Matthieu Bulté

From Points to Objects: Statistical Inference Beyond Euclidean Spaces

PHD THESIS

this thesis has been submitted to
the phd school of the faculty of science, University of Copenhagen
and
the faculty of business administration and economics, university of bielefeld

Department of Mathematical Sciences
University of Copenhagen

January 2025

Matthieu Bulté
mb@math.ku.dk
Department of Mathematical Sciences
University of Copenhagen
2100 Copenhagen, Denmark
matthieu.bulte@uni-bielefeld.de
Bielefeld Graduate School of Economics and Management
University of Bielefeld
33501 Bielefeld, Germany

Thesis title:	From Points to Objects: Statistical Inference Beyond Euclidean Spaces
Supervisor:	Professor Helle Sørensen University of Copenhagen
	Professor Christiane Fuchs University of Bielefeld
Assessment Committee:	Professor Bo Markussen (chair) University of Copenhagen
	Professor Dietmar Bauer University of Bielefeld
	Professor Victor Panaretos École Polytechnique Fédérale de Lausanne
Date of Submission:	January 31, 2025
Date of Defense:	May 16, 2025

ISBN:

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen and Faculty of Business Administration and Economics, University of Bielefeld. It received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 956107, "Economic Policy in Complex Environments (EPOC)".

À vous, Maman, Margaux.

Preface

Foreword

This PhD thesis presents research conducted as part of the Marie Skłodowska-Curie Actions Innovative Training Networks, Economic Policy in Complex Environments (EPOC). The thesis contains the results of two years of research at the Department of Mathematical Sciences at the University of Copenhagen (KU) and one year at the Faculty of Business Administration and Economics at the University of Bielefeld (UNIBI) as part of a double degree program between the two institutions. The research in this thesis was conducted under the supervision of Helle Sørensen (KU) as the main supervisor and Christiane Fuchs (UNIBI) as co-supervisor. Additionally, EPOC involved a two-month non-academic secondment which I spent at the Greenland Institute of Natural Resources (GINR) under the supervision of Mads-Peter Heide-Jørgensen (GINR).

Acknowledgments

It's a little bit funny to look back, thinking about all the people who have been part of my life in meaningful ways. I have been lucky to have met so many great people, and I am grateful to all of them.

I would like to start by thanking my supervisor Helle Sørensen. I will miss our weekly meetings, which have been a fun mix of chatting about our lives and intense project discussions. Helle has been a fantastic supervisor, allowing me to grow by trusting and encouraging my intuitions while providing the right amount of guidance.

The Statistics and Probability Section at the University of Copenhagen has been a fantastic place to do my PhD. This department is filled with fun and interesting people, many of whom have become friends. Thank you for being there: Anton, Cecilie, Christian, Frederik, Leonard, Margherita, Myrto, and Ulises. Particular mentions to Alex, Pedja, and Shimeng with whom I had the chance to share this entire experience, starting with our early days in the corner office 4.3.21.

I feel immensely grateful for having met so many amazing people in my life, with whom I connected not only intellectually but also on a much more personal level. To Flo, Johannes, Laurent, Lea, Lucy, Martin, Nicola, and Pedja: thank you for always being there for me, for being an inspiration, and for being such great people to be around. And to all the many others that I am not able to mention, thank you.

This wouldn't be an acknowledgments section without thanking my family which is and has always been a central part of my life, even in times when I haven't been able to show it. I will start with my father Jérôme, my step mother Hortense, and her parents Alain and Florence. I also want to thank my army of siblings, Colombe, Olympe, Lancelot, and Calypso — each of whom I love in their own special way. I am grateful to my uncle Philip, who has shaped me so much and without whom I wouldn't have studied mathematics, and to the rest of the family, Susanne, Caspar, and Medea for all the love and laughter during my time in Munich. I am thankful for my brother in-law Morgan, and my nephew Gabriel, who both have a special place in my heart. And to my teta, whose warmth and strength I admire, and who brings spices to our tawlé games.

To you, Ksusha, my partner, my friend. I am grateful for the love, support and fun you have brought to my life. I cannot wait for the many years we have ahead of us.

Finally, I would like to thank my mother, Jennifer, and my sister, Margaux, to whom I dedicate this thesis. I couldn't put into words how much you mean to me, but you already know.

Matthieu Bulté
January, 2025

Abstract

Random variables taking values in metric spaces, called *random objects*, have recently received additional attention in the statistical litterature. The abstraction, only requiring the definition of a notion of *distance* between data points, allows for the development of statistical methodology that can be applied to a wide range of complex data types. In particular, this includes types of data not typically covered by existing works. Just like in classical statistics, practitioners encountering such complex data might be interested in answering a wide range of questions. In this thesis, we present the results of three independent projects, each addressing a typical statistical task.

The first chapter provides an introduction to the thesis; it consists of a motivation for the work carried out, together with an attempt at providing the necessary background in metric spaces and their statistical study in order to make this manuscript self-contained. This includes a brief introduction to metric spaces and to relevant geometric concepts. Then, we introduce the notion of random objects together with a generalization of expected value central to this thesis: the *Fréchet mean* of a random object. Each of the remaining chapters corresponds to one of the projects of the thesis, to which we provide a brief introduction consisting of a motivation and presentation of the main contribution.

In the second chapter, we are concerned with regressing a random object on a vector of real numbers, that is, we attempt to learn from an independent and identically distributed sample the conditional expectation of a random object given a vector of real numbers. To that end, we present an adaptation of random forest together with an approximate tree construction algorithm. Our approximation algorithm allows to perform regression in metric spaces where Fréchet means are expensive to compute. We show that the proposed method, the Metric Random Forest (MRF), is pointwise consistent and we provide a simulation study to illustrate its performance.

Time series data, where observations are collected sequentially over time, naturally raise questions about temporal dependence between successive measurements. In the third chapter, we use the additional structure present in Hadamard spaces to extend the classical autoregressive process of order one, introducing the Geodesic Autoregressive Model (GAR). We provide estimators for its parameters, and propose a statistical test for the absence of temporal dependence. Furthermore, we study the asymptotic properties of the estimators and the test. We illustrate the methodology in a simulation study and in an application to economic surveys of consumer expectations.

Finally, in the fourth chapter, we consider the more elementary problem of testing the Fréchet mean of a random object. We introduce a test based on the empirical Fréchet variance of the sample and introduce a randomization methodology exploiting the symmetries of the metric space. We illustrate the methodology in various kinds of metric spaces and investigate the performance of the test under various conditions. We provide a simulation study to illustrate the performance of the test and apply it to a data set of wind data in Denmark.

Sammenfatning

Stokastiske variable, der antager værdier i metriske rum, kaldet *stokastiske objekter*, har for nylig fået øget opmærksomhed i den statistiske litteratur. Abstraktionen, der kun kræver definitionen af en *afstand* mellem datapunkter, muliggør udviklingen af statistisk metodologi, der kan anvendes på en bred vifte af komplekse datatyper. Dette omfatter især datatyper, der typisk ikke dækkes af eksisterende værker. Ligesom i klassisk statistik kan praktikere, der møder sådanne komplekse data, være interesserede i at besvare en bred vifte af spørgsmål. I denne afhandling præsenterer vi resultaterne af tre uafhængige projekter, der hver behandler en typisk statistisk opgave.

Det første kapitel giver en introduktion til afhandlingen; det består af en motivation for det udførte arbejde sammen med et forsøg på at give den nødvendige baggrund i metriske rum og deres statistiske studie for at gøre dette manuskript selvstændigt. Dette inkluderer en kort introduktion til metriske rum og relevante geometriske koncepter. Derefter introducerer vi begrebet stokastiske objekter sammen med en generalisering af forventet værdi, der er central for denne afhandling: *Fréchet-middelværdien* af et stokastisk objekt. Hvert af de resterende kapitler svarer til et af afhandlingens projekter, hvortil vi giver en kort introduktion bestående af en motivation og præsentation af hovedbidraget.

I det andet kapitel beskæftiger vi os med regression af et stokastisk objekt på en vektor af reelle tal, det vil sige, vi forsøger at lære fra en i.i.d. stikprøve den betingede forventning af et stokastisk objekt givet en vektor af reelle tal. Til dette formål præsenterer vi en tilpasning af random forest sammen med en approksimativ træ-konstruktionsalgoritme. Vores approksimationsalgoritme gør det muligt at udføre regression i metriske rum, hvor Fréchet-middelværdier er beregningsmæssigt krævende. Vi viser, at den foreslåede metode, Metric Random Forest (MRF), er punktvist konsistent, og vi præsenterer et simulationsstudie for at illustrere dens ydeevne.

Tidsseriedata, hvor observationer indsamles sekventielt over tid, rejser naturligt spørgsmål om tidsmæssig afhængighed mellem på hinanden følgende målinger. I det tredje kapitel bruger vi den yderligere struktur i Hadamard-rum til at udvide den klassiske AR(1)-model og introducerer den Geodætiske Autoregressive Model (GAR). Vi leverer estimatorer for dens parametre og foreslår en statistisk test for fravær af tidsmæssig afhængighed. Desuden studerer vi estimatorernes og testens asymptotiske egenskaber. Vi illustrerer metodologien i et simulationsstudie og i en anvendelse på økonomiske undersøgelser af forbrugerforventninger.

Endelig behandler vi i det fjerde kapitel det mere elementære problem med at teste Fréchet-middelværdien af et stokastisk objekt. Vi introducerer en test baseret på stikprøvens empiriske Fréchet-varians og introducerer en randomiseringsmetodologi, der udnytter det metriske rums symmetrier. Vi illustrerer metodologien i forskellige typer af metriske rum og undersøger testens ydeevne under forskellige betingelser. Vi præsenterer et simulationsstudie for at illustrere testens ydeevne og anvender den på et datasæt med vinddata i Danmark.

Contributions and Structure

The first chapter provides an introduction to the thesis; it consists of a motivation to the work presented in this thesis together with an attempt at providing the necessary background in metric spaces and their statistical study in order to make the thesis self-contained. Each remaining chapter corresponds to one of the projects of the thesis, to which we provide a brief introduction, its status and the main contributions, followed by the paper itself. Chapter 2 presents our work in [**MRF**] on adapting random forest to the problem of regression in metric spaces. Chapter 3 contains the paper [**GAR**] which introduces assumptions on the metric space allowing to define a simple autoregressive model for random objects. Finally, Chapter 4 consists of the paper [**ISO**] which is concerned with the problem of testing the mean of a random object.

Chapter 2 (Constructing regressors in metric spaces)

[**MRF**] [Bulté and Sørensen, 2024b]. Bulté, M. and Sørensen, H. (2024b). Medoid splits for efficient random forests in metric spaces. *Computational Statistics & Data Analysis*, 198:107995.

Chapter 3 (Modeling time series without linearity)

[**GAR**] [Bulté and Sørensen, 2024a] Bulté, M. and Sørensen, H. (2024a). An Autoregressive Model for Time Series of Random Objects. arXiv:2405.03778.

Paper status: Under review at *Electronic Journal of Statistics*.

Chapter 4 (Exploiting symmetries for testing the mean)

[**ISO**] [Bulté and Sørensen, 2025] Bulté, M. and Sørensen, H. (2025). Isotropic randomization for one-sample testing in metric spaces. arXiv:2501.15945 [stat].

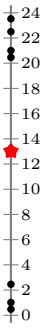
Paper status: working paper.

Contents

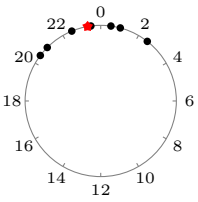
1 Introduction

1.1 A Playful Tale

Consider Ben[1], a PhD student working hard in the last months prior to submitting his thesis. Worried about the impact of his work on his sleep, he decides to start tracking the time at which he goes to bed over a week. On Monday, Ben went to bed at 10:30 PM, followed by 11:30 PM on Tuesday and 12:30 AM on Wednesday. As the week progressed and his deadline loomed closer, Ben's bedtime gradually shifted: 1:00 AM on Thursday and 2:30 AM on Friday when he decided to finish one last simulation. Exhausted but not done, he decides to recover over the weekend and ends the week by going to bed at 8:30 PM and 9:00 PM on Saturday and Sunday.



The first question Ben tries to answer in order to better understand his sleep is *"At what time do I go to bed on average?"*. He starts by converting the recorded bedtimes to a 24-hour format on a continuous scale (e.g., 11:45 PM becomes 23.75) giving the numbers $22.5, 23.5, 0.5, 1.0, 2.5, 20.5, 21.0 \in \mathbb{R}$. As he proceeds to taking the arithmetic mean, giving approximately 13.07 — that is, 1:04 PM — Ben pauses, "*This can't be right*", he thinks. Looking at these times on a line, he sees the problem clearly: while all his bedtimes cluster within a few night hours, their arithmetic mean falls on the opposite side of the clock.



Seeking illumination, Ben shares his dilemma with Mauricette, a fellow graduate student. Instead of offering a solution, she asks him a simple question: "*When you tell someone how much time passes between two events, how do you calculate it?*" Ben thinks about it. "*Well, if I went to bed at 11:00 PM yesterday and 1:00 AM today, I'd say that's... two hours apart.*" "*Exactly!*" Mauricette responds, "*You naturally measure time differences by following the circle of the clock, taking the shorter path. Maybe your average should respect this same structure?*".

Following this insight, they compute a new kind of average — one that respects how we naturally measure distances in time. They find the point on the circle that minimizes the average squared distance to all bedtimes, measuring distance along the clock's arc. This time, their result lands at 11:22 PM, marked as a red star in the illustration, exactly where intuition suggests — in the middle of Ben's nighttime sleep schedule.

[1] A similar version of this example was presented by C. Schötz in his talk *A Law of Large Numbers for Fréchet Mean Sets*

1.2 Metric Spaces and Elementary Properties

While Ben's dilemma might seem specific to circular data, it illustrates a fundamental challenge in modern statistics: how to properly analyze data that lives in non-Euclidean spaces. The bedtime example shows that even seemingly simple statistical tasks, like computing an average, require careful consideration of the underlying space's geometry. To develop a systematic approach to such problems, we need a mathematical framework that can accommodate various types of non-Euclidean data while maintaining enough structure for meaningful statistical analysis. This leads us to the theory of metric spaces which we will present in this section.

The fundamental mathematical abstraction underlying our work is that of metric spaces. This structure provides a general framework that can accommodate a wide variety of data types while retaining enough mathematical expressiveness to develop meaningful statistical procedures. For an in-depth treatment of metric spaces, we refer the reader to [Burago et al., 2001].

Definition 1 (Metric Space)**.** A metric space is a pair (Ω, d) where Ω is a set and $d : \Omega \times \Omega \to \mathbb{R}_+$ is a distance function satisfying three properties: for all $x, y, z \in \Omega$,

1. $d(x, y) \geq 0$ with equality if and only if $x = y$ (positivity)
2. $d(x, y) = d(y, x)$ (symmetry)
3. $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality)

In the remaining of the thesis, we loosely use the terms distance and metric interchangeably. The generality of this framework is to be noted: metric spaces can be anything from finite discrete sets to infinite-dimensional function spaces, from smooth manifolds to fractal sets. This flexibility allows us to develop statistical methods that can handle diverse types of data while maintaining mathematical rigor. However, this generality comes at a cost: without additional structure, the tools available for statistical analysis are limited. This motivates the study of specific assumptions and properties of the metric spaces. One common assumption is that of boundedness which is defined via the diameter of the space.

Definition 2 (Diameter of a metric space)**.** Let (Ω, d) be a metric space. The diameter of Ω is defined as:

$$\operatorname{diam}(\Omega) = \sup_{x,y \in \Omega} d(x, y).$$

We now show that boundedness is a useful property that allows us to derive stronger results. In particular, the following lemmas show that boundedness implies Lipschitz continuity of the distance function. This plays a central role in many of our theoretical developments and for which proofs can be found in the Appendix of Chapter 2

Lemma 1 (Lipschitz Continuity of Distance)**.** *In any metric space (Ω, d), the distance function $d(\cdot, \cdot)$ is 1-Lipschitz in both arguments with respect to itself. That is, for all $x, y, z \in \Omega$:*

$$|d(x, z) - d(y, z)| \leq d(x, y).$$

When working with bounded metric spaces, we obtain stronger properties for the squared distance:

Lemma 2 (Lipschitz Continuity of Squared Distance)**.** *In a bounded metric space (Ω, d) with diameter $D = diam(\Omega)$, the squared distance function $d^2(\cdot, \cdot)$ is $2D$-Lipschitz in both arguments. That is, for all $x, y, z \in \Omega$:*

$$\left|d(x,z)^2 - d(y,z)^2\right| \leq 2Dd(x,y).$$

1.2.1 Familiar Metric Spaces

Before presenting more abstract metric spaces, it is worthwhile to recall the structure of Hilbert spaces, as they provide a familiar setting where many geometric intuitions from $\mathbb{R}^n$ remain valid. A Hilbert space $(H, \langle \cdot, \cdot \rangle)$ is a complete inner product space, where the distance between two points is induced by the inner product:

$$d(x,y) = \sqrt{\langle x - y, x - y \rangle}.$$

Remark 1. The key feature that makes Hilbert spaces particularly amenable to statistical analysis is that they allow for vector space operations while preserving geometric structure. For instance, the notion of mean can be defined through standard arithmetic averaging:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

This operation is well-defined and coincides with the geometric notion of center that we will later generalize to metric spaces.

Another familiar example of a metric space is the $d-$dimensional sphere equipped with its intrinsic metric.

Example 1 (Sphere)**.** *Let $S^d = \left\{x \in \mathbb{R}^{d+1} : \|x\|_2 = 1\right\}$ be the $d-$dimensional sphere. The distance between two points $x, y \in S^d$ is given by the great circle distance:*

$$d_{S^d}(x,y) = \arccos(x^\top y).$$

A space of particular interest in statistics is the space of probability measures equipped with the Wasserstein metric [Ambrosio et al., 2005, Panaretos and Zemel, 2020]. This metric, which measures the "cost" of transforming one probability distribution into another, has found numerous applications in statistics, machine learning. Apart from its theoretical interest, it also finds practical applications in modeling situation where each individual presents a distribution, see for instance [**GAR**] for an application in time series analysis.

Example 2 (Wasserstein Spaces)**.** *Let (Ω, d) be a metric space and $\mathcal{P}(\Omega)$ be the set of Borel probability measures on Ω. The $p-$Wasserstein distance between two measures $\mu, \nu \in \mathcal{P}(\Omega)$ is defined as:*

$$d_{W_p}(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\Omega \times \Omega} d(x, y)^p d\pi(x, y) \right)^{1/p},$$

where $\Pi(\mu, \nu)$ is the set of all joint distributions on $\Omega \times \Omega$ with marginals μ and ν. The $p-$Wasserstein distance is a metric on $\mathcal{P}(\Omega)$. Most commonly, the $2-$Wasserstein distance is used and is referred to as the Wasserstein distance. The special case of probability measures on the real line equipped with the Wasserstein distance is of particular interest, both in terms of application as in terms of theoretical properties. There, for two probability measure μ and ν with quantile functions Q_μ and Q_ν, the Wasserstein distance is given by:

$$d_{W_2}(\mu, \nu) = \left(\int_0^1 (Q_\mu(u) - Q_\nu(u))^2 du \right)^{1/2}$$

1.2.2 Building Metric Spaces

While Hilbert spaces provide a comfortable setting for statistical analysis, many datasets naturally live in spaces with more complex geometric structure. We now present several fundamental ways to construct such spaces while preserving metric properties.

A first simple way to construct a valid distance on an arbitrary set Ω is by combining an existing metric space (Ω', d) and an injective map $\phi : \Omega \to \Omega'$.

Example 3. *Let Ω be a set, (Ω', d) be a metric space and $\phi : \Omega \to \Omega'$ be an injective map. The distance d_ϕ on Ω induced by ϕ is defined as:*

$$d_\phi(x, y) = d(\phi(x), \phi(y)).$$

Then, (Ω, d_ϕ) is a metric space.

This way of constructing metric spaces corresponds to the notion of *induced metric*. We can see that the Wasserstein distance on probability measures on the real line can be seen as a metric induced by the L^2 distance and function mapping distributions to their quantile function. Furthermore, metric spaces can be combined to form new spaces. The following example provides two common ways of doing this.

Example 4 (Product Metric Space)**.** *Given two metric spaces (Ω_1, d_1) and (Ω_2, d_2), the following distances on the product $\Omega_1 \times \Omega_2$ are valid distances,*

1. *The product metric: $d_p((x_1, x_2), (y_1, y_2)) = \sqrt{d_1(x_1, y_1)^2 + d_2(x_2, y_2)^2}$*
2. *The maximum metric: $d_{\max}((x_1, x_2), (y_1, y_2)) = \max\{d_1(x_1, y_1), d_2(x_2, y_2)\}$*

Another useful way of constructing a metric space is through the use of cones. A cone over a metric space (Ω, d) is the quotient $\Omega \times \mathbb{R}_+$ obtained by gluing together all the points in $\Omega \times \{0\}$.

Example 5 (Cone Metric)**.** *Let (Ω, d) be a bounded metric space diam$(\Omega) < \pi$, define the cone over Ω as Con$(\Omega) = \Omega \times \mathbb{R}_+ / \sim$ where $(\omega, s) \sim (\omega', t)$ if and only if $s = t = 0$. It is a metric space equipped with the distance*

$$d_c((\omega, s), (\omega', t)) = \sqrt{s^2 + t^2 - 2st\cos(d(\omega, \omega'))}.$$

Interestingly, it can be shown [Takatsu and Yokota, 2012] that the 2-Wasserstein space of distributions over a space $\mathcal{X}$ is a cone if and only if $\mathcal{X}$ is a cone.

1.2.3 Geodesics and Curvature

A large class of metric spaces for which a geometrical intuition is easier to develop are those in which points can be connected by special curves, called *geodesics*. Roughly speaking, geodesics generalize the notion of straight lines in Euclidean space to more general metric spaces. More formally, geodesics can be defined via the notion of length-minimization

Definition 3 (Length and Geodesics)**.** Let (Ω, d) be a metric space and $\gamma : [a, b] \to \Omega$ be a continuous map from an interval $[a, b] \subset \mathbb{R}$ to Ω. The length of γ is defined as

$$L(\gamma) = \sup \sum_{i=1}^{n} d(\gamma(t_i), \gamma(t_{i-1})), \tag{1.1}$$

where the supremum is taken over all partitions $a = t_0 < t_1 < \ldots < t_n = b$. A curve γ is a geodesic if for all $t_1, t_2 \in [a, b], d(\gamma(t_1), \gamma(t_2)) = L(\gamma|_{[t_1,t_2]})$ where $\gamma|_{[t_1,t_2]}$ is the restriction of γ to the interval $[t_1, t_2]$.

The endpoints $\gamma(a)$ and $\gamma(b)$ of a curve $\gamma : [a, b] \to \Omega$ are said to be *connected* by γ. By the definition of length provided in Equation (1.1), a curve cannot be shorter than the distance between the two points it connects, that is, $L(\gamma) \geq d(\gamma(a), \gamma(b))$. Spaces in which the distance between two points is equal to the infimum of the lengths of all curves connecting them are called *length spaces*. If the distance between two points is achieved by a unique curve, called *geodesic*, the space is then a *geodesic space*.

Geodesics can be used to provide a generalized notion of convexity in geodesic spaces which is used throughout the thesis.

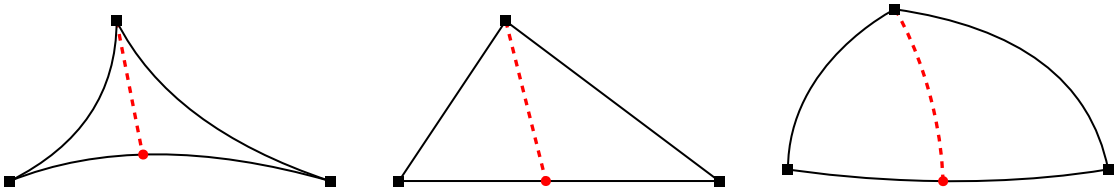
Definition 4 (Convexity)**.** Let (Ω, d) be a geodesic space. A function $f : \Omega \to \mathbb{R}$ is convex if for every two elements $x, y \in \Omega$ with connecting geodesic $\gamma : [0, 1] \to \Omega$, the function $f \circ \gamma : [0, 1] \to \mathbb{R}$ is convex, that is for all $t \in [0, 1]$,

$$f(\gamma(t)) \leq (1 - t)f(x) + tf(y). \tag{1.2}$$

If the inequality is strict, the function is said to be strictly convex. If furthermore there exists a strictly increasing function $\eta : \mathbb{R} \to \mathbb{R}$ such that the following always holds

$$f(\gamma(t)) \leq (1 - t)f(x) + tf(y) - t(1 - t)\eta(d(x, y)), \tag{1.3}$$

then f is called *uniformly convex*.



(a) nonpositive curvature (b) zero curvature (c) nonnegative curvature

Figure 1.1: Illustration of triangles in spaces of different curvatures. For triangles with same side lengths in spaces of negative, zero, and positive curvature, the distance from each vertex to the opposite side is respectively smaller, equal, and larger than in their comparison triangle.

To better understand the behavior of random objects in geodesic spaces, it is helpful to examine a geometric notion of *curvature* which provide a characterization of the convexity or concavity of the distance function. The notion of curvature can be defined in several equivalent ways, see Chapter 4 of [Burago et al., 2001]. While a numerical notion of curvature can be introduced, it is often sufficient to consider this qualitative definition. Here, we chose to present the concept through *comparison triangles*. Let (Ω, d) be a geodesic space and $x, y, z \in \Omega$ be three points. A *triangle* $\triangle xyz$ is the set formed by the three geodesics connecting the pairs of points (x, y), (y, z), and (z, x), which we denote by $[xy], [yz]$ and $[xz]$. The *comparison triangle* $\triangle \bar{x}\bar{y}\bar{z}$ is the triangle in $\mathbb{R}^2$ formed by the points $\bar{x}, \bar{y}, \bar{z} \in \mathbb{R}^2$ such that

$$\|\bar{x} - \bar{y}\|_2 = d(x, y), \quad \|\bar{y} - \bar{z}\|_2 = d(y, z), \quad \|\bar{z} - \bar{x}\|_2 = d(z, x).$$

The existence of this triangle in $\mathbb{R}^2$ follows for the triangle inequality. Considering the distance between vertices of a triangle to their opposite sides leads to the classification of geodesic spaces in three categories: nonpositive curvature, zero curvature, and positive curvature.

Definition 5 (Curvature)**.** A geodesic space (Ω, d) is a space of nonpositive (respectively nonnegative) curvature if the following holds in every neighborhood of Ω: for every $\triangle xyz$ in Ω and $p \in [xy]$, the equivalent point $\bar{p}$ with $d(x, p) = \|\bar{x} - \bar{p}\|_2$ in the comparison triangle $\triangle \bar{x}\bar{y}\bar{z}$ satisfies

$$d(z, p) \leq \|\bar{z} - \bar{p}\|_2 \quad (\text{respectively } d(z, p) \geq \|\bar{z} - \bar{p}\|_2).$$

The space is said to have zero curvature if the above inequality is an equality for all $\triangle xyz$ and $p \in [xy]$.

This motivates the intuition of "thickness" of triangles in a space: in spaces of nonpositive curvature, triangles are "thin" in the sense that the distance from a vertex to the opposite side is smaller than it would be in the plane $\mathbb{R}^2$. In spaces of nonnegative curvature, triangles are "fat" in the sense that the distance is larger than in the plane. This intuition is illustrated in Figure 1.1.

The concept of curvature, plays an important role in determining the behavior of statistical procedures. For instance, as mentioned earlier, one can show that the notion of curvature relates to the convexity of the distance and squared distance functions as we show in the following lemma.

Lemma 3. *Let (Ω, d) be a geodesic space and $x \in \Omega$. If Ω has nonpositive curvature then the function $d_x(y) = d(x, y)^2$ is uniformly convex with $\eta(t) = t^2$.*

Proof. Let $x, y, z \in \Omega$ be distinct and γ be the geodesic connecting x and y. It is sufficient to show that Equation (1.3) holds for $t = 1/2$. Let $p = 1/2$ and $\bar{x}, \bar{y}, \bar{z}, \bar{p} \in \mathbb{R}^2$ be defined as the definition of the comparison triangle. We then have $\bar{p} = (\bar{x} + \bar{y})/2$ and hence since Ω has nonpositive curvature, we obtain

$$\begin{aligned} d(z, p)^2 &\leq \|\bar{z} - \bar{p}\|_2^2 = \|\bar{z}\|_2^2 + \|\bar{p}\|_2^2 - 2\langle \bar{z}, \bar{p} \rangle \\ &= \frac{1}{2}\|\bar{x} - \bar{z}\|_2^2 + \frac{1}{2}\|\bar{y} - \bar{z}\|_2^2 - \frac{1}{4}\|\bar{x} - \bar{y}\|_2^2 \\ &= \frac{1}{2}d(x, z)^2 + \frac{1}{2}d(y, z)^2 - \frac{1}{4}d(x, y)^2. \end{aligned}$$

□

We will see in the next section how this property can be useful in the study of random objects.

In Euclidean minimization problems, convex functions are particularly advantageous due to their desirable properties. Similarly, in the context of geodesic spaces, numerous results — such as existence and uniqueness of a minimizer — are available for minimizing geodesically convex functions; see Section 1 of [Sturm, 2003]. In particular, [Sturm, 2003, Proposition 1.7] establishes that in a complete geodesic space, lower-semicontinuous and uniformly convex functions admit a unique minimizer.

1.3 Statistical Analysis in Metric Spaces

1.3.1 Population quantities

Now that we have defined the theoretical quantities defining the types of objects we will study, we can approach the question: What does it mean to compute an average? When working with real numbers, the arithmetic mean provides a natural measure of central tendency—a single value that represents an entire distribution. However, when working with data in more general metric spaces, the notion of an average becomes more subtle. How does one average shapes, probability distributions, or, as seen in the illustrative example in Section 1.1, times on a clock?

In his seminal work in the late 1940s [Fréchet, 1948], Maurice Fréchet lays the theoretical foundations for the study of random variables in metrics spaces. This includes the question of finding representative elements in these spaces. Building on the observation that the arithmetic mean minimizes the sum of squared deviations, Fréchet proposed a

general framework for defining means in metric spaces through a minimization principle. This approach has become the foundation of statistical analysis in metric spaces.

To extend classical statistical concepts to metric spaces, we must first address a fundamental question: What does it mean to be a "random element" in such spaces? In Euclidean space, random vectors are defined through their coordinates, but metric spaces generally lack such a coordinate system. Instead, we need to build the theory from the metric structure itself. This leads us to the following definition:

Definition 6 (Random objects)**.** Let (Ω, d) be a metric space equipped with its Borel σ-algebra $\mathcal{B}(\Omega)$, that is, the σ-algebra generated by open sets in the metric topology. A *random object* X taking values in Ω is a measurable function from some probability space $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ to $(\Omega, \mathcal{B}(\Omega))$.

To develop meaningful statistical theory, we need to ensure that certain moments of our random object exist. This leads to the definition of L^p spaces via the moments of the distance function.

Definition 7 (L^p spaces of random objects)**.** For $p \geq 1$, we define $L^p(\Omega)$ as the space of random objects X in Ω such that

$$\mathbb{E}[d(X, \omega)^p] < \infty \tag{1.4}$$

for some (and hence all) $\omega \in \Omega$. The expectation in (1.4) is called *the p-th distance moment of* X.

The parenthetical "and hence all" in the above definition is justified by the following. If Equation (1.4) holds for some $\omega \in \Omega$, then it holds for all $\omega' \in \Omega$. Indeed, by the triangle inequality and Jensen's inequality:

$$\mathbb{E}\big[d(X, \omega')^p\big] \leq \mathbb{E}\big[(d(X, \omega) + d(\omega, \omega'))^p\big] \leq 2^{p-1}\mathbb{E}[d(X, \omega)^p] + 2^{p-1}d(\omega, \omega')^p < \infty.$$

With these foundations in place, we can now consider how to define a notion of centrality for random objects. In Euclidean spaces, the expectation $\mathbb{E}[X]$ of a random vector X can be characterized as the vector minimizing the expected squared distance to X. This minimization principle, recognized by Maurice Fréchet in [Fréchet, 1948], provides a natural path to generalizing means to metric spaces.

Definition 8 (Fréchet function)**.** Let (Ω, d) be a metric space and X be a random object in $L^2(\Omega)$. The Fréchet function $F_X : \Omega \to \mathbb{R}_+$ is defined as

$$F_X(\omega) = \mathbb{E}[d(X, \omega)^2]. \tag{1.5}$$

Using this function, we can define the generalized notion of mean of a random object X, called the *Fréchet mean*.

Definition 9 (Fréchet mean and Fréchet Variance)**.** Let (Ω, d) be a metric space and X be a random object in $L^2(\Omega)$. Any element $\mu \in \Omega$ minimizing the Fréchet function,

$$\mu \in \operatorname*{arg\,min}_{\omega \in \Omega} F_X(\omega), \tag{1.6}$$

is called a Fréchet mean of X and is denoted by $\mathbb{E}[X]$. The Fréchet variance of X is defined as the minimum value of the Fréchet function,

$$\mathrm{Var}[X] = \min_{\omega \in \Omega} F_X(\omega). \tag{1.7}$$

Note that this approach can be readily extended to the problem of regression, as introduced in [Petersen and Muller, 2019]. Given a random object $Y \in L^2(\Omega)$ and a random vector $X \in \mathbb{R}^d$, the Fréchet regression of Y on X is the Fréchet mean of Y conditional on $X = x$. This quantity is defined as the minimizer of the conditional Fréchet function

$$F_{Y|X}(x;\omega) = \mathbb{E}\big[d(Y,\omega)^2 \mid X = x\big].$$

The minimizer of this function is called the Fréchet regression of Y on X and is denoted by $\mathbb{E}[Y \mid X = x]$. This is the setting we consider in [**MRF**].

While assuming $X \in L^2(\Omega)$ and defining the Fréchet function via the second distance moment is convenient, this assumption can be relaxed without altering the resulting quantities.

Remark 2. Assume that the random object X is now in $L^1(\Omega)$ and fix an arbitrary point $o \in \Omega$. Then define the function

$$G_X(\omega) = \mathbb{E}\big[d(X,\omega)^2 - d(X,o)^2\big]. \tag{1.8}$$

We have that G_X is well-defined and finite for all $\omega \in \Omega$ since by Jensen's inequality together with the Lipschitz continuity of distances mentioned in Lemma 1, we have

$$\begin{aligned} |G_X(\omega)| &= \big|\mathbb{E}\big[d(X,\omega)^2 - d(X,o)^2\big]\big| \\ &= |\mathbb{E}[(d(X,\omega) - d(X,o))(d(X,\omega) + d(X,o))]| \\ &\leq \mathbb{E}[|d(X,\omega) - d(X,o)|(d(X,\omega) + d(X,o))] \\ &\leq d(\omega,o)\mathbb{E}[d(X,\omega) + d(X,o)] < \infty. \end{aligned}$$

Furthermore, when $X \in L^2(\Omega)$, both G_X and F_X are well-defined and one can see that they both achieve their minimum in the same points. This definition of the Fréchet mean can be found for instance in Proposition 4.3 of [Sturm, 2003].

With these definitions in place, one can come back to the illustrative example used in Section 1.1 and compute the Fréchet mean of the random object X representing Ben's bedtime. There, the metric space is the circle S^1 and the distance is the angular distance defined in Example 1. The Fréchet mean of X is then the point on the circle minimizing the expected squared angular distance to X, as defined in Definition 9. A careful reader could have asked: what if we had chosen a different reference point on the circle and taken the Euclidean mean of the angles, wouldn't that have solved the problem? The following remark addresses this question.

Remark 3. A natural approach would be to construct a mapping $\phi_\mu : S^1 \to \mathbb{R}$ that "*unwraps*" the circle into the real line by measuring oriented distances from a reference point μ. If X has Fréchet mean μ, one has that $\phi_\mu^{-1}(\mathbb{E}[\phi_\mu(X)]) = \mu$. However, this approach fails for two fundamental reasons:

1. The construction of ϕ_μ requires prior knowledge of the Fréchet mean μ, which is precisely what one aims to estimate.

2. No single unwrapping can correctly handle all distributions. Consider, for instance, a mixture of two point masses:

$$X \sim \frac{1}{2}(\delta_{\mu+\pi-\varepsilon} + \delta_{\mu+\pi+\varepsilon}), \quad \varepsilon \in (0, \pi/2)$$

   While X has Fréchet mean $\mu + \pi$, the unwrapping gives $\phi_\mu^{-1}(\mathbb{E}[\phi_\mu(X)]) = \mu$ instead.

As mentioned in the previous section, the behavior of the Fréchet mean can be better understood by considering the geometry of the space in which the random object lives. In particular, the Fréchet mean in spaces of nonpositive curvature has been extensively studied. In these spaces, as shown in Lemma 3, the squared distance is a convex function, which is inherited by the Fréchet function. This convexity property is sufficient to prove the existence and uniqueness of the Fréchet mean, see [Sturm, 2003] for a detailed presentation and review of results.

While spaces of nonpositive curvature provide a rich and convenient setting for studying Fréchet means through convexity, the situation in positively curved spaces is more delicate. A simple example illustrating this phenomenon is the uniform distribution on the circle S^1. In this case, every point on the circle is a Fréchet mean of the distribution.

In some of the spaces presented in Section 1.2.2, the Fréchet mean has a more explicit form, in terms of the underlying metric spaces used. For instance in the case of Example 3, if the map $\phi : \Omega \to \Omega'$ is invertible, it is straightforward to show that the Fréchet mean of a random object $X \in \Omega$ can be computed as $\mathbb{E}[X] = \phi^{-1}(\mathbb{E}[\phi(X)])$. Similarly, in the case of the product metric d_p in Example 4, the Fréchet mean of a random object $X = (X_1, X_2) \in \Omega_1 \times \Omega_2$ can be computed as $\mathbb{E}[X] = (\mathbb{E}[X_1], \mathbb{E}[X_2])$.

A useful way to understand how the Fréchet mean ties to the geometry of the space is through its connection to geodesics. Indeed, let (Ω, d) be a geodesic space, $x, y \in \Omega$ and $X_p \sim (1-p)\delta_x + p\delta_y$ for $p \in (0, 1)$. The following lemma shows that the curve of Fréchet means $p \mapsto \mathbb{E}[X_p]$ coincides with the geodesic γ connecting x to y.

