EMPIRICAL LIKELIHOOD FOR MANIFOLDS

DAISUKE KURISU AND TAISUKE OTSU

Abstract. There has been growing interest in statistical analysis on random objects taking values in a non-Euclidean metric space. One important class of such objects consists of data on manifolds. This article is concerned with inference on the Fréchet mean and related population objects on manifolds. We develop the concept of nonparametric likelihood for manifolds and propose general inference methods by adapting the theory of empirical likelihood. In addition to the basic asymptotic properties, such as Wilks' theorem of the empirical likelihood statistic, we present several generalizations of the proposed methodology: two-sample testing, inference on the Fréchet variance and local Fréchet regression, quasi Bayesian inference, and estimation of the Fréchet mean set. Simulation and real data examples illustrate the usefulness of the proposed methodology and advantage against the conventional Wald test.

1. Introduction

With increasing availability of more complex data as a background, there has been growing interest in statistical analysis on random objects taking values in a non-Euclidean metric space which may not have algebraic structures; see e.g. Marron and Alonso (2014) for a survey. Examples include data on a circle or sphere, directional data, functional data, and correlation matrices, among others, and one of the most important and well-studied classes of such random objects consists of data on Riemannian manifolds. We refer to Patrangenaru and Ellingson (2015) for an overview of statistical methods on manifolds.

Since Fréchet (1948), statistical theory on manifolds has been widely studied, and perhaps one of the most fundamental concepts in this theory is the Fréchet mean, which is a direct generalization of the standard population mean to a non-Euclidean metric space. To conduct statistical inference on the Fréchet mean or its variants, various methodologies analogous to the conventional Euclidean data analysis have been developed in the literature; see Bhattacharya and Patrangenaru (2014) for a survey.

In this article, we develop a nonparametric likelihood concept for the Fréchet mean and related population objects for data on manifolds, and propose general inference methods (for hypothesis testing and confidence set estimation) by adapting the methodology of empirical likelihood (Owen, 2001). In particular, by exploiting the locally Euclidean structure of Riemannian manifolds, we characterize estimating equations for generalized sample Fréchet means via the associated exponential maps, and construct the empirical likelihood function based on those equations. We study the asymptotic properties of the empirical likelihood statistic and establish

We thank Hans-Georg Müller, Jane-Ling Wang, Fushing Hsieh, Spencer Frei, Shogo Kato, Ryo Okui, Yoshikazu Terada, and Takuya Ura for their helpful comments and discussion. We also thank the participants of the seminars at University of California, Davis, The Institute of Statistical Mathematics, Kyoto University, and Osaka University. Our research is supported by JSPS KAKENHI (JP23K12456) (Kurisu).

Wilks' theorem, convergence of the empirical likelihood statistic to a chi-squared distribution. Furthermore, we propose a plug-in empirical likelihood statistic to deal with composite null hypotheses and describe how to compute critical values by the bootstrap.

A conventional approach for inference (i.e., hypothesis testing and confidence set estimation) of the Fréchet mean on manifolds is the Wald test based on the central limit theorem for the sample Fréchet mean (Bhattacharya and Patrangenaru, 2003 and 2005, Bhattacharya and Lin, 2017). Eltzner and Huckemann (2019) generalized the central limit theorem to a smeary case, where the Hessian of the Fréchet function (i.e., the criterion function to compute the sample Fréchet mean) may be singular. As argued in Eltzner and Huckemann (2019) and Eltzner (2022), the smeary case is practically relevant in manifold data analysis since it is difficult to determine with certainty whether the underlying variable exhibits smeary behavior in certain datasets, such as paleomagnetic data used in our real data example. In such situations, inference on the Fréchet mean has to be treated with great care. Additionally, in the presence of smeariness, the convergence rate and limiting distribution of the sample Fréchet mean depend on (non)singularity of the Hessian of the Fréchet function so that the Wald inference becomes non-trivial. A notable feature of our empirical likelihood approach is as follows. When we test a simple null hypothesis or construct a confidence region for the Fréchet mean, it does not involve any condition on (non)singularity of the Hessian of the Fréchet function, so is asymptotically valid regardless of the degree of smeariness. Moreover, If researchers are interested in the inference of the Fréchet median (a generalization of the standard population median and an important population object as well as the Fréchet means), it can be observed that the Wald test may lack theoretical validity even when the underlying distribution does not exhibit smeariness. Indeed, in our simulation study, we compare the finite sample performance of the empirical likelihood and Wald tests for the Fréchet mean and median when the observations are generated from a von Mises-Fisher distribution on the two-dimensional sphere and find that for the Fréchet median, the Wald test shows severer size distortion but the empirical likelihood test works well. This result is attributed to the singularity of the Hessian matrix of the Fréchet function corresponding to the Fréchet medians. See Section 4 for details on the simulation results.

Based on these benchmark results, we generalize our empirical likelihood approach to several contexts of manifold data analysis. First, we extend the plug-in empirical likelihood statistic to two-sample testing of the Fréchet means. This extension is useful to compare different samples on manifolds. Second, our method can accommodate other population objects on manifolds. In particular, we propose a plug-in empirical likelihood statistic for the Fréchet variance, which is also of interest to investigate random objects including manifold data (see, e.g., Dubey and Müller, 2019 and 2020). Third, we argue that our empirical likelihood can serve as a quasi likelihood function to conduct quasi Bayesian inference on the Fréchet mean, and we provide a consistency result of the proposed quasi posterior. Fourth, the notion of the Fréchet mean has been extended to linear or nonparametric regression contexts (Petersen and Müller, 2019) and we can also construct a localized version of empirical likelihood to conduct inference on the conditional Fréchet mean at a value of Euclidean predictors, which complements the estimation method of local linear Fréchet regression by Petersen and Müller (2019). Fifth, we demonstrate that our empirical likelihood function can be employed as a criterion function to construct a set estimator even when uniqueness of the Fréchet mean is not guaranteed. Limit theorems and estimation for the Fréchet mean set have been studied by Evans and Jaffe (2020), Blanchard and Jaffe (2022), and Schötz (2022). This paper provides an alternative estimation strategy. See also Eltzner (2020) for a test of uniqueness of the Fréchet mean. Finally, we note that all these results contribute to the literature of empirical likelihood (see, Owen, 2001, for a survey) to broaden its scope and applicability.

This article is organized as follows. Section 2 introduces the basic setup and presents our empirical likelihood inference methods for the Fréchet mean. Section 3 discusses several extensions of the empirical likelihood approach for wider applicability. In Sections 4 and 5, simulation results and real data examples are provided, respectively, to illustrate the proposed method. In Appendix, we present popular examples of Riemannian manifolds, a description of the Wald test for simulation, and proofs of the theorems.

2. Empirical likelihood

We first introduce our basic setup. Let $\mathcal{P}$ and $\mathcal{Q}$ be separable topological spaces and $\tilde{\rho}:\mathcal{P}\times\mathcal{Q}\to[0,\infty)$ be a continuous map reflecting distance between a data descriptor $p\in\mathcal{P}$ and a datum $q\in\mathcal{Q}$. Consider an independent and identically distributed sample $\{X_i\}_{i=1}^n$ such that all the X_i have the same distribution as the random object $X\in\mathcal{Q}$. Based on the map $\tilde{\rho}$, the generalized population and sample Fréchet functions are defined as $\tilde{F}(p)=\mathbb{E}[\tilde{\rho}(p,X)]$ and $\tilde{F}_n(p)=n^{-1}\sum_{i=1}^n\tilde{\rho}(p,X_i)$ for $p\in\mathcal{P}$, respectively. Then the generalized population and sample Fréchet means are defined as

$$\tilde{E}=\left\{p\in\mathcal{P}:\tilde{F}(p)=\inf_{q\in\mathcal{P}}\tilde{F}(q)\right\},\qquad \tilde{E}_n=\left\{p\in\mathcal{P}:\tilde{F}_n(p)=\inf_{q\in\mathcal{P}}\tilde{F}_n(q)\right\},\tag{1}$$

respectively. For example, when $\mathcal{P}=\mathcal{Q}$ is a Riemannian manifold and $\tilde{\rho}$ is the squared geodesic intrinsic distance, $\tilde{E}$ and $\tilde{E}_n$ are the population and sample Fréchet means, respectively, originally studied in Fréchet (1948). The population and sample L^q Fréchet means are covered by setting $\tilde{\rho}=d^q$ for some distance d in $\mathcal{P}$. In Appendix A.1, we provide some popular examples of Riemannian manifolds.

Let $\|\cdot\|$ be the Euclidean norm. In this section, we impose the following assumptions.

Assumption 1.

(i): $\{X_i\}_{i=1}^n$ *is independent and identically distributed.* $\tilde{E}$ *is non-empty and contains a unique* $\mu\in\mathcal{P}$ *such that for every measurable selection* $\mu_n\in\tilde{E}_n$*, it holds* $\mu_n\xrightarrow{p}\mu$.

(ii): *For an integer* $r\geq 2$*, there exists a neighborhood* $\tilde{U}$ *of* μ *that is an* m*-dimensional Riemannian manifold, i.e., for a neighborhood* U *of* $0\in\mathbb{R}^m$*, the exponential map* $\exp_\mu:U\to\tilde{U}$ *is a* C^r*-diffeomorphism satisfying* $\exp_\mu(0)=\mu$.

(iii): $g(X,\mu):=\left.\frac{\partial\tilde{\rho}(\exp_\mu(x),X)}{\partial x}\right|_{x=0}$ *exists almost surely, and* $\mathbb{E}[||g(X,\mu)||^2]<\infty$.

Assumptions 1 (i) and (ii), which are identical to Assumptions 2.2 and 2.3 of Eltzner and Huckemann (2019), respectively, describe our basic setup. See Section A.1 for some examples of Riemannian manifolds and their exponential maps. Although uniqueness of μ is commonly assumed, it is also of interest to conduct inference on the generalized Fréchet mean set $\tilde{E}$ (Blanchard and Jaffe, 2022). In Section 3.5 below, we relax this uniqueness assumption and study consistent estimation of $\tilde{E}$ based on our empirical likelihood approach. Assumption 1 (iii) is on the derivative $\left.\frac{\partial \tilde{\rho}(\exp_\mu(x), X)}{\partial x}\right|_{x=0}$, and is weaker than Eltzner and Huckemann (2019, Assumption 2.4). In particular, we do not require the Lipschitz condition for $\tilde{\rho}(\exp_\mu(x), X)$ (Assumption 2.4 (ii) of Eltzner and Huckemann, 2019) nor certain smoothness condition for the Fréchet function $\tilde{F}(p)$ (Assumptions 2.5 and 2.6 of Eltzner and Huckemann, 2019) to establish a general central limit theorem for the empirical likelihood statistics allowing smeariness of the descriptor.

If the generalized sample Fréchet mean μ_n satisfies the first-order condition $n^{-1}\sum_{i=1}^n g(X_i, \mu_n) = 0$, then $g(X, \mu)$ can be interpreted as estimating functions for μ_n. Also note that the origin is the preimage of the generalized population Fréchet mean μ. Therefore, the empirical likelihood function for μ can be constructed as

$$\ell(\mu) = -2 \max_{p_1,\dots,p_n} \sum_{i=1}^n \log(np_i),$$

$$\text{s.t.} \quad p_i \geq 0, \qquad \sum_{i=1}^n p_i = 1, \qquad \sum_{i=1}^n p_i g(X_i, \mu) = 0,$$

and its dual form obtained by the Lagrange multiplier method is (see, e.g., Ch. 2.9 of Owen, 2001)

$$\ell(\mu) = 2 \max_{\lambda} \sum_{i=1}^n \log(1 + \lambda' g(X_i, \mu)). \tag{2}$$

As shown in Theorem 1 (i) below, the empirical likelihood statistic $\ell(\mu)$ can be used to test a simple null hypothesis on μ and to construct a confidence set for μ. In order to test a composite null hypothesis on μ, say $H_0 : \mu \in \mathcal{P}^* \subset \mathcal{P}$, we add the following assumptions.

Assumption 2.

(i): *$\tilde{E}^* = \left\{p \in \mathcal{P}^* : \tilde{F}(p) = \inf_{p^* \in \mathcal{P}^*} \tilde{F}(p^*)\right\}$ is non-empty and contains a unique $\mu^* \in \mathcal{P}^*$ such that for every measurable selection $\mu_n^* \in \tilde{E}_n^* = \left\{p \in \mathcal{P}^* : \tilde{F}_n(p) = \inf_{p^* \in \mathcal{P}^*} \tilde{F}_n(p^*)\right\}$, it holds $\mu_n^* \xrightarrow{p} \mu^*$. For an integer $r^* \geq 2$, there exists a neighborhood $\tilde{U}^*$ of μ^* that is an m^*-dimensional Riemannian manifold, i.e., for a neighborhood U^* of $0 \in \mathbb{R}^{m^*}$, the exponential map $\exp^*_{\mu^*} : U^* \to \tilde{U}^*$ is a C^{r^*}-diffeomorphism (onto a neighborhood of $\mathcal{P}^*$) satisfying $\exp^*_{\mu^*}(0) = \mu^*$. Furthermore, $g^*(X, \mu^*) := \left.\frac{d\tilde{\rho}(\exp^*_{\mu^*}(x), X)}{dx}\right|_{x=0}$ exists almost surely, and $\mathbb{E}[||g^*(X, \mu^*)||^2] < \infty$.*

(ii): *$g^*(X, \exp_\mu(\cdot))$ is continuously differentiable in a neighborhood $\mathcal{N}$ of $0 \in \mathbb{R}^{m^*}$ almost surely. $\mathbb{E}\left[\sup_{x \in \mathcal{N}} \left\|g^*(X, \exp_\mu(x))\right\|^2\right] < \infty$ and $\mathbb{E}\left[\sup_{x \in \mathcal{N}} \left\|\frac{\partial g^*(X, \exp_\mu(x))}{\partial x'}\right\|^2\right] < \infty$. Furthermore,*

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n g^*(X_i, \mu^*) \\ \sqrt{n} x_n \end{pmatrix} \xrightarrow{d} N(0, \Sigma),$$

for some positive semi-definite matrix Σ. Note that x_n is the preimage of the generalized sample Fréchet mean $\mu_n \in \mathcal{P}$ under $\exp_\mu(\cdot)$.

