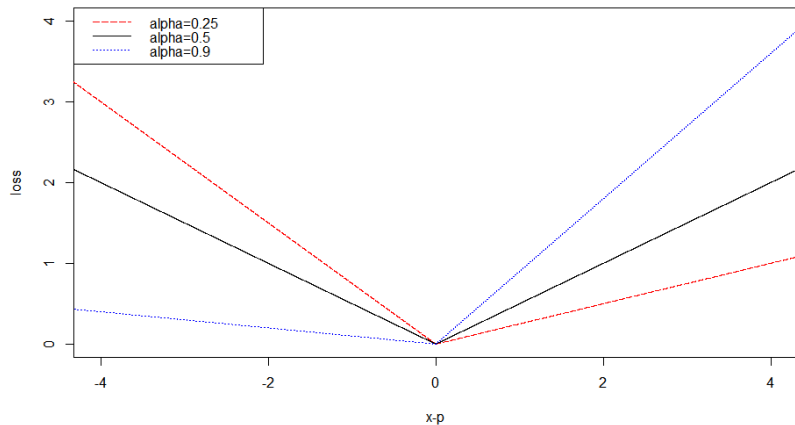


Figure 9: Univariate quantile loss function for various values of α .

Quantiles

- How do we define multivariate quantiles? There is no notion of ordering in multivariate space.
- Generalize the loss function!
- Instead of indexing the univariate quantile loss functions by $\alpha \in (0, 1)$, we can equivalently index them by $u := 2\alpha - 1 \in (-1, 1)$, which is the unit ball in $\mathbb{R}$.
- Then $\rho(x, p) = |x - p| + u(x - p)$.
- Then for $u < 0$, we tilt the median loss function to the right, and for $u > 0$ to the left. The larger $|u|$ is, the more we tilt.
- We can define multivariate geometric quantiles using the following loss function $\rho : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+$:

$$\rho(x, p) = \|x - p\| + \langle u, x - p \rangle$$

where u is in the unit ball in $\mathbb{R}^n$.

Quantiles

- The sample u -quantile becomes

$$\arg \min_{p \in \mathbb{R}^n} \frac{1}{N} \sum_{i=1}^N \rho(X_i, p)$$

and the u -quantile of a random vector X is $\arg \min_{p \in \mathbb{R}^n} E[\rho(X, p)]$.

- Again, think of u as tilting the loss function in the direction of u , by an amount determined by $\|u\|$.

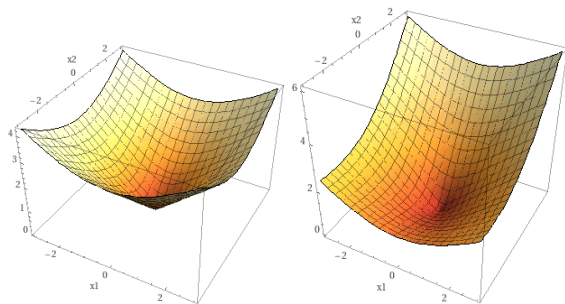


Figure 10: Multivariate median and quantile loss functions.

Quantiles on Hadamard spaces

- How can we generalize this definition further to a Hadamard space M ?
- With the mean and median, it's easy: for an M -random element X , they are $\arg \min_{p \in M} E[d(X, p)^2]$ and $\arg \min_{p \in M} E[d(X, p)]$, respectively (d is the distance function).
- (Actually, this definition works for all metric spaces).
- For quantiles, it's difficult to find an equivalent of the vector u on manifolds/metric spaces in general. (e.g. on Riemannian manifolds, in general we can't canonically identify tangent vectors from different tangent spaces).
- Using the boundary at infinity, we can define quantiles on Hadamard spaces.