Lemma 4. *In the above scenario, the Fréchet mean of X_p is given by $\gamma(p)$.*

Proof. Let $F_p(\omega) = F_{X_p}(\omega) = (1-p)d(\omega, x)^2 + pd(\omega, y)^2$ be the Fréchet function of X_p. For any $\omega \in \Omega$, the triangle inequality yields $d(y, \omega) \geq d(x, y) - d(x, \omega)$. This inequality leads to a lower bound $F_p(\omega) \geq h(\omega)$, where $h(s) = (1-p)s^2 + p(d(x, y) - s)^2$ for $s > 0$. The quadratic function h attains its minimum at $s = pd(x, y) = d(x, \gamma(p))$. Substituting

this value into the lower bound of the Fréchet function gives

$$F_p(\omega) \geq h(d(x, \gamma(p))) = (1-p)d(x, \gamma(p))^2 + p(d(x,y) - d(x, \gamma(p)))^2 = F_p(\gamma(p)).$$

Thus, $\gamma(p)$ minimizes F_p and is therefore the Fréchet mean of X_p. □

While this approach does not fully capture the behavior of the Fréchet mean in a given space, it provides insights into its behavior in geodesic spaces. In practice, one can explore the impact of the choice of metric on the analysis by a visual exploration of how geodesics interpolate between points in the space.

1.3.2 Estimation of the Fréchet mean

The theoretical framework developed in the previous sections naturally leads to questions of statistical estimation. Given observations $X_1, ..., X_n \overset{iid}{\sim} P$ in a metric space (Ω, d), the goal is to estimate the Fréchet mean $\mathbb{E}[X]$ of the underlying distribution P. This is typically done by considering the empirical counterparts to the quantities described in the previous section. This estimation approach falls within the broader statistical framework of M-estimation, see for instance [van der Vaart and Wellner, 2023, Chapter 3.2].

Definition 10 (Empirical Fréchet Function)**.** Let (Ω, d) be a metric space and $X_1, ..., X_n$ be random objects in Ω. Define the *empirical Fréchet function* $\hat{F}_n$, the empirical version of the Fréchet function in Equation (1.5), as

$$\hat{F}_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} d(X_i, \omega)^2. \tag{1.9}$$

Minimization of the empirical Fréchet function $\hat{F}_n$ yields the empirical Fréchet mean $\hat{\mu}_n$ and the empirical Fréchet variance $\hat{V}_n$.

Definition 11 (Empirical Fréchet mean and variance)**.** Let (Ω, d) be a metric space and $X_1, ..., X_n$ be random objects in Ω. The *empirical Fréchet mean* $\hat{\mu}_n \in \Omega$ is defined as any element minimizing the empirical Fréchet function,

$$\hat{\mu}_n \in \arg\min_{\omega \in \Omega} \hat{F}_n(\omega). \tag{1.10}$$

Similarly, the *empirical Fréchet variance* $\hat{V}_n$ is defined as the minimum value of the empirical Fréchet function,

$$\hat{V}_n = \min_{\omega \in \Omega} \hat{F}_n(\omega). \tag{1.11}$$

While the minimizer of the Fréchet function in Euclidean spaces has a closed-form solution, the arithmetic average, computing the Fréchet mean generally requires resorting to numerical optimization. Furthermore, in positively curved spaces, the empirical Fréchet mean may not be unique, even when the population Fréchet mean is. This

phenomenon of non-uniqueness poses challenges both for computation and theoretical analysis.

The consistency of empirical Fréchet means has been extensively studied in the literature, with results of increasing generality. In the most structured case of Riemannian manifolds, [Eltzner and Huckemann, 2018] and [Bhattacharya and Lin, 2017] establish both consistency and asymptotic normality of empirical Fréchet means, with limiting distributions characterized via the tangent space. For bounded metric spaces, [Petersen and Muller, 2019] proves consistency under conditions controlling the space complexity — via entropy integrals — while [Dubey and Müller, 2019] establishes a central limit theorem for the empirical Fréchet variance.

The extension to unbounded spaces requires additional care since the squared distance might not be Lipschitz anymore, see the proof of Lemma 1. In [Schötz, 2019], a Lipschitz-type condition on the squared distance — which is satisfied for instance in Hadamard spaces — to prove convergence of empirical Fréchet means. The most general setting, where Fréchet means may not be unique, is addressed in [Schötz, 2022] through a strong law of large numbers for Fréchet mean sets in general metric spaces. Some of these results form the basis for the theoretical analysis in [**MRF**] and [**ISO**].

1.4 Conclusion

In perhaps our boldest cross-disciplinary leap yet, let us examine the opening story of Ben in Section 1.1 through Dante's four levels of interpretation [Alighieri, 1307] – a reading that, while deliberately overwrought, reveals the unifying principles and motivation of the research presented in this thesis.

Literal. Our tale follows a graduate student tracking his bedtimes over a week. Attempting to make sense of his shifting sleep schedule, he naturally computes their arithmetic mean, only to arrive at the nonsensical result of 1:04 PM. Through a conversation with a fellow student about how we intuitively measure time differences, he discovers that measuring distances along the circle as a basis for his average calculation yields 11:22 PM, capturing the center of his nighttime schedule.

Allegorical. Ben's struggle exemplifies a broader phenomenon: as measurement techniques and data collection become increasingly sophisticated, we encounter observations that naturally live in non-Euclidean spaces. From shape analysis in medical imaging to probability distributions in the study of populations, straightforward application of classical statistical methods — designed with Euclidean geometry in mind — can lead to misleading or meaningless results.

Moral. This story emphasizes a fundamental principle in statistical thinking: the importance of aligning methodology with the structure of our data. Just as Ben discovered that meaningful averages emerge from respecting the circle's geometry, the path to valid statistical methodology lies in careful consideration of mathematical structure, whether this manifests in explicit geometric constraints or in the examination of model assumptions.

Anagogical. At its most abstract level, this work points to the rich mathematical structures that emerge when we step beyond Euclidean spaces. While our methods are motivated by practical statistical problems, they lead us to explore geometric structures and their interactions with randomness that hold intrinsic mathematical interest beyond their immediate practical applications.

This thesis contributes to this broader program by developing statistical methodology in metric spaces across three fundamental tasks: testing, time series analysis, and regression. Each chapter considers how classical statistical tools can be adapted while preserving their essential properties and computational feasibility. Throughout, we attempt to maintain a balance between theoretical guarantees and practical applicability, demonstrating the utility of our proposed methods through both theoretical results and empirical studies.

2 Constructing regressors in metric spaces

This chapter contains the following paper:

[**MRF**] [Bulté and Sørensen, 2024b]. Bulté, M. and Sørensen, H. (2024b). Medoid splits for efficient random forests in metric spaces. *Computational Statistics & Data Analysis*, 198:107995.

The initial motivation for this paper stemmed from an attempt to extend the asymptotic normality results for random forests of [Wager and Athey, 2018] to the Wasserstein space setting. However, as detailed in Section 2.2, we identified an error in their central limit theorem proof. This led us to broaden our investigation to address random forests in general metric spaces, building upon the weight-combination framework presented in [Athey et al., 2019] and adapting the pointwise convergence results of [Wager and Athey, 2018] for the Fréchet regression setting. During this process, we discovered the recent work of [Qiu et al., 2024], which had developed similar methodology and theoretical results. This discovery prompted us to refocus our contribution on the computational aspects of the problem. Specifically, we identified that the calculation of Fréchet means within the splitting procedure created a significant computational bottleneck. Our key contribution thus became the development of a medoid-based splitting criterion that maintains theoretical guarantees while substantially improving computational efficiency.

Medoid splits for efficient random forests in metric spaces

Matthieu Bulté[a], Helle Sørensen[a]

[a]*Department of Mathematical Sciences, University of Copenhagen,*

Abstract

This paper revisits an adaptation of the random forest algorithm for Fréchet regression, addressing the challenge of regression in the context of random objects in metric spaces. Recognizing the limitations of previous approaches, we introduce a new splitting rule that circumvents the computationally expensive operation of Fréchet means by substituting with a medoid-based approach. We validate this approach by demonstrating its asymptotic equivalence to Fréchet mean-based procedures and establish the consistency of the associated regression estimator. The paper provides a sound theoretical framework and a more efficient computational approach to Fréchet regression, broadening its application to non-standard data types and complex use cases.

Keywords: Least squares regression, Medoid, Metric spaces, Random forest, Random objects

1. Introduction

We study the extension of random forest to regression situations where the response takes values in a metric space. The usual expectation is replaced by the Fréchet mean, which can be be very computationally intensive with restrictions on possible applications to follow. Emphasis is therefore on a new medoid-based splitting rule used in the individual trees. Application of this new splitting rule speeds up computations without compromising consistency

Email addresses: mb@math.ku.dk (Matthieu Bulté), helle@math.ku.dk (Helle Sørensen)

or finite-sample prediction quality and makes it possible to analyzer larger datasets in terms of sample size and dimension.

Random forest is a statistical learning method introduced by Breiman (2001) for classification and regression. It constructs predictions by averaging the predictions from a randomized ensemble of regression trees fitted on a training dataset. Random forests have proven to perform well in practice on complex and high-dimensional datasets with little parameter tuning, making them an ubiquitous method in the machine learning toolset.

With the success of statistical analysis, more domains are being analyzed, leading to the study of more complex use cases and non-standard data types. A first generalization that has emerged from this need is the study of random variables taking values in manifolds. More recently, a further generalization has started to be studied, in which the random variables, referred to as random objects, take values in a metric space. This general setting does not require anything more than a distance function, and does not require common algebraic structures such as those found in vector spaces or on manifolds. A core quantity in the study of random objects is the Fréchet mean, as defined by Fréchet (1948), which generalizes the concept of the expected value to metric spaces by defining it as the minimizer of the expected squared loss. This notion allows to generalize the definition of the variance and further construct tools to study the variability of random objects.

The lack of algebraic structure in the study of random objects makes the development of regression algorithms particularly challenging. Efforts in adapting regression algorithms to non-Euclidean objects have mainly been done for the case where the response takes values in finite-dimensional differentiable Riemannian manifolds. This has led to the development and analysis of parametric models exploiting the local structure of the space, such as geodesic regression approaches (Thomas Fletcher, 2013), but also fully non-parametric approaches via adaptations of the Nadaraya-Watson estimator (Davis et al., 2010; Hinkle et al., 2012; Pelletier, 2006; Yuan et al., 2012; Hein, 2009). More recently, Petersen and Müller (2019) proposed a general framework for regression in metric spaces with Euclidean covariates in which the regression method specifies a weighting scheme. For a given input, the training samples are reweighted following that scheme and the prediction is formed by computing the weighted Fréchet mean. They then define a global weighting scheme generalizing the classical linear regression and a local scheme which generalizes local linear regression. This original work paved the way to the construction and study of more flexible regression

techniques adapted to Fréchet regression.

Random forest is a well-suited candidate for adaptation to Fréchet regression. Several authors have already used random forests outside the classical setup of Euclidean response, for example for function spaces (Nerini and Ghattas, 2007; Fu et al., 2021) by either working directly in the Hilbert space L_2 or via a transformation to L_2, but these methods lack a theoretical analysis and their adaptation to metric space data is unclear. Another attempt was made by Capitaine et al. (2020) by direct translation of the random forest algorithm via a two-stage approach in which the forest prediction is constructed as the Fréchet mean of the prediction of the trees. Their approach has been successfully applied to general settings in which both the covariates and response are allowed to be in a metric space and proposes a novel splitting rule that reduces the computational cost of the tree-fitting procedure. Unfortunately, this two-stage procedure complicates theoretical analysis, and to our knowledge, no consistency result for their method has been developed.

Another route for adapting random forest to Fréchet regression is to fit a random forest and aggregate the trees into weights for the observations, which can be used to construct a prediction. This approach was developed by Meinshausen (2006) for quantile regression, and Athey et al. (2019) for conditional Z-estimation under the name of Generalized Random Forest (GRF). It has been exploited by Qiu et al. (2022) to construct a random forest adapted to Fréchet regression. The authors combined this with the splitting rule proposed by Capitaine et al. (2020) to alleviate some of the computational burden of tree fitting. The paper presents results about asymptotic properties of the estimator, but proofs are not provided. Moreover, the results regarding rates of convergence and a central limit theorem depend on strong assumptions regarding the random forest algorithm itself.

One essential ingredient in the adaptation of a random forest algorithm is the construction of a suitable splitting rule used to recursively partition the covariate space in the fitting of individual trees. In their work, Athey et al. (2019) propose a general splitting rule for Z-estimation relying on gradients of the objective function to build splits capturing the heterogeneity of the response. This approach cannot be used for Fréchet regression since it requires a differentiability structure which is not available. A possible approach in Fréchet regression is to use the standard splitting rule when the response is in $\mathbb{R}$, namely by finding the splitting position that results in the highest reduction in variance in the subgroups. However, for each split, this requires

computing a large number of Fréchet variances and hence Fréchet means. In many cases, computing the Fréchet mean is expensive, making this approach unusable. The approach proposed by Capitaine et al. (2020) and used in Qiu et al. (2022) based on 2-means clustering for efficient splitting reduces the number of Fréchet means required to a great extent but, as highlighted by the authors, is still computationally intensive.

In this article, we revisit the algorithm proposed by Qiu et al. (2022), which implements GRF for Fréchet regression. Our main contribution is a new splitting rule which replaces the computation of Fréchet means with medoids, i.e., sample members. This, after the computation of a distances matrix, allows to find splits as efficiently as in the Euclidean case. We denote the adapted procedure Metric Random Forest (MRF). We prove that the new splitting rule is asymptotically equivalent to using the Fréchet mean-based procedure. Furthermore, we prove the consistency of the MRF based on classical M-estimation results from van der Vaart (1998) and following the proof strategy presented in Wager and Athey (2018), and illustrate the benefit of the new splitting rule with numerical experiments.

The paper is organized as follows: Section 2 includes background material on Fréchet regression and existing work on random forest in metric spaces. The new medoid-based splitting rule is introduced in Section 3, and the consistency result for the MRF estimator is stated in Section 4. We present simulations from three scenarios in Section 5 and conclude in Section 6. Proofs are presented in the appendix.

2. Background

2.1. Fréchet regression

Let (Ω, d) be a compact metric space. Consider a pair of random variables (X, Y), where X is a vector of predictors taking value in $[0,1]^d$ with density bounded away from 0 and ∞, and Y is a response taking value in Ω. Following the idea of Fréchet (1948), the notions of mean and variance can be generalized to random objects to obtain the Fréchet mean and Fréchet variance,

$$\omega_\oplus = \operatorname*{argmin}_{\omega \in \Omega} \mathbb{E}\left[d(Y,\omega)^2\right], \qquad V_\oplus = \mathbb{E}\left[d(Y,\omega_\oplus)^2\right]. \tag{1}$$

Building upon these generalized concepts of mean and variance, Müller (2016) define the Fréchet regression function, denoted m, in terms of the conditional

distribution of Y given $X = x$,

$$m(x) = \underset{\omega \in \Omega}{\operatorname{argmin}}\, M(\omega; x), \qquad M(\omega; x) = \mathbb{E}\left[d(Y, \omega)^2 \mid X = x\right], \tag{2}$$

where M is called the conditional Fréchet function.

We are interested in fitting a regression model with response Y and predictors X. Given a dataset $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ of independent random pairs following the same distribution as the prototypical pair (X, Y) and a point $x \in [0, 1]^d$ for which a prediction is to be obtained, the goal is to construct an estimator $m_n(x)$ of the Fréchet regression function $m(x)$. One generally applicable solution to this problem is to estimate the conditional Fréchet function $M(\cdot; x)$ with a data-dependent objective function $M_n(\cdot; x)$ converging to $M(\cdot; x)$ in such a way that the maximizer $m_n(x)$ of $M_n(\cdot; x)$ also converges to $m(x)$. A natural approach consists in using a re-weighting strategy to approximate the conditional expectation in (2), yielding an estimator of the form

$$M_n(\omega; x) = \sum_{i=1}^n w_i(x) d(\omega, Y_i)^2, \tag{3}$$

where the weights $\{w_i(x)\}_{i=1}^n$ depend on the data and the point x for which the prediction is being made. The construction of such weights has been the subject of a large corpus of literature. A common and well studied method is the Nadaraya-Watson estimator weights which defines weights through a kernel function K and a bandwidth parameter $h \in (0, \infty)$, giving $w_i(x) = K_h(X_i - x)$ with $K_h(\cdot) = h^{-1}K(\cdot/h)$. This method is well understood and easy to implement, but it requires a crucial tuning of the bandwidth parameter h, suffers from the curse of dimensionality and is not adaptive, in the sense that the weights are invariant to the part of the covariate space where the prediction is made. This lack of adaptivity makes the Nadaraya-Watson estimator practically ill-suited in case of heteroskedasticity of the response for different values of the predictors.

2.2. Random forest in metric spaces

The original and commonly used random forest algorithm for regression proposed by Breiman (2001) relies on constructing an ensemble of randomized regression trees. A prediction at a given point is then formed by averaging the predictions across the ensemble of trees. The fitting of the trees is randomized by bootstrap aggregation, meaning that each regression tree is

fitted on a random subset of size s of the training data. Each tree is built greedily from a recursion of splits, each attempting to partition the training sample available at the node into two subsets of minimal prediction error (Breiman et al., 1984). The procedure results in a partitioning of the space of predictors into a collection of rectangles called *leaves*. The resulting random forest estimator then takes the form

$$\mathrm{RF}(x) = \frac{1}{B}\sum_{b=1}^{B} \mathrm{T}_b(x) \qquad \text{with} \qquad \mathrm{T}_b(x) = \sum_{i=1}^{n} \frac{\mathbb{1}\{X_i \in L_b(x)\}}{\#\{j : X_j \in L_b(x)\}} Y_i, \quad (4)$$

where B is the size of the ensemble, T_b is the b-th regression tree and $L_b(x)$ is the subset of training points used for fitting T_b falling in the same leaf as x.

In order to generalize this idea to the context of Fréchet regression, Capitaine et al. (2020) directly translate the average-of-averages construction of the random forest by replacing averages with Fréchet means. The resulting estimator, called the Fréchet Random Forest (FRF), is constructed as the Fréchet mean of an ensemble of randomized Fréchet trees (FTs) in which each tree is built to minimize the prediction error, as done in the Euclidean regression setup.

In this manuscript, a different approach, first proposed in Qiu et al. (2022), is followed to constructing a random forest prediction. It does not aggregate the prediction of each individual tree; instead the results of the fitted ensemble are aggregated following the idea presented in Meinshausen (2006) and Athey et al. (2019). For a given x, each tree contributes with a weighting of the observations reflecting the partitioning of the predictor space fitted by the tree. Consider a source of auxiliary randomness ξ and a subset of the data $D \subset \mathcal{D}$, we denote by $\{w_i(\cdot; \xi, D)\}_{i=1}^n$ the set of weights resulting from fitting a single tree on D. These weights take the form

$$w_i(x; \xi, D) = \frac{\mathbb{1}\{X_i \in L(x; \xi, D)\}}{N(L(x; \xi, D))}, \quad (5)$$

where $N(C) = \#\{j : X_j \in C\}$ for any $C \subset [0,1]^d$, and $L(x; \xi, D)$ represents the observations in D falling in the same leaf as x. Note that for observations $(X_i, Y_i) \in \mathcal{D} \backslash D$, the associated weight function $w_i(\cdot; \xi, D)$ is zero. Given B random subsets $\{D_b\}_{b=1}^B$ of $\mathcal{D}$ and auxiliary randomness $\{\xi_b\}_{b=1}^B$, the weights constructed by the individual trees are averaged within the ensemble to construct an adaptive set of weights $\{w_i\}_{i=1}^n$ for a given prediction point x given

by

$$w_i(x) = \frac{1}{B}\sum_{b=1}^{B} w_i(x; \xi_b, D_b). \tag{6}$$

These weights are then used to form an estimator of the conditional Fréchet function defined in (2),

$$M_n(\omega; x) = \sum_{i=1}^{n} w_i(x) d(\omega, Y_i)^2. \tag{7}$$

Minimizing this adaptively weighted sum yields our Metric Random Forest (MRF) estimator of the Fréchet regression function in (2),

$$m_n(x) = \operatorname*{argmin}_{\omega \in \Omega} M_n(\omega; x). \tag{8}$$

It is important to note that, in general, this optimization problem needs not have a unique solution. In practice, a user of the MRF must provide a solver for (8) that can return a unique solution based on any preferred heuristic implemented by the user (e.g., minimal norm, first in lexicographic order). Since this technicality does not influence the implementation or analysis of the MRF algorithm, we can safely assume that $m_n(x)$ exists and is unique. Note that we later assume in Assumption 1 that the population version of this quantity exists and is unique.

3. Splitting criterion in metric spaces

The recursive partitioning used to construct each individual tree is driven by the choice of a splitting criterion. Considering an arbitrary cell $C \subset [0,1]^d$, the goal at each partitioning step is to find the feature index $j \in \{1, \ldots, d\}$ and cut position $z \in [0,1]$ defining an axis-aligned plane along which to split the cell into two subcells, C_l and C_r, given by

$$C_l = \left\{x \in A : x^{(j)} \leq z\right\} \qquad \text{and} \qquad C_r = \left\{x \in A : x^{(j)} > z\right\},$$

where $x^{(j)}$ is the jth coordinate of x. Intuitively, each splitting pair (j, z) should be chosen as to improve the estimation of the Fréchet regression function m. This is done by minimizing the CART splitting criterion of Breiman

et al. (1984). Let us denote by $\hat{\omega}_{\oplus}(C)$ and $\hat{V}_{\oplus}(C)$ the empirical Fréchet mean and variance computed on the observations with X contained in C,

$$\hat{\omega}_{\oplus}(C) = \underset{\omega \in \Omega}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbb{1}\{X_i \in C\} d(Y_i, \omega)^2,$$
$$\hat{V}_{\oplus}(C) = \sum_{i=1}^{n} \frac{\mathbb{1}\{X_i \in C\}}{N(C)} d(Y_i, \hat{\omega}_{\oplus}(C))^2.$$

The CART splitting criterion is then given by

$$\operatorname{err}(j, z) = \frac{N(C_l)}{N(C)} \hat{V}_{\oplus}(C_l) + \frac{N(C_r)}{N(C)} \hat{V}_{\oplus}(C_r). \tag{9}$$

Note that minimizing this criterion corresponds to finding a split that maximizes the decrease in the weighted average of Fréchet variance accross the splitted subcells. A splitting pair $(j^\star, z^\star)$ can be found by iterating over the feature index $j \in \{1, \ldots, d\}$ and possible threshold values $z \in \{X_i^{(j)} : X_i \in C\}$ determined from values of coordinate j taken by the samples present in the cell C.

As highlighted by Athey et al. (2019) and Capitaine et al. (2020), this greedy search algorithm may be computationally prohibitively expensive since it requires for each split and candidate threshold to compute Fréchet means $\hat{m}_l$ and $\hat{m}_r$, which overall requires $O(dn^2)$ computations of Fréchet means. In many cases, computing a Fréchet mean can only be done approximately through computationally expensive algorithms. This is the case for computation of Fréchet means in Wasserstein spaces (Panaretos and Zemel, 2020), in function spaces equipped with the amplitude and phase variation distances (Srivastava and Klassen, 2016) or in Riemannian manifolds.

To overcome this computational limitation, Athey et al. (2019) introduce the *gradient tree algorithm*. This algorithm utilizes an alternative splitting criterion based on a linearization of the target optimization problem in regression and enables faster calculation of the relevant quantities in each candidate subcell. However, this approach cannot be applied to general metric spaces where no vector space structure is available for differentiability. In Capitaine et al. (2020), an alternative splitting criterion is proposed which offers computational advantages over the CART splitting criterion and is also usable in metric spaces. This approach, which is also used in Qiu et al. (2022), reduces the number of Fréchet mean computations by testing a single

possible partitioning for each coordinate, where the partitioning is found by performing a 2-means clustering of the observations in the cell based on the tested coordinate. This allows to compute only $2d$ Fréchet means per split. However, this still results in $O(dn)$ computations of Fréchet means which renders the algorithm unusable in situations where the Fréchet mean is expensive to compute. Notice that trees are fitted to subsamples of $\mathcal{D}$ of size s, but the abovementioned computational complexities are still valid with n replaced by the size of the subsamples.

We propose a solution that is usable in any metric space and completely avoids the need to compute Fréchet means during the fitting of the tree. To do that, we replace the minimization of the Fréchet function over the entire metric space by a minimization over the available sample. Denoting by Ω_n the set of observed responses, $\Omega_n = \{Y_1, \dots, Y_n\}$, we define an approximation to the empirical Fréchet mean $\hat{\omega}_\oplus(C)$, the *Fréchet medoid estimator* $\tilde{\omega}_\oplus(C) \in \Omega_n$, by

$$\tilde{\omega}_\oplus(C) = \operatorname*{argmin}_{\omega \in \Omega_n} \sum_{i=1}^{n} d(\omega, Y_i)^2 \mathbb{1}\{X_i \in C\}.$$

Computation of distances between all pairs of elements in the training data is requires, but once these distances have been computed, the Fréchet medoid is straightforward to find. Using the Fréchet medoid hence provides an efficient alternative to the empirical Fréchet mean in the computation of the CART splitting criterion (9).

The algorithm we propose to fit the random forest is based on replacing the computation of the empirical Fréchet mean by the Fréchet medoid in the split-finding procedure. As a first step, we pre-compute the pairwise distances matrix $\Delta \in \mathbb{R}^{n \times n}$ with $\Delta_{ij} = d(Y_i, Y_j)$. Then, at each step, we find the split (j, z) minimizing the approximate splitting criterion

$$\tilde{\text{err}}(j, z) = \frac{1}{N_n(C)} \sum_{i=1}^{n} \Delta^2_{i,i^\star(C_l)} \mathbb{1}\{X_i \in C_l\} + \Delta^2_{i,i^\star(C_r)} \mathbb{1}\{X_i \in C_r\},$$

where $i^\star$ is the index of the Fréchet medoid, $\tilde{\omega}_\oplus(C) = Y_{i^\star(C)}$.

Provided that the discrete set Ω_n grows dense in Ω and further technical conditions, the Fréchet medoid $\tilde{\omega}_\oplus(C)$ consistently estimates the true Fréchet mean (Proposition 1), and furthermore the associated splitting criterion $\tilde{\text{err}}(j, z)$ inherits this consistency. Notice that Assumption 1 is stated in Section 4, and that a proof for the proposition is given in the appendix.

Proposition 1. Let $Y_1, \ldots, Y_n$ be independent copies of a prototypical random variable Y taking values in a compact metric space (Ω, d) with Fréchet mean $\omega_\oplus$. Assume that the Fréchet mean is well-separated as defined in Assumption 1, and that the Fréchet mean is a *possible value* for Y, in the sense that for every $\varepsilon > 0$,

$$\mathbb{P}\left[d(Y, \omega_\oplus) < \varepsilon\right] > 0.$$

Then, the Fréchet medoid estimator is a consistent estimator of the Fréchet mean, that is, $d(\tilde{\omega}_\oplus, \omega_\oplus) = o_P(1)$.

Together with the Lipschitz continuity of the squared distance function (Lemma 1 in the appendix), one can see that for some constant $K > 0$ independent of n,

$$|\tilde{\text{err}}(j, z) - \text{err}(j, z)| \leq K\{d(\hat{\omega}_\oplus(C_l), \tilde{\omega}_\oplus(C_l)) + d(\hat{\omega}_\oplus(C_r), \tilde{\omega}_\oplus(C_r))\}.$$

Since $\tilde{\omega}_\oplus$ and $\hat{\omega}_\oplus$ are both consistent estimators of $\omega_\oplus$, this implies that the proposed splitting criterion converges to the CART splitting criterion.

4. Asymptotic Theory

We now consider the pointwise consistency of the MRF. We show that for a fixed vector of covariates $x \in [0, 1]^d$, the MRF (8) converges in probability to the output of the regression function (2). Consistency of the random forest estimators have been established in various settings: for simplified forest models Breiman (2004); Biau et al. (2008); Scornet (2016), forest with real number responses (Wager and Athey, 2018; Scornet, 2016) and also in the more general Z-estimator setting (Athey et al., 2019). However, due to the reliance on the outcome taking values in an Euclidean space, both in the assumptions and in the proofs of their theoretical results, existing asymptotic analysis, in particular that of the GRF, cannot be directly applied to the MRF setting.

Instead, we develop a proof combining results from standard M-estimation theory and existing results of pointwise consistency of the classical Euclidean random forest algorithm. Based on standard results in M-estimation theory (see van der Vaart (1998)), the first step to establishing pointwise consistency of the MRF is showing the convergence of the approximate objective function M_n given in (7) to the conditional Fréchet function M from (2) uniformly in $\omega \in \Omega$. We show this based on results from Newey (1991) and the pointwise convergence of the MRF objective function based on the proof presented in

Wager and Athey (2018). Their proof combines the Lipschitz continuity of the objective function with an assumption on the construction of the trees which implies that the diameter of the leaves of the tree shrinks as more datapoints become available (Specification 1 from Athey et al. (2019), see below).

Similarly to Wager and Athey (2018), we study in our analysis a theoretical construction of the random forest in which the bagging procedure is replaced by fitting trees on each possible subsets D of size s of the dataset $\mathcal{D}$. The effect of using an approximation of (11) based on a finite number of subsamples of $\mathcal{D}$ is ignored here, but a detailed investigation by Mentch and Hooker (2016) and Wager et al. (2014) suggest using $B = O(n)$.

The set of weights $\{w_i(\cdot; \xi, D)\}_{i=1}^n$ given in (5) can be used to define the contribution of a single tree to the forest objective function

$$T(\omega, x; \xi, D) = \sum_{i=1}^{n} w_i(x; \xi, D) d(\omega, Y_i)^2. \tag{10}$$

The tree-level weights are then combined to form the theoretical random forest weights

$$w_i^\star(x) = \binom{n}{s}^{-1} \sum_{\substack{D \subset \mathcal{D} \\ |D|=s}} \mathbb{E}_\xi \left[w_i(x; \xi, D) \right]. \tag{11}$$

Based on these weights, we can define the theoretical objective function and regression function studied in this section

$$M_n^\star(\omega; x) = \sum_{i=1}^{n} w_i^\star(x) d(\omega, Y_i)^2 \qquad \text{and} \qquad m_n^\star(x) = \operatorname*{argmin}_{\omega \in \Omega} M_n^\star(\omega; x). \tag{12}$$

We further consider that the forest is implemented in a way that satisfies Specification 1 of Athey et al. (2019): All trees in the forest are symmetric, meaning that the order of the input is not relevant to the fitting of each tree. The splits are balanced in that each split separates the observations on two subsets, each with a proportion of at least $\alpha > 0$ of the parent. Additionally, each tree is grown to depth k, for some $k \in \mathbb{N}$ and each leaf contains between k and $2k - 1$ observations. The forest is honest as described in Wager and Athey (2018), and each tree is built on a subsample of the data of size $s < n$ satisfying $s/n \to 0$ and $s \to \infty$.

The following assumption is common in M-estimation, see for instance Assumption (P0) in Müller (2016), and is used in most proofs of consistency of Fréchet regression methods.

Assumption 1 (Existence of a population solution)**.** For every $x \in [0,1]^d$, the conditional Fréchet regression function $m(x)$ given in (2) exists and is a *well-separated* solution to the conditional Fréchet function $M(\cdot; x)$, that is for every $\varepsilon > 0$,

$$\inf_{\omega : d(\omega, m(x)) > \varepsilon} M(\omega; x) > M(m(x); x).$$

We further require the conditional Fréchet function to be Lipschitz continuous in x. This assumption is also common in Fréchet regression, and is used here to exploit the shrinkage of the tree leaves to show the convergence of the approximate Fréchet function defined by the random forest.

Assumption 2 (x-Lipschitz Fréchet function)**.** For any fixed $\omega \in \Omega$, the function $x \mapsto M(\omega; x)$ is Lipschitz continuous with Lipschitz contant $C(\omega)$ possibly depending on ω.

As indicated, we first state a theorem saying that the objective function $M_n^\star(\cdot; x)$ defined by the random forest algorithm converges uniformly to the true Fréchet conditional Fréchet function M. This is a necessary element in consistency proofs of M-estimators and thus used to prove consistency of the random forest algorithm (Theorem 2). Proofs are given in the appendix.

Theorem 1. Under Assumption 2, the random forest objective function $M_n^\star(\cdot; x)$ converges uniformly in ω to $M(\cdot; x)$, that is, for all $x \in [0,1]^d$

$$\sup_{\omega \in \Omega} |M_n^\star(\omega; x) - M(\omega; x)| = o_P(1).$$

Theorem 2. Under Assumptions 1 and 2, the random forest regressor defined in (12) is a pointwise consistent estimator of the conditional Fréchet regression function, that is, for each $x \in [0,1]^d$,

$$d(m_n^\star(x), m(x)) = o_P(1).$$

5. Numerical experiments

In the following, we perform a simulation study in order to evaluate the benefit of our medoid approach (MRF) as compared to the random forest

weighted local constant Fréchet regression (RFWLCFR) algorithm presented in Qiu et al. (2022). The main difference between the two algorithms is the choice of splitting criterion, cf. Section 3, and we will compare them in terms of computational efficiency and prediction quality.

We study three scenarios with similar data generation processes in three different metric spaces with different computational costs associated with the computation of empirical Fréchet means. The first example is that of one-dimensional density functions over the real line $\mathbb{R}$ in the Wasserstein space. This space is isomorphic to a convex subspace of L_2 in which computations of distances and Fréchet means are straightforward. In a second example, we study data lying on a Riemannian manifold, the sphere $S^2 \subset \mathbb{R}^3$. In this example distances are straightforward to compute but Fréchet means can only be computed through a more expensive gradient descent algorithm. In a last example, we study functional data presenting phase variability through the study of their warping functions as described in Srivastava and Klassen (2016), in which both the evaluation of distances and of the Fréchet mean are computationally demanding.

In each of the three metric spaces, we generate datasets of different training sample sizes $N \in \{100, 200, 400\}$ and covariate dimension $d \in \{2, 5, 10, 20\}$. For each combination of a metric space, a training sample size and a covariate dimension, we generate 100 random datasets with the desired properties. In order to generate a dataset with covariate dimension d, two parameters $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^d$ are sampled with independent standard normal components. The Fréchet regression function is then generated from a single index regression model

$$m(x) = g(\eta) \qquad \text{with} \qquad \eta = \alpha + (x - 0.5)^\top \frac{\beta}{\sqrt{d}} \tag{13}$$

for some function $g : \mathbb{R} \to \Omega$, which is fixed and specific to each metric space. The scaling $1/\sqrt{d}$ has been chosen to ensure that the distribution of η does not change too much as the number of parameters increases, thus keeping the different simulation settings comparable. Each observation (X_i, Y_i) is finally constructed by first sampling $X_i \sim \text{Unif}[0, 1]^d$, computing the conditional expectation $m(X_i)$, and applying an independently sampled noise function $T_i : \Omega \to \Omega$, $Y_i = T_i(m(Y_i))$. Note that m varies is fixed within each dataset but varies between datasets.

In a given experiment, we fit the MFR and RFWLCFR algorithms to the same training dataset and evaluate them on an independently generated

test dataset of size $N_{\text{test}} = 100$ with the same parameters α, β as in the training set. We measure the training time of each algorithm and evaluate the performance of the algorithm using the mean squared error (MSE). For a test set $\{(X_i, Y_i)\}_{i=1}^{N_{\text{test}}}$ and a fitted regressor $\hat{m}$, the MSE of the algorithm is given by

$$\text{MSE}(\hat{m}) = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} d(\hat{m}(X_i), m(X_i))^2.$$

The focus of our comparison is the runtime of the algorithms, while mean squared errors are used to assess whether the approximation in MRF has a negative impact on the quality of the predictions. The simulations show the MFR provides significant performance gains over RFWLCFR, yet without compromising the quality of the predictions, and therefore enables us to analyze larger datasets in terms of sample size and dimension.

All simulations and analysis are done in Python. The implementation of both methods, together with the code to reproduce the results is available at https://github.com/matthieubulte/MetricRandomForest.

5.1. Univariate distributions

We consider the 2-Wasserstein metric space $\Omega = W_2 = W_2(\mathbb{R})$, the space of probability distributions over the real line with finite second moment endowed with the 2-Wasserstein metric d_{W_2}. Specifically, for $\mathbb{P}, \mathbb{Q} \in W_2$ with cumulative distribution functions $F_{\mathbb{P}}$ and $F_{\mathbb{Q}}$, the 2-Wasserstein distance between P and Q is given by

$$d_{W_2}(\mathbb{P}, \mathbb{Q}) = d_{L_2[0,1]}(F_{\mathbb{P}}^{-1}, F_{\mathbb{Q}}^{-1}) = \sqrt{\int_0^1 \left|F_{\mathbb{P}}^{-1}(u) - F_{\mathbb{Q}}^{-1}(u)\right|^2 \mathrm{d}u}.$$

As shown in existing literature on Wasserstein regression (Petersen and Müller, 2019; Ghodrati and Panaretos, 2022), the solution to the weighted Fréchet problem given via the sample $\{\mathbb{P}_i\}_{i=1}^n \subset W_2$ and weights $\{w_i\}_{i=1}^n$ can be found in a two-steps procedure: a pointwise weighted average of the sample inverse cumulative distribution functions is first computed and then projected onto the space of non-decreasing functions. This results in the following optimization problem

$$\underset{\mathbb{Q} \in W_2}{\text{argmin}} \sum_{i=1}^n w_i d_{W_2}(\mathbb{Q}, \mathbb{P}_i)^2 = \underset{\mathbb{Q} \in W_2}{\text{argmin}} \left\| F_{\mathbb{Q}}^{-1}(\cdot) - \sum_{i=1}^n w_i F_{\mathbb{P}_i}^{-1}(\cdot) \right\|_{L_2[0,1]}^2$$

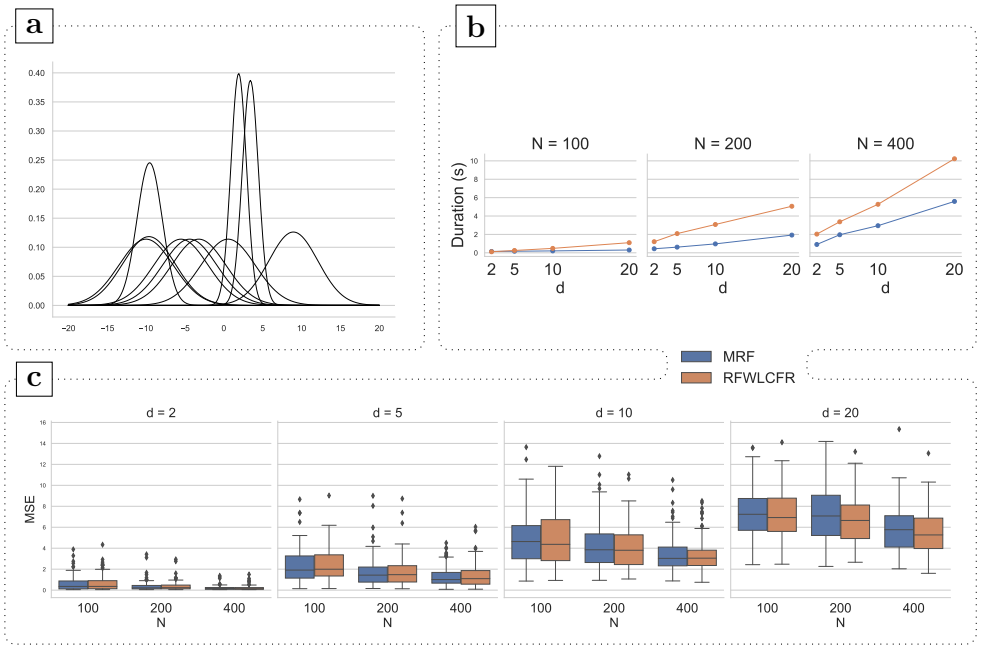


Figure 1: (a) shows the mean function $m(X)$ for 10 sampled values of X. (b) and (c) compare the runtime and MSE, respectively, of the MRF and RFWLCFR algorithms in Wasserstein regression for $N \in \{100, 200, 400\}$ and $d \in \{2, 5, 10, 20\}$.