Based on the moment function g^* defined in Assumption 2 (i), the (dual) empirical likelihood function for the generalized population Fréchet mean μ^* on the subspace $\mathcal{P}^*$ is obtained as

$$\ell^*(\mu^*) = 2 \max_{\gamma} \sum_{i=1}^{n} \log(1 + \gamma' g^*(X_i, \mu^*)), \tag{3}$$

and the plug-in empirical likelihood statistic for testing the composite null hypothesis $H_0 : \mu \in \mathcal{P}^*$ is defined as $\ell^*(\mu_n)$. Note that μ_n is the generalized sample Fréchet mean for $\mathcal{P}$. Under H_0, it holds $\mu^* = \mu$ and the test statistic $\ell^*(\mu_n)$ will converge to a limiting distribution shown in Theorem 1 (ii) below. On the other hand, under the alternative hypothesis, the moment condition $\mathbb{E}[g^*(X, \mu)] = 0$ is violated and $\ell^*(\mu_n)$ will diverge to infinity.[1]

The asymptotic properties of the empirical likelihood statistics are obtained as follows.

Theorem 1.

(i): *Under Assumption 1, it holds*

$$\ell(\mu) \xrightarrow{d} \chi^2_m.$$

(ii): *Suppose Assumptions 1 and 2 hold true. Then under $H_0 : \mu \in \mathcal{P}^*$, it holds*

$$\ell^*(\mu_n) \xrightarrow{d} \mathcal{Z}' V^{*-1} \mathcal{Z},$$

where $\mathcal{Z} \sim N\left(0, [I_{m^*} : G^{*\prime}]\Sigma \begin{bmatrix} I_{m^*} \\ G^* \end{bmatrix}\right)$ *with* $G^{*\prime} = \mathbb{E}\left[\frac{\partial g^*(X, \exp_\mu(x))}{\partial x'}\Big|_{x=0}\right]$ *and* $V^* = \mathbb{E}[g^*(X, \mu^*) g^*(X, \mu^*)']$.

Based on Theorem 1 (i), hypothesis testing for the simple null hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ can be implemented by the rejection rule $\{\ell(\mu_0) > \chi^2_{m,1-\alpha}\}$ with the $(1-\alpha)$-th quantile of the χ^2_m distribution. Also the empirical likelihood confidence set for the generalized population Fréchet mean μ is obtained as $ELCS_{1-\alpha} = \{p \in \mathcal{P} : \ell(p) \leq \chi^2_{m,1-\alpha}\}$. When $ELCS_{1-\alpha}$ yields disjoint sets, one can select a subset containing the generalized sample Fréchet mean μ_n, or more cautiously investigate the values of the sample Fréchet function $\tilde{F}_n(p)$ in $ELCS_{1-\alpha}$ to avoid local maxima.

Remark 1. [Bootstrap calibration for $\ell^*(\mu_n)$] Theorem 1 (ii) says that the plug-in empirical likelihood statistic for testing the composite null hypothesis $H_0 : \mu \in \mathcal{P}^*$ is not asymptotically pivotal. Based on the quadratic approximation of $\ell^*(\mu_n)$ presented in (8) in Appendix, its null distribution can be approximated by the bootstrap counterpart

$$\ell^{\#} = n \left(\frac{1}{n} \sum_{i=1}^{n} \{g^*(X_i^{\#}, \mu_n) - \bar{g}_n\} \right)' V_n^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \{g^*(X_i^{\#}, \mu_n) - \bar{g}_n\} \right),$$

where $\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g^*(X_i, \mu_n)$, $V_n = \frac{1}{n}\sum_{i=1}^{n} g^*(X_i, \mu_n) g^*(X_i, \mu_n)'$, and $\{X_i^{\#}\}_{i=1}^{n}$ is a bootstrap resample drawn with equal weights from the original sample. By the conventional bootstrap

[1] An alternative idea is to construct a test statistic $\ell^*(\mu_n^*)$ by plugging-in the estimator μ_n^* under the constraint of $\mu \in \mathcal{P}^*$. However, the computation of μ_n^* is more involved than μ_n, and the derivation of statistical properties of μ_n^* (especially the limiting distribution) is not trivial. Thus, we focus on the statistic $\ell^*(\mu_n)$ and leave the analysis on $\ell^*(\mu_n^*)$ for future research.

theory, we can see that the $(1-\alpha)$-th quantile $q^{\#}_{1-\alpha}$ of the bootstrap resamples of $\ell^{\#}$ yields an asymptotically valid rejection rule $\{\ell^*(\mu_n) > q^{\#}_{1-\alpha}\}$ of $H_0 : \mu \in \mathcal{P}^*$ under certain additional requirements, such as $\mathbb{E}[\sup_{\mu\in\mathcal{N}_0} ||g(X,\mu)||^3] < \infty$ for some neighborhood $\mathcal{N}_0$ around μ^*.

Remark 2. [Robustness of $\ell(\mu)$ against smeariness] It should be noted that Theorem 1 (i) holds true even if Bhattacharya and Patrangenaru's (2005) central limit theorem on the preimage $\sqrt{n}x_n$ does not hold true due to so-called smeariness, singularity of the Hessian of the Fréchet function $\tilde{F}(p)$ (see Assumptions 2.5 and 2.6 in Eltzner and Huckemann, 2019). Hundrieser, Eltzner and Huckemann (2021) proposed a bootstrap approach under their finite sample smeariness framework. On the other hand, our empirical likelihood inference based on $\ell(\mu)$ is asymptotically valid regardless of the degree of smeariness.

Remark 3. [Composite hypothesis testing for smeary case] In contrast to $\ell(\mu)$, the plug-in statistic $\ell^*(\mu_n)$ is not robust against smeariness, but can be applied if the degree of smeariness is known or can be consistently estimated. Suppose x_n is k-th order smeary with $k > 0$ in the sense of Eltzner and Huckemann (2019, Definition 3.3), i.e., $n^{\frac{1}{2(k+1)}} x_n \xrightarrow{d} \mathcal{X}$, where $\mathcal{X}$ has a non-trivial limiting distribution. Then the limiting distribution of $\ell^*(\mu_n)$ under $H_0 : \mu \in \mathcal{P}^*$ becomes

$$n^{-\frac{k}{k+1}} \ell^*(\mu_n) \xrightarrow{d} \mathcal{X}' G^* V^{*-1} G^{*\prime} \mathcal{X}.$$

On the other hand, under $H_1 : \mu \notin \mathcal{P}^*$, we have $\frac{1}{n}\ell^*(\mu_n) \xrightarrow{p} \mathbb{E}[g^*(X,\mu)]' V^{*-1} \mathbb{E}[g^*(X,\mu)] > 0$. Thus, the test with the critical region $\{n^{-\frac{k}{k+1}} \ell^*(\mu_n) > q^s_{1-\alpha}\}$ is asymptotically valid and still consistent, where $q^s_{1-\alpha}$ is an estimator of the $(1-\alpha)$-th quantile of $\mathcal{X}' G^* V^{*-1} G^{*\prime} \mathcal{X}$.

Remark 4. [Goodness-of-fit testing] Suppose the researcher specifies a parametric distribution $X \sim p(x,\theta)$ with finite dimensional parameters $\theta \in \mathbb{R}^{d_\theta}$, which implies the Fréchet mean $\mu(\theta)$. Then we can adapt the plug-in empirical likelihood approach to construct a goodness-of-fit test statistic, that is $\ell(\mu(\hat{\theta}))$ with a $\sqrt{n}$-consistent estimator $\hat{\theta}$ of θ. An analogous argument to the proof of Theorem 1 (ii) (by replacing "$\sqrt{n}x_n$" with the influence function of $\sqrt{n}(\hat{\theta}-\theta)$ combined with suitable smoothness conditions on $\mu(\theta)$) yields the limiting distribution of $\ell(\mu(\hat{\theta}))$ and validity of the bootstrap inference.

Remark 5. [Higher-order refinement] Under additional conditions that require Cramér's condition and higher moments of $g(X,\mu)$, an analogous argument to DiCiccio, Hall and Romano (1991) implies that the empirical likelihood statistic $\ell(\mu)$ in (2) admits the Bartlett correction to achieve the coverage error of order $O(n^{-2})$.

3. Generalizations

The empirical likelihood approach proposed in the last section can be generalized to various statistical inference problems. Here we discuss extensions for two-sample testing (Section 3.1), inference on the Fréchet variance (Section 3.2), Bayesian empirical likelihood inference (Section 3.3), inference on the local Fréchet mean (Section 3.4), and estimation of the generalized population Fréchet mean set (Section 3.5).

3.1. **Two-sample testing.** The plug-in empirical likelihood statistic presented in Theorem 1 (ii) can be naturally extended to two-sample testing problems. Suppose we have two independent random samples $\{X_i\}_{i=1}^n$ and $\{X_{1j}\}_{j=1}^{n_1}$ on the space $\mathcal{Q}$ with the generalized Fréchet means μ and μ_1, respectively, and wish to test the equivalence null hypothesis $H_0 : \mu = \mu_1$ against $H_1 : \mu \neq \mu_1$. The two-sample plug-in empirical likelihood statistic can be constructed as

$$L = \ell(\mu_n) + \ell_1(\mu_n),$$

where μ_n is the generalized sample Fréchet mean based on the merged sample $\{X_i, X_{1j}, : i = 1, \ldots, n, j = 1, \ldots, n_1\}$, the empirical likelihood $\ell(\mu)$ based on $\{X_i\}_{i=1}^n$ is defined as in (2), and the empirical likelihood $\ell_1(\mu_1)$ based on $\{X_{1j}\}_{j=1}^{n_1}$ is defined as $\ell_1(\mu_1) = 2\max_{\lambda_1} \sum_{j=1}^{n_1} \log(1 + \lambda_1' g(X_{1j}, \mu_1))$.

The asymptotic property of the two-sample statistic L is obtained as follows.

Assumption 3. *Random samples $\{X_i\}_{i=1}^n$ and $\{X_{1j}\}_{j=1}^{n_1}$ are independent and satisfy Assumption 1. $g(X, \exp_\mu(\cdot))$ is continuously differentiable in a neighborhood $\mathcal{N}$ of $0 \in \mathbb{R}^m$ almost surely. $\mathbb{E}\left[\sup_{x\in\mathcal{N}} \left\|g(X, \exp_\mu(x))\right\|^2\right] < \infty$ and $\mathbb{E}\left[\sup_{x\in\mathcal{N}} \left\|\frac{\partial g(X,\exp_\mu(x))}{\partial x'}\right\|^2\right] < \infty$. Furthermore, as $n, n_1 \to \infty$ with $n_1/n \to \rho \in (0, \infty)$,*

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_i, \mu) \\ \frac{1}{\sqrt{n_1}} \sum_{j=1}^{n_1} g(X_{1j}, \mu_1) \\ \sqrt{n + n_1} x_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathcal{Z} \\ \mathcal{Z}_1 \\ \mathcal{Z}_x \end{pmatrix} \sim N(0, \Sigma_L), \tag{4}$$

for some positive semi-definite matrix Σ_L where x_n is the preimage of μ_n.

The conditions on the moment functions are concerned to control the local behavior of the plug-in empirical likelihood statistics. The condition in (4) requires that the preimage x_n is asymptotically normal, which excludes the case where the two samples stem from distributions with different orders of smeariness but possibly equal mean. It is possible to weaken the condition in (4) to allow smeariness but we focus on the standard situation to simplify our theoretical analysis. The two-sample test would be more sensitive to the presence of smeariness compared to the one-sample test.

Theorem 2. *Suppose Assumption 3 holds true. Then under $H_0 : \mu = \mu_1$, it holds*

$$\begin{aligned} L \ \xrightarrow{d} \ & (\mathcal{Z} + (1+\rho)^{-1/2} G'\mathcal{Z}_x)' V^{-1} (\mathcal{Z} + (1+\rho)^{-1/2} G'\mathcal{Z}_x) \\ & + (\mathcal{Z}_1 + \rho^{1/2}(1+\rho)^{-1/2} G'\mathcal{Z}_x)' V^{-1} (\mathcal{Z}_1 + \rho^{1/2}(1+\rho)^{-1/2} G'\mathcal{Z}_x), \end{aligned}$$

as $n, n_1 \to \infty$ with $n_1/n \to \rho \in (0, \infty)$, where $G' = \mathbb{E}\left[\left.\frac{\partial g(X,\exp_\mu(x))}{\partial x'}\right|_{x=0}\right]$ and $V = \mathbb{E}[g(X, \mu) g(X, \mu)']$.

Although the limiting distribution of L is not pivotal, it can be approximated by a bootstrap procedure. The bootstrap counterpart of L is obtained as

$$
\begin{aligned}
L^{\#} &= n\left(\frac{1}{n}\sum_{i=1}^{n}\{g(X_i^{\#},\mu_n)-\bar{g}_n\}\right)' V_{n+n_1}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\{g(X_i^{\#},\mu_n)-\bar{g}_n\}\right)\\
&\quad +n_1\left(\frac{1}{n_1}\sum_{j=1}^{n_1}\{g(X_{1j}^{\#},\mu_n)-\bar{g}_{1n_1}\}\right)' V_{n+n_1}^{-1}\left(\frac{1}{n_1}\sum_{j=1}^{n_1}\{g(X_{1j}^{\#},\mu_n)-\bar{g}_{1n_1}\}\right),
\end{aligned}
$$

where $\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i,\mu_n)$, $\bar{g}_{1n_1} = \frac{1}{n_1}\sum_{j=1}^{n_1} g(X_{1j},\mu_n)$,
$V_{n+n_1} = \frac{1}{n+n_1}\left\{\sum_{i=1}^{n} g(X_i,\mu_n)g(X_i,\mu_n)' + \sum_{j=1}^{n_1} g(X_{1j},\mu_n)g(X_j,\mu_n)'\right\}$, and $\{X_i^{\#}\}_{i=1}^{n}$ and $\{X_{1j}^{\#}\}_{j=1}^{n_1}$ are bootstrap resamples drawn with equal weights from the merged original sample $\{X_i, X_{1j},: i=1,\ldots,n, j=1,\ldots,n_1\}$. Then the two-sample test for $H_0:\mu=\mu_1$ can be implemented by the rejection rule $\{L > q_{L,1-\alpha}^{\#}\}$, where $q_{L,1-\alpha}^{\#}$ is the $(1-\alpha)$-th quantile of the bootstrap resamples of $L^{\#}$.

3.2. **Inference on Fréchet variance.** Our empirical likelihood approach on the Fréchet mean can be extended to conduct inference on other population objects for manifolds. For example, researchers might be also interested in the Fréchet variance $\phi = \mathbb{E}[\tilde{\rho}(\mu, X)]$ in addition to the Fréchet mean (e.g., Dubey and Müller, 2019). In this case, by incorporating the estimating function $\tilde{\rho}(\mu, X) - \phi$ for the Fréchet variance, the empirical likelihood function for the pair (μ, ϕ) can be constructed as

$$
\ell_J(\mu,\phi) = -2\max_{p_1,\ldots,p_n}\sum_{i=1}^{n}\log(np_i),
$$

$$
\text{s.t.}\quad p_i \geq 0,\qquad \sum_{i=1}^{n} p_i = 1,\qquad \sum_{i=1}^{n} p_i g(X_i,\mu) = 0,\qquad \sum_{i=1}^{n} p_i\tilde{\rho}(\mu, X_i) = \phi,
$$

whose dual form is $\ell_J(\mu,\phi) = 2\max_{\lambda_J}\sum_{i=1}^{n}\log(1+\lambda_J' g_J(X_i,\mu,\phi))$ with
$g_J(X_i,\mu,\phi) = (g(X_i,\mu)', \tilde{\rho}(\mu, X_i) - \phi)'$.