Loss function

- For a magnitude $\beta \in [0, 1]$ and direction $\xi \in \partial M$ (the boundary at infinity), define $\rho : M \times M \rightarrow \mathbb{R}^+$ by

$$\rho(x, p) = d(p, x) + \beta d(p, x) \cos(\angle_p(x, \xi)).$$

- The parallel with multivariate quantiles is easier to see in the case of Hadamard manifolds:

$$\rho(x, p) = \|\log_p(x)\| + \langle \beta \xi_p, \log_p(x) \rangle,$$

where ξ_p is the *unique* unit vector in the tangent space at p that faces the direction ξ .

- Compare this to the multivariate ρ :

$$\rho(x, p) = \|x - p\| + \langle u, x - p \rangle.$$

- As in the multivariate case, we can think of this quantile loss function as being a tilted version of the median loss function, with the direction of the tilt defined by ξ and the magnitude by β ($\beta = 0$ gives the median loss function).

Definition and properties

Definition

The (β, ξ) -quantile of a sample $X_1, \dots, X_N \in M$ is

$$\arg \min_{p \in M} \frac{1}{N} \sum_{i=1}^N \rho(X_i, p).$$

Definition

The (β, ξ) -quantile of an M -valued random element X is $\arg \min_{p \in M} E[\rho(X, p)]$.

- We have been able to prove various theoretical properties these quantiles, including strong consistency, asymptotic normality, and an expression for the gradient.

Permutation test

- Very theoretical so far; let's give a concrete example of the use of these quantiles.
- Given two samples X^1, X^2 coming from two distributions F_1, F_2 on some space, the permutation test is an exact test for $H_0 : F_1 = F_2$ vs $H_1 : F_1 \neq F_2$.
- It uses some test statistic $T(X^1, X^2)$ that measures the disparity between the two samples X^1, X^2 , and argues that under H_1 , T should be large.
- For example, a common test statistic is the distance between the sample means or medians: $T(X^1, X^2) = d(\bar{X}^1, \bar{X}^2)$.
- The problem with this is that two distributions could be very different but have similar centers; this permutation test would be unable to reject H_0 in this case; it has *low power*.
- Let's do a permutation test on Hadamard spaces using quantiles!

Permutation test

- For $(\beta_1, \xi_1), \dots, (\beta_K, \xi_K) \in [0, 1) \times \partial M$, and weights $w_1, \dots, w_K$ that add to 1, define

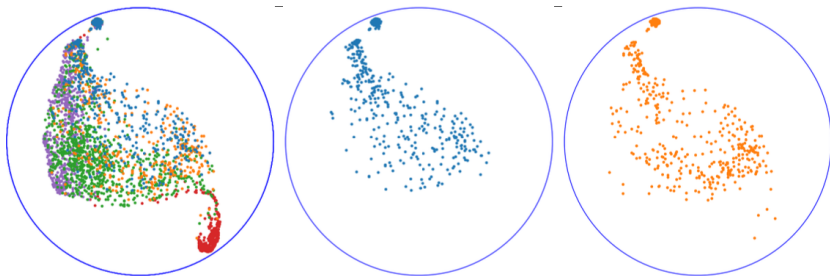
$$T(X^1, X^2) = \sum_{k=1}^K w_k d(q_k(X^1), q_k(X^2)),$$

where $q_k(X^j)$ refers to the (β_k, ξ_k) -quantile of X^j ($j = 1, 2, k = 1, \dots, K$).

- This is the weighted mean of the distances between the quantiles.
- This should have greater power, as it can distinguish between distributions that have similar centers as long as they are sufficiently different at one of these quantiles.
- You can increase the “resolution” of this test by increasing K .

Permutation test

- Consider single-cell gene expression data used to map early blood development.
- This kind of genetic data have clear hierarchical structures, and so are ideal for embedding into hyperbolic space.
- Each dot in the below figures represents a cell, and each cell in the data set belongs to one of five stages, each given a different color below. I looked at the two earliest stages, primitive streak (PS) and neural plate (NP), shown in the middle and right, respectively.
- At a glance, it seems that the two distributions are quite similar with similar centers, and yet certainly not identical.