In practice, each distribution is represented by a vector $y_i = (y_{i1}, \ldots, y_{iM})$ of evaluations of its inverse cumulative distribution function on an equispaced grid $\{u_m\}_{m=1}^M$ on $[0, 1]$. The weighted Fréchet mean $\bar{y}$ is then given by

$$F_{\bar{y}}^{-1} = \underset{\substack{v \in \mathbb{R}^M \\ v_1 \leq \cdots \leq v_M}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \sum_{m=1}^{M} |v_m - y_{im}|^2 .$$

In our application, we chose $M = 100$. As highlighted in Ghodrati and Panaretos (2022), this is an isotonic regression problem which can be efficiently solved using scikit-learn's implementation based on Pool Adjacent Violators Algorithm (PAVA), see Pedregosa et al. (2011); Best and Chakravarti (1990).

Following the example in Petersen and Müller (2019) for data generation, the conditional mean in this experiment maps a covariate vector to a normal

distribution with variable mean and variance parameters,

$$F^{-1}_{m(x)}(u) = \mu(\eta) + \sigma(\eta)\Phi^{-1}(u).$$

Specifically, we use the following functions for the mean and variance components,

$$\mu(\eta) = \eta \qquad \text{and} \qquad \sigma(\eta) = \sigma_0 + \gamma \text{logit}^{-1}(\eta),$$

with $\sigma_0 = 1$ and $\gamma = 2.5$. The random responses are generated by adding random noise to the normal quantiles as described in Panaretos and Zemel (2016). Given a continuous and non-decreasing random noise function $\varepsilon : \mathbb{R} \to \mathbb{R}$, the noise map $T : \Omega \to \Omega$ is defined via the composition of ε with the quantile function of $\mathbb{P}$, giving $F^{-1}_{T(\mathbb{P})} = \varepsilon \circ F^{-1}_{\mathbb{P}}$. The noise functions are given by $\varepsilon(x) = x - \sin(\pi k x) / |\pi k|$, where k is an integer frequency uniformly sampled from $\{-4, \ldots, 4\} \backslash \{0\}$. The observation Y given X are thus given by $F^{-1}_{Y} = \varepsilon \circ F^{-1}_{m(X)}$. Note that adding randomness via composition with a non-linear function results in random distributions outside the class of normal distributions.

Figure 1 compares the time performance and MSE of the methods. Due to the isomorphism of W_2 to the subspace of $L_2[0, 1]$ of quantile functions, the computation of Fréchet means is rather straightforward. However, the 2-means procedure adds an overhead to the RFWLCFR methodology which makes the overall computation slower than our MRF. The methods have similar MSE distributions in all setups showing that the medoid approximation in the MRF does not affect the quality of the predictions.

5.2. Spherical data

We consider the example where $\Omega = \mathbb{S}^q \subset \mathbb{R}^{q+1}$ is the q-dimensional sphere equipped with the geodesic distance

$$d_{\mathbb{S}^q}(x, y) = \arccos\langle x, y \rangle.$$

In this case, computing a weighted Fréchet mean is less straightforward and one must resort to optimization method on manifolds. We do this using the trust region algorithm implemented in Pymanopt (Townsend et al., 2016).

We focus on data lying on the sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ generated as follows. First, the mean function is given by the single index model described earlier, where we first transform η to $\nu \in (0, 1)$ via $\nu = \text{logit}^{-1}(\eta)$, and define

$$g(\eta) = \left(\sqrt{1 - \nu^2} \cos(\pi\nu), \sqrt{1 - \nu^2} \sin(\pi\nu), \nu \right). \tag{14}$$

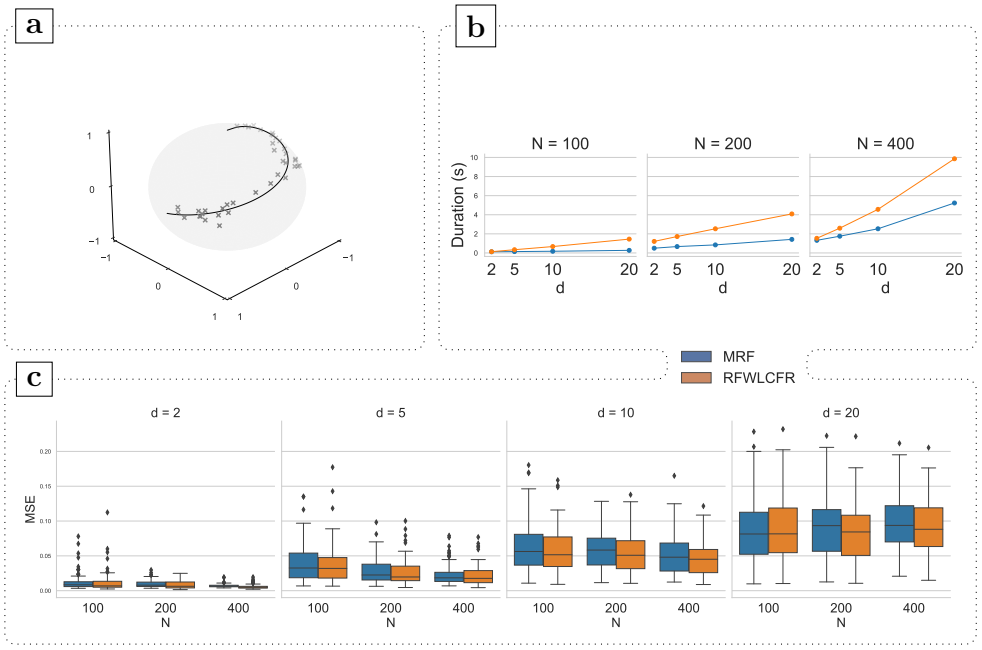


Figure 2: (a) shows the mean curve on the sphere (solid line) as a function of ν in (14) together with random observations (crosses). (b) and (c) compare the runtime and MSE, respectively, of the MRF and RFWLCFR algorithms for regression on the sphere $\mathbb{S}^2$ for $N \in \{100, 200, 400\}$ and $d \in \{2, 5, 10, 20\}$.

Given a random $X \sim \text{Unif}[0,1]^d$ and its associated mean function $m(X)$, the response $Y \in \mathbb{S}^2$ is generated by transforming a bivariate random vector U from the tangent space $T_{m(X)}\mathbb{S}^2$ at $m(X)$ through the corresponding exponential map,

$$Y = \text{Exp}_{m(X)}(U) = \cos(\|U\|)m(X) + \sin(\|U\|)\frac{U}{\|U\|}.$$

The random noise U is sampled from a bivariate normal distribution with independent components and variance $\sigma^2 = 0.1$, that is $U \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_{2\times 2})$.

The results of the experiment are displayed in Figure 2. The time performance profile is similar to the one in the Wasserstein experiment with MRF still providing a clear advantage over the RFWLCFR in terms of runtime. In this example the higher cost of computing Fréchet means does not seem to translate into further runtime improvements of MFR over RFWLCFR. The two methods again yield estimators with similar error distribution.

5.3. Space of warping functions

We now consider the study of functional data displaying phase variability, with observations lying in a non-linear subset of $L_2[0,1]$. As described in Tucker et al. (2013), and more generally Srivastava and Klassen (2016), the variability of functional data can be decomposed in phase and amplitude components. Intuitively, phase variation corresponds to variation on the x axis and amplitude variation corresponds to variation along the y axis.

Let Γ be the set of boundary preserving diffeomorphisms on $[0,1]$, that is, $\Gamma = \{\gamma : [0,1] \to [0,1] \mid \gamma(0) = 0, \gamma(1) = 1, \gamma \text{ is a diffeomorphism}\}$. For a given function $y_0 \in L^2[0,1]$, we define the orbit of y_0 as the set of functions that only differ from y_0 in their phase component, $[y_0] = \{y_0 \circ \gamma \mid \gamma \in \Gamma\}$. While y_0 is not unique in this notation, whenever the starting point y_0 of an orbit $[y_0]$ is fixed in a context, we call it *template function*.

We consider a distribution for Y in which the observations are all part of the same orbit $\Omega = [y_0]$. Such spaces of equivalence classes, which are all equivalent to the space of warping functions Γ, form a non-linear infinite-dimensional manifold which is difficult to analyze and work with when considering the L_2 distance.

Instead, a warping function can be represented by its square-root velocity function $\psi_\gamma = \sqrt{\dot{\gamma}}$. Since the warpings γ are boundary preserving diffeomorphisms on $[0,1]$ we have that $\dot{\gamma}$ exists and is positive, making ψ_γ a well defined quantity. Furthermore $\gamma(0) = 0$ and $\gamma(1) = 1$, hence $\|\psi\|_{L_2}^2 = \int_0^1 \psi(t)^2 \mathrm{d}t = \int_0^1 \dot{\gamma}(t) \mathrm{d}t = 1$ making ψ an element of the $L_2[0,1]$ sphere, $\mathbb{S}^\infty = \left\{ f \in L_2[0,1] \mid \|f\|_{L_2[0,1]} = 1 \right\}$, and more specifically on the positive hemisphere of $\mathbb{S}^\infty$ since $\psi > 0$. This allows to define a distance on Γ via the intrinsic distance on $\mathbb{S}^\infty$. Let γ_1, γ_2 be two warpings with square-root velocity functions ψ_1, ψ_2, then the distance between γ_1 and γ_2 is given by

$$d_\Gamma(\gamma_1, \gamma_2) = d_{\mathbb{S}^\infty}(\psi_1, \psi_2) = \arccos\langle \psi_1, \psi_2 \rangle_{L_2}.$$

See Chapter 4 of Srivastava and Klassen (2016) for more details on this metric space. Let now $y_1, y_2 \in [y_0]$ be two functions within a same orbit, then these functions can be represented by their warpings from the template y_0, denoted γ_1, γ_2, with $y_i = y_0 \circ \gamma_i$ and we can define

$$d_{[y_0]}(y_1, y_2) = d_\Gamma(\gamma_1, \gamma_2).$$

The computation of quantities required to evaluate distances and the Fréchet

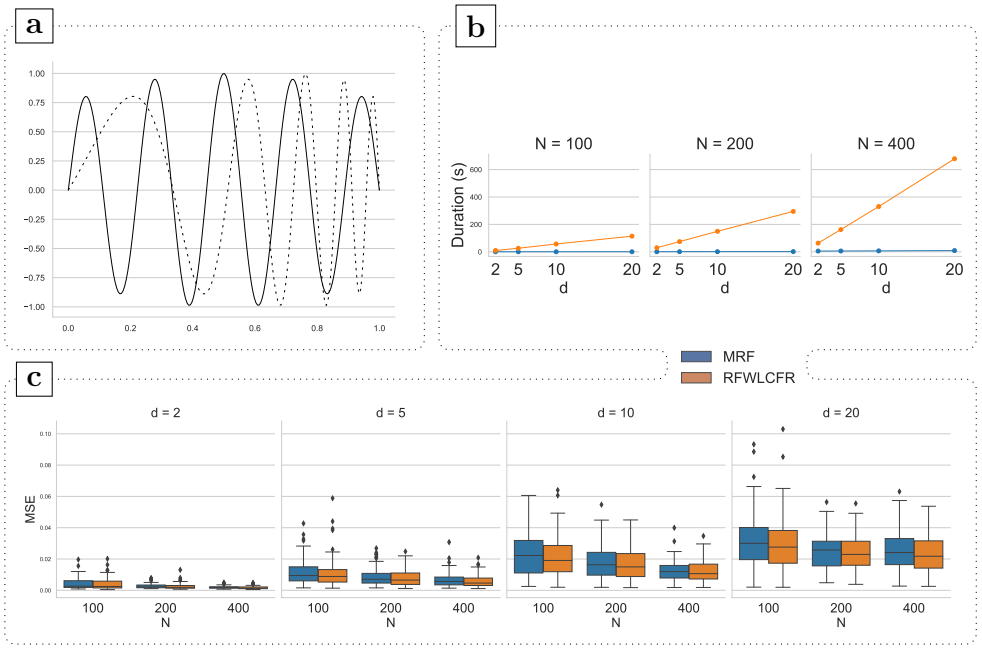


Figure 3: (a) shows the template curve y_0 (solid) and a randomly sampled curve (dashed). (b) and (c) compare the runtime and MSE, respectively, of the MRF and RFWLCFR algorithms for regression in Γ for $N \in \{100, 200, 400\}$ and $d \in \{2, 5, 10, 20\}$.

mean are based on implementations provided in scikit-fda (Ramos-Carreño et al., 2019) and fdasrsf (Tucker et al., 2013).

Next, we turn to the data generating process for our experiment. The template function is chosen as $y_0(u) = (1 - (u - 0.5)^2)\sin(9\pi u)$, see the solid curve in panel (a) of Figure 3. For a given random vector of covariates X, one obtains the linear component η as defined in (13) and the mean warping response is γ is given for all $u \in [0, 1]$ by

$$\gamma(u) = (\exp(4au) - 1)/(\exp(4a) - 1),$$

where $a = 3(\text{logit}^{-1}(\eta) - 0.5)$. The warping is perturbed by adding noise to its square-root velocity function $\psi \in \mathbb{S}^\infty$ by sampling a random Gaussian process V with exponential covariance kernel from the tangent space $T_\psi \mathbb{S}^\infty$, and applying the exponential map at ψ,

$$\text{Exp}_\psi V = \cos(\|V\|)\psi + \sin(\|V\|)\frac{V}{\|V\|}.$$

The response Y is then obtained by warping the template y_0 with the warping function obtained by taking the inverse square-root velocity transform of $\mathrm{Exp}_\psi V$. An example of a sample is displayed in panel (a) of Figure 3 (dashed curve).

The results of the experiment are displayed in Figure 3. Computing a single empirical Fréchet mean can take up to several seconds and computing Fréchet variance is thus a very expensive operation. The benefit of the 2-means approach implemented in RFWLCFR is dominated by the high cost of computing Fréchet variances. On the other hand, the runtime of MFR remains low in all scenarios. Runtime ratios are 1 to 50 in the $N = 100$ and $d = 20$ scenario and 1 to 30 for $N = 200$ and $d = 20$. Similarly to the previous two experiments, the runtime gain does not impact the fitting of the forest and the MSE distributions stay similar for the two methods.

6. Discussion

Defining non-parametric regression method for general metric spaces that are performant and work well in complex setups is a challenge. We have proposed a new version of the random forest Fréchet regression that allows practitioners to perform a random forest regression in high-dimensional problems and in metric spaces where computation of Fréchet means is expensive. We achieved this by replacing the classic computation of the Fréchet means in random forest splits with a computation based on the Fréchet medoid. We proved both the consistency of the random forest Fréchet regression as well as the consistency of our medoid-based approximation. Furthermore, a set of numerical experiments in metric spaces of different characteristics demonstrated important reduction in the time required to fit a random forest without showing any impact on the mean squared error of the method.

One possible extension of this work could be to study rates of convergence, at best based on assumptions on the data generating process only. This, combined with a further investigation of the medoid approximation, could refine the understanding of the asymptotic behaviour of the metric random forest. Our work also paves the way to exploring further approximations to random forest regression in metric spaces. This could be through the use of algorithmic improvements generally applicable to random forests, such as the improvement proposed in Tiwari et al. (2022), or new approximations specific to the use of metric spaces. One could for instance study the use of approximate distance computations or other approximations to the Fréchet

variance, and incorporate such approximations in the theoretical analysis as well as in the implementation.

7. Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 956107, "Economic Policy in Complex Environments (EPOC)".

Appendix A. Proofs

Appendix A.1. Statement of useful theorems

We start by stating two theorems that will be used in the proof of consistency. First, Corollary 2.2 of Newey (1991) provides the conditions under which pointwise convergence of a sequence of random functions can be translated to uniform convergence.

Theorem 3. Let (Ω, d) be a compact metric space, $M_n : \Omega \to \mathbb{R}$ be a sequence of random functions and $M : \Omega \to \mathbb{R}$ be a fixed function such that

1. (Continuity): M is continuous.
2. (Stochastic equicontinuity): There exist some $B_n = O_P(1)$ such that for all $\omega, \omega' \in \Omega, |M_n(\omega) - M_n(\omega')| \leq B_n d(\omega, \omega')$.
3. (Pointwise convergence): For every $\omega \in \Omega$, $M_n(\omega) - M(\omega) = o_P(1)$.

Then

$$\sup_{\omega \in \Omega} |M_n(\omega) - M(\omega)| = o_P(1).$$

Further, Theorem 5.7 of van der Vaart (1998) can be used to translate uniform convergence of a sequence of objective functions to convergence of the minimizers.

Theorem 4. Let (Ω, d) be a metric space, $M_n : \Omega \to \mathbb{R}$ be a sequence of random functions and $M : \Omega \to \mathbb{R}$ be a fixed function such that for every $\varepsilon > 0$,

$$\sup_{\omega \in \Omega} |M_n(\omega) - M(\omega)| = o_P(1)$$
$$\inf_{\omega : d(\omega_0, \omega) > \varepsilon} M(\omega) > M(\omega_0).$$

Then any sequence of estimators $\hat{\omega}_n$ with $M_n(\hat{\omega}_n) \leq M_n(\omega_0) + o_P(1)$ converges in probability to ω_0.

The first condition of this theorem states that M_n converges uniformly to M while the second condition states that there exists a well-separated minimizer ω_0 of M.

Appendix A.2. Consistency of the MRF estimator

We start with a lemma regarding Lipschitz continuity of the distance and squared distance functions.

Lemma 1. Let (Ω, d) be a compact metric space and $\omega_0 \in \Omega$, then the maps $\omega \mapsto d(\omega_0, \omega)$ and $\omega \mapsto d(\omega_0, \omega)^2$ are Lipshitz continuous.

Proof. Starting with the first map, let $\omega, \omega' \in \Omega$, then we have by the triangle inequality that

$$\begin{aligned} d(\omega_0, \omega) \leq d(\omega_0, \omega') + d(\omega', \omega) &\Rightarrow d(\omega_0, \omega) - d(\omega_0, \omega') \leq d(\omega', \omega) \\ d(\omega_0, \omega') \leq d(\omega_0, \omega) + d(\omega', \omega) &\Rightarrow -\left(d(\omega_0, \omega) - d(\omega_0, \omega')\right) \leq d(\omega', \omega), \end{aligned}$$

and hence

$$|d(\omega_0, \omega) - d(\omega_0, \omega')| \leq d(\omega', \omega).$$

For the second map, we have

$$\begin{aligned} \left|d(\omega_0, \omega)^2 - d(\omega_0, \omega')^2\right| &= |d(\omega_0, \omega) + d(\omega_0, \omega')|\,|d(\omega_0, \omega) - d(\omega_0, \omega')| \\ &\leq 2\mathrm{diam}\,\Omega\, d(\omega, \omega'), \end{aligned}$$

where the first term was bounded by the diameter of Ω and we used the Lipshitz continuity of $\omega \mapsto d(\omega_0, \omega)$ for the second term. □

Next, we prove that the random forest objective function converges uniformly in Ω to the conditional variance. Recall definitions (12) of $M_n^\star(\omega; x)$ and (2) of $M(\omega; x)$.

Proof of Theorem 1. This proof is done by verifying that for a fixed $x \in [0,1]^d$, the maps $\omega \mapsto M_n^\star(\omega; x)$ and $\omega \mapsto M(\omega; x)$ respect the conditions of Theorem 3.

Continuity. This follows directly from Lemma 1 since for $\omega, \omega' \in \Omega$,

$$\begin{aligned}
|M(\omega; x) - M(\omega'; x)| &= \left|\mathbb{E}\left[d(\omega, Y)^2 \mid X = x\right] - \mathbb{E}\left[d(\omega', Y)^2 \mid X = x\right]\right| \\
&\leq \mathbb{E}\left[\left|d(\omega, Y)^2 - d(\omega', Y)^2\right| \mid X = x\right] \\
&\leq 2\text{diam}\,\Omega\, d(\omega, \omega')
\end{aligned}$$

Stochastic Equicontinuity. Similarly to the continuity of M, the stochastic equicontinuity of $M_n^\star$ is inherited. Let $\omega, \omega' \in \Omega$,

$$\begin{aligned}
|M_n^\star(\omega; x) - M_n^\star(\omega'; x)| &= \left|\sum_{i=1}^n w_i^\star(x) d(\omega, Y_i)^2 - \sum_{i=1}^n w_i^\star(x) d(\omega', Y_i)^2\right| \\
&\leq \sum_{i=1}^n w_i^\star(x) \left|d(\omega, Y_i)^2 - d(\omega', Y_i)^2\right| \\
&\leq \sum_{i=1}^n w_i^\star(x) 2\text{diam}\,\Omega\, d(\omega, \omega') \\
&= 2\text{diam}\,\Omega\, d(\omega, \omega').
\end{aligned}$$

Pointwise convergence. Let $\omega \in \Omega$ be fixed. We need to show that $M_n^\star - M = o_P(1)$. Recall the contribution of a single fitted tree (10) to the theoretical random forest objective function $M_n^\star$,

$$T(\omega, x; \xi, D) = \sum_{i=1}^n w_i(x; \xi, D) d(\omega; Y_i)^2.$$

We simplify the notation and remove references to x, ω, ξ and D in the remaining of this sub-proof. We now define the Hájek projection $\mathring{M}_n^\star$ of $M_n^\star$,

$$\mathring{M}_n^\star = \mathbb{E}\left[M_n^\star\right] + \sum_{i=1}^n \mathbb{E}\left[M_n^\star \mid (X_i, Y_i)\right] - \mathbb{E}\left[M_n^\star\right].$$

This can be rewritten in terms of expectations of the individual trees as

$$\mathring{M}_n^\star = \mathbb{E}\left[T\right] + \frac{s}{n} \sum_{i=1}^n \mathbb{E}\left[T \mid (X_i, Y_i)\right] - \mathbb{E}\left[T\right].$$

We proceed with the proof of pointwise convergence by showing the stronger result of L_2 convergence of $M_n^\star$, that is $\mathbb{E}\left[\left(M_n^\star - M\right)^2\right] = o(1)$. We do this via

the convergence of the Hájek projection of the forest weights M_n, in a similar fashion to the analysis in Wager and Athey (2018). The mean squared error $\mathbb{E}\left[(M_n^\star - M)^2\right]$ can be upper bounded by the following sum in which each term is more amenable to analysis:

$$\mathbb{E}\left[\left(M_n^\star - \mathring{M}_n^\star\right)^2\right] + \mathbb{E}\left[\left(\mathring{M}_n^\star - \mathbb{E}\left[M_n^\star\right]\right)^2\right] + \left(\mathbb{E}\left[M_n^\star\right] - M\right)^2.$$

By the specification of the construction of the random forest, we have that Lemma 7 in Wager and Athey (2018) applies and we can bound the first summand as

$$\mathbb{E}\left[\left(M_n^\star - \mathring{M}_n^\star\right)^2\right] \leq \frac{s^2}{n^2}\mathrm{Var}\left[T\right].$$

As shown in the proof of Theorem 5 from Wager and Athey (2018), the variance of a single tree is asymptotically bounded

$$\mathrm{Var}\left[T\right] \lesssim k\mathrm{Var}\left[d(\omega, Y)^2 \mid X = x\right] < \infty,$$

where k is the depth at which the trees are grown, as specified in Section 4. Hence

$$\mathbb{E}\left[\left(M_n^\star - \mathring{M}_n^\star\right)^2\right] \lesssim \frac{s^2}{n^2} = o(1)$$

For the second summand, we note that $\mathbb{E}\left[M_n^\star\right] = \mathbb{E}\left[\mathring{M}_n^\star\right]$ and hence the second term is the variance of the Hájek projection. We can again use the fact that the variance of a single tree is asymptotically bounded to obtain

$$\mathrm{Var}\left[\mathring{M}_n^\star\right] = \frac{s^2}{n}\mathrm{Var}\left[\mathbb{E}\left[T \mid Z_1\right]\right] \leq \frac{s}{n}\mathrm{Var}\left[T\right] \lesssim \frac{s}{n} = o(1)$$

Consider at last the third summand. Since $M_n^\star$ is constructed as the mean of identically distributed trees, we have that $\mathbb{E}\left[M_n^\star\right] = \mathbb{E}\left[T\right]$. We thus analyze instead the quantity

$$\mathbb{E}\left[T\right] - \mathbb{E}\left[d(\omega, Y)^2 \mid X = x\right].$$

As done in the proof of Theorem 3 in Wager and Athey (2018), we have by honesty that

$$\begin{aligned}
&\mathbb{E}\left[T\right] - \mathbb{E}\left[d(\omega, Y)^2 \mid X = x\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[d(\omega, Y)^2 \mid X \in L(x)\right] - \mathbb{E}\left[d(\omega, Y)^2 \mid X = x\right]\right],
\end{aligned}$$

where $L(x) = L(x; \xi, D)$. Using Assumption 2, M is Lipschitz continuous with constant $C(\omega)$ and we get

$$\left|\mathbb{E}\left[d(\omega, Y)^2 \mid X \in L(x)\right] - \mathbb{E}\left[d(\omega, Y)^2 \mid X = x\right]\right| \leq C(\omega)\text{diam}\, L(x).$$

Using now that the trees are constructed following Specification 1 of Athey et al. (2019), Lemma 2 of Wager and Athey (2018) applies, implying that the size of the leaves are $o_P(1)$. Hence

$$\left|\mathbb{E}\left[T - \mathbb{E}\left[d(\omega, Y)^2 \mid X = x\right]\right]\right| \leq \mathbb{E}\left[C(\omega)L(x)\right] = \mathbb{E}\left[o_P(1)\right] = o(1)$$

Altogether, we thus have that $\mathbb{E}\left[(M_n^\star - M)^2\right] = o(1)$ and hence we have for any fixed (ω, x) the converge $M_n^\star(\omega; x) - M(\omega; x) = o_P(1)$. □

We can now prove that the random forest estimator is a pointwise constistent estimator of the Fréchet mean.

Proof of Theorem 2. Since Ω is compact and $m_n^\star(x)$ minimizes $M_n^\star(\cdot; x)$, we have by definition that $M_n^\star(m_n^\star(x); x) \leq M_n^\star(m(x); x)$. Furthermore, by Theorem 1, the random forest objective function $M_n^\star$ converges uniformly to M. Together with the assumption that the Fréchet mean is well-separated, we have by Theorem 4 that $m_n^\star(x)$ is a consistent estimator of $m(x)$. □

Appendix A.3. Consistency of the splitting rule

We finally prove consistency of our medoid-based splitting rule. The proof follows a similar structure as the proof of Theorem 2.

Proof of Proposition 1. Let $\Omega_n = \{Y_1, \ldots, Y_n\}$ and $\varepsilon > 0$. By assumption, there exists an $\alpha > 0$ such that $\mathbb{P}\left[d(Y, \omega_\oplus) < \varepsilon\right] > \alpha$. We then have by independence that

$$\mathbb{P}\left[d(\Omega_n, \omega_\oplus) > \varepsilon\right] = \mathbb{P}\left[d(Y, \omega_\oplus) > \varepsilon\right]^n < (1 - \alpha)^n \to 0,$$

showing that $d(\Omega_n, \omega_\oplus) \to 0$ in probability.

Let $M(\omega) = \mathbb{E}\left[d(\omega, Y)^2\right]$ be the unweighted population Fréchet function, minimized by $\omega_\oplus$, and $M_n(\omega) = \frac{1}{n}\sum_{i=1}^n d(\omega, Y_i)^2$ be the unweighted empirical Fréchet functions with a sequence of minimizers $\{\hat{\omega}_n\}$. We start by showing

that M is continuous, which follows directly from the Lipschitz continuity of the squared distance since for $\omega, \omega' \in \Omega$,

$$\begin{aligned}|M(\omega) - M(\omega')| &= \left|\mathbb{E}\left[d(\omega, Y)^2\right] - \mathbb{E}\left[d(\omega', Y)^2\right]\right| \\ &\leq \mathbb{E}\left[\left|d(\omega, Y)^2 - d(\omega', Y)^2\right|\right] \\ &\leq 2\text{diam}\,\Omega\, d(\omega, \omega').\end{aligned}$$

Furthermore, we can also use the Lipschitz continuity of the squared distance to show that M_n is stochastic equicontinuous since for each $\omega, \omega' \in \Omega$,

$$\begin{aligned}|M_n(\omega) - M_n(\omega')| &= \left|\sum_{i=1}^n d(\omega, Y_i)^2 - \sum_{i=1}^n d(\omega', Y_i)^2\right| \\ &\leq \sum_{i=1}^n \left|d(\omega, Y_i)^2 - d(\omega', Y_i)^2\right| \\ &\leq \sum_{i=1}^n 2\text{diam}\,\Omega\, d(\omega, \omega') \\ &= 2\text{diam}\,\Omega\, d(\omega, \omega').\end{aligned}$$

Finally, by the weak law of large numbers, M_n converges pointwise to M, which verifies the conditions of Theorem 3. This shows that $\|M_n - M\|_\Omega = o_P(1)$. Together with the assumption of well-separatedness of $\omega_\oplus$, this shows that M_n fulfills the conditions of Theorem 4, giving that $d(\hat{\omega}_n, \omega_\oplus) = o_P(1)$.

Let $\varepsilon > 0$ and $K = 2\text{diam}\,\Omega$, under the event $E_n = \{d(\Omega_n, \omega_\oplus) \leq \varepsilon/2K\} \cap \{d(\hat{\omega}_n, \omega_\oplus) \leq \varepsilon/2K\}$, there exists for each $n \in \mathbb{N}$ an $\omega_n^\dagger \in \Omega_n$ such that $d(\omega_n^\dagger, \omega_\oplus) \leq \varepsilon/2K$. Hence,

$$\begin{aligned}\min_{\omega\in\Omega_n} M_n(\omega) - \min_{\omega\in\Omega} M_n(\omega) = M_n(\tilde{\omega}_n) - M_n(\hat{\omega}_n) &\leq M_n(\omega_n^\dagger) - M_n(\hat{\omega}_n) \\ &\leq K d(\omega_n^\dagger, \hat{\omega}_n) \leq K(d(\omega_n^\dagger, \omega_\oplus) + d(\omega_\oplus, \hat{\omega}_n)) \leq \varepsilon.\end{aligned}$$

By the two first results in this proof, $\mathbb{P}[E_n] \to 1$ and we obtain that $M_n(\tilde{\omega}_n) - M_n(\hat{\omega}_n) = o_P(1)$, and therefore

$$M_n(\tilde{\omega}_n) = M_n(\hat{\omega}_n) + o_P(1) \leq M_n(\omega_\oplus) + o_P(1).$$

This shows that the sequence of minimizers $\{\tilde{\omega}_n\}$ fulfills the third condition of Theorem 4, hence $d(\tilde{\omega}_n, \omega_\oplus) = o_P(1)$. □

References

Athey, S., Tibshirani, J., Wager, S., 2019. Generalized random forests. The Annals of Statistics 47, 1148 – 1178. URL: https://doi.org/10.1214/18-AOS1709, doi:10.1214/18-AOS1709.

Best, M.J., Chakravarti, N., 1990. Active set algorithms for isotonic regression; A unifying framework. Mathematical Programming 47, 425–439. URL: https://doi.org/10.1007/BF01580873, doi:10.1007/BF01580873.

Biau, G., Devroye, L., Lugosi, G., 2008. On the Performance of Clustering in Hilbert Spaces. IEEE Transactions on Information Theory 54, 781–790. URL: http://ieeexplore.ieee.org/document/4439834/, doi:10.1109/TIT.2007.913516.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. URL: https://doi.org/10.1023/A:1010933404324, doi:10.1023/A:1010933404324.

Breiman, L., 2004. Consistency for a simple model of random forests. University of California at Berkeley. Technical Report 670.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. Taylor & Francis.

Capitaine, L., Bigot, J., Thiébaut, R., Genuer, R., 2020. Fréchet random forests for metric space valued regression with non euclidean predictors URL: http://arxiv.org/abs/1906.01741.

Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S., 2010. Population Shape Regression from Random Design Data. International Journal of Computer Vision 90, 255–266. URL: https://doi.org/10.1007/s11263-010-0367-1, doi:10.1007/s11263-010-0367-1.

Fréchet, M., 1948. Les éléments aléatoires de nature quelconque dans un espace distancié, in: Annales de l'institut Henri Poincaré, pp. 215–310. Volume 10, issue 4.

Fu, G., Dai, X., Liang, Y., 2021. Functional random forests for curve response. Scientific Reports 11, 24159. URL: https://www.nature.com/articles/s41598-021-02265-4, doi:10.1038/s41598-021-02265-4.

Ghodrati, L., Panaretos, V.M., 2022. Distribution-on-distribution regression via optimal transport maps. Biometrika 109, 957–974. URL: https://doi.org/10.1093/biomet/asac005, doi:10.1093/biomet/asac005.

Hein, M., 2009. Robust Nonparametric Regression with Metric-Space Valued Output, in: Advances in Neural Information Processing Systems, Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2009/hash/92977ae4d2ba21425a59afb269c2a14e-Abstract.html.

Hinkle, J., Muralidharan, P., Fletcher, P.T., Joshi, S., 2012. Polynomial Regression on Riemannian Manifolds, in: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (Eds.), Computer Vision – ECCV 2012, Springer, Berlin, Heidelberg. pp. 1–14. doi:10.1007/978-3-642-33712-3_1.

Meinshausen, N., 2006. Quantile Regression Forests. Journal of Machine Learning Research 7, 983–999. URL: http://jmlr.org/papers/v7/meinshausen06a.html.

Mentch, L., Hooker, G., 2016. Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. Journal of Machine Learning Research 17, 1–41. URL: http://jmlr.org/papers/v17/14-168.html.

Müller, H.G., 2016. Peter Hall, functional data analysis and random objects. The Annals of Statistics 44, 1867 – 1887. URL: https://doi.org/10.1214/16-AOS1492, doi:10.1214/16-AOS1492.

Nerini, D., Ghattas, B., 2007. Classifying densities using functional regression trees: Applications in oceanology. Computational Statistics & Data Analysis 51, 4984–4993. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167947306003550, doi:10.1016/j.csda.2006.09.028.

Newey, W.K., 1991. Uniform Convergence in Probability and Stochastic Equicontinuity. Econometrica 59, 1161–1167. URL: https://www.jstor.org/stable/2938179?origin=crossref, doi:10.2307/2938179.

Panaretos, V.M., Zemel, Y., 2016. Amplitude and phase variation of point processes. The Annals of Statistics 44, 771 – 812. URL: https://doi.org/10.1214/15-AOS1387, doi:10.1214/15-AOS1387.

Panaretos, V.M., Zemel, Y., 2020. An Invitation to Statistics in Wasserstein Space. SpringerBriefs in Probability and Mathematical Statistics, Springer International Publishing, Cham. URL: http://link.springer.com/10.1007/978-3-030-38438-8, doi:10.1007/978-3-030-38438-8.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830. URL: http://jmlr.org/papers/v12/pedregosa11a.html.

Pelletier, B., 2006. Non-parametric regression estimation on closed Riemannian manifolds. Journal of Nonparametric Statistics 18, 57–67. doi:10.1080/10485250500504828. publisher: Taylor & Francis.

Petersen, A., Müller, H.G., 2019. Fréchet regression for random objects with Euclidean predictors. The Annals of Statistics 47, 691–719. URL: https://doi.org/10.1214/17-AOS1624, doi:10.1214/17-AOS1624.

Qiu, R., Yu, Z., Zhu, R., 2022. Random Forests Weighted Local Fréchet Regression with Theoretical Guarantee URL: http://arxiv.org/abs/2202.04912.

Ramos-Carreño, C., Suárez, A., Torrecilla, J.L., Carbajo Berrocal, M., Marcos Manchón, P., Pérez Manso, P., Hernando Bernabé, A., García Fernández, D., Hong, Y., Rodríguez-Ponga Eyriès, P.M., Sánchez Romero, A., Petrunina, E., Castillo, A., Serna, D., Hidalgo, R., 2019. GAA-UAM/scikit-fda: Functional Data Analysis in Python. URL: https://doi.org/10.5281/zenodo.7573921, doi:10.5281/zenodo.7573921.

Scornet, E., 2016. On the asymptotics of random forests. Journal of Multivariate Analysis 146, 72–83. URL: https://linkinghub.elsevier.com/retrieve/pii/S0047259X15001542, doi:10.1016/j.jmva.2015.06.009.

Srivastava, A., Klassen, E.P., 2016. Functional and Shape Data Analysis. Springer Series in Statistics, Springer New York, New York, NY. URL: http://link.springer.com/10.1007/978-1-4939-4020-2, doi:10.1007/978-1-4939-4020-2.

Thomas Fletcher, P., 2013. Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds. International Journal of Computer Vision 105, 171–185. URL: http://link.springer.com/10.1007/s11263-012-0591-y, doi:10.1007/s11263-012-0591-y.

Tiwari, M., Kang, R., Lee, J., Piech, C., Shomorony, I., Thrun, S., Zhang, M.J., 2022. Mabsplit: Faster forest training using multi-armed bandits, in: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.. pp. 1223–1237. URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/08857467641ad82f635023d530605b4c-Paper-Conference.pdf.

Townsend, J., Koep, N., Weichwald, S., 2016. Pymanopt: A Python Toolbox for Optimization on Manifolds using Automatic Differentiation. Journal of Machine Learning Research 17, 1–5. URL: http://jmlr.org/papers/v17/16-177.html.

Tucker, J.D., Wu, W., Srivastava, A., 2013. Generative models for functional data using phase and amplitude separation. Computational Statistics & Data Analysis 61, 50–66. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167947312004227, doi:10.1016/j.csda.2012.12.001.

van der Vaart, A.W., 1998. Asymptotic statistics. Cambridge series in statistical and probabilistic mathematics, Cambridge University Press, Cambridge, UK ; New York, NY, USA. doi:10.1017/CBO9780511802256.