An analogous argument to Theorem 1 (i) yields Wilks' theorem, $\ell_J(\mu,\phi) \xrightarrow{d} \chi^2_{m+1}$, which can be used to conduct inference on the pair (μ,ϕ). When the Fréchet mean μ is a nuisance object, we can employ the plug-in statistic $\ell_J(\mu_n,\phi)$ for ϕ, and its asymptotic property is presented as follows.

Theorem 3. *Suppose Assumptions 1 (i) and (ii) hold true. Furthermore, assume that*

- **(i):** *$g_J(X,\mu,\phi)$ exists almost surely, and $\mathbb{E}[||g_J(X,\mu,\phi)||^2] < \infty$.*
- **(ii):** *$g_J(X,\exp_\mu(\cdot),\phi)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of $0 \in \mathbb{R}^m$ almost surely. $\mathbb{E}\left[\sup_{x\in\mathcal{N}}\left\|g_J(X,\exp_\mu(x),\phi)\right\|^2\right] < \infty$ and $\mathbb{E}\left[\sup_{x\in\mathcal{N}}\left\|\frac{\partial g_J(X,\exp_\mu(x),\phi)}{\partial x'}\right\|^2\right] < \infty$. Furthermore,*

$$
\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_J(X_i,\mu,\phi) \\ \sqrt{n}x_n \end{pmatrix} \xrightarrow{d} N(0,\Sigma_J),
$$

for some positive semi-definite matrix Σ_J. Then

$$
\ell_J(\mu_n,\phi) \xrightarrow{d} \mathcal{Z}_J' V_J^{-1}\mathcal{Z}_J,
$$

where $\mathcal{Z}_J \sim N\left(0, [I_{m+1} : G_J'] \Sigma_J \begin{bmatrix} I_{m+1} \\ G_J \end{bmatrix}\right)$ *with* $G_J' = \mathbb{E}\left[\frac{\partial g_J(X, \exp_\mu(x), \phi)}{\partial x'}\Big|_{x=0}\right]$ *and* $V_J = \mathbb{E}[g_J(X, \mu, \phi) g_J(X, \mu, \phi)']$.

Since the proof of this theorem is similar to that of Theorem 1 (ii), it is omitted. The limiting distribution of $\ell_J(\mu_n, \phi)$ can be approximated by an analogous bootstrap method in Remark 1. Based on this theorem, we can conduct inference on the Fréchet variance ϕ.

3.3. **Bayesian empirical likelihood.** This subsection considers quasi Bayesian inference for the Fréchet means; see e.g. Bhattacharya and Dunson (2010) and McCormack and Hoff (2022) for nonparametric and empirical Bayes methods on manifold data, respectively. Since our setup does not assume that X has a distribution in a parametric family, it is not clear how to conduct Bayesian inference for μ. Our empirical likelihood $\ell(\mu)$ in (2) can be employed as a quasi likelihood function for quasi Bayesian inference on the generalized Fréchet mean μ (see Lazar, 2003, for the basic idea of Bayesian empirical likelihood). Suppose the researcher has a prior measure $\pi(\mu)$ on $\mu \in \mathcal{P}$. The empirical likelihood-based quasi posterior can be given by

$$\mathbb{P}\{\mu \in \mathcal{B} | \mathbf{X}\} = \frac{\int_{\mu \in \mathcal{B}} \exp(-\ell(\mu)/2 - \varsigma_n \tilde{F}_n(\mu)) d\pi(\mu)}{\int_{\mu \in \mathcal{P}} \exp(-\ell(\mu)/2 - \varsigma_n \tilde{F}_n(\mu)) d\pi(\mu)}, \tag{5}$$

for any Borel set $\mathcal{B}$, where $\mathbf{X} = (X_1, \ldots, X_n)$ and $\{\varsigma_n\}$ is a non-negative sequence satisfying $\varsigma_n \to \infty$. The additional term $\varsigma_n \tilde{F}_n(\mu)$ is introduced to deal with the situation where the space $\mathcal{P}$ contains multiple solutions for the moment condition $\mathbb{E}_{\mu_0}[g(X, \mu)] = 0$. When this moment condition is satisfied uniquely at μ_0, we can set as $\varsigma_n = 0$. Let $d_\mathcal{P}$ be a metric on $\mathcal{P}$. Concentration of the empirical likelihood-based posterior can be characterized as follows.

Theorem 4. *Let* $\mathbb{P}_{\mu_0}$ *be a probability measure of* X *with the Fréchet mean* μ_0*. Suppose that Assumption 1 holds true,* $\mathcal{P}$ *is compact, and* $\mathbb{E}_{\mu_0}[\sup_{\mu \in \mathcal{P}} ||g(X, \mu)||^a] < \infty$ *for some* $a > 1$*. Then for every* $\epsilon > 0$*,*

(i): *if* $\varsigma_n \to \infty$*, it holds*

$$\mathbb{P}\{\mu \in \mathcal{P} : \tilde{F}(\mu) - \tilde{F}(\mu_0) \geq \epsilon | \mathbf{X}\} \to 0 \quad \text{in } \mathbb{P}_{\mu_0}\text{-probability},$$

(ii): *if* $\varsigma_n = 0$*, it holds*

$$\mathbb{P}\{\mu \in \mathcal{P} : ||\mathbb{E}_{\mu_0}[g(X, \mu)]|| \geq \epsilon | \mathbf{X}\} \to 0 \quad \text{in } \mathbb{P}_{\mu_0}\text{-probability},$$

(iii): *if* $\varsigma_n = 0$ *and* $\mathbb{E}_{\mu_0}[g(X, \mu)] = 0$ *uniquely at* $\mu_0 \in \mathcal{P}$*, it holds*

$$\mathbb{P}\{d_\mathcal{P}(\mu, \mu_0) \geq \epsilon | \mathbf{X}\} \to 0 \quad \text{in } \mathbb{P}_{\mu_0}\text{-probability}.$$

Part (iii) of this theorem says that when the moment condition $\mathbb{E}_{\mu_0}[g(X, \mu)] = 0$ is uniquely satisfied at $\mu = \mu_0$, the empirical likelihood-based quasi posterior without adjustment (i.e., $\mathbb{P}^{EL}\{\mu \in \mathcal{B} | \mathbf{X}\} = \int_{\mu \in \mathcal{B}} \exp(-\ell(\mu)/2) d\pi(\mu) / \int_{\mu \in \mathcal{P}} \exp(-\ell(\mu)/2) d\pi(\mu)$) achieves posterior consistency to μ_0 under the metric $d_\mathcal{P}$. When the moment condition is satisfied at multiple μ's, then the posterior $\mathbb{P}^{EL}\{\mu \in \mathcal{B}|\mathbf{X}\}$ guarantees concentration only to the set of those multiple solutions containing μ_0 (Part (ii) of this theorem). Part (i) of this theorem guarantees posterior concentration to the argmin set of the population Fréchet function $\tilde{F}(\cdot)$ for a general case.

3.4. **Local empirical likelihood.** Petersen and Müller (2019) generalized local linear fitting to the case where the response is a random object and predictors are Euclidean variables, and developed the local Fréchet regression method to estimate a conditional or localized version of the Fréchet mean. Indeed our empirical likelihood approach can be extended to deal with such localized population objects. Let $Z \in \mathbb{R}^k$ be Euclidean predictors. The generalized local population and sample Fréchet means are defined as

$$\tilde{E}(z) = \left\{ p \in \mathcal{P} : \tilde{F}(p; z) = \inf_{q \in \mathcal{P}} \tilde{F}(q; z) \right\}, \qquad \tilde{E}_n(z) = \left\{ p \in \mathcal{P} : \tilde{F}_n(p; z) = \inf_{q \in \mathcal{P}} \tilde{F}_n(q; z) \right\},$$

respectively, where $\tilde{F}(p; z) = \mathbb{E}[\tilde{\rho}(p, X)|Z = z]$ and $\tilde{F}_n(p; z) = \frac{1}{nh^k} \sum_{i=1}^n K\left(\frac{Z_i - z}{h}\right) \tilde{\rho}(p, X_i)$ with a kernel function K and bandwidth h. The object of interest is the generalized local population Fréchet mean at given z. In this case, Assumption 1 is adapted as follows.

Assumption 4.

(i): $\{X_i, Z_i\}_{i=1}^n$ *is independent and identically distributed. For almost every* z, $\tilde{E}(z)$ *is non-empty and contains a unique* $\mu_z \in \mathcal{P}$ *such that for every measurable selection* $\mu_{z,n} \in \tilde{E}_n(z)$, *it holds* $\mu_{z,n} \xrightarrow{p} \mu_z$.

(ii): *For an integer* $r \geq 2$, *there exists a neighborhood* $\tilde{U}$ *of* μ_z *that is an* m*-dimensional Riemannian manifold, i.e., for a neighborhood* U *of* $0 \in \mathbb{R}^m$, *the exponential map* $\exp_\mu : U \to \tilde{U}$ *is a* C^r*-diffeomorphism satisfying* $\exp_{\mu_z}(0) = \mu_z$.

(iii): $g(X, \mu_z) := \left.\frac{d\tilde{\rho}(\exp_{\mu_z}(x), X)}{dx}\right|_{x=0}$ *exists almost surely.*

Based on this assumption, a localized version of the empirical likelihood function for the local Fréchet mean μ_z can be constructed as

$$\ell(\mu_z; z) = -2 \max_{p_1, \ldots, p_n} \sum_{i=1}^n \log(np_i),$$

$$\text{s.t.} \quad p_i \geq 0, \qquad \sum_{i=1}^n p_i = 1, \qquad \sum_{i=1}^n p_i K\left(\frac{Z_i - z}{h}\right) g(X_i, \mu_z) = 0,$$

and its dual form is

$$\ell(\mu_z; z) = 2 \max_{\lambda} \sum_{i=1}^n \log\left(1 + \lambda' K\left(\frac{Z_i - z}{h}\right) g(X_i, \mu_z)\right).$$

The asymptotic property of the local empirical likelihood statistic $\ell(\mu_z; z)$ is obtained as follows.

Theorem 5. *Suppose that Assumption 4 holds true. Additionally assume*

$$\begin{aligned}
&\frac{1}{\sqrt{nh^k}}\sum_{i=1}^n\left\{K\left(\frac{Z_i-z}{h}\right)g(X_i,\mu_z)-\mathbb{E}\left[K\left(\frac{Z_i-z}{h}\right)g(X_i,\mu_z)\right]\right\}\xrightarrow{d}N(0,V_z),\\
&\sqrt{\frac{n}{h^k}}\mathbb{E}\left[K\left(\frac{Z_i-z}{h}\right)g(X_i,\mu_z)\right]\to 0, \qquad (6)\\
&\frac{1}{nh^k}\sum_{i=1}^n K\left(\frac{Z_i-z}{h}\right)^2 g(X_i,\mu_z)g(X_i,\mu_z)'\xrightarrow{p}V_z,
\end{aligned}$$

for some positive definite V_z. Then

$$\ell(\mu_z;z)\xrightarrow{d}\chi^2_m.$$

Standard regularity conditions on the moment function $g(X_i,\mu_z)$ combined with certain requirements on the kernel function K and bandwidth h will be sufficient for (6) to hold. The second condition in (6) requires undersmoothing to ignore the bias component. Based on Theorem 5, we can conduct empirical likelihood inference on the generalized local population Fréchet mean μ_z for each z, which complements the point estimation theory developed in Petersen and Müller (2019).

3.5. **Fréchet mean set.** In this subsection, we consider the case where the generalized population Fréchet mean set $\tilde{E}$ in (1) is not a singleton (i.e., Assumption 1 (i) is violated). One way to adapt our approach in such a scenario is to estimate the subset $\tilde{\mathcal{P}}:=\{p\in\mathcal{P}:\mathbb{E}[g(X,p)]=0\}$, which contains $\tilde{E}$ as a subset, based on the empirical likelihood function $\ell(\cdot)$ in (2).

In particular, by applying the general methodology of Chernozhukov, Hong and Tamer (2007) to estimate set identified statistical models, the empirical likelihood-based set estimator for $\tilde{\mathcal{P}}\supseteq\tilde{E}$ is constructed as a level set:

$$\hat{\mathcal{P}}=\{p\in\mathcal{P}:\ell(p)\le C\log n\}\quad\text{for some } C>0.$$

Let $d_H(A,B)=\max\{\sup_{a\in A}(\inf_{b\in B}d(a,b)),\sup_{b\in B}(\inf_{a\in A}d(a,b))\}$ be the Hausdorff distance of subsets $A,B\subset\mathcal{P}$. Consistency of the set estimator $\hat{\mathcal{P}}$ for $\tilde{\mathcal{P}}$ under d_H is obtained as follows.

Theorem 6. *Suppose that*

- **(i):** *$\{X_i\}_{i=1}^n$ is a collection of independent and identically distributed random variables defined on a complete probability space $(\Omega,\mathcal{F},\mathbb{P})$. For each $p\in\mathcal{P}$, there exists a neighborhood $\tilde{U}$ of p that is an m-dimensional Riemannian manifold, i.e., for a neighborhood U of $0\in\mathbb{R}^m$, the exponential map $\exp_p:U\to\tilde{U}$ is a C^r-diffeomorphism satisfying $\exp_p(0)=p$ with some $r\ge 2$.*
- **(ii):** *For each $p\in\mathcal{P}$, $g(X,p):=\left.\frac{\partial\tilde{\rho}(\exp_p(x),X)}{\partial x}\right|_{x=0}$ exists almost surely, and $\mathbb{E}\left[\sup_{p\in\mathcal{P}}||g(X,p)||^2\right]<\infty$.*
- **(iii):** *$\{g(\cdot,p):p\in\tilde{\mathcal{P}}\}$ is a $\mathbb{P}$-Donsker class, and $\inf_{p\in\mathcal{P}}\lambda_{\min}(\mathbb{E}[g(X,p)g(X,p)'])\ge c$ for some $c>0$, where $\lambda_{\min}(A)$ means the minimum eigenvalue of a matrix A.*

Then

$$d_H(\hat{\mathcal{P}},\tilde{\mathcal{P}})\xrightarrow{p}0.$$

Assumptions (i) and (ii) in Theorem 6 are uniform versions of Assumption 1, but we do not require uniqueness of the generalized population Fréchet mean. Assumption (iii) arises from empirical process theory (c.f. Section 2.1 in van der Vaart and Wellner, 1996). In particular, the requirement that $\{g(\cdot, p) : p \in \tilde{\mathcal{P}}\}$ is a Donsker class is used to control the stochastic order of the criterion function $\ell(p)$ over the identified set $\tilde{\mathcal{P}}$. The proof is an adaptation of Chernozhukov, Hong and Tamer (2007, Theorem 3.1) to the present setup.