Permutation test

- We drew samples of size 120 from the two distributions, and tested H_0 with these samples using four test statistics: mean-based, median-based, quantile-based and also Euclidean quantile-based.
- This process was repeated 1000 times, producing the following table containing the proportion of runs in which the p-value for the given test statistic exceeded the considered value (i.e. estimated Type II error rate, or 1 minus estimated power).

	0.1	0.05	0.01	0.005	0.001
Mean	0.025	0.063	0.167	0.233	0.375
Median	0.060	0.104	0.277	0.359	0.534
Quantiles	0.003	0.006	0.036	0.057	0.118
Euclidean quantiles	0.027	0.056	0.176	0.253	0.409

- See the following paper for more details:
 - Shin, H.-Y. and Oh, H.-S. (2026). Quantiles on global non-positive curvature spaces. *Journal of the Royal Statistical Society, Series B*. Advance online publication. <https://doi.org/10.1093/jrsssb/qkag121>

Table of Contents

1 Introduction

- Geometric statistics
- Hadamard spaces

2 Quantiles on Hadamard spaces

- Euclidean quantiles
- Quantiles on Hadamard spaces
- Application to permutation testing

3 Causal inference on Hadamard spaces

- Causal inference basics
- Treatment effects on Hadamard spaces

Potential outcomes framework

- In the potential outcomes framework, each subject has a potential treatment outcome and potential control outcome (R_T, R_C) ; which is observed depends on some treatment variable Z :

$$R = \begin{cases} R_T & \text{if } Z = 1 \\ R_C & \text{if } Z = 0. \end{cases}$$

- The *average treatment effect* (ATE) is defined in Euclidean space as $E[R_T] - E[R_C]$.
- The *individual treatment effect* (ITE) is $R_T - R_C$; the homogeneous effects model assumes that the ITE is constant.

Randomized experiments and observational studies

- Fundamental problem of causal inference: Only one of the potential outcomes can be observed.
- This can be dealt with using randomization (in experiments) or matching (in observational studies).
- $\mathbf{Z}_n = (Z_1, \dots, Z_n) \in \{0, 1\}^n$, $\mathbf{R}_{Tn} = (R_{T1}, \dots, R_{Tn})$, $\mathbf{R}_{Cn} = (R_{C1}, \dots, R_{Cn})$ and $\mathbf{X}_n = (X_1, \dots, X_n)$; here each X_i is a vector of covariates.
- Completely randomized experiment: $\mathbf{Z}_n$ and $(\mathbf{R}_{Tn}, \mathbf{R}_{Cn})$ are independent.
- Stratified randomized experiment: Covariates are used to form strata, given which $\mathbf{Z}_n$ and $(\mathbf{R}_{Tn}, \mathbf{R}_{Cn})$ are independent.
- Randomized experiments are the gold standard, but are often impossible. A condition called *unconfoundedness*

$$\mathbf{Z}_n \perp\!\!\!\perp (\mathbf{R}_{Tn}, \mathbf{R}_{Cn}) \mid \mathbf{X}_n$$

lets one approximate a stratified randomized experiment via *matching*.

- There are many types of matching; one type makes it so that the distribution of the covariates is similar in the two groups.

Individual treatment effect

- We can use directions (points at infinity) to do statistical inference on Hadamard spaces.
- Whenever you would use vectors in Euclidean space, instead use magnitude/direction pairs $[\beta, \xi] \in A := [0, \infty) \times \partial M / \sim$ (equivalence relation is defined by letting $(0, \xi_1) \sim (0, \xi_2)$ for all $\xi_1, \xi_2 \in \partial M$).
- On Hadamard m -manifolds, this space is homeomorphic to $T_p M$ for all $p \in M$. Let's call the homeomorphism $f_p : A \rightarrow T_p M$.
- Denote by $\gamma_p^\xi : [0, \infty) \rightarrow M$ the unique unit-speed geodesic ray issuing from p pointing at ξ .

Definition

The *individual treatment effect* (ITE) is the unique value $[\beta, \xi] \in A$ for which $R_T = \gamma_{R_C}^\xi(\beta)$.