Wager, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association 113, 1228–1242. URL: https://doi.org/10.1080/01621459.2017.1319839, doi:10.1080/01621459.2017.1319839.

Wager, S., Hastie, T., Efron, B., 2014. Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. Journal of Machine Learning Research 15, 1625–1651. URL: http://jmlr.org/papers/v15/wager14a.html.

Yuan, Y., Zhu, H., Lin, W., Marron, J.S., 2012. Local polynomial regression for symmetric positive definite matrices. Journal of the

Royal Statistical Society: Series B (Statistical Methodology) 74, 697–719. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.01022.x, doi:10.1111/j.1467-9868.2011.01022.x.

2.2 Notes

As mentioned in the introduction of this chapter, the initial goal of this project was to adapt the modified random forest algorithm of [Wager and Athey, 2018] to the case where the target variable Y is a random element of the Wasserstein space $W_2(\mathbb{R})$. The hope being to adapt the central limit theorem result of Theorem 3.4 to a central limit theorem in the tangent space of $W_2(\mathbb{R})$. In the process, we identified a mistake in the proof of Theorem 3.4.

The proof of Theorem 3.4 in [Wager and Athey, 2018] is based on several components: a control of the size of the weights constructed in the procedure, a Hájek projection argument, and an application of Lyapunov's central limit theorem. The mistake lies in verifying the Lyapunov condition where the following inequality is claimed to hold:

$$\mathrm{Var}\left[\mathbb{E}\left[\sum_{i=1}^{n} S_i \mathbb{E}[Y_i \mid X_i] \mid Z_1\right]\right] \leq 2u^2\Big(\mathbb{E}\Big[\mathbb{E}[S_1 \mid Z_1]^2\Big] + \mathrm{Var}[(n-1)\mathbb{E}[S_2 \mid Z_1]]\Big). \quad (2.1)$$

To reach this inequality, the authors implicitly use the following derivation [1]:

$$\begin{aligned}
&\mathrm{Var}\left[\mathbb{E}\left[\sum_{i=1}^{n} S_i \mathbb{E}[Y_i \mid X_i] \mid Z_1\right]\right] \\
&\quad = \mathrm{Var}\left[\mathbb{E}[\mathbb{E}[Y_1 \mid X_1] \mid Z_1] + \mathbb{E}\left[\sum_{i=2}^{n} S_i \mathbb{E}[Y_i \mid X_i] \mid Z_1\right]\right] \\
&\quad \leq 2\mathrm{Var}[\mathbb{E}[\mathbb{E}[Y_1 \mid X_1] \mid Z_1]] + 2\mathrm{Var}\left[\mathbb{E}\left[\sum_{i=2}^{n} S_i \mathbb{E}[Y_i \mid X_i] \mid Z_1\right]\right] \\
&\quad = 2\mathrm{Var}[\mathbb{E}[\mathbb{E}[Y_1 \mid X_1] \mid Z_1]] + 2\mathrm{Var}[(n-1)\mathbb{E}[S_2 \mathbb{E}[Y_2 \mid X_2] \mid Z_1]] \\
&\quad \leq 2u^2 \mathbb{E}\Big[\mathbb{E}[S_1 \mid Z_1]^2\Big] + 2\mathrm{Var}[(n-1)\mathbb{E}[S_2 \mathbb{E}[Y_2 \mid X_2] \mid Z_1]],
\end{aligned}$$

where the last inequality follows from Lipschitz assumptions made in the statement of the theorem. However, to obtain the inequality in Equation (2.1), the authors attempt to use that $|\mathbb{E}[Y \mid X]| \leq u$ to bound the variance term

$$\mathrm{Var}[(n-1)\mathbb{E}[S_2 \mathbb{E}[Y_2 \mid X_2] \mid Z_1]] \leq u^2 \mathrm{Var}[(n-1)\mathbb{E}[S_2 \mid Z_1]].$$

This inequality is akin to applying the bound $\mathrm{Var}[XY] \leq \|Y\|_\infty^2 \mathrm{Var}[X]$, which is not true in general[2]. While the authors suggested a possible alternative approach to the proof based on checking a Lindeberg condition instead, we have not been able to verify this claim.

[1] This was confirmed by the authors in private communications.
[2] A simple counterexample is $\mathbb{P}[Y = 1] = \mathbb{P}[Y = -1] = 1/2$ and $X \sim N(1, 1)$

3 Modeling time series without linearity

This chapter contains the following paper:

[**GAR**] [Bulté and Sørensen, 2024a] Bulté, M. and Sørensen, H. (2024a). An Autoregressive Model for Time Series of Random Objects. arXiv:2405.03778.

This chapter presents our work on modeling time series of random objects. While the nonparametric nature of most existing work on random objects is well-justified given their generality, we identified an opportunity to develop a simple, yet interesting, parametric model for time series in this context. The key insight was to reinterpret the classical autoregressive model as a noisy weighted averaging, or linear interpolation, between the process mean and the previous observation. To enable theoretical analysis, additional assumptions on the structure of the space beyond the geodesic space setting had to be imposed. This led us to focus on Hadamard spaces, where geodesics are well-defined and can naturally replace linear interpolation between points, and benefiting from the convexity properties of squared distances. Our theoretical analysis was grounded in the framework of *iterated random functions* of [Wu and Shao, 2004], which has been used by other recent works to study time series in Wasserstein spaces [Ghodrati and Panaretos, 2024, Zhu and Müller, 2023]. In Section 3.2, we provide a result on the convergence rate for the estimator of the concentration parameter of the model, which was derived after the submission of the paper as well as a note on a mistake in the proof of a result in [Wu and Shao, 2004].

An Autoregressive Model for Time Series of Random Objects

Matthieu Bulté[a,b], Helle Sørensen[a]

[a]*Department of Mathematical Sciences, University of Copenhagen,*
[b]*Faculty of Business Administration and Economics, Bielefeld University,*

Abstract

Random variables in metric spaces indexed by time and observed at equally spaced time points are receiving increased attention due to their broad applicability. The absence of inherent structure in metric spaces has resulted in a literature that is predominantly non-parametric and model-free. To address this gap in models for time series of random objects, we introduce an adaptation of the classical linear autoregressive model tailored for data lying in a Hadamard space. The parameters of interest in this model are the Fréchet mean and a concentration parameter, both of which we prove can be consistently estimated from data. Additionally, we propose a test statistic for the hypothesis of absence of serial correlation and establish its asymptotic normality. Finally, we use a permutation-based procedure to obtain critical values for the test statistic under the null hypothesis. Theoretical results of our method, including the convergence of the estimators as well as the size and power of the test, are illustrated through simulations, and the utility of the model is demonstrated by an analysis of a time series of consumer inflation expectations.

Keywords: Time Series, Random Objects, Autoregressive model, Metric space, Least squares

Email addresses: mb@math.ku.dk (Matthieu Bulté), helle@math.ku.dk (Helle Sørensen)

1. Introduction

Random variables in general metric spaces, also called random objects, have been receiving increasing attention in recent statistical research. The generality of the metric space setup does not require any algebraic structure to exist and is only based on the definition of a distance function. This allows the methods developed to be applied in domains ranging from classical setups to more complex use cases on non-standard data. This includes the study of functional data (Ramsay and Silverman, 2005), data lying on Riemannian manifolds, correlation matrices and applications thereof to fMRI data (Petersen and Müller, 2019) or adjacency matrices and social networks (Dubey and Müller, 2020) among others.

One example of particular interest due to its wide range of applications is that of data comprising of probability density functions. Probability distributions are a challenging example of a space that is both functional, and thus infinite-dimensional, but also non-Euclidean because of the constraints characterizing density functions. This leads to a number of different approaches to studying these objects: they have been studied as the image of Hilbert spaces under transformations (Petersen and Müller, 2016), as specific Hilbert spaces with specific addition and scalar multiplication operators (van den Boogaart et al., 2014), as well as metric spaces with stylized distances constructed to expose certain properties or invariances (Panaretos and Zemel, 2020; Srivastava and Klassen, 2016). See Petersen et al. (2022) for a review of such methodologies. Distributions can be found in many applications, for instance the distribution of socioeconomic factors within a population such as income (Yoshiyuki, 2017), fertility (Mazzuco and Scarpa, 2015) or mortality data (Chen et al., 2021). They are also useful when considering belief distributions of economic factors (Meeks and Monti, 2023), allowing economic analyses to consider entire distributions rather than empirical expectations.

The study of random objects has received recent attention with work on hypothesis testing and inference (Dubey and Müller, 2019, 2020; McCormack and Hoff, 2023, 2022; Köstenberger and Stark, 2023) as well as various approaches to regression (Petersen and Müller, 2019; Bulté and Sørensen, 2024; Hanneke et al., 2021). Since the setup of general metric spaces offers little structure, part of the literature considers additional assumptions on the space in order for standard statistical quantities to be well-defined. This is often done by assuming that the metric space is a Hadamard space, see for instance Sturm (2003) for a detailed review of results in Hadamard spaces

and Bačák (2014) for computation of Fréchet means in such spaces.

In many of the applications mentioned above, the data might be naturally observed repeatedly on a regular time grid, forming a time series. In this case, the observations might not be independent and the models and analyses require additional care to take this dependency into account. The existing literature in this setup has mainly been carried out in a non-parametric setting, with classical weak dependence assumptions. This has been done for instance for testing serial dependence (Jiang et al., 2023) or for proving the consistency of the Fréchet mean estimator (Caner, 2006). While this line of work can be broadly applied, they rely on non-parametric assumptions rather than proposing specific models.

On the other hand, time series models have been developed for specific non-Euclidean random objects by exploiting the structure of the space under study. One popular class of models is that of autoregressive models, which have been defined using the linear structure of functional spaces (Bosq, 2000; Caponera and Marinucci, 2021) or exploiting a tangent space structure of the space (Zhu and Müller, 2024; Xavier and Manton, 2006; Ghodrati and Panaretos, 2024; Zhu and Müller, 2023; Jiang, 2023) to name a few.

Inspired by existing autoregressive models, we propose an autoregressive model for random objects. Relying on an interpretation of iteration in the linear autoregressive model as a noisy weighted sum to the mean, we define a model parametrized by a mean and a concentration parameter. For this to be possible, we assume additional structure and require the space to be a Hadamard space, and exploit the geometry of the space to define the time series iteration through geodesics. We develop the methodology and associated theory for estimation and hypothesis testing in this model. This includes estimators for the mean and the concentration parameter, and we propose a test statistic for testing for no autocorrelation, corresponding to observing an i.i.d. sample, of which we characterize the asymptotic behavior under the null hypothesis and the alternative of a non-zero concentration parameter.

The paper is organized as follows: Section 2 gives a presentation of useful concepts and results in Hadamard spaces for the rest of the article. In Section 3, we present our autoregressive model and present a theorem providing a sufficient condition for the existence of a stationary solution of the iterated system of equations associated with the model, and prove the identifiability of the model parameters. We propose in Section 4 estimators for these parameters and prove convergence results for those estimators. In Section 5,

we present our hypothesis test of independence. Finally, we illustrate our theoretical results with a numerical study in Section 6 and an application to real data in Section 7.

2. Preliminaries

Let (Ω, d) be a metric space and X a random variable, a Borel measurable function from some probability space to Ω. We say that $X \in L^p(\Omega)$ if $\mathbb{E}[d(X,\omega)^p] < \infty$ for some (and hence by the triangle inequality all) $\omega \in \Omega$. In the study of random objects, the concepts of mean and variance are generalized following the ideas of Fréchet (1948). Given a random variable $X \in L^2(\Omega)$, the Fréchet mean and variance of X are defined as

$$\mathbb{E}[X] = \arg\min_{\omega\in\Omega} \mathbb{E}\big[d(X,\omega)^2\big] \qquad \mathrm{Var}[X] = \inf_{\omega\in\Omega} \mathbb{E}\big[d(X,\omega)^2\big]. \tag{1}$$

While the existence of the variance in Euclidean spaces implies the existence of a mean, this is not necessarily the case in general metric spaces. Furthermore, on its own, a metric space offers very little to define parametric models. We now present the additional structure that will be used in this work to construct models for time series of random objects following the presentation in Burago et al. (2001) and Sturm (2003).

We call a map $\gamma : [0,1] \to \Omega$ a *path* if it continuously maps the unit interval to Ω. A path γ such that $d(\gamma(r),\gamma(t)) = d(\gamma(r),\gamma(s)) + d(\gamma(s),\gamma(t))$ for every $r < s < t \in [0,1]$ is called a *geodesic*. Given two elements $\omega, \omega' \in \Omega$, a path γ is said to *connect* ω and ω' if $\gamma(0) = \omega$ and $\gamma(1) = \omega'$. The set of all such paths is denoted by $\Gamma(\omega,\omega')$. The distance function d induces a *length* on the set of paths, defined for each γ by

$$L_p(\gamma) = \sup\left\{\sum_{i=1}^{k} d(\gamma(t_{i-1}),\gamma(t_i)) \mid 0 = t_0 \le \ldots \le t_k = 1, k \ge 1\right\}.$$

By definition, $d(\gamma(0),\gamma(1)) \le L_p(\gamma)$, and hence $d(\omega,\omega') \le \inf_{\gamma\in\Gamma(\omega,\omega')} L_p(\gamma)$ for every $\omega,\omega' \in \Omega$. A metric space in which the previous inequality always holds as an equality is called a *length space*. Furthermore, if there exists a geodesic γ connecting each pair $\omega,\omega' \in \Omega$, then we see that the infimum is attained by γ, and Ω is called a *geodesic space*.

A class of metric spaces of special interest are *Hadamard spaces*. A metric space (Ω, d) is called Hadamard space if it is complete and satisfies the *Non-Positive Curvature (NPC) inequality*: for each pair $\omega_0, \omega_1 \in \Omega$, there exists

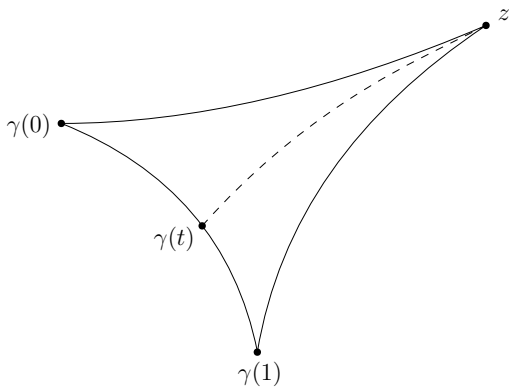


Figure 1: Illustration of the NPC inequality: the distance from any point on a side of the triangle to the opposite vertex is shorter than it would be in an equivalent Euclidean triangle.

an $\omega_{1/2} \in \Omega$ such that for every z,

$$d(z, \omega_{1/2})^2 \leq \frac{1}{2} d(z, \omega_0)^2 + \frac{1}{2} d(z, \omega_1)^2 - \frac{1}{4} d(\omega_0, \omega_1)^2. \tag{2}$$

As illustrated in Figure 1, this inequality means that triangles in Hadamard spaces are *thin*. This interpretation can be used equivalently to define of Hadamard spaces via *comparison triangles*, see Chapter 4 of Burago et al. (2001). The following proposition from Sturm (2003) shows that Hadamard spaces are geodesic spaces, and that (2) holds along geodesics.

Proposition 2.1 (Proposition 2.3 in Sturm (2003))**.** *If (Ω, d) is a Hadamard space then it is a geodesic space. Moreover, for any pair of points $\omega_0, \omega_1 \in \Omega$ there exists a unique geodesic connecting them, denoted $\gamma_{\omega_0}^{\omega_1}$. For $t \in [0, 1]$ the intermediate point $\gamma(t)$ depends continuously on the endpoints ω_0, ω_1. Finally, for any $z \in \Omega$ and $t \in [0, 1]$,*

$$d(z, \gamma(t))^2 \leq (1-t) d(z, \gamma(0))^2 + t d(z, \gamma(1))^2 - t(1-t) d(\gamma(0), \gamma(1))^2. \tag{3}$$

Since (2) is a special case of (3), we will also refer to the latter as the NPC inequality. Hadamard spaces and the NPC inequality provide a rich context for the study of random objects. One important result is that for any $X \in L^1(\Omega)$, the function $\omega \mapsto \mathbb{E}[d(X, \omega)^2 - d(X, z)^2]$ is continuous and uniformly convex, and hence by completeness of the space, has a unique minimizer, see Proposition 4.3 in Sturm (2003). Since z only enters through

an additive term which does not depend on ω, the minimizer of this function is independent of z. This implies an alternative definition of the Fréchet mean for any Hadamard-value random variable in $L^1(\Omega)$ via an arbitrary $z \in \Omega$,

$$\mathbb{E}[X] = \arg\min_{\omega \in \Omega} \mathbb{E}\big[d(X,\omega)^2 - d(X,z)^2\big].$$

We mention some further useful results in Appendix C and refer the reader to Sturm (2003) for a thorough review of the subject.

We now present a few examples from Sturm (2003) of Hadamard spaces and ways of building Hadamard spaces out of existing ones. The most well-known case of Hadamard space are Hilbert spaces. Since Functional Data Analysis (FDA) is typically carried out in the L_2 Hilbert spaces (see Ramsay and Silverman (2005)), considering Hadamard spaces allows to approach FDA tasks from a random object perspective.

Example 2.1 (Hilbert spaces)**.** *Let $\mathcal{H}$ be a Hilbert space with induced norm $\|\cdot\|_{\mathcal{H}}$, then $(\mathcal{H}, d)$ with $d(x,y) = \|x-y\|_{\mathcal{H}}$ is a Hadamard space. In Hilbert spaces, the Fréchet mean of X corresponds to the Bochner integral with respect to the probability measure P of X, $\mathbb{E}[X] = \int X dP$ (see Hsing and Eubank (2015)). In Hilbert spaces, the NPC inequality (3) holds to equality.*

Example 2.2 (Constructed Spaces)**.** *Let (Ω, d) be a Hadamard space. Then*

1. *Any subset $O \subset \Omega$ is a Hadamard space if and only if it is closed and convex.*
2. *Let Θ be an arbitrary set and $\omega : \Theta \to \Omega$ be a bijection. Then, Θ is a Hadamard space equipped with the distance $d_\omega(\theta, \theta') = d(\omega(\theta), \omega(\theta'))$. Furthermore, $\mathbb{E}[X] = \omega^{-1}(\mathbb{E}[\omega(X)])$ holds for any $X \in L^1(\Theta)$.*

One specific example of Hadamard space of particular interest is the space of one-dimensional density functions over an interval equipped with the 2-Wasserstein distance.

Example 2.3 (2-Wasserstein Space)**.** *Let $\mathcal{W}_2(I)$ be the space of probability measures on $I \subset \mathbb{R}$ with finite second moment. This space, endowed with the 2-Wasserstein distance, is a metric space, see Panaretos and Zemel (2020). Consider the subset $\mathcal{D}(I) \subset \mathcal{W}_2(I)$ of distributions having a density with respect to the Lebesgues measure. For two distributions $\mathbb{P}, \mathbb{Q} \in \mathcal{D}(I)$ with quantile functions $F_{\mathbb{P}}^{-1}, F_{\mathbb{Q}}^{-1}$, the 2-Wasserstein distance between $\mathbb{P}$ and $\mathbb{Q}$ is given by*

$$d_{\mathcal{W}_2}(\mathbb{P}, \mathbb{Q}) = \big\|F_{\mathbb{P}}^{-1} - F_{\mathbb{Q}}^{-1}\big\|_{L_2[0,1]}.$$

The space of quantile functions being a closed and convex subspace of $L_2[0,1]$, it is also a Hadamard space. Hence, $(\mathcal{D}(I), d_{\mathcal{W}_2})$ falls under the second case described in Example 2.2 and is also a Hadamard space.

Another useful example of a constructed Hadamard space is that of symmetric positive definite (SPD) matrices together with the *Log-Cholesky* distance.

Example 2.4 (Log-Cholesky distance)**.** *Let $\mathcal{S}_p^+$ be the space of SPD matrices of dimension p and $\mathcal{L}_p^+$ be the space of $p \times p$ lower-triangular matrices with positive diagonal elements. Given a matrix $M \in \mathcal{S}_p^+$, the Cholesky decomposition of M is well defined, meaning that there exists a lower-triangular matrix with positive diagonal elements $L \in \mathcal{L}_p^+$ such that $M = LL^\top$. Let $\lfloor M \rfloor$ be the $p \times p$ matrix such that $\lfloor M \rfloor_{ij} = M_{ij}$ if $i < j$ and 0 otherwise and $D(M)$ be the $p \times p$ diagonal matrix with diagonal entries $D(M)_{ii} = M_{ii}$. While simply using the Froebenius distance between Cholesky factors of SPD matrices yields a valid distance, Lin (2019) argues that it leads to an unwanted swelling effect in geodesics and proposes another distance $d_{\mathcal{S}_p^+}$ treating the diagonal and strictly lower triangular parts of L differently. Let $M_1, M_2 \in \mathcal{S}_p^+$ with Cholesky factors L_1 and L_2, then the distance $d_{\mathcal{S}_p^+}(M_1, M_2)$ is given by*

$$d_{\mathcal{S}_p^+}(M_1, M_2)^2 = \|\lfloor L_1 \rfloor - \lfloor L_2 \rfloor\|_F^2 + \|\log D(L_1) - \log D(L_2)\|_F^2,$$

where $\|\cdot\|_F$ is the Froebenius norm. As a case of Example 2.2, this is a Hadamard space, which is also shown in Lin (2019), together with other properties of the space.

3. The GAR(1) Model

3.1. Model and stationary solution

Let us consider first a time series $\{X_t\}_{t\in\mathbb{N}}$ in $\mathbb{R}$ with constant mean $\mathbb{E}[X_t] = \mu$ for all $t \in \mathbb{N}$. Then, $\{X_t\}_{t\in\mathbb{N}}$ follows a first-order autoregressive model, denoted AR(1), with concentration parameter φ if it satisfies the following relation

$$X_{t+1} - \mu = \varphi(X_t - \mu) + \varepsilon_{t+1}, \tag{4}$$

where the noise terms $\{\varepsilon_t\}_{t\in\mathbb{N}}$ are i.i.d random variables with mean 0. Without the structure of a vector space, this model cannot be directly formulated

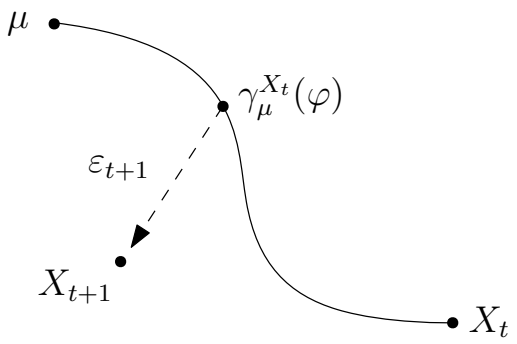


Figure 2: Illustration of the iterated equation (6).

in general metric spaces. A key insight towards a more general definition of AR(1) models is that (4) can be rewritten as

$$X_{t+1} = (1-\varphi)\mu + \varphi X_t + \varepsilon_{t+1}.$$

This shows that each random variable of the time series can be written as a weighted sum of the overall mean of the time series and the previous observation, perturbed by a centered random noise. For $\varphi \in [0,1]$, this weighted sum corresponds to the point along the geodesic from μ to X_t at φ. This interpretation can be used to define an autoregressive process only using geodesics.

Let now (Ω, d) be a Hadamard space. In this context, we consider a broad class of noise models represented by random maps $\varepsilon : \Omega \to \Omega$. We say that a random map ε is *unbiased* if for all $\omega \in \Omega$, the random variable $\varepsilon(\omega)$ is in $L^1(\Omega)$ and $\mathbb{E}[\varepsilon(\omega)] = \omega$. Note that the expectation is the Fréchet mean in Ω, hence by (1), the previous statement can be written as

$$\mathbb{E}\big[d(\varepsilon(\omega),\omega)^2 - d(\varepsilon(\omega),\omega')^2\big] < 0 \qquad \text{for all } \omega,\omega' \in \Omega \,\text{with}\, \omega \neq \omega'. \tag{5}$$

We say that a sequence of random variables $\{X_t\}_{t\in\mathbb{N}} \subset L^1(\Omega)$ with common mean μ follows the *geodesic autoregressive* model of order 1, GAR(1), with concentration parameter $\varphi \in [0,1]$ if it satisfies the following iterated system of equations

$$X_{t+1} = \varepsilon_{t+1}(\gamma_\mu^{X_t}(\varphi)), \tag{6}$$

where $\{\varepsilon_t\}_{t\in\mathbb{N}}$ are i.i.d unbiased noise maps and $\gamma_\mu^{X_t}$, we recall, is the (random) geodesic connecting μ to X_t. The data generating process is illustrated in Figure 2.

This relation opens the question of whether the condition of a shared Fréchet mean $\mathbb{E}[X_t] = \mu$ and equation (6) can mutually be fulfilled. Using that $\{\varepsilon_t\}_{t\in\mathbb{N}}$ are unbiased and assuming that $\mathbb{E}[X_t] = \mu$, the condition becomes

$$\mathbb{E}\left[\gamma^{X_t}_{\mathbb{E}[X_t]}(\varphi)\right] = \mathbb{E}[X_t].$$

Unfortunately, this condition does not hold in every Hadamard space. Thus, we will assume the following.

Assumption 3.1. *For every $X \in L^1(\Omega)$ and $\varphi \in [0,1]$, $\mathbb{E}\left[\gamma^{X}_{\mathbb{E}[X]}(\varphi)\right] = \mathbb{E}[X]$.*

Assumption 3.1 holds for some of the examples mentioned in the previous section. For any Hilbert space $\mathcal{H}$, this condition holds by linearity of the expectation since for any $X \in L^1(\mathcal{H})$,

$$\mathbb{E}\left[\gamma^{X}_{\mathbb{E}[X]}(\varphi)\right] = \mathbb{E}[(1-\varphi)\mathbb{E}[X] + \varphi X] = (1-\varphi)\mathbb{E}[X] + \varphi\mathbb{E}[X] = \mathbb{E}[X].$$

Furthermore, if the condition holds for a Hadamard space (Ω, d), then it also holds for a Hadamard space constructed by taking the image of a bijection ω as described in the second part of Example 2.2.

Lemma 3.1. *Let (Ω, d) be a Hadamard space and (Θ, d_ω) be a constructed Hadamard space based on the bijection ω. If (Ω, d) satisfies Assumption 3.1, then so does (Θ, d_ω).*

Lemma 3.1 shows that Assumption 3.1 holds for a large class of Hadamard spaces, in particular for the subspace $\mathcal{D}(I)$ of $\mathcal{W}_2(I)$ of distributions having a density function, as described in Example 2.3.

To show the existence of a stationary solution to Equation (6), we use the framework of *iterated random function systems* presented in Wu and Shao (2004). Let us first introduce some notation. Given the i.i.d noise maps $\{\varepsilon_t\}_{t\in\mathbb{N}}$, define for all $t \in \mathbb{N}$ the random functions $F_t : \Omega \to \Omega$,

$$F_t(x) = \varepsilon_t(\gamma^x_\mu(\varphi)).$$

Then, Equation (6) can be rewritten as an iterated random function system,

$$X_{t+1} = F_{t+1}(X_t).$$

Further, for any $t \in \mathbb{N}$ and $x \in \Omega$, the following random variable will be useful in expressing the condition of existence of a stationary solution,

$$X_t(x) = F_t \circ F_{t-1} \circ \ldots \circ F_1(x), \tag{7}$$

then, $X_t = X_t(X_0)$, and the construct also allows to study a coupled version $X_t(X_0')$ of X_t for $X_0' \overset{\mathcal{D}}{=} X_0$. The following theorem provides a sufficient condition for the existence of a stationary solution to (7) based on a *geometric-moment contracting* condition on the iteration maps $\{F_t\}_{t\in\mathbb{N}}$.

Theorem 3.1 (Theorem 2 of Wu and Shao (2004))**.** *Suppose there exists* $x_0 \in \Omega, \alpha > 0, r \in (0,1)$ *and* $C > 0$ *such that*

$$\mathbb{E}[d(X_t(x), X_t(x_0))^\alpha] \leq C r^t d(x, x_0)^\alpha \tag{8}$$

holds for all $x \in \Omega, t \in \mathbb{N}$*. Then, for all* $x \in \Omega$

$$X_t^\star = \lim_{m\to\infty} F_t \circ F_{t-1} \circ \ldots \circ F_{t-m+1}(x)$$

exists and does not depend on x*. Moreover,* $\{X_t^\star\}_{t\in\mathbb{N}}$ *is a stationary solution of Equation (6).*

As noted in Wu and Shao (2004), if condition 8 holds for some $\alpha \geq 1$, then Hölder's inequality can be directly use to show that it also holds for any $\beta \in (0, \alpha)$. While the GAR(1) model can be defined for $\varphi = 1$, it is unlikely that the process will have a stationary solution in this case. This is due to the fact that the noise maps $\{\varepsilon_t\}_{t\in\mathbb{N}}$ are unbiased, and the iterated system of equations will not converge to a stationary solution if the noise maps are not contracting.

If we assume that the noise maps $\{\varepsilon_t\}_{t\in\mathbb{N}}$ are Lipschitz with random Lipschitz constants $K_t \in L^\alpha(\mathbb{R}_+)$ with $K := \mathbb{E}[K_t^\alpha]$, we have by using the Geodesic Comparison Inequality (see C.2) on $d(\gamma_\mu^x(\varphi), \gamma_\mu^{x_0}(\varphi))^2$

$$\begin{aligned}
\mathbb{E}[d(X_1(x), X_1(x_0))^\alpha] &= \mathbb{E}\big[d(\varepsilon_1(\gamma_\mu^x(\varphi)), \varepsilon_1(\gamma_\mu^{x_0}(\varphi)))^\alpha\big] \\
&\leq K d(\gamma_\mu^x(\varphi), \gamma_\mu^{x_0}(\varphi))^\alpha \\
&\leq K\varphi^{\alpha/2} d(x, x_0)^\alpha.
\end{aligned}$$

By induction, this implies $\mathbb{E}[d(X_t(x), X_t(x_0))^\alpha] \leq \big(K\varphi^{\alpha/2}\big)^t d(x, x_0)^\alpha$, which shows that condition (8) holds if $r = K\varphi^{\alpha/2} < 1$.

3.2. Identifiability

Under the conditions of Theorem 3.1, Equation (6) has a stationary solution and the model features two quantities of interest: the time-invariant

Fréchet mean of the time series $\mu \in \Omega$, and the concentration parameter $\varphi \in [0, 1]$. Before considering the estimation of these quantities, we show that both are identifiable. The identifiability of the Fréchet mean follows directly from the stationarity of the time series and the definition and existence of the Fréchet mean in a Hadamard space.

Theorem 3.2. *Let $\{X_t\}_{t\in\mathbb{N}} \subset \Omega$ and assume that $\{X_t\}_{t\in\mathbb{N}}$ satisfies the conditions of Theorem 3.1. Then, the Fréchet mean $\mu = \mathbb{E}[X_t]$ is identifiable.*

As for the concentration parameter, we can consider the mean squared error

$$L(u) = \mathbb{E}\Big[d\big(X_{t+1}, \gamma_\mu^{X_t}(u)\big)^2\Big]. \tag{9}$$

Then, assuming that the noise maps are unbiased, we can show that this loss is uniquely minimized by the true concentration parameter φ.

Theorem 3.3. *Let $\{X_t\}_{t\in\mathbb{N}} \subset \Omega$, assume that $\{X_t\}_{t\in\mathbb{N}}$ is in $L^2(\Omega)$ and satisfies Equation (6) with true concentration parameter $\varphi \in [0, 1]$. Assume further that the noise maps $\{\varepsilon_t\}_{t\in\mathbb{N}}$ are unbiased. Then, φ is the unique minimizer of L.*

Proof. Since $\gamma_\mu^{X_t}(\varphi)$ is the Fréchet mean of X_{t+1} given X_t and ε_{t+1} is unbiased, for all $\varphi' \in [0, 1]$, with $\varphi' \neq \varphi$, we have

$$\begin{aligned} L(\varphi') &= \mathbb{E}\Big[d\big(X_{t+1}, \gamma_\mu^{X_t}(\varphi')\big)^2\Big] = \mathbb{E}\Big[d\big(\varepsilon_{t+1}(\gamma_\mu^{X_t}(\varphi)), \gamma_\mu^{X_t}(\varphi')\big)^2\Big] \\ &= \mathbb{E}\Big[\mathbb{E}\Big[d\big(\varepsilon_{t+1}(\gamma_\mu^{X_t}(\varphi)), \gamma_\mu^{X_t}(\varphi')\big)^2 \mid X_t\Big]\Big] \\ &> \mathbb{E}\Big[\mathbb{E}\Big[d\big(\varepsilon_{t+1}(\gamma_\mu^{X_t}(\varphi)), \gamma_\mu^{X_t}(\varphi)\big)^2 \mid X_t\Big]\Big] = L(\varphi) \end{aligned}$$

where the inequality follows from the unbiasedness of ε_{t+1} as defined in Equation (5). □

4. Estimation of model parameters

Now that the identifiability of the Fréchet mean and the concentration parameter have been established, we show that empirical estimation of the associated risks produces consistent estimators. Furthermore, we show that the Fréchet mean can be estimated at a $\sqrt{T}$-rate.

4.1. Fréchet mean

For simplicity, we assume that the $\{X_t\}_{t\in\mathbb{N}}$ are $L^2(\Omega)$. Then, the Fréchet function $M(\omega) = \mathbb{E}[d(X,\omega)^2]$ has a natural empirical version based on the observations $X_1, \dots, X_T$.

$$M_T(\omega) = \frac{1}{T}\sum_{t=1}^{T} d(X_t,\omega)^2. \tag{10}$$

We define the estimator $\hat{\mu}_T$ as the minimizer of M_T, which is well-defined by strict convexity of the squared distance in Hadamard spaces. The asymptotic behavior of $\hat{\mu}_T$ is described by the theory of M-estimation, see for instance van der Vaart and Wellner (1996), where consistency and rates of convergence are readily available for i.i.d data. Here, we adapt results on iterated random function system from Wu and Shao (2004) to verify the general assumptions for M-estimation presented in van der Vaart and Wellner (1996). One assumption which is standard in the study of random objects, concerns the covering number of (Ω, d).

Assumption 4.1. *Let $B(\mu,\delta)$ be the ball in Ω of size δ centered in μ and $N(\varepsilon, B_\delta(\mu))$ be the covering number of $B_\delta(\mu)$ using balls of size ε. Assume*

$$\int_0^1 \sqrt{1+\log N(\varepsilon\delta, B_\delta(\mu))}\, d\varepsilon = O(1) \qquad \text{as } \delta \to 0.$$

In the following theorem, we show that this assumption, together with the assumptions required for stationary of the sequence $\{X_t\}_{t\in\mathbb{N}}$, are enough to obtain the $\sqrt{T}$ consistency of the mean estimator. Note that this result is of more general interest since it does not assume that the data follows our GAR(1) model but only requires control on the dependency of the sequence $\{X_t\}_{t\in\mathbb{N}}$.

Theorem 4.1. *Let (Ω, d) be a Hadamard space and $\{X_t\}_{t\in\mathbb{N}}$ be an $L^2(\Omega)$ sequence of random variables satisfying Equation (8) for some $\alpha \geq 1$. Suppose that Assumption 4.1 holds around the Fréchet mean $\mu = \mathbb{E}[X_t]$. Then, the minimizer $\hat{\mu}_T$ of M_T is a consistent estimator of μ and satisfies*

$$\sqrt{T}d(\mu,\hat{\mu}_T) = O_P(1). \tag{11}$$

The proof of Theorem 4.1 can be found in Appendix B.

4.2. Concentration parameter

Similarly to the Fréchet mean, we construct an estimator of the concentration parameter by minimizing an empirical version of L in Equation (9). We estimate the expectation with the available sample and replace the Fréchet mean μ by the estimator $\hat{\mu}_T$, giving the following risk function

$$L_T(u) = \frac{1}{T-1} \sum_{t=1}^{T-1} d(X_{t+1}, \gamma_{\hat{\mu}_T}^{X_t}(u))^2. \tag{12}$$

We prove the consistency of the resulting estimator based on results from Newey (1991) relying on the compactness of the domain $[0, 1]$ and continuity results about L and L_T. The consistency result is the following.

Theorem 4.2. *Assume that (Ω, d) and $\{X_t\}_{t\in\mathbb{N}}$ satisfy the conditions of Theorem 4.1. Then, the minimizer $\hat{\varphi}_T$ of L_T is a consistent estimator of φ.*

Proof. By Proposition B.2, we have that $\|L_T - L\|_\infty = o_P(1)$. Together with the identifiability result in Theorem 3.3, we have that L has a unique minimizer. By Corollary 3.2.3 in van der Vaart and Wellner (1996), any sequence of minimizers $\hat{\varphi}_T$ of L_T satisfies $|\hat{\varphi}_T - \varphi| = o_P(1)$. □

Furthermore, we show in Lemma C.1 in the Appendix that L, and hence L_T, is strongly convex. This makes it possible to use generic convex solvers to find the minimizer of L_T.

5. Test for serial independence

One hypothesis test of interest is whether the random variables $\{X_t\}_{t\in\mathbb{N}}$ are independent, which corresponds in the GAR(1) model to testing $H_0 : \varphi = 0$ vs. $H_1 : \varphi > 0$. Since no strong results are available about the asymptotic distribution of $\hat{\varphi}_T$, another test statistic must be considered. To that end, let us consider the statistic

$$D_T = \frac{1}{T-1} \sum_{t=1}^{T-1} d(X_t, X_{t+1})^2. \tag{13}$$

We proceed to show that D_T is asymptotically normal with mean and variance depending on the value of the concentration parameter φ, with smaller values of D_T taken for larger values of φ. This allows us to build a test that asymptotically has correct level and power.

Theorem 5.1. *Let X_1, X_2, X_3 be i.i.d copies of $\varepsilon(\mu)$, then, under H_0 and as $T \to \infty$,*

$$\sqrt{T}(D_T - \mathbb{E}\big[d(X_1, X_2)^2\big]) \to N(0, \sigma_0^2),$$

where $\sigma_0^2 = Var[d(X_1, X_2)^2] + 2\,Cov[d(X_1, X_2)^2, d(X_1, X_3)^2]$.

Proof. Under H_0, the sequence $\{X_t\}_{t\in\mathbb{N}}$ is formed of i.i.d random variables. We consider the centered summands, $Y_t = d(X_t, X_{t+1})^2 - \mathbb{E}[d(X_1, X_2)^2]$. The sequence $\{Y_t\}_{t\in\mathbb{N}}$ is then m-dependent with $m = 1$. We obtain the desired result by the Central Limit Theorem for m-dependent sequences, see Theorem 2 in Hoeffding and Robbins (1948). □

To study the behavior of this test statistic under $H_1 : \varphi \neq 0$, we base our analysis on Theorem 3 of Wu and Shao (2004) which provides conditions for the asymptotic normality of sums of the form of D_T. Under the assumptions required for the existence of a stationary solution, we obtain the result.