4. Simulation

In this section, we conduct a simulation study to evaluate the finite sample performance of the proposed method. We consider the case, where $\mathcal{P}$ and $\mathcal{Q}$ are 2-dimensional spheres $\mathbb{S}^2$, and focus on inference for the generalized Fréchet mean with the geodesic intrinsic distance $\tilde{\rho} = d$ (hereafter, called the Fréchet median) and its square d^2 (hereafter, called the Fréchet mean).

We generate an independent and identically distributed sequence $\{X_i\}_{i=1}^n$ from (i) the von Mises-Fisher distribution on $\mathbb{S}^2$ with the mean direction $\mu_0 = (0, 0, 1)'$ and concentration parameter $\kappa \in \{1, 2\}$, or (ii) the two-smeary distribution Q provided in Eltzner and Huckemann (2019), which is defined as follows. Let X be a random variable distributed on $\mathbb{S}^2$ that is uniformly distributed on the lower half sphere $\mathbb{L}^2 = \{p = (p_1, p_2)' \in \mathbb{S}^2 : p_2 \leq 0\}$ with total mass $\alpha = \frac{4}{4+\pi}$ and on the north pole $\mu = (0, 0, 1)'$ with probability $1 - \alpha$. We set $n \in \{200, 500\}$ for Case (i), and $n \in \{200, 500, 1000, 2000\}$ for Case (ii).

First, we consider inference on the mean/median direction $\mu_0 = (0, 0, 1)'$. Figure 1 shows the sample and population Fréchet means (red and orange points, respectively) with the 95% empirical likelihood confidence region (red line) of Case (i) with $\kappa \in \{1, 2\}$ and $n = 200$. Figure 2 shows the sample and population Fréchet medians (purple and orange points, respectively) with the 95% empirical likelihood confidence region (red line) of Case (i) with $\kappa \in \{1, 2\}$ and $n = 200$. The black points are observations.

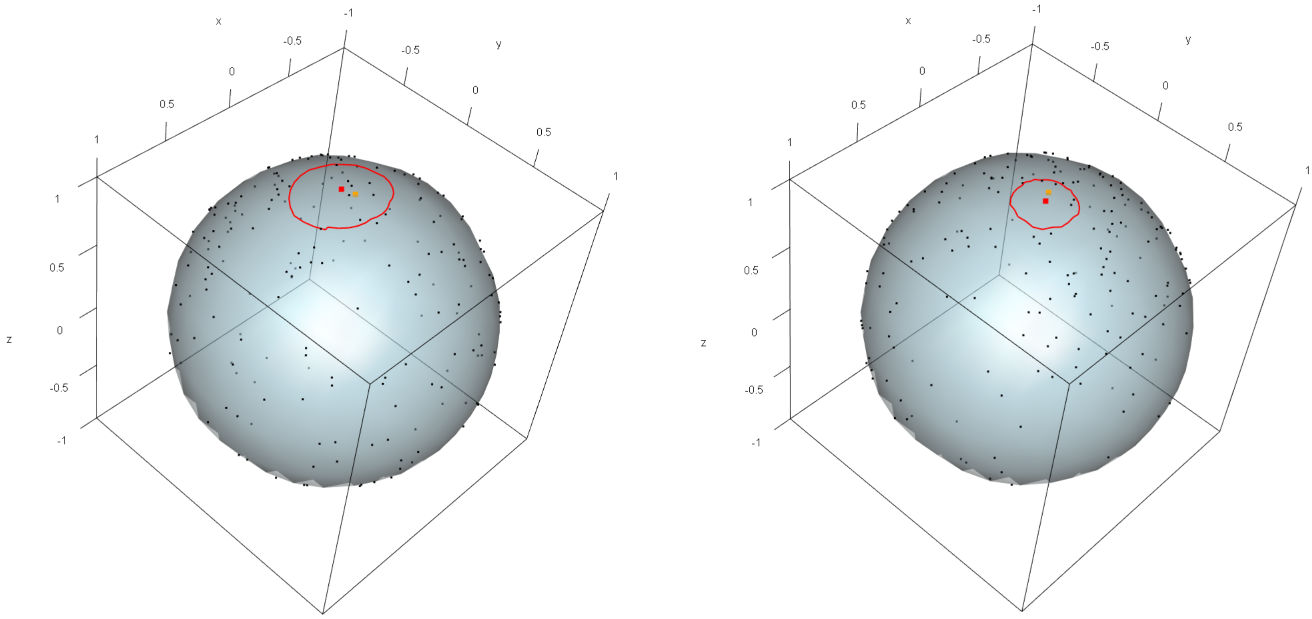

FIGURE 1. Sample Fréchet mean (red point) with 95% empirical likelihood confidence region (red line). The orange point is the true mean direction and the black points are observations of Case (i) with $(\kappa, n) = (1, 200)$ (left) and $(\kappa, n) = (2, 200)$ (right).

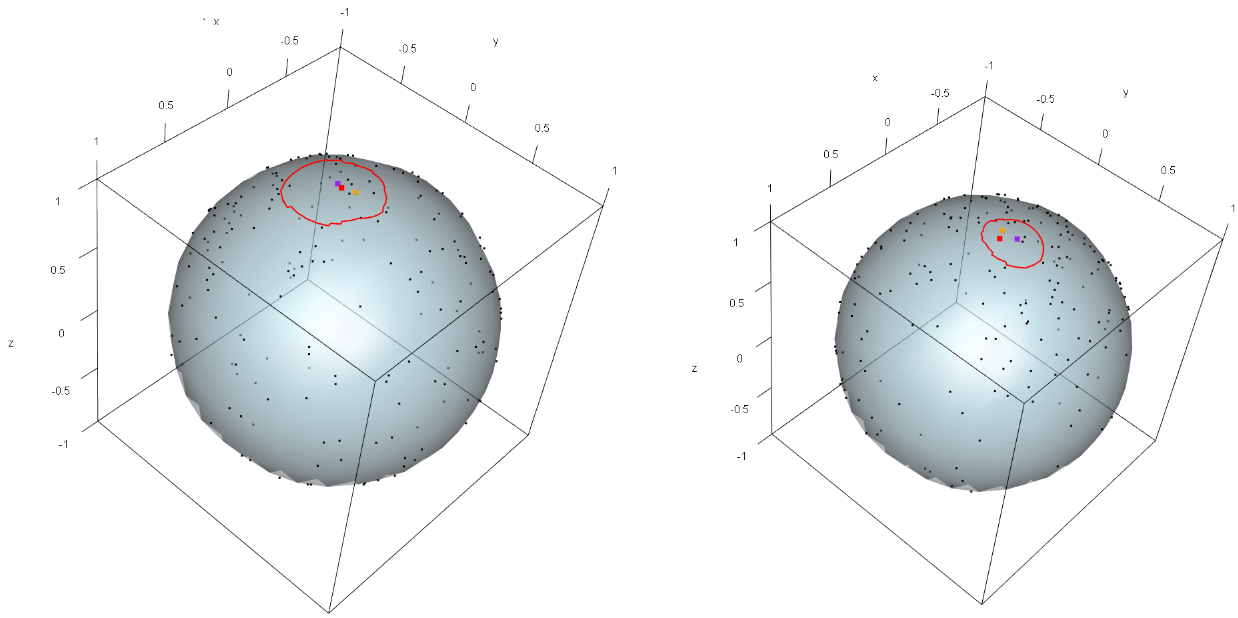

FIGURE 2. Sample Fréchet median (purple point) with 95% EL confidence region (red line). The orange point is the true mean direction, the red point is the sample Fréchet mean, and the black points are observations of Case (i) with $(\kappa, n) = (1, 200)$ (left) and $(\kappa, n) = (2, 200)$ (right).

Second, we consider hypothesis testing for the null $H_0 : \mu = \mu_0 = (0, 0, 1)'$ against the alternative $H_1 : \mu \neq \mu_0$. Note that Case (ii) is an example that exhibits two-smeariness for the Fréchet mean and the Fréchet median is not smeary in this model. We set the significance level at 5%, and compare the empirical likelihood test based on Theorem 1 (i) and the conventional Wald test described in Appendix A.2 below. Figure 3 shows empirical sizes and powers of the empirical likelihood test and Wald test based on the Fréchet mean, and Figure 4 shows empirical sizes and powers of the empirical likelihood test and Wald test based on the Fréchet median for Case (i) with $\mu = (\sin\theta\cos\xi, \sin\theta\sin\xi, \cos\xi)'$, $\theta \in \{0, \frac{\pi}{64}, \frac{2\pi}{64}, \ldots, \frac{16\pi}{64}\}$, $\xi = 0$, and $\kappa \in \{1, 2\}$. Figure 5 shows empirical sizes and powers of the empirical likelihood test on the Fréchet mean and median for Case (ii) with $\theta \in \{0, \frac{\pi}{16}, \frac{2\pi}{16}, \ldots, \pi\}$, $\mu = (\sin\theta\cos\xi, \sin\theta\sin\xi, \cos\xi)'$, $\xi = 0$, and $n \in \{200, 500, 1000, 2000\}$. For Case (ii), we do not report the results of the Wald test due to the erroneous numerical behaviors and lack of theoretical justification. Note that μ_0 corresponds to the case with $(\theta, \xi) = (0, 0)$. The number of Monte Carlo repetitions is 1000. Our findings are summarized as follows:

- For Case (i) on the Fréchet mean, both tests exhibit reasonable size properties, but the empirical likelihood test is more powerful than the Wald test. The power gain is particularly large for the less concentrated case, $\kappa = 1$.
- For Case (i) on the Fréchet median, we find that the Wald test shows severer size distortions (around 40%). This distortion is due to the divergence of the Hessian matrix components. On the other hand, the empirical likelihood test has reasonable size and power. Therefore, in this case, the proposed empirical likelihood inference clearly outperforms the conventional Wald test.
- For Case (ii) on the Fréchet mean and median, we only consider the empirical likelihood test (due to erroneous behaviors of the Wald test), and it can be observed that the empirical likelihood test gradually improves the power in both the Fréchet mean and Fréchet median as the sample size increases. Since the Fréchet mean is smeary in Case (ii), the results provide clear illustrations of the effect of smeariness decreasing the power of the test.

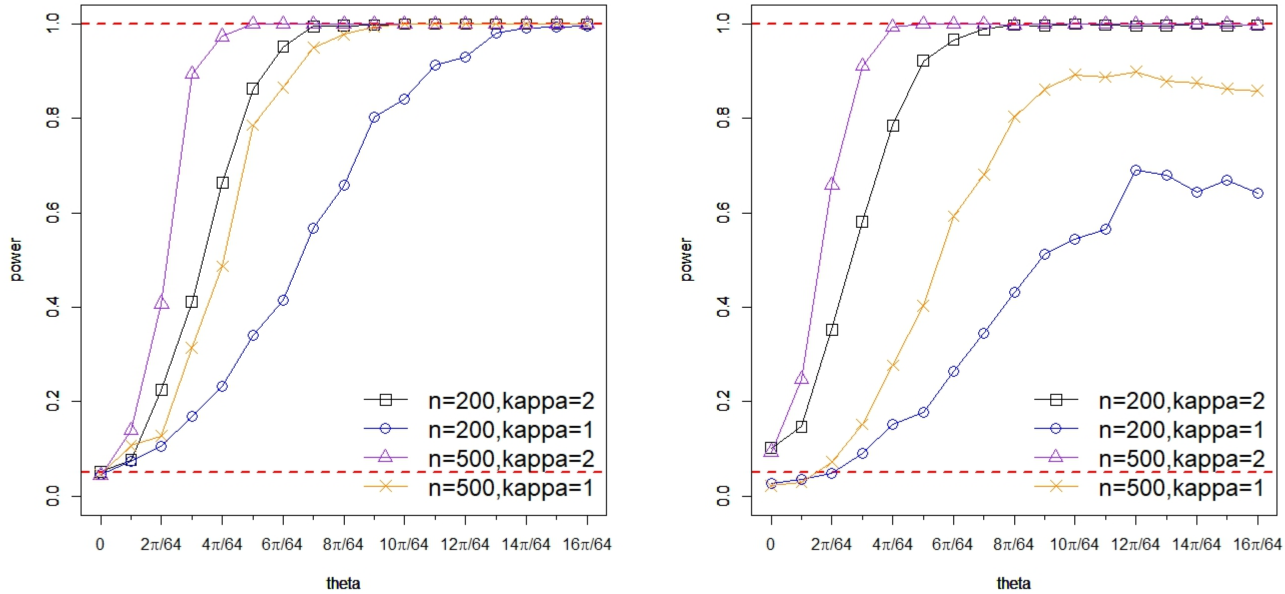

FIGURE 3. Empirical sizes and powers of the EL test (left) and Wald test (right) based on the Fréchet mean for Case (i) with $\theta \in \{0, \frac{\pi}{64}, \frac{2\pi}{64}, \ldots, \frac{16\pi}{64}\}$, $n \in \{200, 500\}$, and $\kappa \in \{1, 2\}$.

