Electronic Journal of Statistics
Vol. 18 (2024) 3274–3309
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2276

Finite sample smeariness of Fréchet means with application to climate

Shayan Hundrieser[1], **Benjamin Eltzner**[2] and **Stephan Huckemann**[1]

[1] *Institute for Mathematical Stochastics, University of Göttingen, Germany,*
e-mail: s.hundrieser@math.uni-goettingen.de; huckeman@math.uni-goettingen.de

[2] *Max Planck Institute for Multidisciplinary Sciences, University of Göttingen, Germany,*
e-mail: benjamin.eltzner@mpinat.mpg.de

Abstract: Fréchet means on non-Euclidean spaces may exhibit nonstandard asymptotic rates rendering quantile-based asymptotic inference inapplicable. We show here that this affects, among others, all circular distributions whose support is not contained in a closed half circle. We exhaustively describe this phenomenon on the circle and introduce a new concept on metric spaces which we call finite samples smeariness (FSS). In the presence of FSS, for which we provide a test, it turns out that quantile-based tests for equality of Fréchet means systematically feature effective levels higher than their nominal level. This effect can persevere asymptotically for increasing sample size, or may not, depending on the type of FSS. In contrast, suitable bootstrap-based tests correct for FSS and asymptotically attain the correct level. For illustration of relevance we apply our method to directional wind data from two European cities. It turns out that quantile based tests, not correcting for FSS, find a multitude of significant wind changes. This multitude condenses to a few years featuring significant wind changes when our bootstrap tests are applied, correcting for FSS.

AMS 2000 subject classifications: Primary 62H11; secondary 60F05, 62G10.
Keywords and phrases: Fréchet means, smeariness, one- and two-sample tests, parametric asymptotic quantile based tests, bootstrap tests, directional data on circles and tori, wind directions.

Received October 2022.

1. Introduction

For a random variable X on a metric space (M, d), generalizing the concept of the Euclidean *expected value*, Fréchet (1948) proposed to consider minimizers of the expected squared distance

$$\mathbb{M}(X) := \operatorname*{argmin}_{p \in M} F(p) \quad \text{where} \quad F(p) = \mathbb{E}\left[d(p, X)^2\right], \tag{1.1}$$

known as the *Fréchet population mean set*. Under mild assumptions, Ziezold (1977) derived for samples $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} X$ strong set-wise asymptotic consistency as $n \to \infty$ of the *Fréchet sample mean set*

$$\mathbb{M}_n(X_1, \dots, X_n) := \operatorname*{argmin}_{p \in M} F_n(p) \quad \text{where} \quad F_n(p) = \frac{1}{n}\sum_{j=1}^{n} d(p, X_j)^2. \tag{1.2}$$

Under stronger conditions, among others that M is a finite dimensional manifold, and under uniqueness of the minimizer μ of (1.1), then called the *Fréchet population mean*, Bhattacharya and Patrangenaru (2005) derived, in a local chart $\phi : U \to \mathbb{R}^m$ near $\mu \in U \subseteq M$, a central limit theorem for a measurable selection $\widehat{\mu}_n \in \mathbb{M}_n$ from (1.2), called a *Fréchet sample mean*, with a Gaussian limiting distribution and the usual rate of $n^{-1/2}$:

$$\sqrt{n}\,(\phi(\widehat{\mu}_n) - \phi(\mu)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)\,. \tag{1.3}$$

While the covariance matrix Σ above reflects the asymptotic rescaled covariance of $\phi(\widehat{\mu}_n)$, it is usually approximated by the covariance matrix $\widehat{\Sigma}_n$ of the sample $\phi(X_1), \dots, \phi(X_n)$, giving rise to the quantile-based one- and two-sample tests proposed by Bhattacharya and Patrangenaru (2005), see also e.g. Bhattacharya and Lin (2017). These tests rest on the approximation

$$(\phi(\widehat{\mu}_n) - \phi(\mu))^T\, \widehat{\Sigma}_n^{-1}\, (\phi(\widehat{\mu}_n) - \phi(\mu)) \xrightarrow{\mathcal{D}} \chi^2_m\,. \tag{1.4}$$