Theorem 5.2. *Assume that $\{X_t\}_{t\in\mathbb{N}}$ satisfies the conditions of Theorem 3.1 with $\varphi > 0$, then there exists a $\sigma_\varphi \geq 0$ such that*

$$\sqrt{T}(D_T - \mathbb{E}\big[d(X_1, X_2)^2\big]) \to N(0, \sigma_\varphi^2).$$

In general, it is not clear whether $\mathbb{E}_{\varphi=0}[D_T] \neq \mathbb{E}_{\varphi=\varphi^\star}[D_T]$ for an arbitrary $\varphi^\star \neq 0$, and whether the test described above has asymptotic power. One possible way to avoid this issue is to require the following monotonicity condition on the noise maps.

Assumption 5.1. *For all $x, y, z \in \Omega$, then the noise maps ε satisfy the following monotonicity condition*

$$d(x, z) < d(y, z) \Rightarrow \mathbb{E}\big[d(\varepsilon(x), z)^2\big] < \mathbb{E}\big[d(\varepsilon(y), z)^2\big]. \tag{14}$$

For any $\varphi > 0$, $d(\gamma_\mu^{X_t}(\varphi), X_t) = (1 - \varphi)d(\mu, X_t) < d(\mu, X_t)$. Together with Assumption 5.1, this gives

$$\mathbb{E}_{\varphi=\varphi^\star}[D_T] = \mathbb{E}\big[d(\varepsilon_{t+1}(\gamma_\mu^{X_t}(\varphi^\star)), X_t)^2\big] < \mathbb{E}\big[d(\varepsilon_{t+1}(\mu), X_t)^2\big] = \mathbb{E}_{\varphi=0}[D_T],$$

which implies that the asymptotic power of the test is 1.

To construct a level α hypothesis test for $H_0 : \varphi = 0$ vs. $H_1 : \varphi > 0$, one could reject H_0 if the absolute deviation of D_T from its asymptotic mean exceeds a certain threshold q_α based on the result in Theorem 5.1.

However, the asymptotic mean and variance of D_T required for this test depend on the underlying data distribution and are unknown. Alternatively, one could attempt to center D_T, for instance by considering the randomized statistic $\tilde{D}_T = \frac{1}{T-1}\sum_{t=1}^{T-1} d(X_t, X_{t+1})^2 - d(X_t, X_{\pi(t)})^2$, where π is a random permutation. Similar theoretical arguments as for D_T shows that $\tilde{D}_T$ is asymptotically normal under H_0, with zero mean and a variance estimable from data. This enables normalization of $\tilde{D}_T$ and the construction of a test based on the asymptotic approximation.

Instead, we use a permutation procedure to compute approximate p-values under H_0 for better finite sample properties. Specifically, let $B \in \mathbb{N}$ be the number of permutations used for constructing the approximate p-value and let $\pi_1, \dots, \pi_B$ be random permutations of $\{1, \dots, T\}$. For each permutation π, we denote by D_T^π the test statistic computed based on the permuted sample $\left(X_{\pi(1)}, \dots, X_{\pi(T)}\right)$ and define the approximated p-value, $\hat{p}_B = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\{D_T \geq D_T^{\pi_b}\}$. Under the null hypothesis of independence, we have that $D_T^\pi \overset{\mathcal{D}}{=} D_T$, and shuffling the observations under the alternative allows to loosen the dependency between consecutive observations, giving an approximate sample under H_0. The resulting level α test is then constructed by rejecting H_0 if $\hat{p}_B \leq \alpha$, see Hemerik and Goeman (2018).

6. Numerical experiments

In the following, we illustrate our theoretical results with numerical experiments taking place in different Hadamard spaces. We empirically verify the convergence rate of $\hat{\mu}_T$ proved in Theorem 4.1, verify the consistency of $\hat{\varphi}_T$ proved in Theorem 4.2 and show that the test constructed via the bootstrapping procedure described in Section 5 has the desired size and increasing power as T grows.

We study three scenarios of time series following the GAR(1) model. The first example is that of the real line $\mathbb{R}$ equipped with the standard Euclidean distance, with a multiplicative noise model. For the second example, we consider the space of density distributions over the real line equipped with the 2-Wasserstein distance, with a geodesic noise model. For the last example, we consider SPD matrices with the Log-Cholesky metric with a noise model based on a Lie group structure defined in Lin (2019).

In each of these scenarios, we generate time series of different lengths $T \in \{40, 80, 160, 320, 640\}$ and for different values of the concentration parameter

$\varphi \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 1\}$. Naturally, $\varphi = 0$ and $\varphi = 1$ are special cases that we will consider with care in the evaluation of our results. For each combination of metric space, number of observations and concentration parameter, we generate 1000 datasets. For each dataset, we compute the estimators $\hat{\mu}_T$ and $\hat{\varphi}_T$, and run the permutation-based hypothesis test at level $\alpha = 0.05$. We report for each combination of a metric space, T and φ, the average estimation errors $d(\hat{\mu}_T, \mu)$ and $|\hat{\varphi}_T - \varphi|$, as well as properties of the hypothesis test, all calculated over the 1000 runs.

Additional results can be found in Appendix A where we compare our permutation test for D_T to three other tests: another permutation test based on the $\hat{\varphi}_T$ estimator and the two tests developed in Jiang et al. (2023). This appendix shows that the test based on D_T is better calibrated and achieves considerably higher power than all other evaluated in all simulation scenarios considered.

All simulations and analyses are done in Python. The code to reproduce the experiments and figures is available online[1].

6.1. $\mathbb{R}$ with multiplicative noise

In the first experimental setup, we investigate the simple case of (Ω, d) being the real line line $\mathbb{R}$ equipped with the Euclidean distance $d(x, y) = |x - y|$. Here, the Fréchet mean corresponds to the ordinary mean and geodesics are given by straight lines, $\gamma_x^y(t) = (1-t)x + ty$. Furthermore, Assumption 4.1 is verified since $N(\varepsilon\delta, B_\delta(\mu)) = \varepsilon^{-1}$, thus the entropy integral is bounded and does not depend on δ.

One can see that the unconstrained minimizer of Equation (12) is the sample autocorrelation. Using the convexity of L_T, the estimator $\hat{\varphi}_T$ of the concentration parameter is still available in closed form by clipping the autocorrelation to positive numbers, giving

$$\hat{\varphi}_T = \max\left\{0, \frac{\sum_{t=1}^{T-1}(X_{t+1} - \bar{X})(X_t - \bar{X})}{\sum_{t=1}^{T-1}(X_t - \bar{X})^2}\right\}, \tag{15}$$

where $\bar{X}$ is the sample mean of the time series.

We consider multiplicative noise maps $\varepsilon_i(x) = (1 + \eta_i)x$ where $\eta_i \sim N(0, \sigma^2)$. Then, the noise maps $\{\varepsilon_t\}_{t\in\mathbb{N}}$ are unbiased and the condition of

[1] https://github.com/matthieubulte/GAR

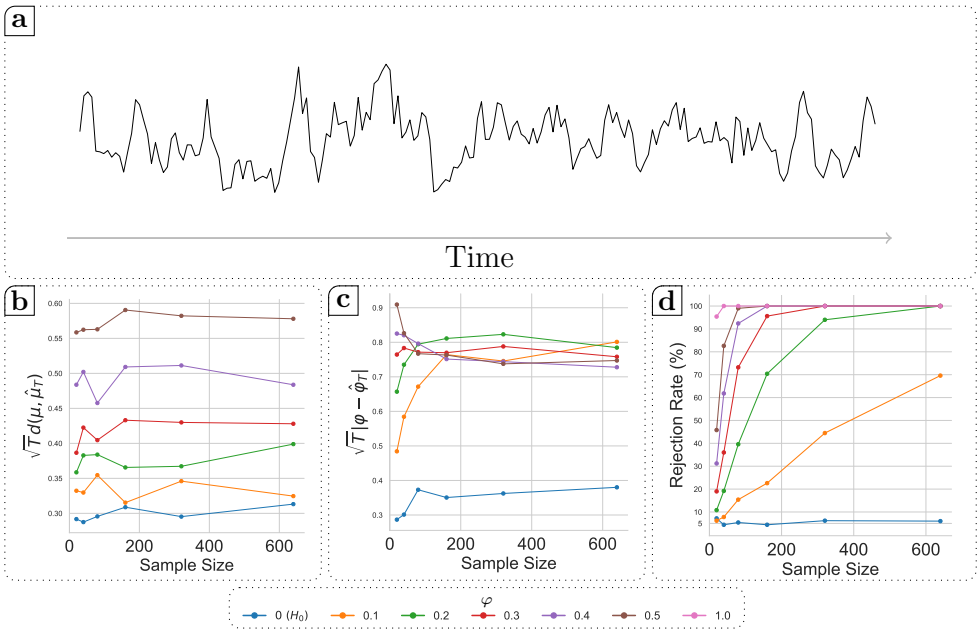


Figure 3: Panel (a) depicts a trajectory of 200 time steps from the GAR(1) process described in Section 6.1. Panel (b) illustrates the $\sqrt{T}$ convergence of the mean estimator $\hat{\mu}_T$. Panel (c) shows the convergence of the concentration parameter estimator $\hat{\varphi}_T$. Panel (d) shows the rejection rate of the independence test described in Section 5 with target level 0.05 for different values of φ.

Theorem 3.1 is satisfied for $\varphi < (1+\sigma^2)^{-1/2}$ since

$$\mathbb{E}\big[(X_t(x_0) - X_t(x))^2\big] = \big[\varphi^2(1+\sigma^2)\big]^t (x_0 - x)^2.$$

In our simulation setup, we work with $\sigma^2 = 0.25^2$ which gives an upper bound $\varphi < (1+\sigma^2)^{-1/2} \approx 0.97$.

The theoretical results presented in the previous sections are illustrated in Figure 3. The $\sqrt{T}$ convergence of the Fréchet mean estimator holds for $\varphi < 1$ with a reasonably stable value at the tested sample sizes, see panel (b). Panel (c) indicates that the estimator $\hat{\varphi}_T$ also converges at the parametric $\sqrt{T}$ rate to φ for $\varphi < 1$. We can see that for $\varphi = 0$, the error is lower which can be explained by the fact the the sample is i.i.d. in this case. In the non-i.i.d. case, the estimation error seems smaller for the larger values of φ considered. Finally, the rejection rates for the independence test presented in Section 5 show that the test is well calibrated and achieves high power for

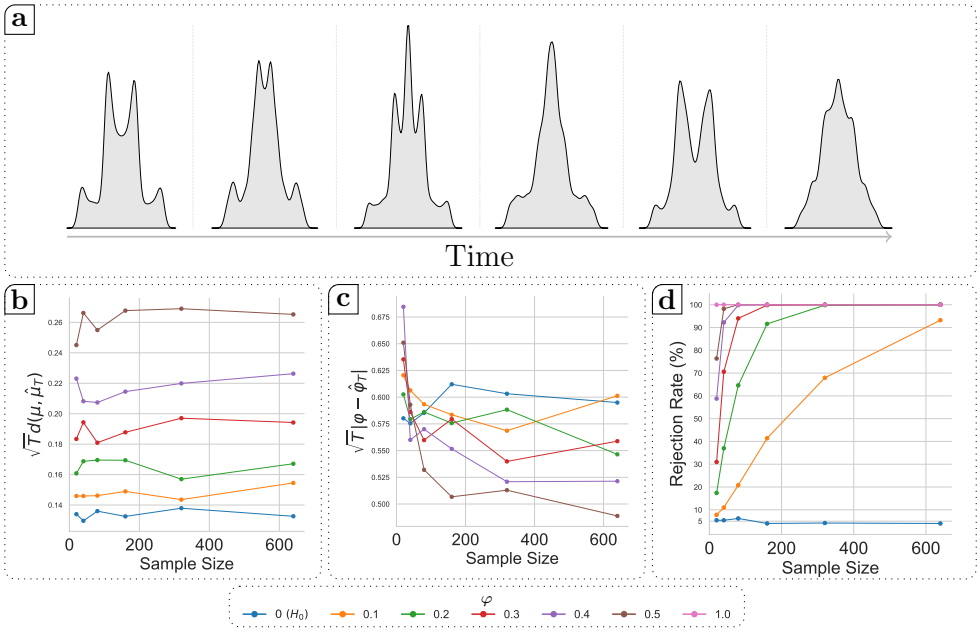


Figure 4: Panel (a) shows six consecutive densities sampled from the GAR(1) process described in Section 6.2. Panels (b), (c) and (d) are generated in the same way as in Figure 3.

moderate sample sizes, see panel (d).

6.2. Univariate distributions with a density

In this second experiment, we consider time series in the space $\mathcal{D}([0,1])$ of density functions over $[0,1]$ equipped with the 2-Wasserstein distance, as described in Example 2.3. Since the support of the distributions is bounded, the space $\mathcal{D}([0,1])$ is bounded as well. Geodesics are given by linear interpolation of the corresponding quantile function: given two distributions $\mathbb{P}, \mathbb{Q} \in \mathcal{D}([0,1])$ with quantile functions $F_{\mathbb{P}}^{-1}, F_{\mathbb{Q}}^{-1}$, the quantile function of any point on the connection geodesic $\gamma_{\mathbb{P}}^{\mathbb{Q}}$ is $F^{-1}_{\gamma_{\mathbb{P}}^{\mathbb{Q}}(t)}(u) = (1-t)F_{\mathbb{P}}^{-1}(u) + tF_{\mathbb{Q}}^{-1}(u)$.

We generate the time series with the standard normal distribution $N(0,1)$ truncated to $[0,1]$ as the Fréchet mean. Then, the data is generated according to Equation (6). The noise sampling is based on sampling a random optimal transport $\eta : [0,1] \to [0,1]$ and applying it by quantile composition, which corresponds to computing the pushforward under η: given a distribution

$\mathbb{P} \in \mathcal{D}([0,1])$ with quantile function $F_{\mathbb{P}}^{-1}$, the noise map ε is then given by $F_{\varepsilon(\mathbb{P})}^{-1} = \eta \circ F_{\mathbb{P}}^{-1}$. To generate the transport maps η, we follow the procedure described in Panaretos and Zemel (2016). First, a random integer frequency is uniformly sampled from $\{-4, \dots, 4\} \backslash \{0\}$, then, the maps are given by $\eta(x) = x - \sin(\pi k x)/|\pi k|$. The random maps η are smooth, strictly increasing and satisfy $\eta(0) = 0$ and $\eta(1) = 1$. By symmetry of the random parameter k, one can see that for any $x \in [0,1]$, we have $\mathbb{E}[\eta(x)] = x$, and this property is inherited by the noise maps ε. While the noise maps have a Lipschitz constant of 2, meaning that the condition of Theorem 3.1 is satisfied for $\varphi < 0.5$, we observe empirically that the estimators seem to still be consistent even for values of $\varphi \in [0.5, 1)$.

The results in Figure 4 match those observed in the previous experiment. The scaled error curves displayed in panel (b) confirm the convergence rate proved in Theorem 4.1. Similarly to the previous experiment, the conditions of Theorem 3.1 are not satisfied for $\varphi = 1$, and the estimator $\hat{\mu}_T$ fails to converge; to improve readability, we did not include the associated curve. Similarly to the previous experiment, the estimator $\hat{\varphi}_T$ seems to converge at a $\sqrt{T}$ rate. The error does not seem to improve for i.i.d. observations, but the monotonicity as a function of φ observed in the previous experiment approximately holds as seen in panel (c). We also observe in panel (d) that the hypothesis test behaves as expected. The blue curve, corresponding to the null hypothesis, demonstrates that the empirical size of the test is correct for all sample sizes considered.

6.3. SPD matrices

In this last experiment we investigate the properties of the GAR(1) model in the space $\mathcal{S}_{10}^{+}$ of 10×10 SPD matrices with the Log-Cholesky distance described in Example 2.4. In this space, matrices $M_0, M_1 \in \mathcal{S}_{10}^{+}$ are uniquely identified by their Cholesky factors L_0, L_1. Points on the geodesic line between these matrices are given by linearly interpolating off-diagonal entries of the Cholesky factors and geometrically interpolating the diagonal elements. That is, for $t \in [0,1]$, the Cholesky factor L_t of $\gamma_{M_0}^{M_1}(t)$ is given via $\lfloor L_t \rfloor = (1-t)\lfloor L_0 \rfloor + t \lfloor L_t \rfloor$ and $D(L_t) = D(L_0)^{1-t} D(L_1)^t$.

We generate time series with the identity matrix 1_{10} as the Fréchet mean. Each noise map in this experiment applies a random congruent transformation of the input with a random lower-triangular matrix $L_\varepsilon \in \mathbb{R}^{10 \times 10}$ with $\varepsilon(X) = L_\varepsilon X L_\varepsilon^\top$. The lower-triangular entries of L_ε are i.i.d. following a normal distribution $\lfloor L_\varepsilon \rfloor_{ij} \sim N(0, 0.5^2)$, and the diagonal entries are

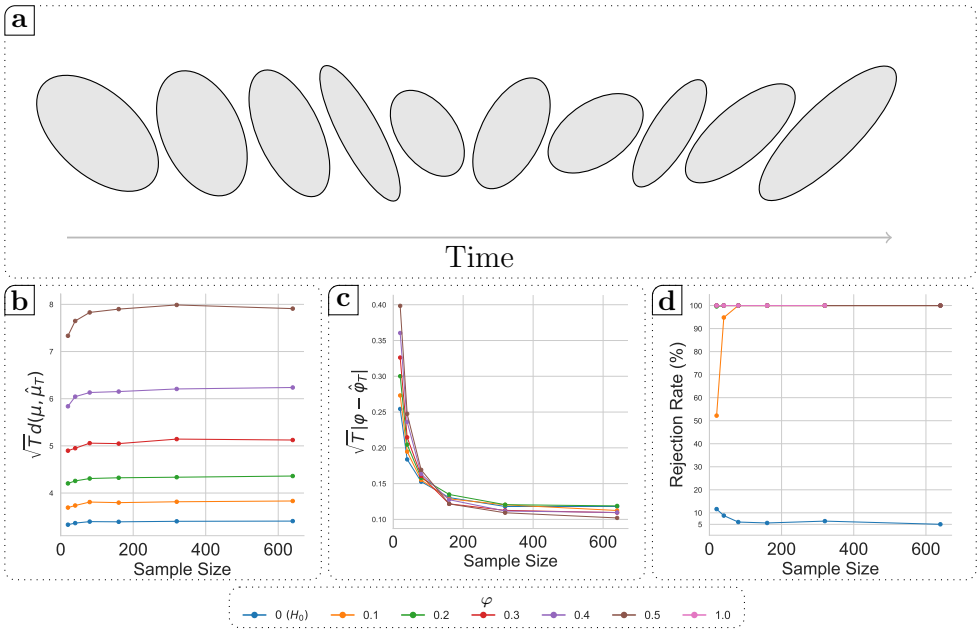


Figure 5: Panel (a) displays 10 consecutive covariance ellipses corresponding to the top-left 2×2 submatrix of the covariances sampled from the GAR(1) process described in Section 6.2. Each ellipse is the application of a covariance submatrix to a unit circle. Panels (b), (c) and (d) are generated in the same way as in Figure 3.

i.i.d. following a log-normal distribution $\log D(L_\varepsilon)_{ii} \sim N(0, 0.2^2)$. For a matrix $X \in \mathcal{S}_{10}^+$ with Cholesky decomposition $X = LL^\top$, the matrix $\varepsilon(X)$ is also $\mathcal{S}_{10}^+$ and has Cholesky decomposition $\varepsilon(X) = L_\varepsilon L$, implying $\mathbb{E}[\varepsilon(X)] = X$.

Figure 5 shows similar results as in the other two experimental settings. The convergence rate proved in Theorem 4.1 is confirmed in panel (b). In this setting, the stability of the error curves indicates an early attainment of the asymptotic regime. Panel (c) suggests a convergence rate of $\hat{\varphi}_T$ faster than $\sqrt{T}$, but additional simulations for larger sample sizes rejects this conjecture. We observe in panel (d) that the test exposes the correct level and high power already at small sample sizes for all non-zero tested values of φ.

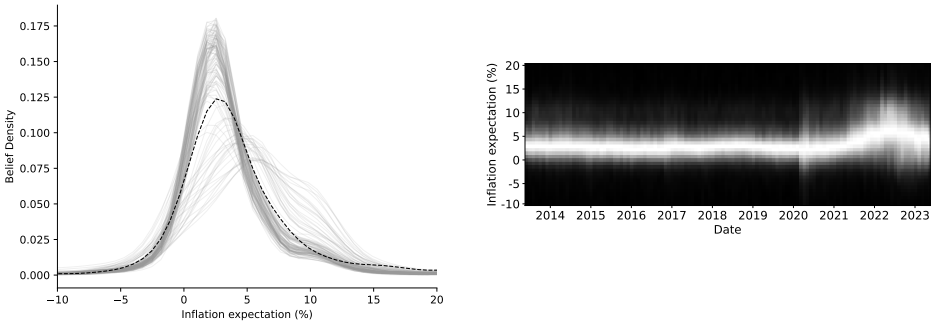


Figure 6: Left: smoothed densities of the monthly 12-months-ahead inflation expectation. The empirical Fréchet mean of these densities is displayed in dashed black. Right: display of the monthly 12-months-ahead inflation expectation densities. In both panels, we only show the $[-10, 20]$ range of the data.

7. Application: Inflation expectation

Analysis of consumer inflation expectations brings insights into how the general populations perceives broader economic trends (Dietrich et al., 2022; Meeks and Monti, 2023). The *Survey of Consumer Expectations* (SCE) is a monthly survey maintained by the Federal Reserve Bank of New York collecting information on households' expectations on a broad variety of economic topics between June 2013 and May 2023, see Armantier et al. (2017).

We focus our attention on the inflation expectation question, in which each consumer is asked to provide a distribution representing their belief for the 12-months ahead inflation. The survey respondents are presented with pre-defined bins over which they can distribute percentage points, defining a histogram of their beliefs. The bins are given by the nodes -36%, -12%, -8%, -4%, -2%, 0%, 2%, 4%, 8%, 12% and 36%. Each month, an average of approximately 1293 response histograms are available. We aggregate the histograms monthly by first taking the median belief of each histogram (which is already present in the dataset) and approximate the monthly median belief density via kernel density estimator with a Gaussian kernel and using Scott's rule (Scott, 1992) for the choice of the bandwidth. This results in a time-series of $T = 114$ elements in $\mathcal{D}([-36, 36])$ displayed in Figure 6.

We fit the parameters of the GAR(1) model as described in Section 4 and obtain an empirical Fréchet mean $\hat{\mu}_T$ displayed in the left panel of Figure 6 (dashed black) and a concentration parameter $\hat{\varphi}_T = 0.85$ indicating a strong sequential dependence of the densities. The hypothesis test presented

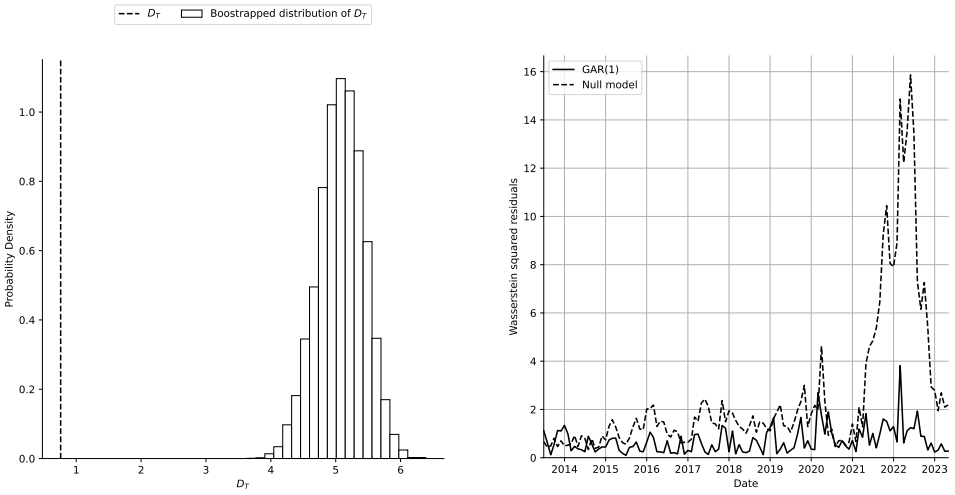


Figure 7: Left: Histogram of the distribution of D_T under the null hypothesis together with the computed value of D_T on the observed data. Right, top: Empirical cumulative distribution function of the residuals for the GAR(1) model (full line) and null model (dashed line). Right, bottom: Residuals of both models over time.

in Section 5 rejects the hypothesis of independence at level at a 5% level with a test statistic $D_T \approx 0.76$ and estimated p-value $\hat{p}_B \approx 10^{-3}$ with $B = 1000$ permutations. The left panel of Figure 7 displays the histogram of the bootstrapped values of D_T, illustrating the approximate normal distribution of D_T under H_0.

To further highlight the auto-regressive aspect of the data, we compare the residuals of the GAR(1) model with those obtained under the null model, $X_t = \varepsilon_t(\mu)$. Using fitted parameters, we generate predictions for each time step under the GAR(1) model as $\hat{X}_{t+1} = \gamma^{X_t}_{\hat{\mu}_T}(\hat{\varphi}_T)$, and under the null model as $\hat{X}^0_{t+1} = \hat{\mu}_T$. We then consider the squared residuals $d(X_{t+1}, \hat{X}_{t+1})^2$ and $d(X_{t+1}, \hat{\mu}_T)^2$, respectively. To asses the model's performance, we compute a metric space adaptation of the coefficient of determination, as proposed by Petersen and Müller (2019). The empirical estimator $R^2_\oplus$ of $R^2_\oplus$ is given by

$$\hat{R}^2_\oplus = 1 - \frac{\sum_{t=1}^{T-1} d(X_{t+1}, \gamma^{X_t}_{\hat{\mu}_T}(\hat{\varphi}_T))^2}{\sum_{t=1}^{T-1} d(X_t, \hat{\mu}_T)^2}.$$

In this analysis, we find an empirical coefficient of determination of $\hat{R}^2_\oplus = 0.72$ indicating that the GAR(1) model is able to explain a significant portion

of the variability present in the data. The right panel of Figure 7 shows that the residuals of the GAR(1) fit are smaller than for the fit under the null model, as shown in the upper graph. Furthermore, as shown in the lower-right panel, the residuals under the null model residuals increase and are high during the years 2021 and 2022 while the GAR(1) residuals stay stable over time. This is consistent with the data as shown in the right panel of Figure 6 where it is visible that there is a shift in inflation expectation from 2021 onwards, possibly due to the economical impact of the COVID 19 pandemic and the escalation of the Russo-Ukranian war in early 2022. While this shift results in unexplained deviations from the mean in the null model starting in 2021, the time dependence of the GAR(1) model allows for a better fit in these years.

8. Discussion

This paper proposed a first-order autoregressive model for time series of random variable residing in metric spaces. The model is parametrized by a Fréchet mean and a concentration parameter which we proved can be consistently estimated under mild assumptions. This paper also presents a test for serial dependence under the GAR(1) model, allowing to test it against a null hypothesis of repeated i.i.d. measurements. Monte Carlo experiments as well as a real data analysis demonstrated the theoretical properties of the model as well as its practical relevance.

Several directions could be taken to extend the results of this paper. First, as observed in the experiments in Section 6, a $\sqrt{T}$ rate of convergence of the estimator of the concentration parameter φ appears to hold. This rate, as well as other stronger results about the asymptotic behavior of $\hat{\varphi}_T$, might be obtained using moment assumptions on the subderivates of L, see Niemiro (1992), or by assuming differentiability, see results from Haberman (1989). On the modeling side, two directions could be interesting to explore. A useful extension would be to adapt the model to allow for a negative relationship to the previous time step, with $\varphi < 0$. This can be done naturally in some specific cases by using an existing tangent space structure of the metric space, as done in Ghodrati and Panaretos (2024); Zhu and Müller (2023), but it is not necessarily clear how to define the notion of a *negative direction* in a more general case. Furthermore, the model presented here only allows for a first-order auto-regressive structure. One possible extension would be to consider higher-order autoregressive models by applying the same principle,

this time replacing the one geodesic update in Equation (6) with multiple updates using previous time steps. However, we expect this approach to be challenging to analyze. Instead, it could be of interest to propose another auto-regressive model sharing the similarity to the AR(1) model on the real line, but which could be more easily extended to a higher number of lags. Finally, one could develop other classical tests found in time series analyses to this model class, for instance a test to detect change-points in the Fréchet mean or concentration parameter, see Jiang et al. (2024).

9. Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 956107, "Economic Policy in Complex Environments (EPOC)".

References

Armantier, O., Topa, G., Van der Klaauw, W., Zafar, B., 2017. An overview of the survey of consumer expectations. Economic Policy Review , 51–72.

Bačák, M., 2014. Computing Medians and Means in Hadamard Spaces. SIAM Journal on Optimization 24, 1542–1566. URL: https://epubs.siam.org/doi/abs/10.1137/140953393, doi:10.1137/140953393. publisher: Society for Industrial and Applied Mathematics.

van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V., 2014. Bayes Hilbert Spaces. Australian & New Zealand Journal of Statistics 56, 171–194. URL: https://onlinelibrary.wiley.com/doi/10.1111/anzs.12074, doi:10.1111/anzs.12074.

Bosq, D., 2000. Linear Processes in Function Spaces. volume 149 of *Lecture Notes in Statistics*. Springer New York, New York, NY. URL: http://link.springer.com/10.1007/978-1-4612-1154-9, doi:10.1007/978-1-4612-1154-9.

Bulté, M., Sørensen, H., 2024. Medoid splits for efficient random forests in metric spaces. Computational Statistics & Data Analysis 198, 107995. URL: https://www.sciencedirect.com/science/article/

pii/S0167947324000793, doi:https://doi.org/10.1016/j.csda.2024.107995.

Burago, D., Burago, Y., Ivanov, S., 2001. A Course in Metric Geometry. volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island. URL: http://www.ams.org/gsm/033, doi:10.1090/gsm/033.

Caner, M., 2006. M-estimators with non-standard rates of convergence and weakly dependent data. Journal of Statistical Planning and Inference 136, 1207–1219. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378375804003647, doi:10.1016/j.jspi.2004.09.004.

Caponera, A., Marinucci, D., 2021. Asymptotics for spherical functional autoregressions. The Annals of Statistics 49. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-49/issue-1/Asymptotics-for-spherical-functional-autoregressions/10.1214/20-AOS1959.full, doi:10.1214/20-AOS1959.

Chen, Y., Lin, Z., Müller, H.G., 2021. Wasserstein Regression. Journal of the American Statistical Association , 1–14URL: https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1956937, doi:10.1080/01621459.2021.1956937.

Dietrich, A.M., Kuester, K., Müller, G.J., Schoenle, R., 2022. News and uncertainty about COVID-19: Survey evidence and short-run economic impact. Journal of Monetary Economics 129, S35–S51. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304393222000216, doi:10.1016/j.jmoneco.2022.02.004.

Dubey, P., Müller, H.G., 2019. Fréchet analysis of variance for random objects. Biometrika 106, 803–821. URL: https://academic.oup.com/biomet/article/106/4/803/5609104, doi:10.1093/biomet/asz052.

Dubey, P., Müller, H.G., 2020. Fréchet change-point detection. The Annals of Statistics 48, 3312–3335. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-48/issue-6/Fr%c3%a9chet-change-point-detection/10.1214/19-AOS1930.full, doi:10.1214/19-AOS1930. publisher: Institute of Mathematical Statistics.

Fréchet, M., 1948. Les éléments aléatoires de nature quelconque dans un espace distancié, in: Annales de l'institut Henri Poincaré, pp. 215–310. Issue: 4.

Ghodrati, L., Panaretos, V.M., 2024. On distributional autoregression and iterated transportation. Journal of Time Series Analysis 45, 739–770. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jtsa.12736, doi:https://doi.org/10.1111/jtsa.12736, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jtsa.12736.

Gordin, M.I., Lifšic, B.A., 1978. The central limit theorem for stationary Markov processes, in: Doklady Akademii Nauk, Russian Academy of Sciences. pp. 766–767. Issue: 4.

Haberman, S.J., 1989. Concavity and Estimation. The Annals of Statistics 17, 1631–1661. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-17/issue-4/Concavity-and-Estimation/10.1214/aos/1176347385.full, doi:10.1214/aos/1176347385. publisher: Institute of Mathematical Statistics.

Hanneke, S., Kontorovich, A., Sabato, S., Weiss, R., 2021. Universal Bayes consistency in metric spaces. The Annals of Statistics 49, 2129–2150. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-49/issue-4/Universal-Bayes-consistency-in-metric-spaces/10.1214/20-AOS2029.full, doi:10.1214/20-AOS2029. publisher: Institute of Mathematical Statistics.

Hemerik, J., Goeman, J., 2018. Exact testing with random permutations. TEST 27, 811–825. URL: https://doi.org/10.1007/s11749-017-0571-1, doi:10.1007/s11749-017-0571-1.

Hoeffding, W., Robbins, H., 1948. The central limit theorem for dependent random variables. Duke Mathematical Journal 15, 773–780. URL: https://projecteuclid.org/journals/duke-mathematical-journal/volume-15/issue-3/The-central-limit-theorem-for-dependent-random-variables/10.1215/S0012-7094-48-01568-3.full, doi:10.1215/S0012-7094-48-01568-3. publisher: Duke University Press.

Hsing, T., Eubank, R., 2015. Theoretical foundations of functional data analysis, with an introduction to linear operators. volume 997. John Wiley & Sons.

Jiang, F., Gao, H., Shao, X., 2023. Testing serial independence of object-valued time series. Biometrika , asad069URL: https://doi.org/10.1093/biomet/asad069, doi:10.1093/biomet/asad069.

Jiang, F., Zhu, C., Shao, X., 2024. Two-sample and change-point inference for non-Euclidean valued time series. Electronic Journal of Statistics 18, 848–894. URL: https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-18/issue-1/Two-sample-and-change-point-inference-for-non-Euclidean-valued/10.1214/24-EJS2218.full, doi:10.1214/24-EJS2218. publisher: Institute of Mathematical Statistics and Bernoulli Society.

Jiang, Y., 2023. Wasserstein multivariate auto-regressive models for modeling distributional time series and its application in graph learning. URL: http://arxiv.org/abs/2207.05442, doi:10.48550/arXiv.2207.05442. arXiv:2207.05442 [cs, stat].

Köstenberger, G., Stark, T., 2023. Robust Signal Recovery in Hadamard Spaces. URL: http://arxiv.org/abs/2307.06057, doi:10.48550/arXiv.2307.06057. arXiv:2307.06057 [math, stat].

Lin, Z., 2019. Riemannian Geometry of Symmetric Positive Definite Matrices via Cholesky Decomposition. SIAM Journal on Matrix Analysis and Applications 40, 1353–1370. URL: https://epubs.siam.org/doi/abs/10.1137/18M1221084, doi:10.1137/18M1221084. publisher: Society for Industrial and Applied Mathematics.

Lyons, R., 2013. Distance covariance in metric spaces. The Annals of Probability 41. URL: https://projecteuclid.org/journals/annals-of-probability/volume-41/issue-5/Distance-covariance-in-metric-spaces/10.1214/12-AOP803.full, doi:10.1214/12-AOP803.

Mazzuco, S., Scarpa, B., 2015. Fitting Age-Specific Fertility Rates by a Flexible Generalized Skew Normal Probability Density Function. Journal

of the Royal Statistical Society Series A: Statistics in Society 178, 187–203. URL: https://doi.org/10.1111/rssa.12053, doi:10.1111/rssa.12053.

McCormack, A., Hoff, P., 2022. The Stein effect for Fréchet means. The Annals of Statistics 50, 3647–3676. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-50/issue-6/The-Stein-effect-for-Fr%c3%a9chet-means/10.1214/22-AOS2245.full, doi:10.1214/22-AOS2245. publisher: Institute of Mathematical Statistics.

McCormack, A., Hoff, P.D., 2023. Equivariant estimation of Fréchet means. Biometrika URL: https://doi.org/10.1093/biomet/asad014, doi:10.1093/biomet/asad014.

Meeks, R., Monti, F., 2023. Heterogeneous beliefs and the Phillips curve. Journal of Monetary Economics URL: https://www.sciencedirect.com/science/article/pii/S0304393223000703, doi:10.1016/j.jmoneco.2023.06.003.

Newey, W.K., 1991. Uniform Convergence in Probability and Stochastic Equicontinuity. Econometrica 59, 1161. URL: https://www.jstor.org/stable/2938179?origin=crossref, doi:10.2307/2938179.

Niemiro, W., 1992. Asymptotics for M-Estimators Defined by Convex Minimization. The Annals of Statistics 20, 1514–1533. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-20/issue-3/Asymptotics-for-M-Estimators-Defined-by-Convex-Minimization/10.1214/aos/1176348782.full, doi:10.1214/aos/1176348782. publisher: Institute of Mathematical Statistics.

Panaretos, V.M., Zemel, Y., 2016. Amplitude and phase variation of point processes. The Annals of Statistics 44. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-44/issue-2/Amplitude-and-phase-variation-of-point-processes/10.1214/15-AOS1387.full, doi:10.1214/15-AOS1387.

Panaretos, V.M., Zemel, Y., 2020. An Invitation to Statistics in Wasserstein Space. SpringerBriefs in Probability and Mathematical Statistics, Springer

International Publishing, Cham. URL: http://link.springer.com/10.1007/978-3-030-38438-8, doi:10.1007/978-3-030-38438-8.

Petersen, A., Müller, H.G., 2019. Fréchet regression for random objects with euclidean predictors. The Annals of Statistics 47, 691–719.

Petersen, A., Müller, H.G., 2016. Functional data analysis for density functions by transformation to a hilbert space. The Annals of Statistics 44, 183–218. URL: http://www.jstor.org/stable/43818904.

Petersen, A., Zhang, C., Kokoszka, P., 2022. Modeling Probability Density Functions as Data Objects. Econometrics and Statistics 21, 159–178. URL: https://linkinghub.elsevier.com/retrieve/pii/S245230622100054X, doi:10.1016/j.ecosta.2021.04.004.

Raginsky, M., Sason, I., 2013. Concentration of Measure Inequalities in Information Theory, Communications, and Coding. Foundations and Trends in Communications and Information Theory 10, 1–246. URL: https://www.nowpublishers.com/article/Details/CIT-064, doi:10.1561/0100000064. publisher: Now Publishers, Inc.