Individual treatment effect

- We can define the homogeneous effects model on Hadamard spaces as that in which the ITE is constant.
- For any specific $[\beta, \xi] \in A$, $H_0^{[\beta, \xi]} : \text{ITE} = [\beta, \xi]$ is equivalent to Fisher's sharp null hypothesis with $\gamma_{RCi}^{\xi}(\beta)$ in the role of the control outcomes.
- Thus this can be tested with randomization inference.
- We can also form a $100(1 - \alpha)\%$ level confidence region for the constant ITE with $\{[\beta, \xi] \in A : H_0^{[\beta, \xi]} \text{ is not rejected at level } \alpha\}$.

Average treatment effect

- How should we define ATE for Hadamard space-valued R_T and R_C ?
- Define it as the magnitude/direction pair $[\beta, \xi]$ satisfying the following: If R_C is moved distance β in direction ξ , its Fréchet mean coincides with the Fréchet mean of R_T .
- Denote the Fréchet mean of an M -valued Y with $\mu(Y)$.
- Denote the weighted sample Fréchet mean by $\hat{\mu}_n(r_1, \dots, r_n; w_1, \dots, w_n)$.

Definition

The *average treatment effect set* (ATE set) for (R_T, R_C) is $\tau := \{[\beta, \xi] \in A : \mu(\gamma_{R_C}^\xi(\beta)) \cap \mu(R_T) \neq \emptyset\}$; any element of this set is called an ATE.

Definition

The *minimal ATE magnitude* for (R_T, R_C) is $\min\{\beta : [\beta, \xi] \text{ is an ATE}\}$.

Average treatment effect

- What alternative generalizations of ATE to Hadamard (and other metric) spaces exist?
- Absolute average treatment effect (AATE): $d(\mu(R_T), \mu(R_C))$.
- Geodesic average treatment effect (GATE): The geodesic segment from $\mu(R_C)$ to $\mu(R_T)$ itself. The length of the GATE is the AATE.

Average treatment effect

- Our ATE may be more intuitive; consider a toy example on a Hadamard space called the 3-spider.
- Unlike the ATE, the AATE and GATE are constant as the blue points (R_C) are moved to infinity at the same speed.

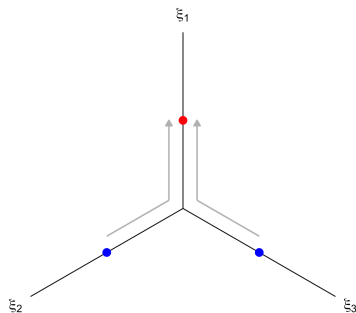


Figure 11: 3-spider. The red point represents R_T , the blue points R_C . Each of the blue points has 0.5 probability.

Average treatment effect

- Interpretability is another issue.
- Except in cases where the quantities are zero, the meaning and clinical significance of any specific value for the AATE and GATE may not be apparent to physicians, and may instead feel like abstract mathematical objects.
- For any control observation $r \in M$, if $[\beta, \xi]$ is an ATE, $\gamma_r^\xi(\beta)$ can be calculated and interpreted as the value of this observation under the average treatment effect (we will see an example later).
- In contrast, the AATE gives no sense of direction, and the GATE only gives a sense of direction at $\mu(R_C)$ specifically.

Mean transport function

- We now define a function that will be crucial in investigating ATEs and their estimators.

Definition

The map $A \rightarrow M$ defined by $[\beta, \xi] \mapsto \mu_{[\beta, \xi]}(Y) := \mu(\gamma_Y^\xi(\beta)) = \mu(\exp_Y(\beta\xi_Y))$ is called the *mean transport function* of Y .

Definition

Given a set of points $y_1, \dots, y_n \in M$ and weights $w_1, \dots, w_n \in [0, 1]$ adding to 1, the map $A \rightarrow M$ defined by

$$[\beta, \xi] \mapsto \hat{\mu}_n^{[\beta, \xi]}(y_1, \dots, y_n; w_1, \dots, w_n) := \hat{\mu}_n(\gamma_{y_1}^\xi(\beta), \dots, \gamma_{y_n}^\xi(\beta); w_1, \dots, w_n)$$

is called the weighted sample mean transport function of $y_1, \dots, y_n$ with associated weights $w_1, \dots, w_n$. When the weights are all equal to $1/n$, the function is simply called the sample mean transport function.