In order to assess the validity of this approximation, we consider under a unique Fréchet sample and population mean the suitably rescaled quotient of Fréchet variances (see Definition 2.2 for a formulation under nonunique means)

$$\mathfrak{m}_n := \frac{n\mathbb{E}[d(\mu, \widehat{\mu}_n)^2]}{\mathbb{E}[d(\mu, X)^2]}\,, \tag{1.5}$$

which has been called by Pennec (2019) the *modulation of the rate of convergence of the variance* for sample size n and which we abbreviate as *variance modulation*. The following phenomena have been observed in the literature:

(A) $\mathfrak{m}_n = 1$ for all $n \in \mathbb{N}$,
(B) $\lim_{n\to\infty} \mathfrak{m}_n = 0$ and there exists a random integer $N \in \mathbb{N}$ such that $\widehat{\mu}_n = \mu$ for $n \geq N$ (stickiness),
(C) $\lim_{n\to\infty} \mathfrak{m}_n = \infty$ (smeariness),

Phenomenon (A) is the case on Euclidean spaces whenever second moments are finite.[1] As we will show here, it is also the case if (M, d) is a flat torus with sufficiently concentrated X. For some nontrivial random variables on non-manifolds, using not a local chart but a suitable embedding, Phenomenon (B) has been observed by Barden, Le and Owen (2013, 2018) on the BHV spaces of Billera, Holmes and Vogtmann (2001) for phylogenetic trees, it has been observed on related spaces by Hotz et al. (2013); Huckemann et al. (2015), further investigated by Lammers et al. (2023) and general central limit theorem have been derived by Mattingly, Miller and Tran (2023). Furthermore, Phenomenon (C) has been observed on the circle by Hotz and Huckemann (2015) with $\mathfrak{m}_n$ of rate $n^{\frac{\gamma}{\gamma+1}}$ with arbitrary $1 \leq \gamma \in \mathbb{N}$, and on spheres of arbitrary dimension

[1]For independent and identically distributed random variables $X_1, \dots, X_n \sim X$ on $\mathbb{R}^d$ the Fréchet sample mean coincides with the average $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Hence, under a finite second moment, $\mathbb{E}[\|X\|^2] < \infty$, it follows that $\mathbb{E}[\|\overline{X}_n - \mathbb{E}[X]\|^2] = \mathbb{E}[\|X - \mathbb{E}[X]\|^2]/n$.

by Eltzner and Huckemann (2019) with $\mathfrak{m}_n$ of rate $n^{2/3}$. Moreover, in case of nonunique means, as discussed in Eltzner (2020), the finite sample behavior is similar to that in case of smeariness, because in all known models with smeary means, one can define a one-parameter family of models with parameter a, such that the mean is unique and non-smeary for $a < 0$, smeary for $a = 0$ and nonunique for $a > 0$. The recent work by Hundrieser, Eltzner and Huckemann (2024) also shows that near the transition of uniqueness to nonuniqueness, the variance modulation $\mathfrak{m}_n$ can be arbitrarily large in the finite sample regime, without necessarily leading to smeariness.

Notably, all but (A) render the approximation (1.4) invalid. At this point we remark that stickiness (B) is conceptually different from the opposite of smeariness (C), namely

(D) $\mathfrak{m}_n > 0$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} \mathfrak{m}_n = 0$ (anti-smeariness),

which has been observed by Schötz (2019a) for so-called *extrinsic means*,[2] see Bhattacharya and Patrangenaru (2003).

In this contribution we bring to attention two new phenomena:

(E) $\mathfrak{m}_n > 1$ for all $n \in \mathbb{N}$ and $1 < \lim_{n\to\infty} \mathfrak{m}_n < \infty$
(F) $\mathfrak{m}_n > 1$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} \mathfrak{m}_n = 1$,

and investigate them on the circle and the torus. Phenomenon (E) affects all (!) parametric distributions on the circle with nowhere vanishing density like the von Mises, the wrapped Gaussian, etc. In particular, it may render the approximation (1.4) invalid as dramatically illustrated in the left display of Table 1. While we introduce here definitions for rather general metric spaces and bootstrap tests for Riemannian manifolds, detailed illustrations of the effects and development of a proper asymptotic theory for the bootstrap beyond circle and torus on general manifolds are out of scope for the present paper. First steps towards theory and effects on spheres $\mathbb{S}^m$ of dimension $m > 1$ have been taken in Eltzner, Hundrieser and Huckemann (2021). In general, as mentioned above, smeariness seems to occur at the boundary of non-identifiability. In this vein the effects (E) and (F) can be viewed as a geometric manifestation of non-identifiability issues reaching well into the domain of identifiability.