Ramsay, J.O., Silverman, B.W., 2005. Functional data analysis. Springer series in statistics. 2nd ed ed., Springer, New York.

Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley. URL: http://dx.doi.org/10.1002/9780470316849, doi:10.1002/9780470316849. iSSN: 1940-6347 Publication Title: Wiley Series in Probability and Statistics.

Srivastava, A., Klassen, E.P., 2016. Functional and Shape Data Analysis. Springer Series in Statistics, Springer New York, New York, NY. URL: http://link.springer.com/10.1007/978-1-4939-4020-2, doi:10.1007/978-1-4939-4020-2.

Sturm, K.T., 2003. Probability measures on metric spaces of nonpositive curvature, in: Auscher, P., Coulhon, T., Grigor'yan, A. (Eds.), Contemporary Mathematics. American Mathematical Society, Providence, Rhode Island. volume 338, pp. 357–390. URL: http://www.ams.org/conm/338/, doi:10.1090/conm/338/06080.

van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer Series in Statistics, Springer New York, New York, NY. URL: http://link.springer.com/10.1007/978-1-4757-2545-2, doi:10.1007/978-1-4757-2545-2.

Wu, W.B., Shao, X., 2004. Limit theorems for iterated random functions. Journal of Applied Probability 41, 425–436. URL: https://www.cambridge.org/core/journals/journal-of-applied-probability/article/abs/limit-theorems-for-iterated-random-functions/9CA522B1DE090C02B7374ED8212739AB#, doi:10.1239/jap/1082999076. publisher: Cambridge University Press.

Xavier, J., Manton, J., 2006. On the Generalization of AR Processes To Riemannian Manifolds, in: 2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings, IEEE, Toulouse, France. pp. V–1005–V–1008. URL: http://ieeexplore.ieee.org/document/1661448/, doi:10.1109/ICASSP.2006.1661448.

Yoshiyuki, A., 2017. A Functional Linear Regression Model in the Space of Probability Density Functions. Discussion papers URL: https://ideas.repec.org//p/eti/dpaper/17015.html. number: 17015 Publisher: Research Institute of Economy, Trade and Industry (RIETI).

Zhu, C., Müller, H.G., 2023. Autoregressive optimal transport models. Journal of the Royal Statistical Society Series B: Statistical Methodology 85, 1012–1033. URL: https://doi.org/10.1093/jrsssb/qkad051, doi:10.1093/jrsssb/qkad051.

Zhu, C., Müller, H.G., 2024. Spherical autoregressive models, with application to distributional and compositional time series. Journal of Econometrics 239, 105389. URL: https://www.sciencedirect.com/science/article/pii/S0304407623000209, doi:https://doi.org/10.1016/j.jeconom.2022.12.008.

Appendix A. Comparison to other approaches

In this section, we compare the test presented in this paper to the serial independence test proposed in Jiang et al. (2023) as well as a test for serial independence constructed by testing $\varphi = 0$ via the estimator $\hat{\varphi}_T$. We compare the rejection rate of the different tests at level $\alpha = 0.05$ for different simulation setups, values of φ and sample sizes in the same three scenarios as in Section 6: $\mathbb{R}$ with multiplicative noise, univariate distributions with the 2-Wasserstein distance and SPD matrices with the Log-Cholesky metric.

The first alternative test uses $\hat{\varphi}_T$ as the test statistic. Similarly to the test constructed via the statistic D_T, to construct a level α hypothesis test for $H_0 : \varphi = 0$ vs. $H_1 : \varphi > 0$, we use a permutation procedure to compute approximate p-values under H_0. With the same notation as in Section 5, let $\hat{\varphi}_T^\pi$ be the estimator of φ computed on the randomly permuted sample $(X_{\pi(1)}, \dots, X_{\pi(T)})$. The approximate p-value using a bootstrapped sample of $B \in \mathbb{N}$ replicas is $\hat{p}_B = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{\hat{\varphi}_T \leq \hat{\varphi}_T^\pi\}$. We reject the null hypothesis if $\hat{p}_B \leq \alpha$ to construct a level α test. While we do not have theoretical results about the asymptotic distribution of $\hat{\varphi}_T$ justifying this test, it is still of interest to empirically analyze the result of this procedure.

We also compare our test to the methodology presented in Jiang et al. (2023). The authors propose a generalization of the spectral density function to metric spaces based on the distance covariance, see Lyons (2013). Based on this, the define two test statistics, CvM_T and KS_T. Similarly to our D_T, their test statistics are non-pivotal and a wild bootstrap is proposed to obtain the critical values of their test statistics.

The summary of the results can be found in Table A.1. We observe that the tests based on D_T or $\hat{\varphi}_T$ seem to almost uniformly outperform the tests proposed by Jiang et al. (2023). This could partially be explained by the fact that our tests are tailored for the data generating process considered, while the tests in Jiang et al. (2023) do not assume any structure on the autoregressive data generating process. We observe that the tests based on $\hat{\varphi}_T$ and D_T expose approximately similar performance in the first scenario of real numbers with multiplicative noise. For the other two scenarios, the two tests seem to be equally well calibrated, up to stochastic difference, while the test based on D_T achieves considerably higher power compared uniformly over every alternative and sample size we considered.

			CvM_T	KS_T	$\hat{\varphi}_T$	D_T
$\mathbb{R}$	$H_0 : \varphi = 0$	$n = 40$	0.112	0.090	0.058	**0.050**
		$n = 80$	0.076	0.076	**0.052**	0.034
		$n = 160$	0.090	0.074	0.044	**0.048**
	$\varphi = 0.1$	$n = 40$	0.120	0.106	**0.166**	0.142
		$n = 80$	0.160	0.106	0.228	**0.236**
		$n = 160$	0.144	0.066	**0.312**	**0.312**
	$\varphi = 0.3$	$n = 40$	0.238	0.108	**0.492**	0.478
		$n = 80$	0.434	0.116	0.800	**0.806**
		$n = 160$	0.664	0.118	0.936	**0.972**
$\mathcal{D}$	$H_0 : \varphi = 0$	$n = 40$	0.156	0.092	**0.056**	0.036
		$n = 80$	0.158	0.092	**0.042**	0.040
		$n = 160$	0.146	0.114	**0.048**	0.044
	$\varphi = 0.1$	$n = 40$	**0.206**	0.114	0.132	0.180
		$n = 80$	0.218	0.096	0.186	**0.276**
		$n = 160$	0.248	0.098	0.150	**0.530**
	$\varphi = 0.3$	$n = 40$	0.336	0.138	0.596	**0.798**
		$n = 80$	0.446	0.146	0.732	**0.982**
		$n = 160$	0.616	0.122	0.818	**1**
$\mathcal{S}_{10}^{+}$	$H_0 : \varphi = 0$	$n = 40$	0.850	0.344	**0.012**	**0.012**
		$n = 80$	0.966	0.426	**0.018**	0.014
		$n = 160$	0.986	0.430	0.016	**0.042**
	$\varphi = 0.1$	$n = 40$	0.882	0.320	0.524	**0.984**
		$n = 80$	0.978	0.368	0.618	**1**
		$n = 160$	0.980	0.444	0.654	**1**
	$\varphi = 0.3$	$n = 40$	0.824	0.332	**1**	**1**
		$n = 80$	0.980	0.390	**1**	**1**
		$n = 160$	0.992	0.396	**1**	**1**

Table A.1: Empirical rejection rate for each test at level $\alpha = 0.05$. Each rejection rate is based on 1000 simulations, as described in the introduction to Section 6. In each line, the number in bold corresponds to the rejection rate closest to the desired test level $\alpha = 0.05$ for the rows with $\varphi = 0$ and with the highest rejection rate for the rows with $\varphi > 0$.

Appendix B. Proofs of theorems

Consistency of the mean estimator

We start by defining the following function which will be useful in the proofs presented in this section. Let $\omega, \omega_0 \in \Omega$, we define for all $x \in \Omega$ the function

$$g^{\omega}_{\omega_0}(x) = d(x,\omega)^2 - d(x,\omega_0)^2.$$

In a Hadamard space, $g^{\omega}_{\omega_0}$ has the following Lipschitz property holding both in x and in the pair (ω, ω_0).

Lemma B.1. *Let (Ω, d) be a Hadamard space and $\omega, \omega_0, x, x' \in \Omega$, then*

$$\left|g^{\omega}_{\omega_0}(x) - g^{\omega}_{\omega_0}(x')\right| \leq 2d(\omega,\omega_0)d(x,x').$$

Proof. By Reshetnyak's Quadruple Comparison (see Proposition C.1),

$$\begin{aligned}
&d(x',\omega_0)^2 + d(x,\omega)^2 \leq d(x,\omega_0)^2 + d(x',\omega)^2 + 2d(x,x')d(\omega,\omega_0)\\
&\Rightarrow d(x,\omega)^2 - d(x',\omega)^2 - \left(d(x,\omega_0)^2 - d(x',\omega_0)^2\right) \leq 2d(x,x')d(\omega,\omega_0)\\
&\Rightarrow g^{\omega}_{\omega_0}(x) - g^{\omega}_{\omega_0}(x') \leq 2d(\omega,\omega_0)d(x,x')
\end{aligned}$$

By inverting the role of x and x' we obtain the same upper bound on $g^{\omega}_{\omega_0}(x') - g^{\omega}_{\omega_0}(x)$ which completes the proof. □

In order to study the asymptotic behavior of the minimizer $\hat{\mu}_T$ of M_T, we need to quantify the deviations of the empirical loss function M_T from its population limit M. To that end, given some $\omega_0 \in \Omega$, we define the process $\omega \mapsto H^{\omega}_{\omega_0}$ by

$$H^{\omega}_{\omega_0} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g^{\omega}_{\omega_0}(X_t) - \mathbb{E}\left[g^{\omega}_{\omega_0}(X)\right]. \tag{B.1}$$

We start with the following proposition showing that $H^{\omega}_{\omega_0}$ is sub-Gaussian.

Proposition B.1. *Under the assumptions of Theorem 4.1, there exists constants $c_1, c_2 > 0$ such that for all $\lambda > 0$,*

$$\mathbb{P}\left[\left|H^{\omega}_{\omega_0}\right| \geq \lambda\right] \leq c_1 \exp\left\{-\frac{\lambda^2}{c_2 d(\omega,\omega_0)^2}\right\}. \tag{B.2}$$

Proof. Following Wu and Shao (2004) and Gordin and Lifšic (1978), we study the asymptotic behavior of the scaled process $\sqrt{T}H^{\omega}_{\omega_0}$ by considering the solution $h \in L_2(\Omega)$ to Poisson's equation

$$h(x) - \mathbb{E}[h(X_{t+1}) \mid X_t = x] = g^{\omega}_{\omega_0}(x) - \mathbb{E}\big[g^{\omega}_{\omega_0}(X)\big]. \tag{B.3}$$

A solution to this equation exists and can be written as

$$h(x) = \sum_{t=0}^{\infty} \big(\mathbb{E}\big[g^{\omega}_{\omega_0}(X_t) \mid X_0 = x\big] - \mathbb{E}\big[g^{\omega}_{\omega_0}(X)\big]\big).$$

Using (B.3), we can decompose $\sqrt{T}H^{\omega}_{\omega_0}$ as

$$\begin{aligned}
\sqrt{T}H^{\omega}_{\omega_0} &= \sum_{t=1}^{T} g^{\omega}_{\omega_0}(X_t) - \mathbb{E}\big[g^{\omega}_{\omega_0}(X)\big] = \sum_{t=1}^{T} h(X_t) - \mathbb{E}[h(X_{t+1}) \mid X_t] \\
&= \mathbb{E}[h(X_1) \mid X_0] - \mathbb{E}[h(X_{T+1}) \mid X_T] + \sum_{t=1}^{T} h(X_t) - \mathbb{E}[h(X_t) \mid X_{t-1}] \\
&:= R_T + \sum_{t=1}^{T} D_t,
\end{aligned}$$

where we introduced $R_T = \mathbb{E}[h(X_1) \mid X_0] - \mathbb{E}[h(X_{T+1}) \mid X_T]$ and $D_t = h(X_t) - \mathbb{E}[h(X_t) \mid X_{t-1}]$. Note that D_t is a martingale-difference and that $R_T = O_P(1)$. To show that this decomposition is valid, we start by showing that h is absolutely summable. That is, we show

$$\sum_{t=0}^{\infty} \big|\mathbb{E}\big[g^{\omega}_{\omega_0}(X_t) \mid X_0 = x\big] - \mathbb{E}\big[g^{\omega}_{\omega_0}(X)\big]\big| < \infty.$$

Using Lemma B.1 and the independence of $\{\varepsilon_t\}_{t\in\mathbb{N}}$, along with the assumption that condition (8) holds for some $\alpha > 1$ (and hence also for $\alpha = 1$), we have

$$\begin{aligned}
\big|\mathbb{E}\big[g^{\omega}_{\omega_0}(X_t) \mid X_0 = x\big] - \mathbb{E}\big[g^{\omega}_{\omega_0}(X)\big]\big| &\leq 2d(\omega, \omega_0)\mathbb{E}_{X_t}[\mathbb{E}_X[d(X_t, X)] \mid X_0 = x] \\
&= 2d(\omega, \omega_0)\mathbb{E}\Big[d(X_t(x), X_t(\tilde{X}_0))\Big] \leq 2d(\omega, \omega_0)Cr^t.
\end{aligned}$$

Using this bound in the infinite sum, we obtain

$$\sum_{t=0}^{\infty} \big|\mathbb{E}\big[g^{\omega}_{\omega_0}(X_t) \mid X_0 = x\big] - \mathbb{E}\big[g^{\omega}_{\omega_0}(X)\big]\big| \leq 2d(\omega, \omega_0)C \sum_{t=0}^{\infty} r^t = \tilde{C}d(\omega, \omega_0).$$

Hence, $\lim_{T\to\infty}\sqrt{T}H^{\omega}_{\omega_0}$ exists and is almost surely bounded implying that the decomposition presented above is valid and that both D_t and R_t are also almost surely bounded with $|D_t| < C$ and $|R_T| < C$. In particular, we have that $|D_t| \leq \tilde{C}d(\omega,\omega_0)$ for some $\tilde{C} > 0$, since

$$\begin{aligned}
D_t &= h(X_t) - \mathbb{E}[h(X_t) \mid X_{t-1}] \\
&= \sum_{k=0}^{\infty} \mathbb{E}\big[g^{\omega}_{\omega_0}(X_k) \mid X_t\big] - \mathbb{E}\big[\mathbb{E}\big[g^{\omega}_{\omega_0}(X_k) \mid X_t\big] \mid X_{t-1}\big] \\
&= \sum_{k=0}^{\infty} \mathbb{E}\big[g^{\omega}_{\omega_0}(X_k) \mid X_t\big] - \mathbb{E}\big[g^{\omega}_{\omega_0}(X_k) \mid X_{t-1}\big] \qquad \text{(Tower rule)} \\
&= \sum_{k=t}^{\infty} \mathbb{E}\big[g^{\omega}_{\omega_0}(X_k) \mid X_t\big] - \mathbb{E}\big[g^{\omega}_{\omega_0}(X_k) \mid X_{t-1}\big]. \qquad (k \leq t-1 \Rightarrow \sigma(X_k) \subset \sigma(X_{t-1}))
\end{aligned}$$

We can use the function notation in Equation (7) to get $X_k = X_{k:t+1}(X_t)$ and rewrite

$$\begin{aligned}
&\mathbb{E}\big[g^{\omega}_{\omega_0}(X_k) \mid X_t\big] - \mathbb{E}\big[g^{\omega}_{\omega_0}(X_k) \mid X_{t-1}\big] \\
&= \mathbb{E}\big[g^{\omega}_{\omega_0}(X_{k:t+1}(X_t)) \mid X_t\big] - \mathbb{E}\big[g^{\omega}_{\omega_0}(X_{k:t+1}(X_t)) \mid X_{t-1}\big]
\end{aligned}$$

where $X_{k:t+1}$ is random in both conditional expectations, but X_t is only random in the second conditional expectation. Taking absolute values and using Lemma B.1 together with (2) in Wu and Shao (2004) gives

$$\big|\mathbb{E}\big[g^{\omega}_{\omega_0}(X_k) \mid X_t\big] - \mathbb{E}\big[g^{\omega}_{\omega_0}(X_k) \mid X_{t-1}\big]\big| \leq 2d(\omega,\omega_0)Cr^{k-(t+1)}.$$

Using this bound in the sum gives

$$|D_t| \leq \sum_{k=t}^{\infty} 2d(\omega,\omega_0)Cr^{k-(t+1)} = 2Cd(\omega,\omega_0)\sum_{k=0}^{\infty} r^k = \tilde{C}d(\omega,\omega_0).$$

We now show the following result, which is equivalent to (B.2)

$$\mathbb{P}\left[\left|R_T + \sum_{t=1}^{T} D_t\right| \geq T\lambda\right] \leq e^2 \exp\left\{-\frac{T\lambda^2}{4Cd(\omega,\omega_0)^2}\right\}. \tag{B.4}$$

To do this, we consider two cases. If $\sqrt{T}\lambda < 4C$, the bound is vacuous. Otherwise, for $\sqrt{T}\lambda \geq 4C$, we have

$$\mathbb{P}\left[\left|R_T + \sum_{t=1}^{T} D_t\right| \geq T\lambda\right] \leq \mathbb{P}\left[|R_T| \geq \frac{T\lambda}{4}\right] + \mathbb{P}\left[\left|\sum_{t=1}^{T} D_t\right| \geq \frac{3T\lambda}{4}\right].$$

Since $\sqrt{T}\lambda \geq 4C$ and $|R_t| \leq C$, we have that $\left\{|R_T| \geq \frac{T\lambda}{4}\right\}$ is a probability zero event. We can thus focus on bounding the martingal difference sum. By Chernoff's bounding technique, we have for all $\lambda > 0$

$$\begin{aligned}\mathbb{P}\left[\sum_{t=1}^{T} D_t \geq \frac{3T\lambda}{4}\right] &= \mathbb{P}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} D_t \geq \frac{3\sqrt{T}\lambda}{4}\right] \\ &\leq \min_{u>0} \exp\left(-u\frac{3\sqrt{T}\lambda}{4}\right)\mathbb{E}\left[\exp\left(u\frac{1}{\sqrt{T}}\sum_{t=1}^{T} D_t\right)\right].\end{aligned}$$

Using the bound $|D_t| \leq \tilde{C}d(\omega, \omega_0)$ and following the proof of Azuma-Hoeffding's inequality (see Theorem 2.2.1 in Raginsky and Sason (2013)) we can bound the martingale-difference sum,

$$\begin{aligned}\mathbb{P}\left[\sum_{t=1}^{T} D_t \geq \frac{3T\lambda}{4}\right] &\leq \min_{u>0} \exp\left(-u\frac{3\sqrt{T}\lambda}{4} + \frac{u^2}{2}\tilde{C}^2 d(\omega, \omega_0)^2\right) \\ &\leq \exp\left(-\frac{\lambda^2 T}{4\tilde{C}^2 d(\omega, \omega_0)^2}\right).\end{aligned}$$

Where the last inequality comes from taking $u = \lambda\sqrt{T}/(\tilde{C}^2 d(\omega, \omega_0)^2)$. By symmetry the arguments can be repeated on the mirrored inequality to obtain a bound on $\mathbb{P}\left[\left|\sum_{t=1}^{T} D_t\right| \geq \frac{3T\lambda}{4}\right]$. Together with the previous argument, this gives

$$\mathbb{P}\left[\left|R_T + \sum_{t=1}^{T} D_t\right| \geq T\lambda\right] \leq e^2 \exp\left(-\frac{\lambda^2 T}{4\tilde{C}^2 d(\omega, \omega_0)^2}\right).$$

□

Using the fact that the empirical process applied to $g^{\omega}_{\omega_0}$ is sub-Gaussian, we can use standard M-estimation theory to provide a proof of Theorem 4.1.

Proof of Theorem 4.1. Noting that $(M_T - M)(\omega) - (M_T - M)(\mu) = T^{-1/2} H^{\omega}_{\mu}$ we have by Proposition B.1 that

$$\mathbb{P}\left[\sqrt{T}|(M_T - M)(\omega) - (M_T - M)(\mu)| \geq \lambda\right] \leq c_1 \exp\left\{-\frac{\lambda^2}{c_2 d(\omega, \omega_0)^2}\right\}.$$

So $\left\{\sqrt{T}(M_T - M)(\omega)\right\}_{\omega\in\Omega}$ is sub-Gaussian. By Corollary 2.2.8 in van der Vaart and Wellner (1996), we have

$$\mathbb{E}\left[\sup_{d(\omega,\mu)<\delta}\sqrt{T}|(M_T - M)(\omega) - (M_T - M)(\mu)|\right] \lesssim \int_0^\delta \sqrt{\log(1 + D(\varepsilon, d))}\mathrm{d}\varepsilon$$
$$= \delta\int_0^1 \sqrt{\log(1 + D(\delta\varepsilon, d))}\mathrm{d}\varepsilon.$$

Since by assumption the entropy integral is bounded and $O(1)$ for $\delta \to 0$, we bound (up to a multiplicative constant) the modulus of continuity by $T^{-1/2}\delta$. Additionally, by the variance inequality in Hadamard spaces (see Proposition 4.4 in Sturm (2003)), we have that the condition $M(\omega) - M(\mu^\star) \geq d(\omega, \mu^\star)^2$ holds. Thus by Theorem 3.2.5 in van der Vaart and Wellner (1996), $d(\hat{\mu}_T, \mu^\star) = O_P(T^{-1/2})$. □

Uniform convergence of L_T

Proposition B.2. *Under the conditions of Theorem 4.2, we have that* $\|L_T - L\|_\infty = o_P(1)$.

Proof. We show this result by verifying the conditions of Corollary 2.2 of Newey (1991). Namely, we need to show that:

1. L is continuous;
2. L_T converges pointwise to L;
3. There exists a sequence $C_t = O_P(1)$ such that for all $\varphi, \varphi' \in (0, 1)$, $|L_T(\varphi) - L_T(\varphi')| \leq C_T|\varphi - \varphi'|$.

We proceed to verify these conditions.

1. Continuity of L. By definition $\omega \mapsto d(\omega_0, \omega)^2$ is continous. Since Ω is a Hadamard space, we also have that geodesics are continuous in $t \in [0, 1]$, hence for all $x_t, x_{t+1} \in \Omega$ and $\varphi_0 \in [0, 1]$, $d(x_{t+1}, \gamma_\mu^{x_t}\varphi_n) \to d(x_{t+1}, \gamma_\mu^{x_t}\varphi_0)$ for any sequence $\varphi_n \to \varphi_0$. Furthermore, by geodesic convexity of the squared distance, we have that $d(x_{t+1}, \gamma_\mu^{x_t}\varphi)^2 \leq (1 - \varphi)d(x_{t+1}, \mu)^2 + \varphi d(x_{t+1}, x_t)^2$ which is integrable with respect to (X_t, X_{t+1}) since X_t and X_{t+1} have second moments. By dominated convergence, this shows that $L(\varphi) \to L(\varphi_0)$ as $\varphi_n \to \varphi_0$ for a ny sequence $\varphi_n \to \varphi_0$, and hence L is continuous.

2. Pointwise convergence. Let $\varphi_0 \in (0, 1)$. Using the fact that $X_{t+1} = \varepsilon_{t+1}(\gamma_\mu^{X_t}(\varphi))$, we can decompose the pointwise deviation of L_T from L as

follows,

$$
\begin{aligned}
&|L_T(\varphi_0) - L(\varphi_0)| \\
&= \left| \frac{1}{T-1} \sum_{t=1}^{T-1} d\big(X_{t+1}, \gamma_{\hat{\mu}_T}^{X_t}(\varphi_0)\big)^2 - \mathbb{E}\Big[d\big(\varepsilon_{t+1}(\gamma_{\mu}^{X_t}(\varphi)), \gamma_{\mu}^{X_t}(\varphi_0)\big)^2 \Big] \right| \\
&\leq \left| \frac{1}{T-1} \sum_{t=1}^{T-1} d\big(X_{t+1}, \gamma_{\hat{\mu}_T}^{X_t}(\varphi_0)\big)^2 - d\big(X_{t+1}, \gamma_{\mu}^{X_t}(\varphi_0)\big)^2 \right| \\
&\qquad + \left| \frac{1}{T-1} \sum_{t=1}^{T-1} d\big(X_{t+1}, \gamma_{\mu}^{X_t}(\varphi_0)\big)^2 - \mathbb{E}\Big[d\big(\varepsilon_{t+1}(\gamma_{\mu}^{X_t}(\varphi)), \gamma_{\mu}^{X_t}(\varphi_0)\big)^2 \Big] \right|.
\end{aligned}
$$

By Lipschitz continuity of the squared distance in a bounded metric space, together with its geodesic convexity in Hadamard spaces and the fact that $d(\mu, \hat{\mu}_T) = O_P(T^{-1/2})$, we have that the first sum in the upper bound is $O_P(T^{-1/2})$.

$$
\begin{aligned}
&\left| d\big(X_{t+1}, \gamma_{\hat{\mu}_T}^{X_t}(\varphi_0)\big)^2 - d\big(X_{t+1}, \gamma_{\mu}^{X_t}(\varphi_0)\big)^2 \right| \\
&\leq C_1 d\big(\gamma_{\hat{\mu}_T}^{X_t}(\varphi_0), \gamma_{\mu}^{X_t}(\varphi_0)\big) \left| d\big(X_{t+1}, \gamma_{\hat{\mu}_T}^{X_t}(\varphi_0)\big) + d\big(X_{t+1}, \gamma_{\mu}^{X_t}(\varphi_0)\big) \right|.
\end{aligned}
$$

By the geodesic comparison inequality, $d\big(\gamma_{\hat{\mu}_T}^{X_t}(\varphi_0), \gamma_{\mu}^{X_t}(\varphi_0)\big) \leq \varphi_0 d(\mu, \hat{\mu}_T)$, and using that $x \mapsto d(x_0, x)$ is geodesically convex, we get

$$
\begin{aligned}
d(X_{t+1}, \gamma_{\hat{\mu}_T}^{X_t}(\varphi_0)) &\leq \varphi_0 d(X_{t+1}, \hat{\mu}_T) + (1 - \varphi_0) d(X_{t+1}, X_t) \\
&\leq d(X_{t+1}, \hat{\mu}_T) + d(X_{t+1}, X_t) \\
&\leq 2d(X_{t+1}, \mu) + d(X_t, \mu) + d(\mu, \hat{\mu}_T).
\end{aligned}
$$

Similarly, $d\big(X_{t+1}, \gamma_{\mu}^{X_t}(\varphi_0)\big) \leq 2d(X_{t+1}, \mu) + d(X_t, \mu)$, giving

$$
\begin{aligned}
&\left| d\big(X_{t+1}, \gamma_{\hat{\mu}_T}^{X_t}(\varphi_0)\big)^2 - d\big(X_{t+1}, \gamma_{\mu}^{X_t}(\varphi_0)\big)^2 \right| \\
&\leq C_2 d(\mu, \hat{\mu}_T)[d(X_{t+1}, \mu) + d(X_t, \mu) + d(\mu, \hat{\mu}_T)]
\end{aligned}
$$

Taking the average over $t = 1, \ldots, T-1$, we get

$$
\begin{aligned}
&\left| \frac{1}{T-1} \sum_{t=1}^{T-1} d\big(X_{t+1}, \gamma_{\hat{\mu}_T}^{X_t}(\varphi_0)\big)^2 - d\big(X_{t+1}, \gamma_{\mu}^{X_t}(\varphi_0)\big)^2 \right| \\
&\dot{\leq} d(\mu, \hat{\mu}_T)^2 + d(\mu, \hat{\mu}_T) \frac{1}{T-1} \sum_{t=1}^{T-1} d(X_{t+1}, \mu) + d(X_t, \mu).
\end{aligned}
$$

We now show that the second term is $O_P(T^{-1/2})$ as well. We do this using Theorem 3 in Wu and Shao (2004) with $Y_t = (X_t, X_{t+1})$ and $g(X_t, X_{t+1}) = d\big(X_{t+1}, \gamma_\mu^{X_t}(\varphi_0)\big)^2 - \mathbb{E}\Big[d\big(\varepsilon_{t+1}(\gamma_\mu^{X_t}(\varphi)), \gamma_\mu^{X_t}(\varphi_0)\big)^2\Big]$. Let ρ be the product metric on $\Omega \times \Omega$, $\rho((x_1, x_2), (y_1, y_2)) = \sqrt{d(x_1, y_1)^2 + d(x_2, y_2)^2}$. Let $Y_t = (X_t, X_{t+1})$ and $\tilde{Y}_t = (\tilde{X}_t, \tilde{X}_{t+1})$ be pairs in $\Omega \times \Omega$ such that $\rho(Y_t, \tilde{Y}_t) \leq \delta$, then

$$
\begin{aligned}
\Big|g(Y_t) - g(\tilde{Y}_t)\Big| &= \Big|d\big(X_{t+1}, \gamma_\mu^{X_t}(\varphi_0)\big)^2 - d\Big(\tilde{X}_{t+1}, \gamma_\mu^{\tilde{X}_t}(\varphi_0)\Big)^2\Big| \\
&\leq \Big|d\big(X_{t+1}, \gamma_\mu^{X_t}(\varphi_0)\big)^2 - d\Big(\tilde{X}_{t+1}, \gamma_\mu^{X_t}(\varphi_0)\Big)^2\Big| \\
&\qquad + \Big|d\Big(\tilde{X}_{t+1}, \gamma_\mu^{X_t}(\varphi_0)\Big)^2 - d\Big(\tilde{X}_{t+1}, \gamma_\mu^{\tilde{X}_t}(\varphi_0)\Big)^2\Big| \\
&\leq C d(X_{t+1}, \tilde{X}_{t+1}) + C d\Big(\gamma_\mu^{X_t}(\varphi_0), \gamma_\mu^{\tilde{X}_t}(\varphi_0)\Big) \\
&= C d(X_{t+1}, \tilde{X}_{t+1}) + C\varphi_0 d(X_t, \tilde{X}_t)
\end{aligned}
$$

Since $\rho(Y_t, \tilde{Y}_t) \leq \delta$, we have that $\max\Big\{d(X_{t+1}, \tilde{X}_{t+1}), d(X_t, \tilde{X}_t)\Big\} \leq \delta$ and hence $\Big|g(Y_t) - g(\tilde{Y}_t)\Big| \leq C\delta$, showing that g is Dini continuous and also stochastically Dini continuous. Theorem 3 in Wu and Shao (2004) gives that the second term in the above equation converges to a Brownian motion when scaled by $\sqrt{T}$ and hence is $O_P(T^{-1/2})$ which completes the proof of pointwise convergence.

3. Stochastic Lipschitz Continuity of L_T. Let $\varphi, \varphi' \in (0, 1)$, then using that Ω is bounded and thus the squared distance is Lipschitz, we have that

$$
\begin{aligned}
|L_T(\varphi) - L_T(\varphi')| &\leq \frac{1}{T-1}\sum_{t=1}^{T-1}\Big|d\big(X_{t+1}, \gamma_{\hat{\mu}_T}^{X_t}(\varphi)\big)^2 - d\big(X_{t+1}, \gamma_{\hat{\mu}_T}^{X_t}(\varphi')\big)^2\Big| \\
&\leq C\frac{1}{T-1}\sum_{t=1}^{T-1} d(\gamma_{\hat{\mu}_T}^{X_t}(\varphi), \gamma_{\hat{\mu}_T}^{X_t}(\varphi')) \\
&= |\varphi - \varphi'| C\frac{1}{T-1}\sum_{t=1}^{T-1} d(\hat{\mu}_T, X_t).
\end{aligned}
$$

Again using that Ω is bounded, the average is also bounded and we obtain the desired result. □

Appendix C. Auxiliary theoretical results

General results in Hadamard spaces

We start by stating results available in Hadamard spaces that will be used in the rest of the Appendix.

Proposition C.1 (Reshetnyak's Quadruple Comparison; Proposition 2.4 in Sturm (2003))**.** *Let (Ω, d) be a Hadamard space. For all $x_1, x_2, x_3, x_4 \in \Omega$,*

$$d(x_1, x_3)^2 + d(x_2, x_4)^2 \leq d(x_2, x_3)^2 + d(x_4, x_1)^2 + 2d(x_1, x_2)d(x_3, x_4).$$

Specializing this inequality to geodesics yields the following

Proposition C.2 (Geodesic Comparison Inequality; Corollary 2.5 in Sturm (2003))**.** *Let (Ω, d) be a Hadamard space, $\gamma, \eta : [0, 1] \to \Omega$ be geodesics and $t \in [0, 1]$. Then*

$$\begin{aligned} d(\gamma(t), \eta(t))^2 \leq &(1-t)d(\gamma(0), \eta(0))^2 + td(\gamma(1), \eta(1))^2 \\ &- t(1-t)[d(\gamma(0), \gamma(1) - d(\eta(0), \eta(1))]^2. \end{aligned}$$

Strong convexity of L

It is possible to extend the identifiability result in Theorem 3.3 and show that L does not only have a unique minimizer, but is also strongly convex. This, as we show in the following lemma is a consequence of the geodesic convexity of the squared distance in Hadamard spaces.

Lemma C.1. *Let $\{X_t\}_{t\in\mathbb{N}} \subset \Omega$, assume that $\{X_t\}_{t\in\mathbb{N}}$ are $L^2(\Omega)$ and satisfies Equation (6) with true concentration parameter $\varphi \in [0, 1]$. Then, the function L is strongly convex.*

Proof. We show that the strong convexity of L is inherited from the geodesic convexity of the squared distance in Hadamard spaces. Indeed, let $\varphi_1, \varphi_2 \in [0, 1]$, wlog $\varphi_1 < \varphi_2$. Let $t \in [0, 1]$ and define $\varphi_t = (1-t)\varphi_1 + t\varphi_2$. Then, $\gamma_\mu^{X_t}$ restricted to $[\varphi_1, \varphi_2]$ and reparametrized on $[0, 1]$ gives the geodesic connecting $\gamma_\mu^{X_t}(\varphi_1)$ to $\gamma_\mu^{X_t}(\varphi_2)$ and hence

$$L(\varphi_t) = \mathbb{E}\Big[d\big(X_{t+1}, \gamma_\mu^{X_t}((1-t)\varphi_1 + t\varphi_2)\big)^2\Big] = \mathbb{E}\left[d\left(X_{t+1}, \gamma_{\gamma_\mu^{X_t}(\varphi_1)}^{\gamma_\mu^{X_t}(\varphi_2)}(t)\right)^2\right].$$

Using Proposition 2.1 we get

$$\begin{aligned} L(\varphi_t) &\leq \mathbb{E}\big[(1-t)d(X_{t+1}, \gamma_\mu^{X_t}(\varphi_1))^2 + td(X_{t+1}, \gamma_\mu^{X_t}(\varphi_2))^2\big] \\ &\quad - \mathbb{E}\big[t(1-t)d(\gamma_\mu^{X_t}(\varphi_1), \gamma_\mu^{X_t}(\varphi_2))^2\big] \\ &= (1-t)L(\varphi_1) + tL(\varphi_2) - t(1-t)|\varphi_1 - \varphi_2|^2\mathbb{E}\big[d(X_{t+1}, \mu)^2\big]. \end{aligned}$$

Since X_{t+1} is $L^2(\Omega)$, we have that $\mathbb{E}[d(X_{t+1}, \mu)^2] < \infty$, showing that L is strongly convex. □

3.2 Notes

A convergence rate for the concentration parameter

Between the submission of [**GAR**] and the writing of this thesis, we were able to derive the convergence rate of the estimator $\hat{\varphi}_T$ of the concentration parameter φ. We provide the proof here for completeness.

Theorem 1. *Assume that (Ω, d) and $\{X_t\}_{t\in\mathbb{N}}$ satisfy the assumptions of Theorem 4.1 in [Bulté and Sørensen, 2024a]. Then, the minimizer $\hat{\varphi}_T$ of L_T is a consistent estimator of φ with rate $O_P(T^{-1/4})$.*

Proof. To start, we introduce the oracle empirical loss that, had the true Fréchet mean μ been known,

$$L_T^o(u) = \frac{1}{T-1}\sum_{t=1}^{T-1} d(\gamma_\mu^{X_t} u, X_{t+1})^2.$$

Further, we introduce the loss function $f_\mu^u(X_t, X_{t+1}) = d(\gamma_\mu^{X_t} u, X_{t+1})^2$. As a function of u, it is strongly convex and thus has a non-empty sub-gradient and we write $g_\mu^u(X_t, X_{t+1}) \in \partial_u f_\mu^u(X_t, X_{t+1})$ for an arbitrary element of the sub-gradient in u. Using a sub-gradient expansion of L_T^o around the true φ and the strong convexity of f_μ^u, see [Nesterov, 2018, Corollary 3.2.1] we have

$$L_T^o(\hat{\varphi}_T) \geq L_T^o(\varphi) + \frac{1}{T-1}\sum_{t=1}^{T-1} g_\mu^\varphi(X_t, X_{t+1})(\hat{\varphi}_T - \varphi) + C(\hat{\varphi}_T - \varphi)^2, \tag{3.1}$$

where $C = \mathrm{Var}\big[d(X,\mu)^2\big]$. Combining the $\sqrt{T}$ convergence of $\hat{\mu}_T$ to the Lipschitz continuity of squared distances and the geodesic comparison inequality on Hadamard spaces, see [Sturm, 2003, Corollary 2.5], we have that $\|L_T^o - L_T\|_\infty = O_P(1/\sqrt{T})$ since for all $u \in [0,1]$

$$\begin{aligned}
|L_T^o(u) - L_T(u)| &\leq \frac{1}{T-1}\sum_{t=1}^{T-1}\Big|d(\gamma_\mu^{X_t} u, X_{t+1})^2 - d(\gamma_{\hat{\mu}_T}^{X_t} u, X_{t+1})^2\Big| \\
&\leq \frac{1}{T-1}\sum_{t=1}^{T-1} d(\gamma_\mu^{X_t} u, \gamma_{\hat{\mu}_T}^{X_t} u) \leq \frac{1}{T-1}\sum_{t=1}^{T-1}(1-u)d(\mu, \hat{\mu}_T) \\
&\leq d(\mu, \hat{\mu}_T) = O_P(1/\sqrt{T}).
\end{aligned}$$

Thus, for all $u \in [0,1], L_T^o(u) = L_T(u) + O_p(1/\sqrt{T})$. Using this in expansion (3.1) together with the fact that $\hat{\varphi}_T$ minimizes L_T, we have

$$\begin{aligned}
&L_T(\hat{\varphi}_T) + O_p(1/\sqrt{T}) \\
&\geq L_T(\varphi) + O_p(1/\sqrt{T}) + \frac{1}{T-1}\sum_{t=1}^{T-1} g_\mu^\varphi(X_t, X_{t+1})(\hat{\varphi}_T - \varphi) + C(\hat{\varphi}_T - \varphi)^2 \\
&\geq L_T(\hat{\varphi}_T) + O_p(1/\sqrt{T}) + Z_T(\mu)(\hat{\varphi}_T - \varphi) + C(\hat{\varphi}_T - \varphi)^2,
\end{aligned}$$

where $Z_T(\mu) = \frac{1}{T-1}\sum_{i=1}^{T-1} g_\mu^\varphi(X_t, X_{t+1})$. Since $\varphi^\star$ is the minimizer of L, we have that $Z_T(\mu)$ is mean-zero and hence by a similar central limit theorem for m-dependent processes as used in the proof of [Bulté and Sørensen, 2024a, Theorem 5.1], $Z_T(\mu) = O_P(1/\sqrt{T})$. Hence, the inequality above yields $O_P(1/\sqrt{T}) \geq O_P(1/\sqrt{T})(\hat{\varphi}_T - \varphi) + C(\hat{\varphi}_T - \varphi)^2$. Matching the convergence rates on both sides of the inequality results in the convergence rate $|\hat{\varphi}_T - \varphi| = O_P(T^{-1/4})$. □

On bounding tails in the framework of iterated random functions

The proof of [Bulté and Sørensen, 2024a, Proposition B.1] in this chapter is in part an adaptation of the following result, which we leave in its original formulation.