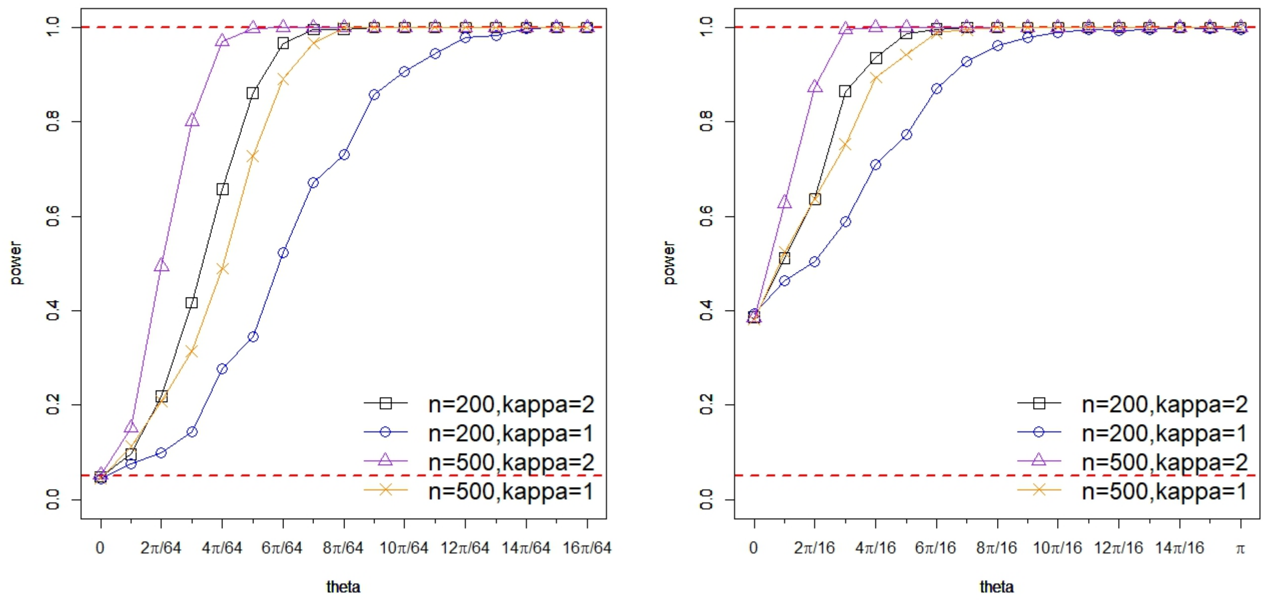

FIGURE 4. Empirical sizes and powers of the EL test (left) and Wald test (right) based on the Fréchet median for Case (i) with $\theta \in \{0, \frac{\pi}{64}, \frac{2\pi}{64}, \ldots, \frac{16\pi}{64}\}$, $n \in \{200, 500\}$, and $\kappa \in \{1, 2\}$.

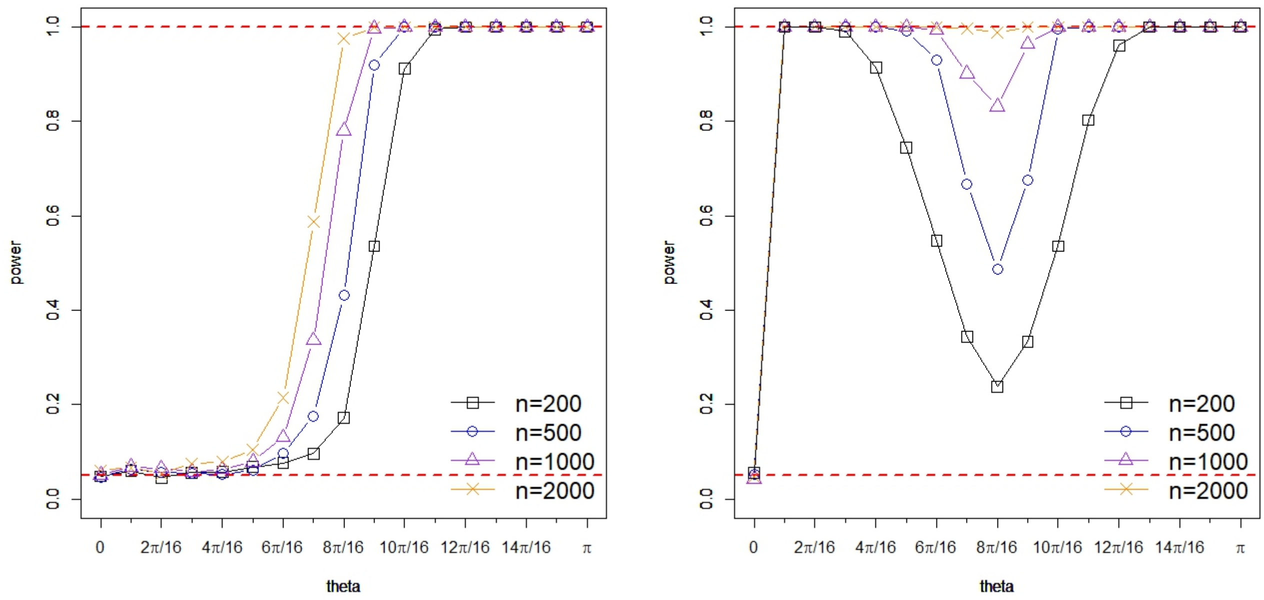

FIGURE 5. Empirical sizes and powers of the EL test of Fréchet mean (left) and Fréchet median (right) for Case (ii) with $\theta \in \{0, \frac{\pi}{16}, \frac{2\pi}{16}, \dots, \pi\}$ and $n \in \{200, 500, 1000, 2000\}$.

5. Real data analysis

In this section, we apply our empirical likelihood approach to conduct inference on the Fréchet mean and median of (i) turtle data (a dataset on circle) and (ii) paleomagnetic data (a dataset on sphere).

5.1. Turtle data.

In this subsection, we apply our method to the dataset of directions of 76 female turtles after laying eggs from Mardia and Jupp (2000). This dataset is a motivating example in Eltzner and Huckemann (2019) for the statistical analysis of generalized Fréchet means in the presence of smeariness.

Figure 6 shows the plot of the turtle data. Figure 7 presents the 95% confidence regions (red line) of Fréchet mean (left) and Fréchet median (right). The orange point is the sample Fréchet mean and the purple point is the sample Fréchet median. In this example, the confidence region of the Fréchet median is included within the confidence region of the Fréchet mean, and the sample Fréchet mean lies within the confidence region of the Fréchet median. Figure 8 shows the plots of the empirical likelihood statistics $\ell(\mu)$ for the Fréchet mean (left) and Fréchet median (right) of the turtle data. We set $\mu = (\cos\theta, \sin\theta)$ for $\theta \in \{0, \frac{\pi}{128}, \frac{2\pi}{128}, \dots, \pi\}$. The red (purple) horizontal and virtual lines correspond to the critical value $q_{0.95}$ and the sample Fréchet mean (median), respectively. One can use these figures for the visualization of confidence regions of the Fréchet mean/median or Fréchet mean/median sets. In Figure 8, we obtain the disjoint confidence sets for the Fréchet mean/median. In these cases, we choose the subsets containing the sample Fréchet mean/median since other sets correspond to local maxima of the sample Fréchet function as shown in Figure 9.

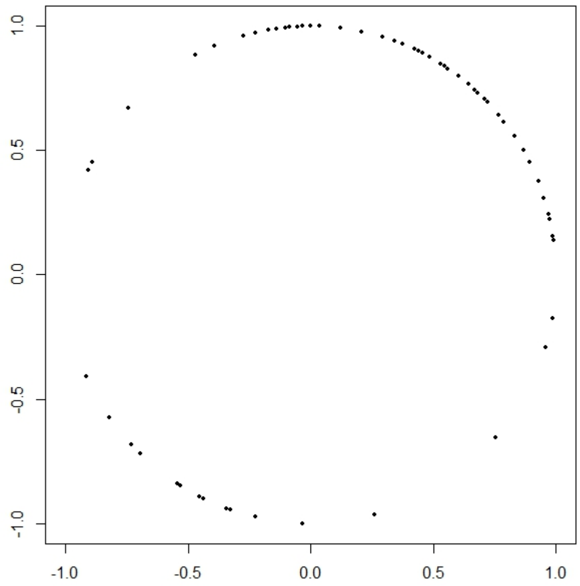

FIGURE 6. Turtle data in Mardia and Jupp (2000).

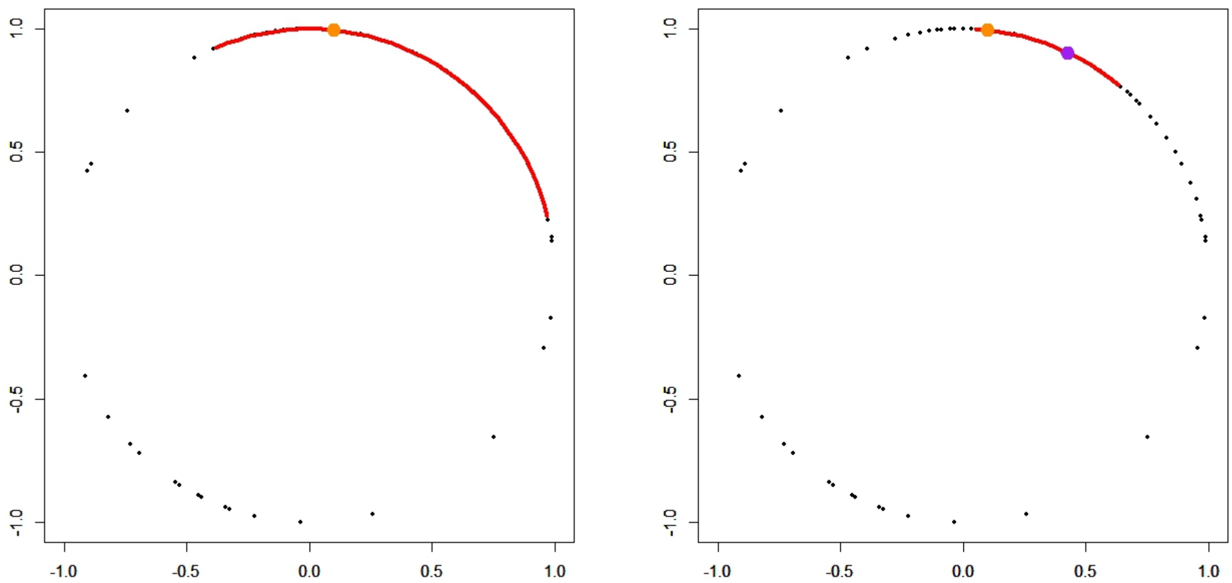

FIGURE 7. 95% confidence regions (red line) of the Fréchet mean (left) and the Fréchet median (right). The orange point is the sample Fréchet mean and the purple point is the sample Fréchet median.

5.2. **Paleomagnetic data.** Paleomagnetism provides highly valuable information in earth sciences, including geology (Butler, 1992). Among them, the analysis of virtual geomagnetic poles (VGPs) serves as crucial data to understand the changes in the positions of geomagnetic poles from the past to the present.

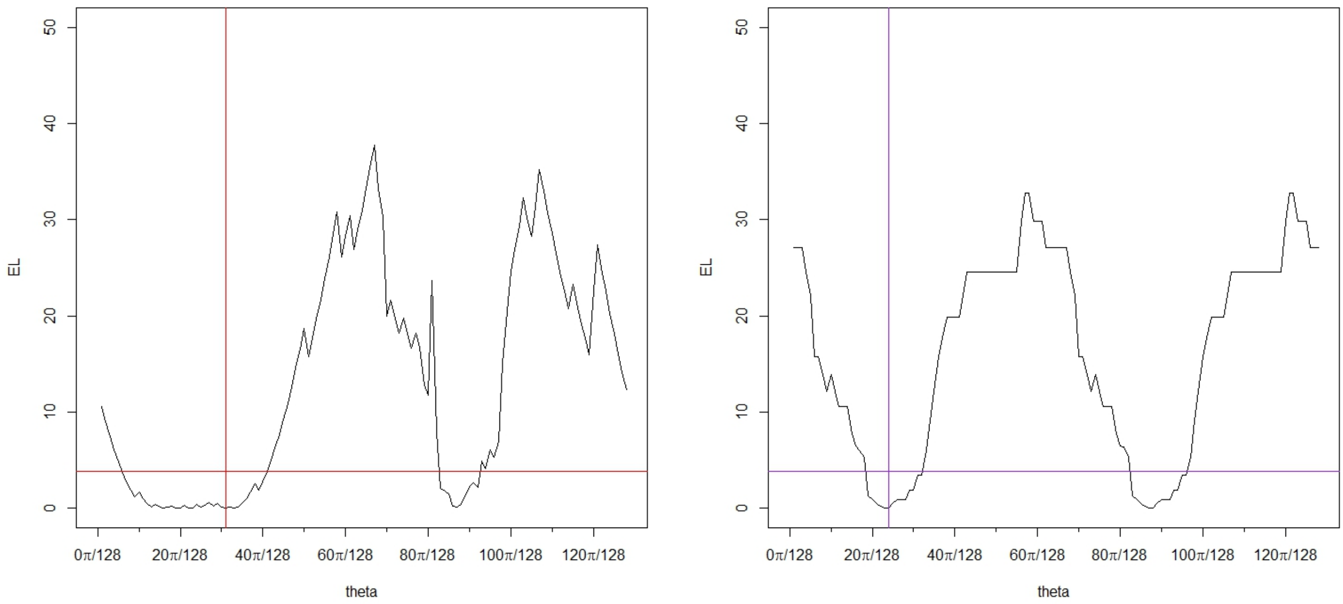

FIGURE 8. Plot of the EL statistics $\ell(\mu)$ for the Fréchet mean (left) and the Fréchet median (right) of turtle data. We set $\mu = (\cos\theta, \sin\theta)$, $\theta \in \{0, \frac{\pi}{128}, \frac{2\pi}{128}, \ldots, \pi\}$. The red (purple) horizontal and virtual lines correspond to the critical value $q_{0.95}$ and the sample Fréchet mean (median), respectively.

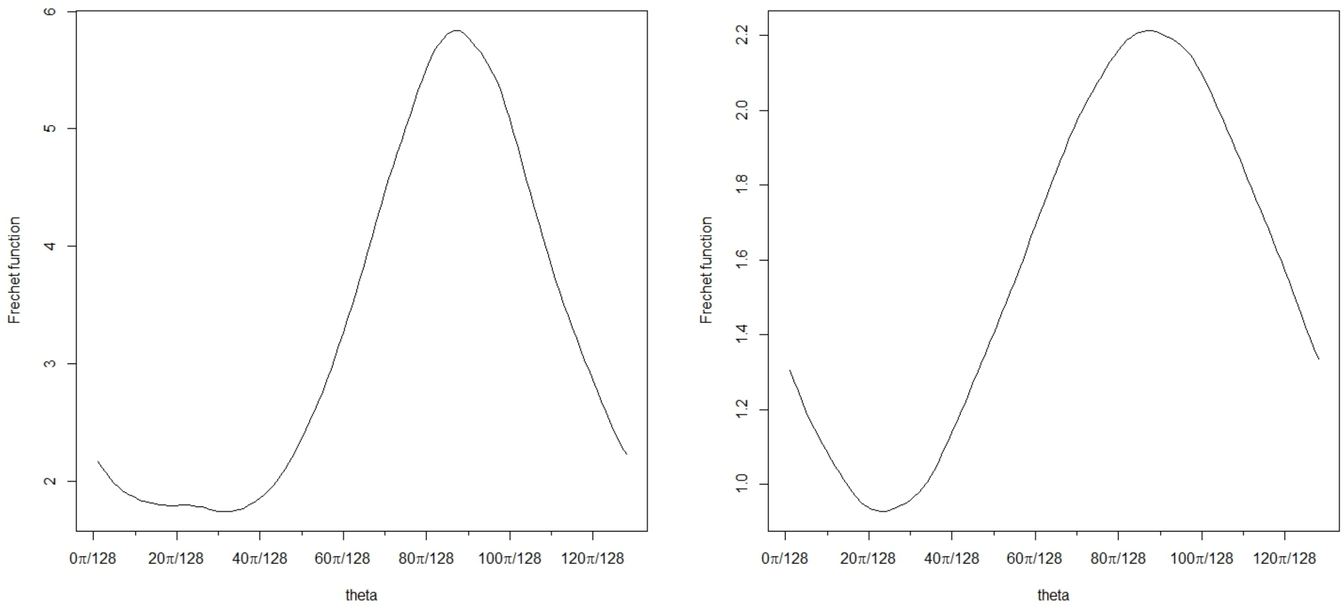

FIGURE 9. Sample Fréchet function of the Fréchet mean (left) and the Fréchet median (right) of turtle data.