Mean transport function

Theorem (Properties of the mean transport function)

The mean transport function of Y satisfies the following properties.

(a) It is well-defined.

(b) It is continuous.

(c) For all $p \in M$, there exists some $B_p > 0$ such that $d(\gamma_p^\xi(\beta), \mu_{[\beta, \xi]}(Y)) < B_p$ for all $[\beta, \xi] \in A$.

(d) Given a sequence $[\beta_1, \xi_1], [\beta_2, \xi_2], \dots \in A$, if $\beta_j \rightarrow \infty$ as $j \rightarrow \infty$, $d(p, \mu_{[\beta_j, \xi_j]}(Y)) \rightarrow \infty$ for all $p \in M$, and if $\xi_j = \xi$ for each j , $\log_p(\mu_{[\beta_j, \xi]}(Y))/d(p, \mu_{[\beta_j, \xi]}(Y)) \rightarrow \xi_p$.

(e) For any $p \in M$, the pre-image of $\{p\}$ is nonempty and compact.

Corollary

(a) The ATE set is nonempty and compact.

(b) The minimal ATE magnitude exists.

An ATE estimator

- Let $\Xi := \{1, \dots, S\}$ be the strata in a stratified randomized experiment.
- Let $\mathbf{S}_n \in \Xi^n$ encode stratum membership for a sample of size n , and $\hat{\lambda}_n^1, \dots, \hat{\lambda}_n^S$ be stratum-wise weights satisfying $\sum_{s=1}^S \hat{\lambda}_n^s = 1$.
- The ATE estimator is based on weighted empirical treatment and control distributions.

$$\hat{\tau}_n(\mathbf{Z}_n, \mathbf{R}_n, \mathbf{S}_n) \\ = \{[\beta, \xi] \in A : \hat{\mu}_n^{[\beta, \xi]}(R_1, \dots, R_n; W_{C1n}, \dots, W_{Cnn}) = \hat{\mu}_n(R_1, \dots, R_n; W_{T1n}, \dots, W_{Tnn})\}$$

where

$$W_{Tin} = Z_i \sum_{s=1}^S \hat{\lambda}_n^s I\{S_i = s\} / N_{Tn}^s \quad \text{and} \quad W_{Cin} = (1 - Z_i) \sum_{s=1}^S \hat{\lambda}_n^s I\{S_i = s\} / N_{Cn}^s$$

and $N_{Tn}^s = \sum_{k=1}^n Z_k I\{S_k = s\}$ and $N_{Cn}^s = \sum_{k=1}^n (1 - Z_k) I\{S_k = s\}$.

- Essentially, $\hat{\tau}_n$ is a weighted sample ATE.
- If M is Euclidean, $\hat{\tau}_n$ reduces to $\sum_{s=1}^S \hat{\lambda}_n^s (\bar{R}_T^s - \bar{R}_C^s)$, where $\bar{R}_T^s$ and $\bar{R}_C^s$ are respectively the sample means of the treatment and control outcomes in stratum s . This is the standard estimator of the ATE in a stratified context.

Asymptotic properties

- Proved a weighted uniform law of large numbers for the mean transport function.
- Uniform laws of large numbers commonly show uniform convergence over some compact set; this theorem does so over all of A .
- Proved the strong consistency of $\hat{\tau}_n$ in
 - stratified randomized experiments, and
 - in covariate balanced observational studies when estimating average treatment effect on the treated (ATT) and average treatment effect on the control (ATC).