Proposition 1 (Proposition 1 in [Wu and Shao, 2004])**.** *Let g be a bounded function, $\mathbb{E}[g(Y_n)] = 0$ and*

$$C := \sup_{x\in\mathcal{X}} \sum_{n=0}^{\infty} |\mathbb{E}[g(Y_n) \mid Y_0 = x]| < \infty.$$

Then, there exists $c_1, c_2 > 0$ which only depend on $\{Y_n\}$ and g such that for all $\lambda > 0$,

$$\mathbb{P}[|S_n(g)| > n\lambda] \leq c_1 \exp\bigl(-nc_2\lambda^2\bigr). \tag{3.2}$$

In the process, we discovered the following mistake. Specifically, based on $|R_n| \leq r$ and the standard exponential inequality for bounded martingales, the authors claim in the first equation of the proof that

$$\mathbb{E}\left[\exp\left\{\beta\left(\sum_{i=1}^{n} D_i + R_n\right)\right\}\right] \leq \exp(r + nI(\beta d)).$$

However, the right hand side should be $\exp(\beta r + nI(\beta d))$. This difference propagates through a standard argument omitted in the original proof, resulting in a bound of the form of

$$\mathbb{P}[|S_n(g)| > n\lambda] \leq c_1 \exp\left\{-\left(n + \frac{1}{2}\right)c_2\lambda^2\right\}.$$

This bound is insufficient for the sub-Gaussian properties needed in the proof of Proposition B.1 of [Bulté and Sørensen, 2024a]. The authors provided the following corrected proof.

Proof. Consider the sum $S_n(g) = M_n + R_n$ where $M_n = \sum_{i=1}^n D_i$ is the martingale part of the sum. Define C such that $|R_n| < 2C$ and $|D_n| \leq 2C$. We will prove that this proposition holds for $c_1 = e^2$ and $c_2 = \frac{1}{8C^2}$. Using Azuma's inequality to bound M_n, we obtain for any $n \in \mathbb{N}$ and $\lambda > 0$,

$$\mathbb{P}\left[|M_n| > \frac{n\lambda}{2}\right] \leq 2\exp\left\{-\frac{n\lambda^2}{8C^2}\right\}.$$

To obtain a concentration inequality for $S_n(g)$, consider the two cases:

Case (A) If $\sqrt{n}\lambda > 4C$, it holds that $2C < \frac{\sqrt{n}\lambda}{2} < \frac{n\lambda}{2}$. Since $|R_n| \leq 2C$, this implies that $\mathbb{P}\big[|R_n| > \frac{n\lambda}{2}\big] = 0$. Together with the concentration inequality for M_n, we get

$$\mathbb{P}[|S_n(g)| > n\lambda] \leq \mathbb{P}\left[|M_n| > \frac{n\lambda}{2}\right] + \mathbb{P}\left[|R_n| > \frac{n\lambda}{2}\right] \leq 2\exp\left\{-\frac{n\lambda^2}{8C^2}\right\}.$$

Case (B) If $\sqrt{n}\lambda \leq 4C$, inequality (3.2) yields a vacuous bound since

$$c_1 \exp\big(-nc_2\lambda^2\big) = e^2 \exp\left(-\frac{n\lambda^2}{8C^2}\right) \geq \exp\left(2 - \frac{16C^2}{8C^2}\right) = 1.$$

□

4 Exploiting symmetries for testing the mean

This chapter contains the paper:

[**ISO**] [Bulté and Sørensen, 2025] Bulté, M. and Sørensen, H. (2025). Isotropic randomization for one-sample testing in metric spaces. arXiv:2501.15945 [stat].

In this project, we aim to introduce ideas, principles, and theoretical results for testing the Fréchet mean of a random object. Rather than providing an explicit test, we focus on presenting a testing methodology based on combining the classical randomization testing ideas from [Lehmann and Romano, 2022] and exploiting symmetries of the metric space through its isotropy groups. The key insight is to exploit the invariance of the Fréchet mean and variance under the action of the isotropy group of the Fréchet mean, which allows us to construct a test statistic that is invariant at the population level under the null hypothesis. This led us to explore the connections between the Fréchet mean and isotropy groups in more detail and consider cases in which intuition from Euclidean spaces does not directly extend to general metric spaces.

Isotropic randomization for one-sample testing in metric spaces

Matthieu Bulté[1,2] and Helle Sørensen[1]

[1]Department of Mathematical Sciences, University of Copenhagen
[2]Faculty of Business Administration and Economics, Bielefeld University

Abstract

We address the fundamental problem of testing hypotheses about the Fréchet mean in metric spaces, extending classical mean testing from Euclidean spaces to more general settings. We extend an Euclidean testing procedure progresively, starting with test construction in Riemannian manifolds, leveraging their natural geometric structure through exponential and logarithm maps, and then extends to general metric spaces through the introduction of admissible randomization techniques. This approach preserves essential geometric properties required for valid statistical inference while maintaining broad applicability. We establish theoretical guarantees for our testing procedure and demonstrate its effectiveness through numerical experiments across different metric spaces and distributional settings. The practical utility of our method is further illustrated through an application to wind data in western Denmark, showcasing its relevance for real-world statistical analysis.

1 Introduction

The statistical analysis of non-standard data types has gained increasing attention as new methods of measurement and data collection emerge across various fields. This has led to the development of methods for analyzing random variables taking values in metric spaces, also called random objects, where only a notion of distance between points is available rather than the rich structure of a vector space.

The study of random objects spans multiple application domains. In functional data analysis, methods have been developed for analyzing curve data [28]. Random objects also appear in neuroimaging through the analysis of correlation matrices from fMRI data [26, 9, 32], and in network science through the study of adjacency matrices representing social networks [11]. The analysis of probability distributions is another important example, where different metrics can be used such as the Wasserstein distance [25].

Probability distributions are a particularly well-studied example of random objects and various approaches have been developed for their analysis. They have been studied as images of Hilbert spaces under transformations [27], as specific Hilbert spaces with tailored addition and scalar multiplication operators [34], as well as metric spaces with distances constructed to expose certain properties or invariances [25, 30].

The statistical theory for random objects has seen substantial development in recent years. Fundamental work has addressed hypothesis testing and inference [10, 11, 24, 19], alongside various approaches to regression [26, 7, 14] and time series models [17, 6]. Since metric spaces offer limited inherent structure, additional assumptions are often introduced to ensure well-defined statistical quantities. A common approach is to assume the metric space is a Hadamard space, which provides a rich geometric framework while maintaining generality [31, 1].

The Fréchet mean, a generalization of the expected value to metric spaces, has been the subject of extensive theoretical investigation. Recent work has examined its concentration properties [5], asymptotic distributional behavior [4, 35], or more fundamental properties of the quantity itself in various scenarios [23, 24, 16]. One phenomenon of interest highlighting the difference with estimating the mean in Euclidean spaces is that of of smeariness [12], where the Fréchet mean can exhibit unusual asymptotic behavior.

In this paper, we focus on the problem of testing whether the Fréchet mean of a distribution on a metric space equals a hypothesized value. This extends the classical problem of testing the mean in Euclidean space to the more general setting of metric spaces. We start by considering the case of Riemannian manifolds, where the exponential and logarithm maps provide natural tools for constructing tests. We then extend our framework to general metric spaces through the introduction of admissible randomization, which preserve key geometric properties needed for valid inference. We demonstrate the practical utility of our approach through numerical experiments and a case study of circular data.

The paper is organized as follows: Section 2 provides an introduction to Fréchet means in metric spaces and to the mean testing problem. It prevents a gradually more general solution to the mean testing problem up to an approach in Riemannian manifolds. Section 3 introduces the main contribution of the paper, a general approach to mean testing in metric spaces. Section 4 illustrates the performance of the test on a series of numerical experiments in various metric spaces and distributional setups. Finally, Section 5 demonstrates the use of the method to a real dataset.

2 Background

2.1 Fréchet mean and metric spaces

Let (Ω, d) be a metric space equipped with the Borel $\sigma-$algebra induced from the metric topology on Ω. A random variable X over Ω is a Borel measurable function from some probability space to Ω. For $p \geq 1$, the space $L^p(\Omega)$ contains all random variables X such that the $p-$th moment of the distance function is defined, that is, $\mathbb{E}[d(X, \omega)^p] < \infty$ for some $\omega \in \Omega$ – and hence for all $\omega' \in \Omega$ since by the triangle inequality and Jensen's inequality $\mathbb{E}[d(\omega', X)^p] \leq 2^{p-1}\mathbb{E}[d(\omega, X)^p] + 2^{p-1}d(\omega, \omega')^p < \infty$. A notion of centrality of random variable $X \in L^2(\Omega)$ can be defined by considering the expected value of the square distance function

$$F_X(\omega) = \mathbb{E}\big[d(X, \omega)^2\big]. \tag{1}$$

Fréchet [13] proposes the minimizer of this quantity as a generalization of the expectation in the Euclidean case. This minimizer, together with the minimal value attained by F_X, are commonly called the *Fréchet mean* and *Fréchet variance* of X

$$\mathbb{E}[X] = \arg\min_{\omega \in \Omega} F_X(\omega) \qquad \mathrm{Var}[X] = F_X(\mathbb{E}[X]). \tag{2}$$

The Fréchet mean generalizes the expected value to metric spaces, and, similarily to Euclidean space, provides a notion of *center of the distribution*. This stems from the fact that for a random variable $X \sim P$, the integral $\int xP(\mathrm{d}x)$ is the minimizer of the Fréchet function. However, unlike the Euclidean case, the Fréchet mean is not guaranteed to exist for any random variable in $L^2(\Omega)$, and when it does, it is not necessarily unique. Consider for example the case

where the space Ω is the $d-$dimensional sphere $S^d = \left\{x \in \mathbb{R}^{d+1} : \|x\| = 1\right\}$ and X being uniformly distributed. Then $F_X(\omega)$ is constant and hence a unique minimizer does not exist. In the rest of the paper, it will be assumed that the Fréchet mean uniquely exists.

In this work, we will be concerned with the problem of using a sample $X_1, \ldots, X_n$ of copies of $X \in L^2(\Omega)$ to test whether Fréchet mean of X takes a specific hypothesized value $\mu \in \Omega$,

$$H_0 : \mathbb{E}[X] = \mu \qquad \text{vs.} \qquad H_1 : \mathbb{E}[X] \neq \mu. \tag{3}$$

The test developed here is based on the Fréchet mean and variance of X and their sample counterparts constructed through minimization of the empirical Fréchet function,

$$\hat{\mu}_n = \arg\min_{\omega \in \Omega} \frac{1}{n} \sum_{i=1}^{n} d(X_i, \omega)^2 \qquad \hat{V}_n = \frac{1}{n} \sum_{i=1}^{n} d(X_i, \hat{\mu}_n)^2. \tag{4}$$

To study the behavior of our final test statistic based on the Fréchet mean and variance, we will rely on asymptotic properties of these quantities. The following assumptions will be made throughout the paper. The first assumption is a standard assumption coming from the study of M-estimators providing a form of control of the complexity of the metric space, see for instance [33]. It is commonly found in various forms in the study of random objects and Fréchet means, see [10, 29].

Assumption 2.1. *Let $N(\varepsilon, U)$ be the covering number of $U \subset \Omega$ with balls of size ε. Then*

1. *For any $\omega \in \Omega$, $\int_0^1 \sqrt{1 + \log N(\varepsilon\delta/2, B_\delta(\mu))}\, d\varepsilon \to 0$ as $\delta \to 0$.*
2. *The entropy integral $\int_0^1 \sqrt{1 + \log N(\varepsilon, \Omega)}\, d\varepsilon$ is finite.*

The second assumption, also common in the study of random objects and more generally M-estimators, requires that the theoretical and empirical Fréchet means uniquely exist. It is a central assumption in proving that the empirical Fréchet mean is consistent, that is that $d(\mu, \hat{\mu}_n) = o_P(1)$, see for instance [33, Corrolary 3.2.3].

Assumption 2.2. *The random variable X has a unique Fréchet mean $\mu \in \Omega$ and its sample estimator $\hat{\mu}_n$ exists almost surely, and for any $\varepsilon > 0$, the population Fréchet mean satisfies $\inf_{d(\omega,\mu)>\varepsilon} F_X(\omega) > F_X(\mu)$.*

Our test relies on the asymptotic behavior of the empirical Fréchet mean and variance under the null hypothesis, based on the results in [10]. The contribution of this work lies in proposing an approach to randomization suitable for the testing problem in general metric spaces and finite sample. We first present the testing problem and randomization approach in the context of Euclidean spaces, then progressively extend the testing methodology to Riemannian manifolds and later to general metric spaces.

2.2 Mean test on the real line

Suppose that we observe $X_1, \dots, X_n \overset{iid}{\sim} P$ be independent and identically distributed random variables in $\mathbb{R}$. Given a $\mu \in \mathbb{R}$, we would like to design a testing procedure for hypothesis (3) without parametric assumptions on P. For simplicity, we restrict the model class to distributions symmetric around μ, which is equivalent to saying that the law of X is invariant under the *reflection map* $g_\mu : x \mapsto 2\mu - x$,

$$\mathbb{P}[X \in A] = \mathbb{P}[g_\mu \cdot X \in A] \qquad \forall A \in \mathcal{B}(\mathbb{R}).$$

Following the approach presented in Chapter 17 of [21], the symmetry of X can be used to construct a randomization that preserves the distribution of X under the null hypothesis. Given a $z \in \{0, 1\}$, define the *randomized variable*

$$g_\mu^z \cdot X := \begin{cases} X & \text{if } z = 0, \\ g_\mu \cdot X & \text{if } z = 1. \end{cases} \tag{5}$$

By symmetry, under the null hypothesis, $g_\mu^z \cdot X$ has the same distribution as X for any $z \in \{0, 1\}$. This properties carries on to the distribution of any test statistic T evaluated on a randomized sample. That is, under the assumption of symmetry and an arbitrary binary vector $z \in \{0, 1\}^n$, the randomized statistic $T(g_\mu^{z_1} \cdot X_1, \dots, g_\mu^{z_n} \cdot X_n)$ has the same distribution as $T(X_1, \dots, X_n)$. Consider now all 2^n binary randomization vectors and let $T_{(1)}, \dots, T_{(2^n)}$ be the associated evaluations of the test statistic, sorted. For a nominal level $\alpha \in (0, 1)$, a level-α hypothesis test for $H_0 : \mathbb{E}[X] = \mu$ can be constructed by rejecting the null hypothesis if the observed test statistic $T(X_1, \dots, X_n)$ is *too large* compared to the randomized sampple, that is, if $T(X_1, \dots, X_n) \geq T_{(k)}$ with $k = 2^n - \lfloor \alpha 2^n \rfloor$. Since the randomization g_μ is self-inverse, the set $\{\mathrm{id}, g_\mu\}$ is a group, and the randomization procedure, together with the symmetry of X, fits into the *Randomization Hypothesis*

framework of [21]. By Theorem 17.2.1 of the same manuscript, this test has the desired level α.

In practice though, applying the 2^n randomizations to construct the test is computationally infeasible for large n. Instead, one recognized that the testing procedure can be linked to an expected value with respect to the uniform distribution over the product group $\{\mathrm{id}, g_\mu\}^n$, which can be approximated via Monte Carlo techniques by sampling a large number of randomizations and computing the test statistic for each of them. For a given number of replicates B and a significance level α, one can sample $Z_1, \ldots, Z_n \overset{iid}{\sim} \mathrm{B}(1/2)$ and compute the test statistic $T_b = T(g_\mu^{Z_1} \cdot X_1, \ldots, g_\mu^{Z_n} \cdot X_n)$. Then, the level-$\alpha$ hypothesis test is constructed following the same procedure as outlined above, applied to the sorted statistics $T_{(1)}, \ldots, T_{(B)}$. A natural choice for the test statistic $T(X_1, \ldots, X_n)$ is the empirical variance of the sample, given by

$$\hat{V}_n(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2. \tag{6}$$

The construction above provides a principled approach to testing symmetry on the real line using randomization techniques, without requiring parametric assumptions beyond symmetry itself. The choice of empirical variance as the test statistic is particularly natural: under symmetry around μ, we expect the spread of observations to be balanced on either side of μ. When the true mean differs from the hypothesized value μ, the empirical variance tends to increase as observations are shifted away from μ, making it sensitive to departures from the null hypothesis. Moreover, the empirical variance has the advantage of being invariant under the reflection map g_μ, ensuring that the randomization procedure maintains consistent power across different samples.

The extension of these ideas to Riemannian manifolds presents several challenges. First, the notion of reflection needs to be appropriately generalized to account for the manifold's geometry. Second, the test statistic must be adapted to capture meaningful deviations from the null hypothesis while respecting the manifold structure. In particular, we will need to carefully consider how to define a variance-like quantity that preserves the desirable properties of sensitivity to asymmetry and invariance under the appropriate generalization of reflection. The following section develops these extensions, showing how the fundamental principles of randomization testing can be preserved in the more general setting.

2.3 Extension to Riemannian manifolds

As a first generalization step, we will see how the intuition of the randomization procedure on the real line extends naturally to directional data on the unit circle $S^1 = \{x \in \mathbb{R}^2 : \|x\| = 1\}$. The distance between two points $x, y \in S^1$ is given by $d_{S^1}(x, y) = \arccos(\langle x, y\rangle)$. Since $\langle x, y\rangle = \|x\|\|y\| \cos(\angle(x, y)) = \cos(\angle(x, y))$, the distance between two points on the circle corresponds to the angle between the two vectors. For a point $x \in S^1$, denote by $\theta_x \in [0, 2\pi)$ its angular representation with $\theta_{(1,0)} = 0$. Directly generalizing the randomization procedure from the real line, define the reflection map in angular representation with

$$g_\mu \cdot \theta_x = 2\theta_\mu - \theta_x \mod 2\pi \tag{7}$$

Without loss of generality, assume that $\mu = (0, 1)$, which can be achieved without affecting distances via a rotation. The reflection map g_μ then corresponds to a sign flip of the angle or equivalently a the reflection of y through the x-axis in the vector representation. This transformation does not change distances between points and is thus an isometry.

To generalize this construct to complete connected $d-$dimensional Riemannian manifolds, we need to introduce the notions of geodesics, exponential map and the logarithm maps. We will illustrate each of these concepts using the circle as a concrete example. Geodesics are curves $\gamma : (a, b) \to M$ that are locally length-minimizing, playing the role that straight lines do in Euclidean space. At any point $p \in M$, the tangent space T_pM is a $d-$dimensional vector space that can be thought of as containing all possible velocities of curves passing through p. For the circle, the tangent space at any point is simply a line – isomorphic to $\mathbb{R}$ – tangent to the circle at that point. The exponential map $\exp_p : T_pM \to M$ takes a tangent vector $v \in T_pM$ and follows the geodesic starting at p with initial velocity v for one unit of time, i.e., $\exp_p(v) = \gamma(1)$ where γ is the unique geodesic such that $\gamma(0) = p$ and $\gamma'(0) = v$. On the circle, $\exp_p(v)$ corresponds to starting at p and moving counterclockwise by an angle of $|v|$ radians if v is positive, or clockwise if v is negative. We call the *cut locus* of p the boundary of the set $\left\{v \in T_pM \mid d_M(\exp_p v, p) = \|v\|\right\}$ and denote it by $\mathrm{cut}(p)$. The distance of a point q to the cut locus is called the *injectivity radius* of p and is denoted by $\mathrm{inj}(p)$. On the circle, the cut locus of any point p is its antipodal point, $\mathrm{cut}(p) = \{-p\}$, and hence the injectivity radius at any point is $\mathrm{inj}(p) = \pi$. Within the injectivity radius of p, the exponential map is a diffeomorphism

between T_pM and a neighborhood of p in M with inverse $\log_p$. That is, for any point $q \in M$ with $d(p,q) < \text{inj}(p)$, it holds that $\exp_p(\log_p q) = q$. For a more thorough and rigorous introduction to Riemannian geometry, we refer the reader to [8].

Given a point $\mu \in S^1$, we can now define the map g_μ in terms of the exponential and logarithm maps. For any point $x \in S^1$, the reflection $g_\mu \cdot x$ provided in (7) as

$$g_\mu \cdot x = \begin{cases} x & \text{if } x \in \text{cut}(\mu), \\ \exp_\mu(-\log_\mu(x)) & \text{otherwise.} \end{cases} \tag{8}$$

This map is also refered to in the literature as the *geodesic symmetry* since $g_\mu \cdot x = \gamma_-(1)$ where γ_- is the geodesic with $\gamma_-(0) = \mu$ and $\gamma_-'(0) = -\log_\mu x$. The randomization procedure in (5) can be written on the circle more generically as first sampling a transformation $\mathbf{g}$ from the set $G_\mu = \{\text{id}, g_\mu\}$ and setting $X^\star = \mathbf{g} \cdot X$. The set of mappings G_μ corresponds to the set of all isometries preserving 0 in the tangent space of S^1 in μ.

For tangent space of dimension $d > 1$, alternative transformations of the vector space can be possible. Consider a $d-$dimensional Riemannian manifold M, then for any point $p \in M$, the tangent space $\text{Tan}_p M$ is isomorphic to $\mathbb{R}^d$ and we can consider any isometry on $\mathbb{R}^d$ mapping the 0 vector to itself as candidates for the randomization maps: this is the orthonormal group $O(d)$ containing the rotations and reflections. This allows to define a set of mapping on M preserving the test mean $\mu \in M$ via $G_\mu = \left\{g_\mu^Q : Q \in O(d)\right\}$ where

$$g_\mu^Q \cdot x = \begin{cases} x & \text{if } x \in \text{cut}(\mu), \\ \exp_\mu(Q \log_\mu(x)) & \text{otherwise.} \end{cases}$$

In general Riemannian manifold, even with further standard regularity conditions, this set of maps is not as well-behaved as on the real line or on the circle. The maps g_μ^Q still map μ to itself but are not always an isometries themselves – even if Q is. Furthermore, even if the random variable X is symmetric around its Fréchet mean μ and is almost surely within the injectivity radius of μ, the maps g_μ^Q do not necessarily preserve the Fréchet mean. However, if the Fréchet function of X is convex, an adjacent concept of mean preservation still holds. Assuming that X is almost surely contained within the injectivity radius of μ and that the Fréchet function F_X defined in Equation (1) is convex, the gradient of F_X in a point $p \in \text{inj}(\mu)$ is given

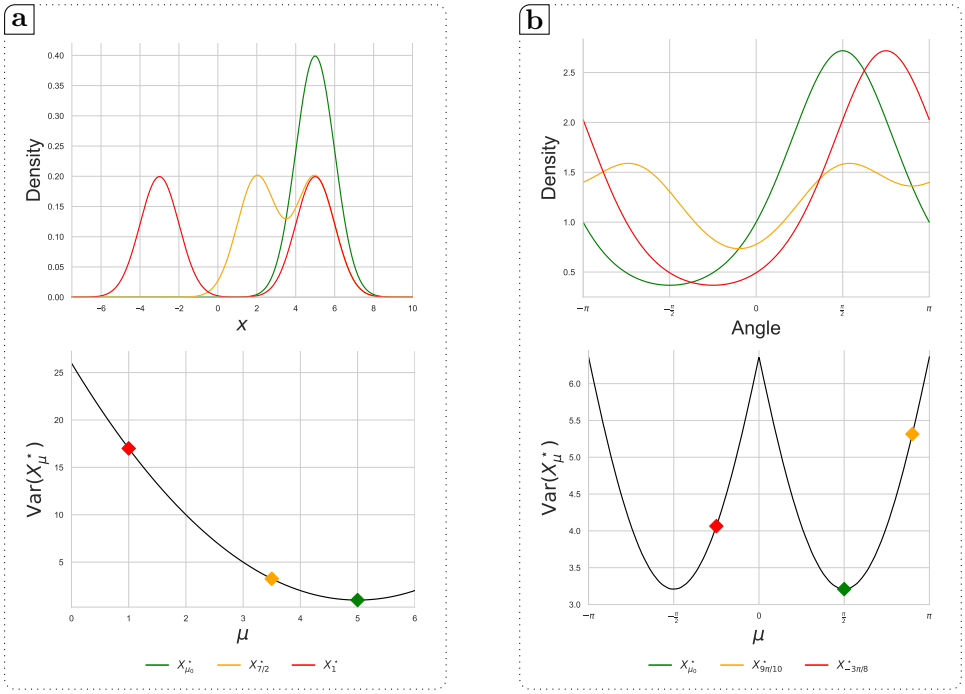


Figure 1: In both panels, the different plots display the behavior of a random variable and its variance under the reflective randomization proposed in (8). In Panel (a), the random variable follows a normal distribution $N(5,1)$ and is randomized around its true mean (green) as well as other hypothesised points μ on the real line (orange, red). Panel (b) considers a random variable on the circle S^1 following a von Mises distribution $VM(\pi/2, 0.3)$. As in panel (a), we show the result of the randomization with respect to the true mean (green) as well as randomizations with respect to other points μ on the circle (orange, red). In both panels, the top plot displays the probability density functions of each variable, while the bottom plot shows their Fréchet variance as a function of the randomization point μ. The colored points on the variance curves correspond to the specific distributions displayed above.

by $\mathrm{grad}F_X(p) = -2\mathbb{E}\big[\log_p(X)\big]$ and the Fréchet mean of X exists and solves the score equation $\mathbb{E}\big[\log_\mu(X)\big] = 0$, see [18, 20, 12]. In this situation, any $g_\mu^Q \in G_\mu$ preserves the Fréchet mean μ by convexity together with

$$\mathbb{E}\big[\log_\mu(g_\mu^Q \cdot X)\big] = \mathbb{E}\big[\log_\mu(\exp_\mu(Q\log_\mu(X)))\big] = Q\mathbb{E}\big[\log_\mu(X)\big] = 0.$$

While this support restriction may appear to be rather restrictive, we will show in the next generalization that for a large class of Riemannian manifolds, the maps g_μ^Q can be used to construct a valid randomization without having to assume that X is almost surely within the injectivity radius of μ.

While the previous discussion focused on manifolds where the cut locus plays a crucial role in defining our randomization maps, there exist important manifolds in statistical applications where the cut locus is empty. This simplifies the construction of the randomization procedure significantly, as illustrated by the following example of the space of symmetric positive definite (SPD) matrices equiped with the Bures-Wasserstein distance.

Example 2.1 (Bures-Wasserstein)**.** *Let $\mathcal{D}_2(\mathbb{R}^p)$ be the set of probability measures on $\mathbb{R}^p$ with finite second moment and admit a density. The Wasserstein distance [25] of order 2 between two measures $\mu, \nu \in \mathcal{D}_2(\mathbb{R}^p)$ is defined as*

$$d_{W_2}^2(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \mathbb{E}_{(X,Y)\sim\gamma}\left[\|x-y\|_2^2\right]$$

where $\Pi(\mu,\nu)$ is the set of all joint distributions on $\mathbb{R}^p \times \mathbb{R}^p$ with marginals μ and ν. The Wasserstein distance can be used to define a distance $d_{\mathcal{B}}$ on the set of covariance matrices $\mathbb{S}_+^p$ via centered Gaussian distributions, resulting in the Bures-Wasserstein distance $d_{\mathcal{B}}(A,B) = d_{W_2}(\mathcal{N}(0,A),\mathcal{N}(0,B))$. The metric space $(\mathbb{S}_+^p, d_{\mathcal{B}})$ is a Riemannian manifold called the Bures-Wasserstein space, and the distance function has the following closed form expression

$$d_{\mathcal{B}}(A,B)^2 = trA + trB - 2tr(A^{1/2}BA^{1/2})^{1/2}.$$

See [3, 9] for more details on Bures-Wasserstein spaces. In this space, the cut loci are empty sets and the exponential map is a non-isomorphic diffeomorphism over the whole space.

This generalization demonstrates how the basic reflection principle from $\mathbb{R}$ extends naturally to Riemannian manifolds through three key steps: (1) first extending to the circle S^1 where reflections correspond to angle reversals, (2) generalizing to arbitrary Riemannian manifolds using the exponential and logarithm maps to define geodesic symmetries, and (3) incorporating the full orthogonal group $O(d)$ as a richer example of transformations to capture all possible isometries of $\mathbb{R}^d$. While additional technical conditions are needed compared to the Euclidean case - particularly regarding the cut locus and convexity of the Fréchet function - the fundamental principle of constructing randomization maps that preserve both the test mean and the metric structure remains the same.

3 Isotropic randomizaton

The randomization tests developed for Euclidean and Riemannian manifolds rely heavily on the existence of exponential and logarithm maps to transport data points. However, many interesting metric spaces, such as graphs, trees, or stratified spaces, lack these differential geometric tools. In this section, we develop a general framework for randomization tests based on isometries that preserves the key properties of the Euclidean approach while being applicable to any metric space with sufficient symmetries. This generalization allows us to extend mean testing procedures to a broader class of spaces while maintaining theoretical guarantees on test validity and power.

Let (Ω, d) be a metric space and denote by $\mathrm{Iso}(\Omega)$ the set of all bijective isometries from Ω onto itself. This set is a group under composition, called the *isometry group* of Ω, and acts on Ω by $g \cdot x = g(x)$. For any $\mu \in \Omega$, the *isotropy group* of μ is the subgroup G_μ of $\mathrm{Iso}(\Omega)$ of isometries mapping μ to itself, $G_\mu = \{g \in \mathrm{Iso}(\Omega) : g \cdot \mu = \mu\}$. A relevant property of the Fréchet mean is that it is invariant under the action of an isometry and as the next proposition shows, the isotropy group of the Fréchet mean of a random variable X preserves all the distance moments of X.

Proposition 3.1. *Let (Ω, d) be a metric space, $p \geq 2$ and $X \in L^p(\Omega)$ be a random variable in Ω with Fréchet mean μ. Then, for any random isometry $\mathbf{g}$ independent of X with support equal to a subgroup of G_μ, the variable $\mathbf{g} \cdot X$ has the same Fréchet mean and distance moments as X, that is*

$$\mathbb{E}[\mathbf{g} \cdot X] = \mu \qquad \text{and} \qquad \mathbb{E}\big[d(\mathbf{g} \cdot X, \mu)^k\big] = \mathbb{E}\big[d(X, \mu)^k\big],$$

for all $k \leq p$.

Proof. We will first prove that μ is the minimizer of the Fréchet function of $\mathbf{g} \cdot X$. Let $\omega \in \Omega$, we have using that $\mathbf{g}$ is an isometry that

$$\begin{aligned}
\mathbb{E}\big[d(\mathbf{g} \cdot X, \omega)^2\big] &= \mathbb{E}\big[d((\mathbf{g}^{-1} \circ \mathbf{g}) \cdot X, \mathbf{g}^{-1} \cdot \omega)^2\big] \\
&= \mathbb{E}\big[d(X, \mathbf{g}^{-1} \cdot \omega)^2\big] \\
&= \mathbb{E}_{\mathbf{g}}\big[\mathbb{E}_X\big[d(X, \mathbf{g}^{-1} \cdot \omega)^2\big]\big]
\end{aligned}$$

Since μ minimizes $\omega \mapsto \mathbb{E}[d(X, \omega)^2]$, we have that the inner expectation can be lower bounded by $\mathbb{E}[d(X, \mu)^2]$. Using this bound and reversing the above

computations we get

$$\begin{aligned}\mathbb{E}\big[d(\mathbf{g}\cdot X,\omega)^2\big] &\geq \mathbb{E}\big[d(X,\mu)^2\big] \\ &= \mathbb{E}\big[d(\mathbf{g}\cdot X,\mathbf{g}\cdot\mu)^2\big] && \text{(by isometry)} \\ &= \mathbb{E}\big[d(\mathbf{g}\cdot X,\mu)^2\big] && \text{(since } \mathbf{g}\in G_\mu\text{)}\end{aligned}$$

Therefore, $\mathbb{E}[d(\mathbf{g}\cdot X,\mu)^2] \leq \mathbb{E}[d(\mathbf{g}\cdot X,\omega)^2]$ for all $\omega\in\Omega$ proving the mean invariance $\mathbb{E}[\mathbf{g}\cdot X]=\mu$. The claim concerning the distance moments is also a direct consequence of the fact that $\mathbf{g}\in G_\mu$,

$$\mathbb{E}\big[d(\mathbf{g}\cdot X,\mu)^k\big] = \mathbb{E}\big[d((\mathbf{g}^{-1}\circ\mathbf{g})\cdot X,\mathbf{g}^{-1}\cdot\mu)^k\big] = \mathbb{E}\big[d(X,\mu)^k\big].$$

□

This suggest that the isotropy group of the Fréchet mean of X is a natural candidate for the set maps to use for the randomization procedure. Given a random variable $\mathbf{g}\sim P_{\mathbf{g}}$ over G_μ, possibly with support equal to a subgroup of G_μ, we define the *isotropic randomized variable* $X^\star = \mathbf{g}\cdot X$. An example of metric spaces with an isotropy group allowing for a randomization procedure similar to the one presented previously are the globally symmetric spaces, a certain kind of Riemmanian manifolds in which the reflection map acts as an isomorphism.

Example 3.1 (Globally symmetric spaces)**.** *Let M be a Riemannian manifold; M is called Riemannian globally symmetric if each $p\in M$ is an isolated fixed point of an involutive isometry s_p, see [15, Chapter IV]. Hence the definition of a symmetric space directly corresponds to assuming that for any $\mu\in M$, the set $R_\mu=\{id,s_\mu\}$ is a subgroup of the isotropy group G_μ. Examples of globally symmetric spaces include the sphere S^d and the hyperbolic space $\mathbb{H}^d$. In these spaces, a randomization procedure can thus be constructed by sampling $\mathbf{g}$ uniformly from R_μ.*

Proposition 3.1 ensures that under the null hypothesis $H_0:\mathbb{E}[X]=\mu$, the randomized random variable $X^\star$ has the same Fréchet mean and distance moments as X. Under the alternative hypothesis $H_1:\mathbb{E}[X]\neq\mu$, we can expect that transforming X via the isotropy group of μ will not preserve the Fréchet mean and distance moments of X. This, together with the following theorem, suggests that the empirical Fréchet variance of $X^\star$ can be used as a test statistic for the hypothesis (3).

Theorem 3.1 (Theorem 1 in [10]). *Suppose that the metric space (Ω, d) is bounded and satisfies Assumption 2.1. If additionally the random variable $X \in \Omega$ satisfies Assuption 2.2, then*

$$\sqrt{n}(\hat{V} - Var[X]) \to \mathcal{N}(0, \sigma_d^2) \tag{9}$$

where $\sigma_d^2 = Var[d(X, \mathbb{E}[X])^2]$

Since our goal is to construct a test based on the empirical variance of $X^\star$, we also need to make sure that it will have power against alternatives to the null hypothesis.

Let us consider two simple cases to illustrate where the variance alone might not be sufficient. First, let Ω be the Euclidean space $\mathbb{R}^2$; here, for any $x \in \mathbb{R}^2$ the isotropy group G_x is given by the rotations and reflections, up to a conjugation with the translation to x. For any direction $x \in \mathbb{R}^2$, the reflection r_x with respect to the line spanned by x is $r_\mu(x) = S_\mu x$ with $S_\mu = 2\frac{\mu\mu^\top}{\mu^\top\mu} - \mathrm{id}_2$. As a reflection r_μ is an involutive isotropy of x, and hence the set $R_x = \{\mathrm{id}, r_x\}$ is a subgroup of G_x. For ease of computation and implementation we might consider restricting the support of the randomization group action to R_μ where μ is the tested mean. However now $\mu \in \mathbb{R}^2$ and $X \sim \mathcal{N}(0, \mathrm{id}_2)$ be the standard normal distribution in $\mathbb{R}^2$. By elementary properties of the multivariate normal distribution, $\mathrm{Var}[g \cdot (X + \lambda\mu)] = \mathrm{Var}[g \cdot (X + \mu)]$ for all $\lambda \in \mathbb{R}$ and $g \in R_\mu$. Hence, the variance alone does not have power against colinear alternatives when restricting the support of $\mathbf{g}$ too much.

While the previous example might seem artificial, there exist spaces in which the entire isotropy group in a point does not identiify the point itself. For instance on the example of the circle S^1 considered above, the isotropy group in any point $x \in S^1$ is $G_x = \{\mathrm{id}, g_x\}$, where g_μ is the angular reflection given in Equation (7). There, we have that the antipodal point x^- of x has the exact same isotropy group $G_{x^-} = G_x$. Thus, whether the randomization is done to test the null $H_0 : \mathbb{E}[X] = x$ or $H_0 : \mathbb{E}[X] = x^-$, the testing procedure will randomize the data in the exact same way making the two situations indistiguishable. This shows that for the circle, and the same argument can be extended to higher dimensional spheres, only considering the variance of the distribtuion randomized via the isotropy group will have no power against certain alternatives.

These two examples motivate the notion of an *admissible randomization* for constructing a test based only on the empirical Fréchet variance. In simple

Algorithm 1 Isotropic Randomization Test for Fréchet Mean

Require: Sample $X_1, \ldots, X_n$, hypothesized mean μ, randomization distribution $P_{\mathbf{g}}$, number of replicates B, significance level α

1: **for** $b = 1, \ldots, B$ **do**
2: Sample independently $\mathbf{g}_1, \ldots, \mathbf{g}_n \overset{iid}{\sim} P_{\mathbf{g}}$
3: Construct randomized sample $X_i^\star = \mathbf{g}_i \cdot X_i$ for $i = 1, \ldots, n$
4: Compute test statistic $V_b = \hat{V}_n(X_1^\star, \ldots, X_n^\star)$
5: **end for**
6: Sort the test statistics: $V_{(1)} \leq \cdots \leq V_{(B)}$
7: **return** Reject H_0 if $\hat{V}_n(X_1, \ldots, X_n) > V_{(\lfloor \alpha B \rfloor)}$

words, an admissible randomization should make it possible to distinguish between the null hypothesis and any other point in the space.