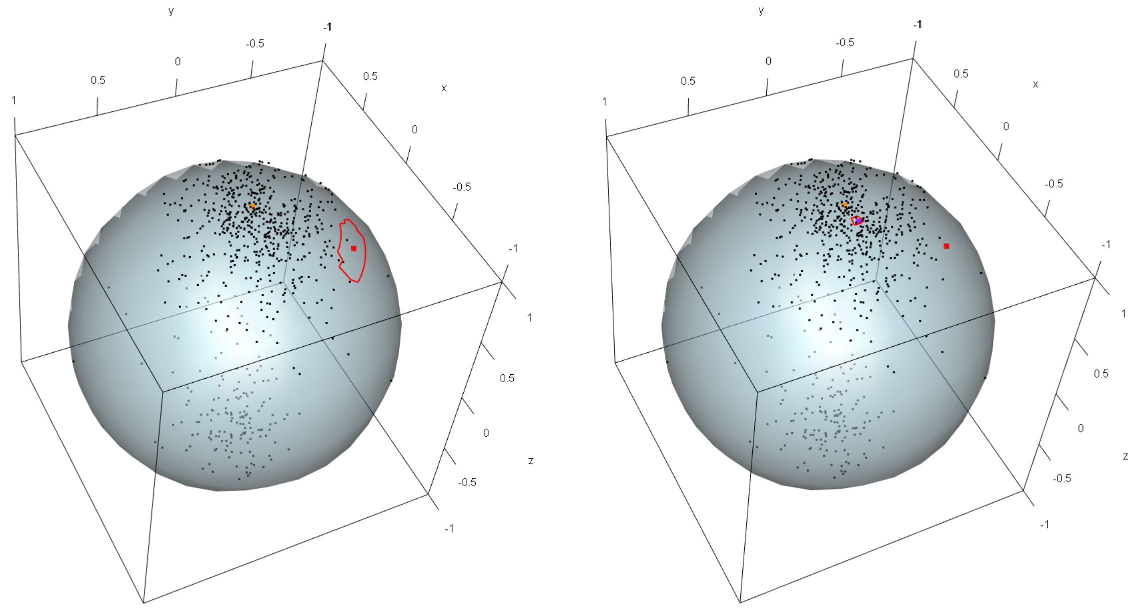

FIGURE 10. 95% confidence regions (red line) of the Fréchet mean (left) and the Fréchet median (right). The red (purple) point is the sample Fréchet mean (median), the orange point is the north pole, and the black points are VGP locations.

In this subsection, we utilize the dataset of VGP positions analyzed in Gallo et al. (2023) with a sample size $n = 689$. We estimate the average positions of geomagnetic poles from 60 million years ago to the present using the sample Fréchet mean and the sample Fréchet median. Additionally, we also computed 95% confidence regions of the population Fréchet mean and median. Given the observation in Eltzner (2022) that the dataset of VGP positions tends to exhibit smeariness, it seems prudent to construct a confidence region using our empirical likelihood method, which is robust to the smeariness.

Figure 10 shows 95% confidence regions (red line) of the Fréchet mean (left) and the Fréchet median (right). The red (purple) point is the sample Fréchet mean (median), the orange point is the north pole, and the black points are VGP locations. In this example, the confidence region of the Fréchet mean and the confidence region of the Fréchet median are not in an inclusion relationship. Specifically, while the sample Fréchet mean is somewhat distant from the north pole, the sample Fréchet median is close to it, and its confidence region is very narrow. Figure 11 shows the plots of the EL statistics $\ell(\mu)$ for the Fréchet mean (left) and the Fréchet median (right) of VGP data and we obtain the disjoint confidence sets for the Fréchet mean/median. In these cases, we selected the subsets containing the sample Fréchet mean/median since other sets correspond to local maxima or minima of the sample Fréchet functions as shown in Figure 12. We set $\mu = (\sin\theta\cos\xi, \sin\theta\sin\xi, \cos\theta)'$ for $\theta \in \{0, \frac{\pi}{64}, \frac{2\pi}{64}, \ldots, \pi\}$ and $\xi \in \{0, \frac{\pi}{128}, \frac{2\pi}{128}, \ldots, 2\pi - \frac{\pi}{128}\}$. As these figures show, the shapes of the empirical likelihood confidence regions are flexibly determined by the data.

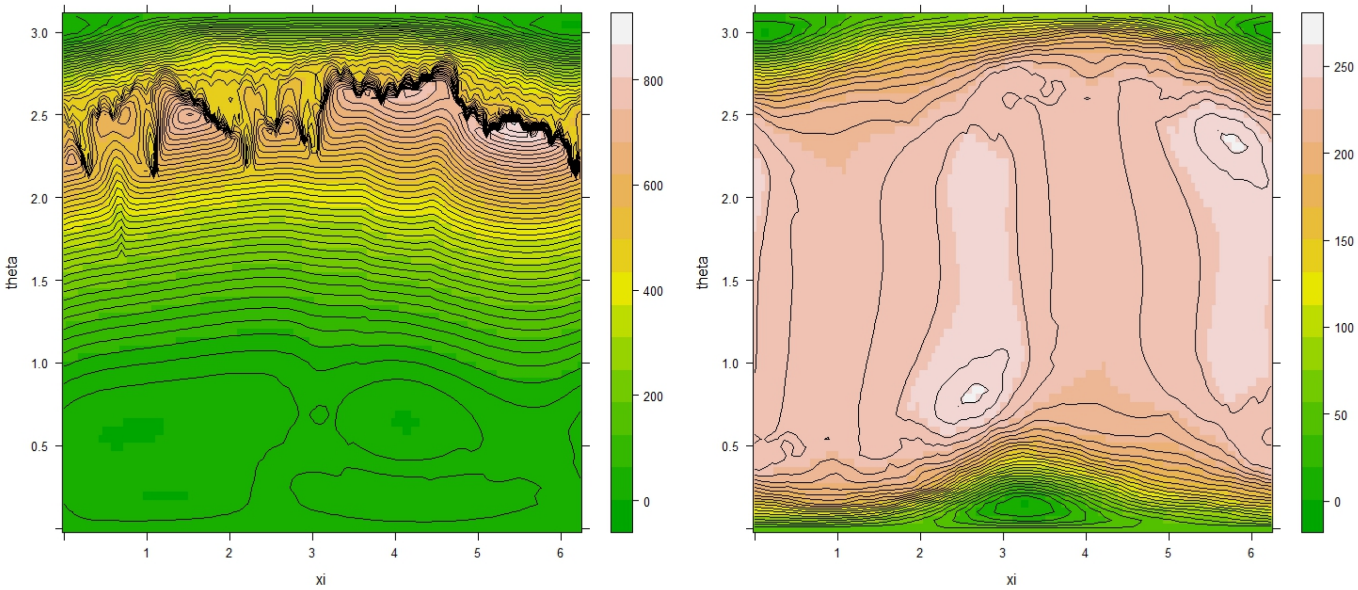

FIGURE 11. Plots of the EL statistics $\ell(\mu)$ for the Fréchet mean (left) and the Fréchet median (right) of VGP data. We set $\mu = (\sin\theta\cos\xi, \sin\theta\sin\xi, \cos\theta)'$, $\theta \in \{0, \frac{\pi}{64}, \frac{2\pi}{64}, \ldots, \pi\}$, $\xi \in \{0, \frac{\pi}{128}, \frac{2\pi}{128}, \ldots, 2\pi - \frac{\pi}{128}\}$.

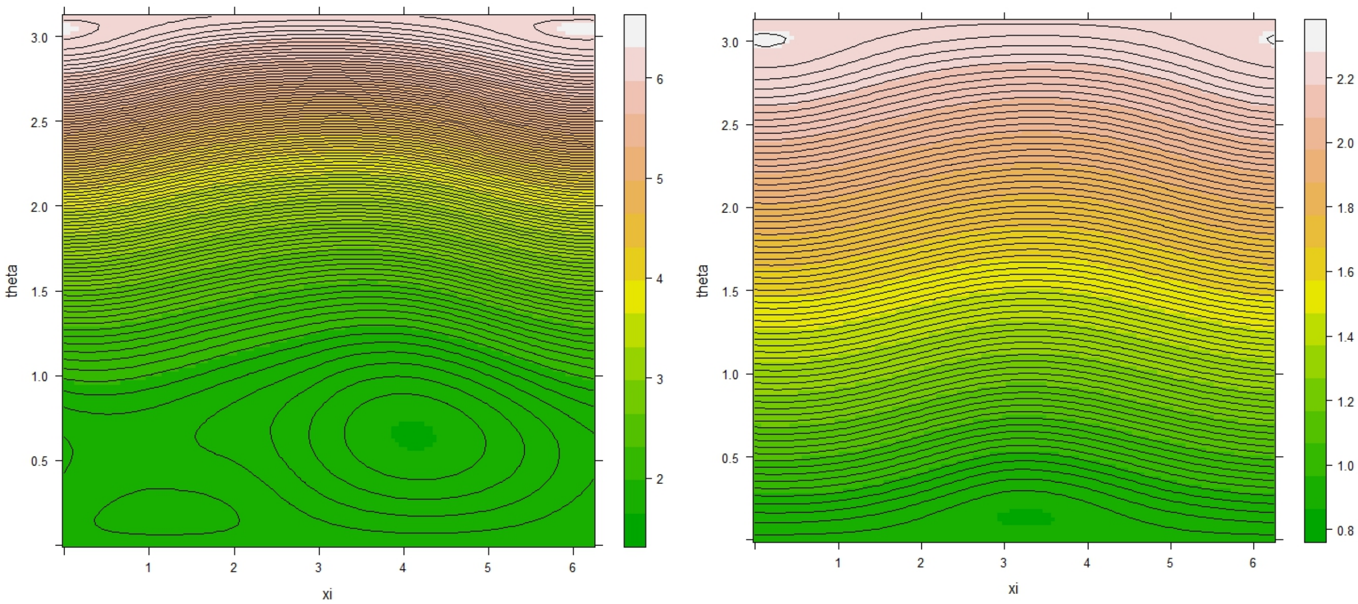

FIGURE 12. Sample Fréchet function of the Fréchet mean (left) and the Fréchet median (right).

6. Conclusion

This paper introduces an empirical likelihood approach to conduct inference on the Fréchet mean and related population objects used to characterize distributions of data on Riemannian manifolds. We investigate the asymptotic properties of the empirical likelihood statistic to test the simple and composite null hypotheses, and present several generalizations including two-sample test, inference on the Fréchet variance and local Fréchet regression, quasi Bayesian inference, and estimation of the Fréchet mean set. Our numerical studies via Monte Carlo simulations and real data examples show that our empirical likelihood approach can be a useful complement to the existing inference approach for Riemannian manifolds.

Appendix A. Appendix

In this appendix, we will obey the following notation. For any positive sequences a_n and b_n, we write $a_n \lesssim b_n$ if there is a positive constant $C > 0$ independent of n such that $a_n \leq Cb_n$ for all n, $a_n \sim b_n$ if $a_n \lesssim b_n$and $b_n \lesssim a_n$.

A.1. **Examples.** Here we provide some popular examples of Riemannian manifolds. See Bhattacharya and Patrangenaru (2003, 2005, 2014) for other examples and their applications.

Example 1. [m-dimensional sphere] Let $\mathbb{S}^m = \{p \in \mathbb{R}^{m+1} : \|p\| = 1\}$ be the m-dimensional sphere with the geodesic distance $\arccos(p'q)$ for $p, q \in \mathbb{S}^m$. The tangent space at a point $p \in \mathbb{S}^m$ is $T_p\mathbb{S}^m = \{x \in \mathbb{R}^{m+1} : x'p = 0\}$. In this case, the exponential map $\exp_p(\cdot) : T_p\mathbb{S}^m \to \mathbb{S}^m$ is given by $\exp_p(x) = \cos(\|x\|)p + \sin(\|x\|)\frac{x}{\|x\|}$. Spherical data arise in many research fields such as astrophysics, biology, geology, material science, meteorology, and political science. For those applications, we refer to Watson (1983), Briggs (1993), Mardia and Jupp (2000), Franke *et al.* (2015) and Ley and Verdebout (2017). In our numerical illustrations, we apply our empirical likelihood methods to the analysis of several real datasets on the circle $\mathbb{S}^1$ and two-dimensional sphere $\mathbb{S}^2$.

Example 2. [Planar shape space] Consider the Kendall's planer shape space Σ_2^k, where k and 2 denote the number of landmarks and the Euclidean dimension on which landmarks lie, respectively (Kendall, 1984). An element of Σ_2^k is a set of k points in the plane (not all equal), modulo similarity transformation in $\mathbb{R}^2$, i.e., translation, rotation and scaling. Let $S_2^k = \{u = (u_1, \ldots, u_k)' \in \mathbb{C}^k : \sum_{i=1}^k u_i = 0, u'\bar{u} = 1\}$ be the pre-shape sphere which is the unit sphere in the k-dimensional complex space. Here $\bar{u}$ is the complex conjugate of u. The tangent space of S_2^k is $T_zS_2^k = \{v \in \mathbb{C}^k : v'1_k = 0, \mathrm{Re}(z'\bar{v}) = 0\}$, where 1_k is the column vector of ones of size k and $\mathrm{Re}(w)$ is the real part of the complex number w. The elements of the planer shape space Σ_2^k can be represented as equivalence classes $\pi(z)$ where $\pi(z) := [z] = \{e^{i\theta}z : 0 \leq \theta < 2\pi\}$ is a map from S_2^k to Σ_2^k. Note that π is a Riemannian submersion and so the tangent space $T_{[z]}\Sigma_2^k$ is isometric with the subspace of $T_zS_2^k$ called the horizontal subspace $H_z = \{v \in \mathbb{C}^k : z'\bar{v} = 0, v'1_k = 0\}$. Let $\iota_{[z]} : T_{[z]}\Sigma_2^k \to H_z$ denote the isometric map. Then the exponential map $\exp_{[z]}(\cdot) : T_{[z]}\Sigma_2^k \to \Sigma_2^k$ is given by $\exp_{[z]}(x) = \pi \circ \exp_z \circ \iota_{[z]}(x)$ where $\exp_z$ is the exponential map of S_2^k. The geodesic distance d_g between $[x], [y] \in \Sigma_2^k$ is given by $d_g([x], [y]) = \arccos(|x'\bar{y}|)$. For applications of (general) shape space to archaeology, astronomy, geography, morphometrics, medical diagnostics, and physical chemistry, we refer to Kendall (1989), Small (1996), Bookstein (1997), Bhattacharya and Patrangenaru (2014) and Dryden and Mardia (2016).