Computation

- How do we actually calculate the ATE?
- Finding the ATE can be framed as a root-finding problem for a function $\mathbb{R}^m \rightarrow \mathbb{R}^m$.
- Choose some smooth chart $\phi : M \rightarrow \mathbb{R}^m$ and $p \in M$ and identify $T_p M$ with $\mathbb{R}^m$. Denote by $h : \mathbb{R}^m \rightarrow \mathbb{R}^m$ the map

$$v \mapsto \phi(\mu^{[\beta, \xi]}(R_C)) - \phi(\mu(R_T)),$$

where $[\beta, \xi] = f_p^{-1}(v)$.

- Then an ATE is any solution of $h(v) = 0$.
- Newton's method would require differentiating; this is not trivial.
- Instead, use Broyden's method, which is similar to Newton's method but uses secants instead of tangents.
- Requires no differentiation.

Consistency and confidence regions with simulations

- Generated response data on hyperbolic space, with two covariates.
- Generated both randomized experiment data and observational study data, which we fully matched using propensity scores based on the Mahalanobis distance.
- We have already proven consistency in randomized experiments. What about observational studies using matching?

Table 1: Estimated MAE for observational study simulations.

n	125	250	500	1000
MAE	0.265	0.184	0.130	0.093

- MAE was calculated in some tangent space via f_p^{-1} .

Consistency and confidence regions with simulations

- Used basic (i.e. pivotal) bootstrap to form confidence regions.
- $\hat{\theta}_n - V_1$ corresponds to a spherical region, $\hat{\theta}_n - V_2$ an elliptical region based on estimated covariance.

Table 2: Estimated basic bootstrap confidence region coverage.

Setting	Region	$n = 125$	$n = 250$	$n = 500$	$n = 1000$
Experiments	$\hat{\theta}_n - V_1$	0.948	0.944	0.959	0.952
	$\hat{\theta}_n - V_2$	0.934	0.945	0.956	0.954
Observational studies	$\hat{\theta}_n - V_1$	0.933	0.942	0.928	0.948
	$\hat{\theta}_n - V_2$	0.933	0.941	0.922	0.936

Consistency and confidence regions with simulations

- Two different confidence intervals have been proposed for the AATE(=magnitude of the GATE).
- However, their coverage is suspect when $AATE=0$, making their use for testing $H_0 : AATE = 0$ dubious.

Interpretability and Euclidean comparisons with DTI data

- Observational study with 27 Parkinson's patients and 26 matched control subjects.
- Took their diffusion tensors from a region of the corpus callosum; these can be visualized as ellipsoids.
- A DT measures the diffusion of water molecules at a location in the brain. They are used to map the structure of white matter tracts of the brain.
- Diffusion tensors can be identified with 3×3 SPD matrices. SPD spaces are Hadamard.

Interpretability and Euclidean comparisons with DTI data

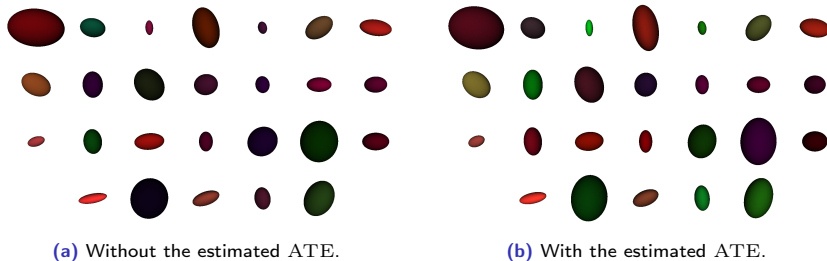


Figure 12: Average DTs for 26 control subjects with and without the estimated ATE.

Interpretability and Euclidean comparisons with DTI data

- Using the aforementioned confidence regions, we tested $H_0 : \text{ATE} = 0$, and tested against Euclidean alternatives.
- Euclidean alternatives: Summarize each DT with a scalar (fractional anisotropy FA, mean diffusivity MD) and do a two-sample t -test.
- ATE p -value: 0.007.
- FA p -value: 0.171.
- MD p -value: 0.542.
- A paper titled “Treatment effects on Hadamard spaces” based on this research is currently in review.