Definition 3.1. *Let (Ω, d) be a metric space and $\mu \in \Omega$. A random group action $\mathbf{g}$ over Ω is called $\mu-$admissible randomization if the support of $\mathbf{g}$ is a subgroup of isotropy group G_μ and there exists no points $x \in \Omega \backslash \{\mu\}$ such that $\mathbf{g} \cdot x$ is almost surely constant.*

With this definition, the following proposition shows that the admissible randomizations result in a test with power against alternatives.

Proposition 3.2. *Let $X \in L^2(\Omega)$ be a random variable satisfying Assumption 2.2. For any $\mu \in \Omega$ and $\mu-$admissible randomization $\mathbf{g}$, let $X^\star = \mathbf{g} \cdot X$. Then, $Var[X^\star] \geq Var[X]$ with equality if and only if $\mathbb{E}[X] = \mu$.*

Proof. Let $\mu^\star = \mathbb{E}[X^\star]$ and $\mu_X = \mathbb{E}[X]$. By definition of $X^\star$ and the admissibility of $\mathbf{g}$, we have that

$$\begin{aligned}\mathrm{Var}[X^\star] &= \mathbb{E}\big[d(\mathbf{g} \cdot X, \mu^\star)^2\big] = \mathbb{E}\big[d((\mathbf{g}^{-1} \circ \mathbf{g}) \cdot X, \mathbf{g}^{-1} \cdot \mu^\star)^2\big] \\ &= \mathbb{E}\big[d(X, \mathbf{g}^{-1} \cdot \mu^\star)^2\big] = \mathbb{E}_{\mathbf{g}}\big[\mathbb{E}_X\big[d(X, \mathbf{g}^{-1} \cdot \mu^\star)^2\big]\big]\end{aligned}$$

By definition of the Fréchet mean, we have that $\mathbb{E}[d(X, \omega)^2] \geq \mathbb{E}[d(X, \mu_X)^2]$ for all $\omega \in \Omega$ and hence

$$\mathrm{Var}[X^\star] \geq \mathbb{E}\big[d(X, \mu_X)^2\big] = \mathrm{Var}[X],$$

Since $\mathbb{E}[d(X, \mu_X)^2]$ lower bounds the Fréchet function and by Assumption 2.2 the Fréchet mean is well separated, equality holds if and only if $\mathbf{g}^{-1} \cdot \mu^\star = \mu_X$, or equivalently $\mu^\star = \mathbf{g} \cdot \mu_X$ holds almost surely. Since $\mathbf{g}$ is $\mu-$admissible, this can only hold if $\mu_X = \mu$. □

Hence, if an admissible randomization exists, the empirical variance $\hat{V}_n$ defined in (6) can be used as a test statistic for the hypothesis (3). We formalize this testing procedure in Algorithm 1.

The practical implementation of Algorithm 1 requires constructing isometries that form an admissible randomization. While these arise naturally in globally symmetric spaces, more complex spaces that lack global symmetry require careful consideration of the local geometry. We now examine an example of such a space - a stratified space formed by gluing together Euclidean halfspaces, which naturally arises in applications involving branching structures or networks.

Example 3.2 (Booklets)**.** *Let $k \in \mathbb{N}$, we define the d-dimensional booklet B_d^k as k copies of halfspaces $\mathbb{R}_+ \times \mathbb{R}^{d-1}$ glued together along $\{0\} \times \mathbb{R}^{d-1}$. Each point is then represented by a tuple (z, x, y) where $z \in \{1, \dots, k\}, x \in \mathbb{R}_+$ and $y \in \mathbb{R}^{d-1}$. The distance between two points (z, x, y) and (z', x', y') is given by*

$$d((z,x,y),(z',x',y'))^2 = \begin{cases} (x-x')^2 + \|y-y'\|_2^2 & \text{if } z = z', \\ (x+x')^2 + \|y-y'\|_2^2 & \text{otherwise.} \end{cases}$$

For a point $\mu = (z, x, y)$, one can construct a reflection map with two ingredients: an $y-$isotropy $g_y \in G_y$; an involutive permutation π_z on $\{1, \dots, k\}$ with $\pi_z z = z$. The reflection map g_μ is then defined as

$$g_\mu(z', x', y') = (\pi_z z', x', g_y y').$$

The framework developed above provides a general approach for testing hypotheses about Fréchet means in metric spaces. A natural question is whether we can identify classes of distributions where this approach is particularly well-suited. Radially symmetric distributions, which we examine next, form such a class - their inherent symmetry properties align naturally with the isometric randomization procedure, making them an ideal setting for applying these tests. Moreover, studying these distributions helps us better understand the relationship between geometric symmetry and statistical inference in metric spaces.

3.1 Application to radially symmetric distributions

The framework is particularly well-suited for *radially symmetric distributions* on metric spaces. We define a metric space radially symmetric distribution

with the goal of generalizing the definition in Euclidean sapces. Recall that for $\mu \in \mathbb{R}^d$, a distribution P_μ with density f with respect to the Lebesgue measure is called radially symmetric around μ if its density can be written as $f(x) = h(d(x, \mu))$ for some $h : \mathbb{R} \to \mathbb{R}$. This definition naturally extends to Riemmanian manifolds by assuming a radial density with respect to the Riemannian volume measure. A common example of such a distribution is the von Mises-Fisher distribution on the $d-$dimensional sphere S^d with location and concentration parameters $\mu \in S^d$ and $\kappa > 0$. The density of this distribution with respect to the volume measure is $f_{\mu,\kappa}(x) \propto \exp(\kappa x^\top \mu) = \exp(\kappa d(x, \mu))$, which is invariant under isotropies of μ. It can then be shown [23, Theorem 2] that in homogeneous Riemannian manifolds, the Fréchet mean is the same as the point of symmetry μ. In order to avoid having to work with the technicalities of generalizing the measure volume, we define the notion of radially symmetric distributions on metric spaces via the group action of the isometry group.

Definition 3.2. *A random variable X on Ω is called radially symmetric around $\mu \in \Omega$ if it is invariant under the action of any $g \in G_\mu$. That is, $g \cdot X \overset{\mathcal{D}}{=} X$ for every $g \in G_\mu$.*

Consider a random variable $X \sim P_\mu$, radially symmetric around $\mu \in \mathbb{R}^d$, with density $f(x) = h(d(\mu, x))$ with respect to the Lebegues measure. For any $g \in G_\mu$ a change of variable argument shows that the random variable $g \cdot X$ has density $f_g(y) = h(d(g^{-1} \cdot y, \mu))$. By the fact that $g \in G_\mu$, this gives $f_g(y) = h(d(y, g(\mu))) = h(d(y, \mu))$ and hence $g \cdot X \overset{\mathcal{D}}{=} X$. Hence the new definition of radially symmetric distributions is consistent with the previous one in Euclidean spaces.

Example 3.3 (Normal distribution on the spider)**.** *Consider the booklet of 1 dimension with k branches defined in Example 3.2, also called spider, constructed by gluing together $k \in \mathbb{N}$ copies of the real line through the origin. There, given a point $\mu = (j, x)$, a simple radially symmetric distribution can be constructed by considering mixture of normal distributions P_μ on the real line. Sample the branch index J uniformly from $\{1, \dots, k\}$ and sample $X \mid J = j \sim N(x, 1)$ and $X \mid J \neq j \sim N(0, 1)$. The isotropy group of μ is only composed of transformations that map the branch j to itself and, when applied to the other branches, applies a permutation to the branch index and a reflection on the position. By construction, P_μ is thus invariant under all isotropies of μ and is thus symmetric in μ.*

It is not clear whether the Fréchet mean of a radially symmetric distribution is the point of symmetry as in the case of homogeneous Riemannian manifolds. This stems from the same reason as why some randomizations are not admissible: if two points share the same isotropy group, a random variable will either be radially symmetric around both points or neither. However, the following lemma shows that the Fréchet variance of a radially symmetric distribution is invariant under the action of the isotropy group. However, the following lemma shows that if this problem doesn not occur, the Fréchet mean of a radially symmetric distribution is the point of symmetry.

Lemma 3.1. *Let $X \in L^2(\Omega)$ be a radially symmetric distribution around $\mu \in \Omega$ with Fréchet mean μ_X. If the isotropy group G_μ is not contained in the isotropy group of any other point in Ω, then $\mu_X = \mu$.*

Proof. This is a direct consequence of the equivariance of the Fréchet mean described in [23]: for any $g \in \mathrm{Iso}(\Omega)$, we have that $\mathbb{E}[g \cdot X] = g \cdot \mathbb{E}[X]$. Hence, by symmetry of X around μ we have for all $g \in G_\mu$ that $g \cdot X \overset{\mathcal{D}}{=} X$ and hence $\mathbb{E}[g \cdot X] = \mu_X$. Therefore by equivariance $\mu_X = \mathbb{E}[g \cdot X] = g \cdot \mu_X$ implying that $g \in G_{\mu_X}$ and hence $G_\mu \subset G_{\mu_X}$. Since G_μ is not contained in the isotropy group of any other point, this implies that $\mu_X = \mu$. □

4 Numerical experiments

In this section, we illustrate our theoretical results with numerical experiments taking place in different metric spaces. We consider the example of the circle, with a both a symmetric distribution and a non-symmetric mixture distribution. In the second example, we explore the behavior of the test on a bounded subspace of the 2−dimensional booklet described in Example 3.2. Finally, we consider the space of symmetric positive definite (SPD) matrices equipped with the Bures-Wasserstein distance. Each space present different metric space properties and challenges for the test.

In each of these scenarios, we consider a fixed $\mu_0 \in \Omega$ to test for and generate datasets of size $n = 200$ with true Fréchet mean μ, where the values of μ are elements on the geodesic ray connecting μ_0 and a chosen μ_1. This allows us to consider the performance of the test as a function of the distance between μ_0 and μ. Additionally, we evaluate the power of the test as a function of n for $n \in \{100, 200, 400, 600\}$. For each metric space, we

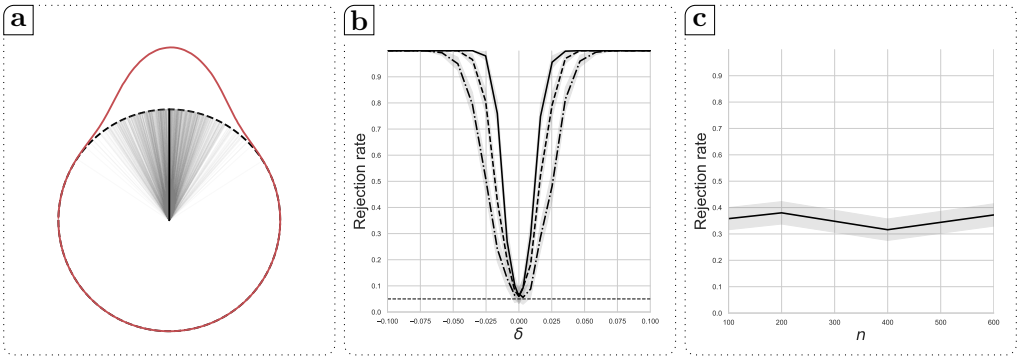


Figure 2: Panel (a) shows a sample (light gray rays) from the VM$(\pi/2, 1)$ distribution. The density of the mixture (red curve) and the Fréchet mean (solid black ray) are also represented. Panels (b) and (c) are generated in the same way as in figure 3

generate 500 datasets. For each dataset, we run the isotropic test at level $\alpha = 0.05$ with $B = 1000$ resampling replicates and record the rejection rate.

All simulations and analyses are done in Python. The code to reproduce the experiments and figures is available online[1].

4.1 Directional data

In the first experimental setup, we consider the space of directions on the circle $\Omega = S^1$. We generate data from two different distributions: a von Mises distribution $VM(0, 0.3)$ and a mixture $P = (1-p)VM(\pi/2, 0.3) + pVM(0, 0.3)$ with mixing proportion $p = 1/3$. In both cases, for a given tested Fréchet mean μ_0, we construct the randomization based on the subgroup of G_{μ_0} consisting of the identity and the reflection map described in Equation (7) and sample the isotropy $\mathbf{g}$ uniformly over this subgroup. In terms of the randomized sample, this corresponds with equal probability to either the original sample or the sample reflected around μ_0.

In the first case, the distribution is radially symmetric around its mean $\pi/2$, as discussed in Section 3.1. The alternativce hypotheses H_1 are expressed in terms of the angle difference δ with $\mu_\delta = \mu_0 + \delta$. For testing the power of the test against local alternative, we sample according to the von

[1] https://github.com/matthieubulte/meantesting

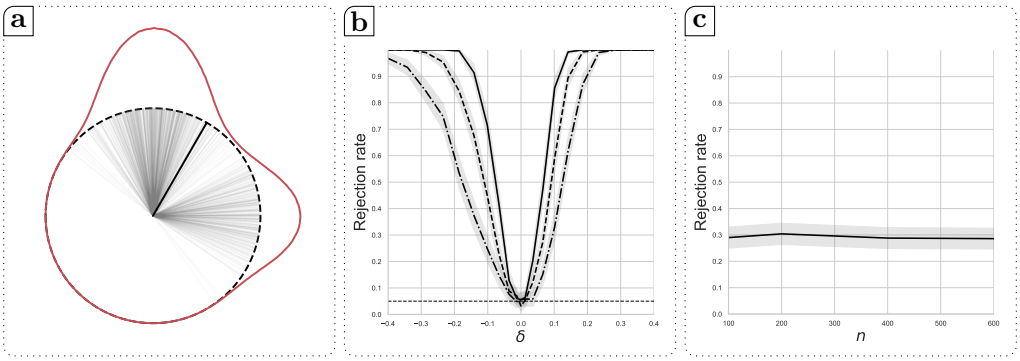


Figure 3: Panel (a) shows a sample (light gray rays) from the mixture described in Section 4.1. The density of the mixture (red curve) and the Fréchet mean (solid black ray) are also represented. Panel (b) displays the rejection rate of the mean test based on the randomization described in Section 4. The rejection rate is shown as a function of the deviation δ from the null hypothesis for $n = 100$ (dash-dotted), $n = 200$ (dashed) and $n = 400$ (solid). Finally, Panel (c) illustrates the power of the test against $\sqrt{n}-$alternatives.

Mises distribution with mean $\mu_0 + 0.2n^{-1/2}$ for different n considered. The visualization of the data, sample and resulting size and power are shown in Figure 2. Panel (b) shows that the test has correct size 0.05 under the null hypothesys ($\delta = 0$) and rapidely converges to 1 as $|\delta|$ increases. While not proved in the previous sections, Panel (c) shows that the test has power against $\sqrt{n}-$alternatives in this scenario.

In the second case of the mixture of von Mises, the true Fréchet mean of is $\mu = \pi/3$ but the unequal weighting of the two mixture components renders the distribution non-symmetric around its mean. Here, alternative distributions are generated by shifting the mean of both component of the mixture, giving $P_\delta = (1-p)VM(\pi/2 + \delta, 0.3) + pVM(\delta, 0.3)$. The data space and distribution along with experiment results are displayed in Figure 3. While the distribution is not symmetric, the experimental results show a similar behavior as in the symmetric case, and the test has both correct size under the null and power against local alternatives.

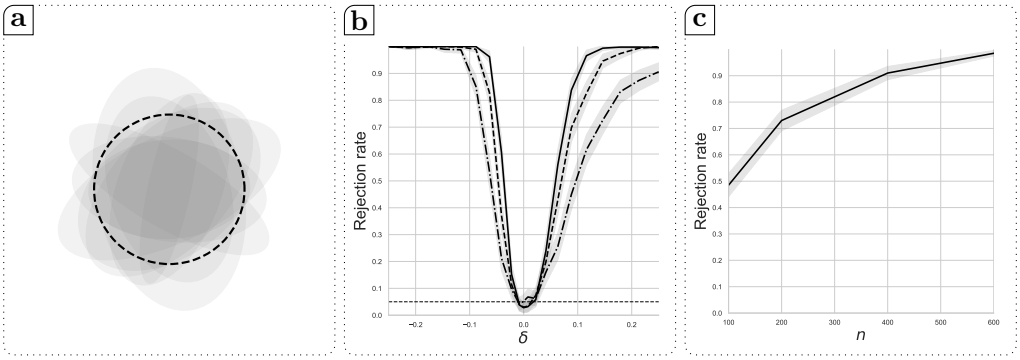


Figure 4: Panel (a) displays a visualization of 20 covariance matrices samples according to the sampling process described in Section 4.2. Each ellipse is the application of a covariance matrix to the unit circle. The Fréchet mean is also displayed (dashed unit circle). Panel (b) and (c) are generated similarly to Figure 3. However, since the test does not have power against $\sqrt{n}$-alternatives Panel (c) here illustrates the asymptotic power of the test against a fixed alternatives.

4.2 Bures-Wasserstein distance on SPD matrices

We now consider the space of 2x2 SPD matrices $\Omega = \mathcal{S}_+^2$ equipped with the Bures-Wasserstein distance, described in Example 2.1. To construct a distribution with a given a mean $\mu \in \mathcal{S}_+^2$ on the Bures-Wasserstein manifold, we first sample a random element in the tangent space Tan_μ and push it forward through $\exp_\mu$. Using that the tangent space at a point $\mu \in \mathcal{S}_+^2$ is the set of symmetric matrices [2], we sample the element of the tangent space in μ by first sampling a 2x2 matrix M with i.i.d. standard normal entries and symmetrize it via $V = (M + M.T)/2$. The sample point is then constructed by applying the exponential map, giving $X = \exp_\mu V$. An explicit form of the exponential map can be written in terms of the Lyapunov operator [22]. In the examples, we chose $\mu = \mathbb{1}_2$ giving closed forms for the exponential and logarithm maps, $\exp_\mu V = (V/2 + \mathbb{1}_2)(V/2 + \mathbb{1}_2)$ and $\log_\mu M = 2M^{1/2} - 2\mathbb{1}_2$. Similarly to the previous example, the randomization is done by sampling $\mathbf{g}$ uniformly over $\{\mathrm{id}, g_\mu\}$ where g_μ is the reflection defined in Equation (8) in terms of the exponential and logarithm maps.

To assess the power and size of our test, we sample under alternative distributions generated by sampling from the same process, with a Fréchet mean

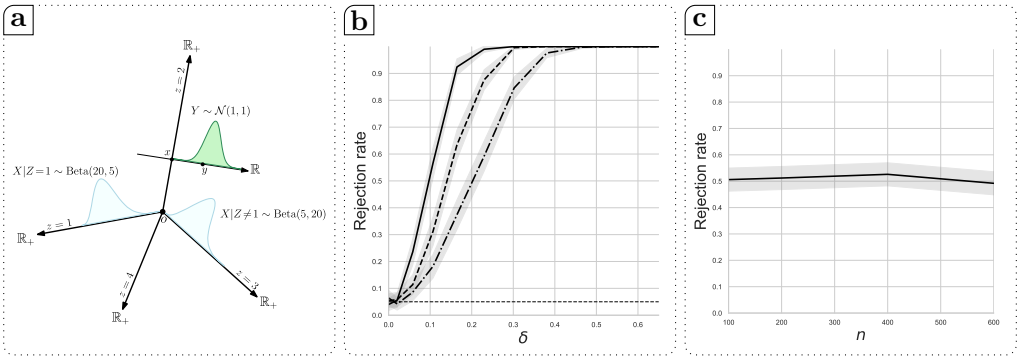


Figure 5: Panel (a) displays a visualization of the booklet space B_1^4 along with the sampling mechanism described in Equation (10). Panel (b) and (c) are generated similarly to Figure 3.

μ_δ lying on the geodesic ray passing through $\mu_0 = \mathbb{1}_2$ and $\mu_1 = \left(\begin{smallmatrix} 4 & 1 \\ 1 & 3 \end{smallmatrix}\right)$, where $\delta \in [-1, 1]$. A similar display of the experiment as in the previous numerical experiments is found in Figure 4. The experiment shows that the isotropic test is able to detect the change in the mean of the distribution for $|delta| > 0$ and maintain level, corresponding to $\delta = 0$. While the power increases with sample size for a fixed alternative, shown in Panel (c), our numerical results suggest that the test lacks power against $\sqrt{n}$-local alternatives.

4.3 Booklet

In this final experiment, we consider the space of booklets B_1^4 described in Example 3.2. As a reminder, the space B_1^4 is constructed by gluing together 4 *branches* of which are copies of the positive line $\mathbb{R}^+$ via the origin 0, a 4−spider, and attaching to each point on this structure a copy of the real line $\mathbb{R}$. Hence each point in B_1^4 can be represented via three coordinates (z, x, y): the index of the branch $z \in \{1, 2, 3, 4\}$, the position on the branch $x \in \mathbb{R}^+$, and the position on the real line $y \in \mathbb{R}$. A visualization of this space can be found in Panel (a) of Figure 5.

To generate a distribution on B_1^4, we consider the following hierarchical model where the distribution of X is determined the branch Z on which the

point lies and Y is independent of X and Z,

$$\begin{aligned} X \mid Z &\sim \begin{cases} \text{Beta}(20,5) & \text{if } Z = 1, \\ \text{Beta}(5,20) & \text{if } Z \in \{2,3,4\} \end{cases} \\ Y &\sim \mathcal{N}(1,1) \\ Z &\sim \text{Categorical}\left(\frac{4}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}\right). \end{aligned} \tag{10}$$

Minimizing the Fréchet function over these distributions yields that the Fréchet mean of this distribution has $\mu_0^z = 1$, $\mu_0^x = 0$ and $\mu_0^y = 1$. The random isotropy $\mathbf{g}$ is chosen as described in Example 3.2. First, a random permutation $\mathbf{g_z}$ over $\{1,2,3,4\}$ is drawn uniformly over the set of randomizations with fixed point μ_z. Given the geometry of the spider, there is no isotropy on the spider that modifies the x component. The y component is changed via $\mathbf{g_y}$ uniformly sampled over $\{\text{id}, g_{\mu_y}\}$ where g_{μ_y} is the reflection on the real line defined in Section 2.2. All in all, we get that for a random $X \in B_1^4$, it's randomization is given by $\mathbf{g} \cdot X = (\mathbf{g_z} \cdot X_z, X_x, \mathbf{g_y} \cdot X_y)$.

In this scenario, the data sampling process is not changed but rather the null hypothesis considered and the test is run for $H_0 : \mu = \mu_\delta$ where μ_δ is on the geodesic between the true μ_0 of the data generating process and $\mu_1 = (2,1,0)$. The results of the experiment are displayed in Figure 5. Similarly to the previous experiments, we observe that the isotropic test is able to detect the change in the mean of the distribution for $\delta > 0$ and maintain level. Panel (c) suggests that the test here has power against $\sqrt{n}-$alternatives.

5 Application: Wind in Western Denmark

As an illustrative example of how the methodology developed here might be used, we analyze wind direction data from the Danish Meteorological Institute (DMI). We are interesting in testing whether the mean wind direction aligns with the documented south-westerly pattern in the region, $\mu_0 = 5\pi/4 = 225°$. The data is obtained at the Blåvandshuk Fyr station, on the western coast of Denmark. The station records the wind direction hourly, with direction measured in degrees from North, represented by elements on the circle S^1. Each measurement consists of a single reading from a wind vane, with values ranging from 0 to 359 degrees, representing the direction from which the wind is blowing, meaning that a wind blowing from North to

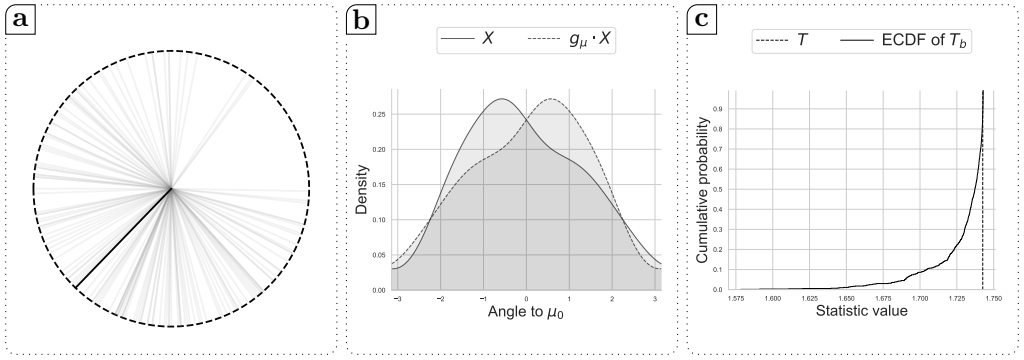


Figure 6: Panel (a) displays the dataset of wind directions collected from the weather station (gray rays) and their empirical Fréchet mean (black ray). Panel (b) shows displays a kernel estimator of the density of the angles $\angle(\mu_0, X)$ (solid) and of $\angle(\mu_0, g^{\mu_0} \cdot X)$ (dashed), after reflection around μ_0, hinting that the wind directions are not radially symmetrical around μ_0. Panel (c) displays the result of isotropic test with the empirical CDF of the test statistic T_b under randomization displayed along with the observed value of the statistic on the original sample T (vertical, dashed).

the South would correspond to a measurement of $\pi/2$. In order to extract a sample as close as possible to being identically distributed with independent observations, we only consider a subset of the data consisting of measurements taken at 12pm every day in from June to November 2024, resulting in 152 observations.

The dataset is displayed in Panel (a) of Figure 6, where the wind directions appear to be evenly distributed around the empirical Fréchet mean $\hat{\mu}_n \approx 3.944 \approx 225.98°$. The proximity of the empirical Fréchet mean to the hypothesized mean μ_0 provides initial qualitative evidence supporting the hypothesis. We further investigate whether the data is symmetrically distributed around μ_0. Panel (b) shows kernel density estimates of the signed angles from μ_0 to X and μ_0 to $g^{\mu_0} \cdot X$ (after reflection around μ_0). The non-overlap of these densities indicates that X is not invariant under isotropies of μ_0, suggesting a lack of radial symmetry around μ_0.

We proceed to test the null hypothesis $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu \neq \mu_0$. We use the isotropic test specialized to the circle S^1 described in Section 2.3 with $B = 1000$ randomizations. The result of the test is displayed in Panel (c) of Figure 6. The empirical CDF of the test

statistic under randomization is displayed, along with the observed value of the statistic T on the original sample. The observed value of the test statistic is $T \approx 1.743$, corresponding to an approximated p-value of $\hat{p}_n \approx 0.64$ which is not significant at the 5% level. This result provides evidence against the alternative hypothesis, suggesting that the wind directions at Blåvandshuk Fyr are not significantly different from the south-westerly pattern.

6 Conclusion

In this paper, we have developed a general framework for testing whether a hypothesized value is the Fréchet mean of a distribution on a metric space. Our approach relies on constructing isotropic randomizations that preserve key geometric properties under the null hypothesis while having power to detect deviations from it. We began by examining the case of the real line, generalizing to the circle S^1 and further to Riemannian manifolds, where the exponential and logarithm maps provide natural tools for constructing such randomizations. Building on these insights, we extended the methodology to general metric spaces through the introduction of admissible reflections.

A key contribution is the characterization of admissible randomizations, which ensures that the resulting test has both correct size and power against alternatives. We have shown that for radially symmetric distributions, our test is particularly well-suited as the randomization preserves distributional properties under the null hypothesis. Our numerical experiments across different metric spaces demonstrate that the test maintains the desired level while achieving good power against alternatives, even in finite samples. The application to wind direction data illustrates the practical utility of our approach in a real-world setting where traditional Euclidean methods are not applicable.

Several directions for future work emerge from this study. First, it remains to prove that the randomization scheme is consistent under the null hypothesis, which would provide a theoretical guarantee that the test has the correct size. Additionally, the development of optimal randomization schemes, particularly for spaces where the isotropy group is rich, presents an interesting theoretical challenge. Finally, further investigation could be brought to the connection between the power of our test and the geometric properties of the underlying metric space.

7 Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 956107, "Economic Policy in Complex Environments (EPOC)".

References

[1] Miroslav Bačák. Computing Medians and Means in Hadamard Spaces. *SIAM Journal on Optimization*, 24(3):1542–1566, January 2014. Publisher: Society for Industrial and Applied Mathematics.

[2] Rajendra Bhatia and John Holbrook. Riemannian geometry and matrix geometric means. *Linear Algebra and its Applications*, 413(2):594–618, March 2006.

[3] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, June 2019.

[4] Rabi Bhattacharya and Lizhen Lin. Omnibus Clts for Fréchet Means and Nonparametric Inference on Non-Euclidean Spaces. *Proceedings of the American Mathematical Society*, 145(1):413–428, 2017. Publisher: American Mathematical Society.

[5] Victor-Emmanuel Brunel and Jordan Serres. Concentration of empirical barycenters in metric spaces, March 2023. arXiv:2303.01144 [math, stat].

[6] Matthieu Bulté and Helle Sørensen. An Autoregressive Model for Time Series of Random Objects, September 2024. arXiv:2405.03778.

[7] Matthieu Bulté and Helle Sørensen. Medoid splits for efficient random forests in metric spaces. *Computational Statistics & Data Analysis*, 198:107995, October 2024.

[8] Isaac Chavel. *Cambridge studies in advanced mathematics: Riemannian geometry: A modern introduction series number 98*. Cambridge studies in advanced mathematics. Cambridge University Press, Cambridge, England, 2 edition, April 2006.

[9] Sinho Chewi, Tyler Maunu, Philippe Rigollet, and Austin J. Stromme. Gradient descent algorithms for Bures-Wasserstein barycenters. In *Proceedings of Thirty Third Conference on Learning Theory*, pages 1276–1304. PMLR, July 2020. ISSN: 2640-3498.

[10] Paromita Dubey and Hans-Georg Müller. Fréchet analysis of variance for random objects. *Biometrika*, 106(4):803–821, December 2019.

[11] Paromita Dubey and Hans-Georg Müller. Fréchet change-point detection. *The Annals of Statistics*, 48(6):3312–3335, December 2020. Publisher: Institute of Mathematical Statistics.

[12] Benjamin Eltzner and Stephan F. Huckemann. A Smeary Central Limit Theorem for Manifolds with Application to High Dimensional Spheres, January 2018. arXiv:1801.06581.

[13] Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, volume 10, pages 215–310, 1948. Issue: 4.

[14] Steve Hanneke, Aryeh Kontorovich, Sivan Sabato, and Roi Weiss. Universal Bayes consistency in metric spaces. *The Annals of Statistics*, 49(4):2129–2150, August 2021. Publisher: Institute of Mathematical Statistics.

[15] Sigurdur Helgason. *Differential geometry, lie groups and symmetric spaces*. Graduate studies in mathematics. American Mathematical Society, Providence, RI, June 2001.

[16] Shayan Hundrieser, Benjamin Eltzner, and Stephan F. Huckemann. A Lower Bound for Estimating Fréchet Means, February 2024. arXiv:2402.12290 [math, stat].

[17] Feiyu Jiang, Hanjia Gao, and Xiaofeng Shao. Testing serial independence of object-valued time series. *Biometrika*, page asad069, November 2023.

[18] Wilfrid S. Kendall. Probability, Convexity, and Harmonic Maps with Small Image I: Uniqueness and Fine Existence. *Proceedings of the London Mathematical Society*, s3-61(2):371–406, 1990. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1112/plms/s3-61.2.371.

[19] Georg Köstenberger and Thomas Stark. Robust Signal Recovery in Hadamard Spaces, July 2023. arXiv:2307.06057 [math, stat].

[20] Huiling Le. On the Consistency of Procrustean Mean Shapes. *Advances in Applied Probability*, 30(1):53–63, 1998. Publisher: Applied Probability Trust.

[21] E L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer texts in statistics. Springer Nature, Cham, Switzerland, 4 edition, June 2022.

[22] Luigi Malagò, Luigi Montrucchio, and Giovanni Pistone. Wasserstein Riemannian geometry of Gaussian densities. *Information Geometry*, 1(2):137–179, December 2018.

[23] A McCormack and P D Hoff. Equivariant estimation of Fréchet means. *Biometrika*, page asad014, February 2023.

[24] Andrew McCormack and Peter Hoff. The Stein effect for Fréchet means. *The Annals of Statistics*, 50(6):3647–3676, December 2022. Publisher: Institute of Mathematical Statistics.

[25] Victor M. Panaretos and Yoav Zemel. *An Invitation to Statistics in Wasserstein Space*. SpringerBriefs in Probability and Mathematical Statistics. Springer International Publishing, Cham, 2020.

[26] Alexander Petersen and Hans-Georg Muller. Frechet regression for random objects with Euclidean predictors. page 29, 2019.

[27] Alexander Petersen and Hans-Georg Müller. Functional data analysis for density functions by transformation to a Hilbert space. *The Annals of Statistics*, 44(1), February 2016. arXiv: 1601.02869.

[28] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer series in statistics. Springer, New York, 2nd ed edition, 2005.

[29] Christof Schötz. Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electronic Journal of Statistics*, 13(2):4280–4345, January 2019. Publisher: Institute of Mathematical Statistics and Bernoulli Society.

[30] Anuj Srivastava and Eric P. Klassen. *Functional and Shape Data Analysis*. Springer Series in Statistics. Springer New York, New York, NY, 2016.

[31] Karl-Theodor Sturm. Probability measures on metric spaces of nonpositive curvature. In Pascal Auscher, Thierry Coulhon, and Alexander Grigor'yan, editors, *Contemporary Mathematics*, volume 338, pages 357–390. American Mathematical Society, Providence, Rhode Island, 2003.

[32] Yann Thanwerdas and Xavier Pennec. Theoretically and Computationally Convenient Geometries on Full-Rank Correlation Matrices. *SIAM Journal on Matrix Analysis and Applications*, 43(4):1851–1872, December 2022. Publisher: Society for Industrial and Applied Mathematics.

[33] A. W. van der Vaart. *Asymptotic statistics*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, UK ; New York, NY, USA, 1998.

[34] Karl Gerald van den Boogaart, Juan José Egozcue, and Vera Pawlowsky-Glahn. Bayes Hilbert Spaces. *Australian & New Zealand Journal of Statistics*, 56(2):171–194, June 2014.

[35] Takumi Yokota. Law of large numbers in CAT(1)-spaces of small radii. *Calculus of Variations and Partial Differential Equations*, 57(2):35, February 2018.

Bibliography

[Alighieri, 1307] Alighieri, D. (1307). *Il Convivio*. Book 2.

[Ambrosio et al., 2005] Ambrosio, L., Gigli, N., and Savaré, G. (2005). *Gradient flows: in metric spaces and in the space of probability measures*. Lectures in mathematics ETH Zürich. Birkhäuser, Boston.

[Athey et al., 2019] Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2).

[Bhattacharya and Lin, 2017] Bhattacharya, R. and Lin, L. (2017). Omnibus Clts for Fréchet Means and Nonparametric Inference on Non-Euclidean Spaces. *Proceedings of the American Mathematical Society*, 145(1):413–428. Publisher: American Mathematical Society.

[Bulté and Sørensen, 2024a] Bulté, M. and Sørensen, H. (2024a). An Autoregressive Model for Time Series of Random Objects. arXiv:2405.03778.

[Bulté and Sørensen, 2024b] Bulté, M. and Sørensen, H. (2024b). Medoid splits for efficient random forests in metric spaces. *Computational Statistics & Data Analysis*, 198:107995.

[Bulté and Sørensen, 2025] Bulté, M. and Sørensen, H. (2025). Isotropic randomization for one-sample testing in metric spaces. arXiv:2501.15945 [stat].

[Burago et al., 2001] Burago, D., Burago, Y., and Ivanov, S. (2001). *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island.

[Dubey and Müller, 2019] Dubey, P. and Müller, H.-G. (2019). Fréchet analysis of variance for random objects. *Biometrika*, 106(4):803–821.

[Eltzner and Huckemann, 2018] Eltzner, B. and Huckemann, S. F. (2018). A Smeary Central Limit Theorem for Manifolds with Application to High Dimensional Spheres. arXiv:1801.06581.

[Fréchet, 1948] Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, volume 10, pages 215–310. Issue: 4.

[Ghodrati and Panaretos, 2024] Ghodrati, L. and Panaretos, V. M. (2024). On distributional autoregression and iterated transportation. *Journal of Time Series Analysis*, 45(5):739–770. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jtsa.12736.

[Lehmann and Romano, 2022] Lehmann, E. L. and Romano, J. P. (2022). *Testing statistical hypotheses*. Springer texts in statistics. Springer Nature, Cham, Switzerland, 4 edition.

[Nesterov, 2018] Nesterov, Y. (2018). *Lectures on convex optimization*. Springer optimization and its applications. Springer International Publishing, Cham, Switzerland, 2 edition.

[Panaretos and Zemel, 2020] Panaretos, V. M. and Zemel, Y. (2020). *An Invitation to Statistics in Wasserstein Space*. SpringerBriefs in Probability and Mathematical Statistics. Springer International Publishing, Cham.

[Petersen and Muller, 2019] Petersen, A. and Muller, H.-G. (2019). Frechet regression for random objects with Euclidean predictors. page 29.

[Qiu et al., 2024] Qiu, R., Yu, Z., and Zhu, R. (2024). Random forest weighted local fréchet regression with random objects. *Journal of Machine Learning Research*, 25(107):1–69.

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

[Schötz, 2019] Schötz, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electronic Journal of Statistics*, 13(2):4280–4345. Publisher: Institute of Mathematical Statistics and Bernoulli Society.

[Sturm, 2003] Sturm, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In Auscher, P., Coulhon, T., and Grigor'yan, A., editors, *Contemporary Mathematics*, volume 338, pages 357–390. American Mathematical Society, Providence, Rhode Island.

[Takatsu and Yokota, 2012] Takatsu, A. and Yokota, T. (2012). Cone structure of l2-wasserstein spaces. *Journal of Topology and Analysis*, 04(02):237–253.

[van der Vaart and Wellner, 2023] van der Vaart, A. W. and Wellner, J. A. (2023). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer International Publishing.

[Wager and Athey, 2018] Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242. Publisher: Taylor & Francis.

[Wu and Shao, 2004] Wu, W. B. and Shao, X. (2004). Limit Theorems for Iterated Random Functions. *Journal of Applied Probability*, 41(2):425–436. Publisher: Applied Probability Trust.

[Zhu and Müller, 2023] Zhu, C. and Müller, H.-G. (2023). Autoregressive optimal transport models. *Journal of the Royal Statistical Society Series B: Statistical*

Methodology, 85(3):1012–1033. _eprint: https://academic.oup.com/jrsssb/article-pdf/85/3/1012/50859838/qkad051.pdf.