Example 3. [Real projective space] Consider the real projective space $\mathbb{R}P^m$. The elements of $\mathbb{R}P^m$ can be represented as equivalence classes $[x] = [x_1 : x_2 : \cdots : x_{m+1}] = \{\lambda x : \lambda \neq 0\}$ where $x = (x_1, \ldots, x_{m+1})' \in \mathbb{R}^{m+1}\backslash\{0\}$. Since any line through the origin in $\mathbb{R}^{m+1}$ is uniquely determined by its points of intersection with the unit sphere $\mathbb{S}^m$, one may identify $\mathbb{R}P^m$ with $\mathbb{S}^m/G$, with G comprising the identity map and the antipodal map $p \mapsto -p$. The geodesic distance d_g between $[x], [y] \in \mathbb{R}P^m$ is given by $d_g([x], [y]) = \arccos(|x'y|)$. Let $T_{[z]}\mathbb{R}P^m$ be the tangent space of $\mathbb{R}P^m$. The exponential map of $\mathbb{R}P^m$ at $[z]$ is $\exp_{[z]}(x) = \pi \circ \exp_z \circ \iota_{[z]}(x)$, where $\iota_{[z]} : T_{[z]}\mathbb{R}P^m \to T_z\mathbb{S}^m$ is an isometric map, $\exp_z$ is the exponential map of $\mathbb{S}^m$, and $\pi : \mathbb{S}^m \ni z \mapsto [z] \in \mathbb{R}P^m$ is a Riemannian submersion. For applications of the real projective space to computer vision, geology, paleomagnetism, robotics, and sociology, we refer to Beran and Fisher (1998), Mardia and Jupp (2000), Haines and Wilson (2008) and Glover et al.(2012).

A.2. **Description of Wald test for simulation .** Let $p = (\sin\theta\cos\xi, \sin\theta\sin\xi, \cos\theta)'$. Following Bhattacharya and Patrangenaru (2005, Theorem 2.1), we have $\sqrt{n}\log_p(\mu_n) \xrightarrow{d} N(0, \Lambda^{-1}\Sigma\Lambda^{-1})$ where μ_n is a sample Fréchet mean, $\log_p : \mathbb{S}^2 \to \mathbb{R}^2$ is the logarithmic map defined as

$$\log_p(x) = (e_1, e_2)' C_\theta' \arccos(p'x) \frac{x - (p'x)p}{\|x - (p'x)p\|},$$

where $e_1 = (1,0,0)'$, $e_2 = (0,1,0)'$, $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^3$, C_θ is a 3×3 matrix defined as

$$C_\theta = \begin{pmatrix} \cos\theta\cos\xi & -\sin\xi & \sin\theta\cos\xi \\ \cos\theta\sin\xi & \cos\xi & \sin\theta\sin\xi \\ -\sin\theta & 0 & \cos\theta \end{pmatrix},$$

and Λ and Σ are defined as

$$\begin{aligned} \Lambda &= \mathbb{E}\left[\frac{\partial^2}{\partial x \partial x'} \arccos^2((\exp_p(x))'X) \Big|_{x=(0,0)'} \right], \\ \Sigma &= \mathbb{E}[g(X_1, p) g(X_1, p)'] \\ &= \mathbb{E}\left[\left(\frac{\partial}{\partial x} \arccos^2((\exp_p(x))'X) \Big|_{x=(0,0)'} \right) \left(\frac{\partial}{\partial x} \arccos^2((\exp_p(x))'X) \Big|_{x=(0,0)'} \right)' \right]. \end{aligned}$$

Then we define the Wald statistic as

$$W_n(p) := n(\log_p(\mu_n))'(\hat\Lambda\hat\Sigma^{-1}\hat\Lambda)\log_p(\mu_n),$$

where $\hat\Lambda$ and $\hat\Sigma$ are sample counterparts of Λ and Σ, respectively.

A.3. **Proof of Theorem 1.**

Proof of (i). Under the assumption $\mathbb{E}[||g(X,\mu)||^2] < \infty$ (Assumption 1 (iii)), the Borel-Cantelli argument as in Owen (1988) imply $\max_{1\le i\le n}\|g(X_i,\mu)\| = o_p(n^{1/2})$. An analogous argument as in Owen (2001, Chapter 11) yields the quadratic expansion

$$\ell(\mu) = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_i,\mu) \right)' \left(\frac{1}{n} \sum_{i=1}^n g(X_i,\mu) g(X_i,\mu)' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_i,\mu) \right) + o_p(1).$$

Therefore, under Assumption 1 (iii), the central limit theorem $\frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i,\mu) \xrightarrow{d} N(0, \mathbb{E}[g(X,\mu)g(X,\mu)'])$ and the law of large numbers $\frac{1}{n}\sum_{i=1}^n g(X_i,\mu)g(X_i,\mu)' \xrightarrow{p} \mathbb{E}[g(X,\mu)g(X,\mu)']$ yields the conclusion.

Proof of (ii). Let $G(X,x) = \left(\frac{\partial g^*(X, \exp_\mu(x))}{\partial x'}\right)'$. An expansion around $x_n = 0$ yields

$$g^*(X_i, \mu_n) = g^*(X_i, \exp_\mu(x_n)) = g^*(X_i, \mu) + G(X_i, \tilde{x})' x_n, \tag{7}$$

where $\tilde{x}$ is a point on the line joining x_n and 0. Thus, the Borel-Cantelli argument as in Owen (1988) and $x_n = O_p(n^{-1/2})$ (Assumption 2 (ii)) imply $\max_{1 \leq i \leq n} \|g^*(X_i, \mu_n)\| = o_p(n^{1/2})$, and an analogous argument as in Owen (2001, Chapter 11) yields the quadratic expansion

$$\ell^*(\mu_n) = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g^*(X_i, \mu_n) \right)' \hat{V}^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g^*(X_i, \mu_n) \right) + o_p(1), \tag{8}$$

where $\hat{V} = \frac{1}{n} \sum_{i=1}^n g^*(X_i, \mu_n) g^*(X_i, \mu_n)'$.

A uniform law of large numbers (Lemma 2.4 of Newey and McFadden, 1994) implies

$$\sup_{x \in \mathcal{N}} \left\| \frac{1}{n} \sum_{i=1}^n g^*(X_i, \exp_\mu(x)) g^*(X_i, \exp_\mu(x))' - \mathbb{E}[g^*(X_i, \exp_\mu(x)) g^*(X_i, \exp_\mu(x))'] \right\| \xrightarrow{p} 0,$$

$$\sup_{x \in \mathcal{N}} \left\| \frac{1}{n} \sum_{i=1}^n G(X_i, x) - \mathbb{E}[G(X_i, x)] \right\| \xrightarrow{p} 0.$$

Since $\mu_n \xrightarrow{p} \mu = \mu^*$ under H_0, we obtain

$$\hat{V} \xrightarrow{p} \mathbb{E}[g^*(X_i, \mu^*) g^*(X_i, \mu^*)'], \qquad \frac{1}{n} \sum_{i=1}^n G(X_i, \tilde{x}) \xrightarrow{p} \mathbb{E}[G(X, 0)]. \tag{9}$$

Thus, by (7), it holds that under H_0,

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n g^*(X_i, \mu_n) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g^*(X_i, \mu^*) + \left(\frac{1}{n} \sum_{i=1}^n G(X_i, \tilde{x}) \right)' \sqrt{n} x_n \\ &\xrightarrow{d} N \left(0, [I_{m^*} : \mathbb{E}[G(X, 0)]'] \Sigma \begin{bmatrix} I_{m^*} \\ \mathbb{E}[G(X, 0)] \end{bmatrix} \right). \end{aligned} \tag{10}$$

The conclusion follows by (9) and (10).

A.4. **Proof of Theorem 2.** The proof is analogous to the one for Theorem 1 (ii). As in (8), the statistic L can be expanded as

$$\begin{aligned}
L &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i,\mu_n)\right)' \hat{V}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i,\mu_n)\right) \\
&\quad + \left(\frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1} g(X_{1j},\mu_n)\right)' \hat{V}_1^{-1}\left(\frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1} g(X_{1j},\mu_n)\right) + o_p(1) \\
&= \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i,\mu) + \sqrt{\frac{1}{1+\rho}}\left(\frac{1}{n}\sum_{i=1}^{n} G(X_i,\tilde{x})\right)' \sqrt{n+n_1}x_n\right\}' \hat{V}^{-1} \\
&\qquad \times \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i,\mu) + \sqrt{\frac{1}{1+\rho}}\left(\frac{1}{n}\sum_{i=1}^{n} G(X_i,\tilde{x})\right)' \sqrt{n+n_1}x_n\right\} \\
&\quad + \left\{\frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1} g(X_{1j},\mu) + \sqrt{\frac{\rho}{1+\rho}}\left(\frac{1}{n_1}\sum_{j=1}^{n_1} G(X_{1j},\tilde{x})\right)' \sqrt{n+n_1}x_n\right\}' \hat{V}_1^{-1} \\
&\qquad \times \left\{\frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1} g(X_{1j},\mu) + \sqrt{\frac{\rho}{1+\rho}}\left(\frac{1}{n_1}\sum_{j=1}^{n_1} G(X_{1j},\tilde{x})\right)' \sqrt{n+n_1}x_n\right\} + o_p(1),
\end{aligned}$$

where $\hat{V} = \frac{1}{n}\sum_{i=1}^{n} g(X_i,\mu_n)g(X_i,\mu_n)'$, $\hat{V}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} g(X_{1j},\mu_n)g(X_{1j},\mu_n)'$, $G(X,x) = \left(\frac{\partial g(X,\exp_\mu(x))}{\partial x'}\right)'$, and the second equality follows from the expansions $g(X_i,\mu_n) = g(X_i,\exp_\mu(x_n))$ and $g(X_{1j},\mu_n) = g(X_{1j},\exp_\mu(x_n))$ around $x_n = 0$.

Therefore, the conclusion follows by the asymptotic normality assumed in (4) and the uniform law of large numbers for $\frac{1}{n}\sum_{i=1}^{n} G(X_i,x)$ and $\frac{1}{n_1}\sum_{j=1}^{n_1} G(X_{1j},x)$ over a neighborhood $x \in \mathcal{N}$.

A.5. **Proof of Theorem 4.**

Proof of (i). Let $\lambda(\mu) = \mathbb{E}_{\mu_0}[g(X,\mu)]$, where $\mathbb{E}_{\mu_0}[\cdot]$ is expectation under $X \sim \mathbb{P}_{\mu_0}$. Pick any $\epsilon, \epsilon_1 > 0$, and define $\mathcal{B}_0^c = \{\mu \in \mathcal{P} : \tilde{F}(\mu) - \tilde{F}(\mu_0) \geq \epsilon\}$, $\mathcal{A} = \{\mu \in \mathcal{P} : \lambda(\mu)'\lambda(\mu) \leq \epsilon_1\}$, and $\mathcal{A}^c = \mathcal{P} \setminus \mathcal{A}$. Since we have

$$\mathbb{P}\{\mu \in \mathcal{B}_0^c|\mathbf{X}\} \leq \mathbb{P}\{\mu \in \mathcal{A}^c|\mathbf{X}\} + \mathbb{P}\{\mu \in \mathcal{B}_0^c \cap \mathcal{A}|\mathbf{X}\},$$

it is sufficient to show that (I) $\mathbb{P}\{\mu \in \mathcal{A}^c|\mathbf{X}\} \to 0$ and (II) $\mathbb{P}\{\mu \in \mathcal{B}_0^c \cap \mathcal{A}|\mathbf{X}\} \to 0$ in $\mathbb{P}_{\mu_0}$-probability.

First, we show (I). Let $\bar{g}(\mu) = n^{-1}\sum_{i=1}^{n} g(X_i,\mu)$ and $\hat{\mathcal{A}}^c = \{\mu \in \mathcal{P} : \lambda(\mu)'\bar{g}(\mu) > \epsilon_1/2\}$. Decompose

$$\begin{aligned}
\mathbb{P}\{\mu \in \mathcal{A}^c|\mathbf{X}\} &\sim \int_{\mu\in\mathcal{A}^c} \exp(-\ell(\mu)/2)\exp(-\varsigma_n\{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_n)\})d\pi(\mu) \\
&\leq \int_{\mu\in\mathcal{A}^c} \exp(-\ell(\mu)/2)d\pi(\mu) \\
&\leq \int_{\mu\in\hat{\mathcal{A}}^c} \exp(-\ell(\mu)/2)d\pi(\mu) + \int_{\mu\in\mathcal{A}^c\cap\hat{\mathcal{A}}} \exp(-\ell(\mu)/2)d\pi(\mu) \\
&=: T_1 + T_2,
\end{aligned}$$

where the second inequality follows from $\tilde{F}_n(\mu) - \tilde{F}_n(\mu_n) \geq 0$ for every $\mu \in \mathcal{P}$. For T_2, the uniform law of large numbers ($\sup_{\mu\in\mathcal{P}} ||\bar{g}(\mu) - \lambda(\mu)|| \to 0$ in $\mathbb{P}_{\mu_0}$-probability) guarantees $T_2 \to 0$ in $\mathbb{P}_{\mu_0}$-probability. For T_1, note that

$$\begin{aligned}
-\frac{1}{2}\ell(\mu) &= \min_{\lambda} -\sum_{i=1}^{n}\log(1+\lambda' g(X_i,\mu)) \leq -\sum_{i=1}^{n}\log(1+n^{-1/a}\lambda(\mu)' g(X_i,\mu)) \\
&\leq -n^{-1/a}\lambda(\mu)'\sum_{i=1}^{n} g(X_i,\mu) + \frac{1}{2}n^{-2/a}\lambda(\mu)'\sum_{i=1}^{n}\frac{g(X_i,\mu)g(X_i,\mu)'}{(1+r(X_i,\mu))^2}\lambda(\mu) \\
&\lesssim -n^{1-1/a}\lambda(\mu)'\bar{g}(\mu) + \frac{1}{2}n^{1-2/a}\lambda(\mu)'\mathbb{E}_{\mu_0}[g(X,\mu)g(X,\mu)']\lambda(\mu) \\
&\lesssim -n^{1-1/a}\lambda(\mu)'\bar{g}(\mu), \qquad (11)
\end{aligned}$$

where the second inequality follows from an expansion and $r(X_i,\mu)$ is a point on the line joining $\lambda(\mu)' g(X_i,\mu)$ and 0, the first wave inequality follows from the uniform law of large numbers ($\sup_{\mu\in\mathcal{P}} ||n^{-1}\sum_{i=1}^{n} g(X_i,\mu)g(X_i,\mu)' - \mathbb{E}_{\mu_0}[g(X,\mu)g(X,\mu)']|| \to 0$ in $\mathbb{P}_{\mu_0}$-probability) and $\max_{1\leq i\leq n}\sup_{\mu\in\mathcal{P}} ||g(X_i,\mu)|| = o_p(n^{1/a})$. Thus, it holds

$$T_1 \lesssim \int_{\mu\in\hat{\mathcal{A}}^c} \exp(-n^{1-1/a}\lambda(\mu)'\bar{g}(\mu))d\pi(\mu) \leq \exp(-n^{1-1/a}\epsilon_1/2) \to 0, \qquad (12)$$

in $\mathbb{P}_{\mu_0}$-probability, and we obtain (I).

Next, we show (II). Observe that

$$\begin{aligned}
\mathbb{P}\{\mu \in \mathcal{B}_0^c \cap \mathcal{A}|\mathbf{X}\} &\sim \int_{\mu\in\mathcal{B}_0^c\cap\mathcal{A}} \exp(-\ell(\mu)/2)\exp(-\varsigma_n\{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_n)\})d\pi(\mu) \\
&\lesssim \int_{\mu\in\mathcal{B}_0^c\cap\mathcal{A}} \exp(-n^{1-1/a}\bar{g}(\mu)'\bar{g}(\mu))\exp(-\varsigma_n\{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_n)\})d\pi(\mu) \\
&\leq \int_{\mu\in\mathcal{B}_0^c\cap\mathcal{A}} \exp(-\varsigma_n\{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_n)\})d\pi(\mu),
\end{aligned}$$

where the wave inequality follows from the same argument as in (11) by replacing "$\lambda(\mu)$" with "$\bar{g}(\mu)$", and the inequality follows from $\bar{g}(\mu)'\bar{g}(\mu) \geq 0$ for every $\mu \in \mathcal{P}$. Now we have

$$\begin{aligned}
&\int_{\mu\in\mathcal{B}_0^c\cap\mathcal{A}} \exp(-\varsigma_n\{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_n)\})d\pi(\mu) \\
\leq\ & \int_{\mu\in\mathcal{B}_0^c} \exp(-\varsigma_n\{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_0)\})d\pi(\mu) \\
=\ & \int_{\mu\in\mathcal{P}:\tilde{F}(\mu)-\tilde{F}(\mu_0)\geq\epsilon} \exp(-\varsigma_n\{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_0)\})d\pi(\mu) \\
\leq\ & \int_{\mu\in\mathcal{P}:\tilde{F}(\mu)-\tilde{F}(\mu_0)\geq\epsilon,|\tilde{F}_n(\mu)-\tilde{F}(\mu)|<\epsilon/4,|\tilde{F}_n(\mu_0)-\tilde{F}(\mu_0)|<\epsilon/4} \exp(-\varsigma_n\{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_0)\})d\pi(\mu) \\
&+ \int_{\mu\in\mathcal{P}:|\tilde{F}_n(\mu_0)-\tilde{F}(\mu_0)|\geq\epsilon/4} \exp(-\varsigma_n\{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_0)\})d\pi(\mu) \\
&+ \int_{\mu\in\mathcal{P}:|\tilde{F}_n(\mu)-\tilde{F}(\mu)|\geq\epsilon/4} \exp(-\varsigma_n\{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_0)\})d\pi(\mu) \\
=:\ & T_3 + T_4 + T_5,
\end{aligned}$$

where the first inequality follows from $\tilde{F}_n(\mu_n) \leq \tilde{F}_n(\mu_0)$. Since $\sup_{\mu \in \mathcal{P}} |\tilde{F}_n(\mu) - \tilde{F}(\mu)| \to 0$ in $\mathbb{P}_{\mu_0}$-probability, it holds $T_4 \to 0$ and $T_5 \to 0$ in $\mathbb{P}_{\mu_0}$-probability. Furthermore,

$$T_3 \leq \int_{\mu \in \mathcal{P}: \tilde{F}_n(\mu) - \tilde{F}_n(\mu_0) \geq \epsilon/2,} \exp(-\varsigma_n \{\tilde{F}_n(\mu) - \tilde{F}_n(\mu_0)\}) d\pi(\mu) \leq \exp(-\varsigma_n \epsilon/2) \to 0,$$

in $\mathbb{P}_{\mu_0}$-probability. Combining these results, we obtain (II). Therefore, the conclusion follows.

Proof of (ii). It follows from the proof of Part (i) of this theorem. In particular, the results in (11) and (12) yield the conclusion.

Proof of (iii). If $\mathbb{E}_{\mu_0}[g(X, \mu)] = 0$ uniquely at $\mu_0 \in \mathcal{P}$, then the inequality $d_{\mathcal{P}}(\mu, \mu_0) \geq \epsilon$ implies $||\mathbb{E}_{\mu_0}[g(X, \mu)]|| \geq \epsilon_1$ for some $\epsilon_1 > 0$. Thus Part (ii) of this theorem yields the conclusion.

A.6. **Proof of Theorem 5.** Pick any z. As in (8), the statistic $\ell(\mu_z; z)$ can be expanded as

$$\ell(\mu_z; z) = \left\{ \frac{1}{\sqrt{nh^k}} \sum_{i=1}^{n} K\left(\frac{Z_i - z}{h}\right) g(X_i, \mu_z) \right\}' \hat{V}_z^{-1} \left\{ \frac{1}{\sqrt{nh^k}} \sum_{i=1}^{n} K\left(\frac{Z_i - z}{h}\right) g(X_i, \mu_z) \right\} + o_p(1),$$

where $\hat{V}_z = \frac{1}{nh^k} \sum_{i=1}^{n} K\left(\frac{Z_i - z}{h}\right)^2 g(X_i, \mu_z) g(X_i, \mu_z)'$. Then the conclusion follows from the assumptions in (6).

A.7. **Proof of Theorem 6.** As in (8), the empirical likelihood function $\ell(p)$ can be uniformly approximated as

$$\sup_{p \in \mathcal{P}} |n^{-1} \ell(p) - Q_n(p)| \xrightarrow{p} 0,$$

where $Q_n(p) = \left\{n^{-1} \sum_{i=1}^{n} g(X_i, p)\right\}' \left(n^{-1} \sum_{i=1}^{n} g(X_i, p) g(X_i, p)'\right)^{-1} \left\{n^{-1} \sum_{i=1}^{n} g(X_i, p)\right\}$. Thus, it is sufficient for the conclusion to verify the conditions in Chernozhukov, Hong and Tamer (2007, Theorem 3.1) providing a generic consistency result for a level set estimator:

$$\sup_{p \in \mathcal{P}} |Q_n(p) - Q(p)| \xrightarrow{p} 0, \qquad \sup_{p \in \tilde{\mathcal{P}}} nQ_n(p) = O_p(1), \qquad \tilde{\mathcal{P}} = \arg\min_{p \in \mathcal{P}} Q(p), \tag{13}$$

where $Q(p) = \mathbb{E}[g(X, p)]' \left(\mathbb{E}[g(X, p) g(X, p)']\right)^{-1} \mathbb{E}[g(X, p)]$.

The first condition in (13) is verified by applying the uniform law of large numbers

$$\sup_{p \in \mathcal{P}} \left| \frac{1}{n} \sum_{i=1}^{n} g(X_i, p) - \mathbb{E}[g(X, p)] \right| \xrightarrow{p} 0, \qquad \sup_{p \in \mathcal{P}} \left| \frac{1}{n} \sum_{i=1}^{n} g(X_i, p) g(X_i, p)' - \mathbb{E}[g(X, p) g(X_i, p)'] \right| \xrightarrow{p} 0,$$

under Assumptions (i)-(ii) in Theorem 6. The second condition in (13) is verified by Assumptions (iii) in Theorem 6. Finally, the third condition in (13) is satisfied because $Q(p) = 0$ if and only if $p \in \tilde{\mathcal{P}}$. Therefore, the conclusion follows by Chernozhukov, Hong and Tamer (2007, Theorem 3.1).

References

[1] Beran, R. and N. I. Fisher (1998) Nonparametric comparison of mean directions or mean axes, *Annals of Statistics*, 26, 472-493.

[2] Bhattacharya, A. and D. B. Dunson (2010) Nonparametric Bayesian density estimation on manifolds with applications to planar shapes, *Biometrika*, 2010, 851-865.

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

[4] Bhattacharya, R. and V. Patrangenaru (2005) Large sample theory of intrinsic and extrinsic sample means on manifolds II, *Annals of Statistics*, 33, 1225-1259.

[5] Bhattacharya, R. and V. Patrangenaru (2014) Statistics on manifolds and landmarks based image analysis: A nonparametric theory with applications, *Journal of Statistical Planning and Inference*, 145, 1-22.

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

[7] Blanchard, M. and A. Q. Jaffe (2022) Fréchet mean set estimation in the Hausdorff metric, via relaxation, arXiv:2212.12057.

[8] Bookstein, F. L. (1997) *Morphometric Tools for Landmark Data: Geometry and Biology*, Cambridge University Press.

[9] Briggs, M. S. (1993) Dipole and quadrupole tests of the isotropy of gamma-ray burst locations, *Astrophysical Journal*, 407, 126-134.

[10] Butler, R. F. (1992) *Paleomagnetism: Magnetic Domains to Geologic Terranes,* Blackwell Scientific Publications.

[11] Chernozhukov, V., Hong, H. and E. Tamer (2007) Estimation and confidence regions for parameter sets in econometric models, *Econometrica*, 75, 1243-1284.

[12] DiCiccio, T. J., Hall, P. and J. P. Romano (1991) Empirical likelihood is Bartlett-correctable, *Annals of Statistics*, 19, 1053-1061.

[13] Dryden, I. L. and K. V. Mardia (2016) *Statistical Shape Analysis with Applications in R* (2nd ed.), Wiley.

[14] Dubey, P. and H-.G. Müller (2019) Fréchet analysis of variance for random objects, *Biometrika*, 106, 803-821.

[15] Dubey, P. and H-.G. Müller (2020) Fréchet change-point detection, *Annals of Statistics*, 48, 3312-3335.

[16] Eltzner, B. (2020) Testing for uniqueness of estimators, arxiv:2011.14762.

[17] Eltzner, B. (2022) Geometrical smeariness - a new phenomenon of Fréchet means, *Bernoulli*, 28, 239-254.

[18] Eltzner, B. and S. F. Huckemann (2019) A smeary central limit theorem for manifolds with application to high-dimensional spheres, *Annals of Statistics*, 47, 3360-3381.

[19] Evans, S. N. and A. Q. Jaffe (2020) Strong law of large numbers for Fréchet means, arXiv:2012.12859, forthcoming in *Bernoulli*.

[20] Franke, J., Redenbach, C. and N. Zhang (2016) On a mixture model for directional data on the sphere, *Scandinavian Journal of Statistics*, 43, 139-155.

[21] Fréchet, M. (1948) Les éléments aléatoires de nature quelconque dans un espace distancié, *Annales de l'institut Henri Poincaré*, 10, 215-310.

[22] Gallo, L. C., Domeier, M., Sapienza, F., Swanson-Hysell, N. L., Vaes, B., Zhang, Y., Arnould, M., Eyster, A., Gürer, D., A. Kiräly, Robert, B., Rolf, T., Shephard, G., and A. Van der Boon (2023) Embracing uncertainty to resolve polar wander: A case study of Cenozoic North America, *Geophysical Research Letters,* 50, e2023GL103436.

[23] Glover, J., Bradski, G. and R. B. Rusu (2012) Monte Carlo pose estimation with quaternion kernels and the distribution, *Robotics: Science and Systems VII*, 97-104.

[24] Haines, T. S. and R. C. Wilson (2008) Belief propagation with directional statistics for solving the shape-from-shading problem, *European Conference on Computer Vision*, pp. 780-791, Springer.

[25] Hundrieser, S. Eltzner, B. and S. F. Huckemann (2021) Finite sample smeariness of Fréchet means and application to climate, arXiv:2005.0231.

[26] Kendall, D. G. (1984) Shape manifolds, procrustean metrics, and complex projective spaces, *Bulletin of the London Mathematical Society*, 16, 81-121.

[27] Kendall, D. G. (1989) A survey of the statistical theory of shape, *Statistical Science*, 4, 87-120.

[28] Lazer, N. A. (2003) Bayesian empirical likelihood, *Biometrika*, 90, 319-326.

[29] Ley, C. and T. Verdebout (2017) *Modern Directional Statistics*, CRC Press.

[30] Mardia, K. V. and P. E. Jupp (2000) *Directional Statistics*, Wiley.

[31] Marron, J. S. and A. M. Alonso (2014) Overview of object oriented data analysis, *Biometrical Journal*, 56, 732-753.

[32] McCormack, A. and P. Hoff (2022) The Stein effect for Fréchet means, *Annals of Statistics*, 50, 3647-3676.

[33] Newey, W. K. and D. McFadden (1994) Large sample estimation and hypothesis testing, *Handbook of Econometrics*, vol. IV, Chapter 36.

[34] Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237-249.

[35] Owen, A. B. (2001) *Empirical likelihood*, CRC Press.

[36] Patrangenaru, V. and L. Ellingson (2015) *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*, CRC Press.

[37] Schötz, C. (2022) Strong laws of large numbers for generalizations of Fréchet mean sets, *Statistics*, 56, 34-52.

[38] Small, C. G. (1996) *The Statistical Theory of Shape*, Springer.

[39] van der Vaart, A. W. and J. A. Wellner (1996) *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer.

[40] Watson, G. S. (1983) *Statistics on Spheres*, University of Arkansas Lecture Notes in Mathematical Sciences, 6, Wiley.

Center for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan.

Email address: `daisukekurisu@csis.u-tokyo.ac.jp`

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

Email address: `t.otsu@lse.ac.uk`