Phenomenon (F) affects all distributions on the circle whose support strictly exceeds a closed half circle as long as a neighborhood of the antipodal of their Fréchet mean carries no probability mass. While this phenomenon renders the approximation (1.4) asymptotically valid, for surprisingly high sample sizes, the approximation may still be far off.

In simulations, we see that in cases (E) and (F), $\mathfrak{m}_n$ initially has a rate comparable to smeariness as in Phenomenon (C), in particular, the rate of $\mathbb{E}[d(\widehat{\mu_n}, \mu)^2]$ is strictly lower than the classical n^{-1}. Equivalently, $n \mapsto \mathfrak{m}_n$ starts off nonhorizontally as illustrated in Figures 1 and 2. Since in these cases, $n \mapsto \mathfrak{m}_n$ is only asymptotically horizontal, we give these new phenomena the name *finite*

[2] Extrinsic means have also been introduced by Hendriks and Landsman (1998) under the name of *mean locations* and encompass *mean directions* on spheres (Mardia and Jupp, 2000).

TABLE 1

Empirical rejection probability based on 100 000 simulation runs of two-sample tests for equality of Fréchet means with nominal significance level 0.05 *under the null hypothesis of a mixture of two equally weighted ($\beta = 1/2$) von Mises distributions on the circle with antipodal means (the first with concentration parameter $\kappa = 3$ and the second with varying concentration parameter λ) as defined in (3.1). Left: the quantile-based Test 5.1(ii) resting on (1.4). Right: the bootstrap-based Test 5.4(ii) with $B = 1000$.*

	quantile-based				bootstrap-based			
λ	0	1/4	1/2	3/4	0	1/4	1/2	3/4
$n = 100$	0.320	0.447	0.582	0.689	0.045	0.041	0.039	0.035
$n = 1000$	0.330	0.474	0.656	0.818	0.046	0.045	0.044	0.042
$n = 10000$	0.331	0.477	0.666	0.876	0.050	0.049	0.049	0.051

sample smeariness (FSS) and distinguish between Type I (E) and Type II (F). Alternatively for these phenomena, the term *lethargic means* has been proposed (Pennec, 2020). FSS has also been observed for *diffusion means*, which seem, however, less affected by actual smeariness, see Hansen et al. (2021).

The fact that FSS encompasses nonasymptotic phenomena that depend on sample size is crucial view of practical applications: For a given sample size n we speak of the *presence* of FSS, if the variance modulation $\mathfrak{m}_n$ is considerably larger than one. While under the presence of FSS, empirical levels of quantile-based tests can deviate strongly from their nominal level, we show for the circle that suitably designed bootstrap tests approximately keep their level, as illustrated in the right display of Table 1. In fact, we show in this work that the deviation of the quantile-based test perseveres as the sample size n tends to infinity whereas the bootstrap-based test asymptotically attains the correct level. In addition to simulations, in application to wind direction data we see that asymptotic quantile based two-sample tests, on the circle and a torus, due to the presence of FSS in the data, give a wrong impression of the extent of extreme wind change events for two continental European cities. Using bootstrap tests which preserve the nominal level, extreme wind changes reduce to a few concise events, only making the transitions from the years 2002 to 2003, 2005 to 2006, 2017 to 2018 and 2018 to 2019 exceptional, hinting towards an effect of recent climate change.

In the following Section 2 we introduce FSS on rather general metric spaces in terms of the variance modulation and give a test for the presence of FSS in terms of a large variance modulation. In Section 3 we exhaustively characterize FSS for random variables on the circle and torus. In Section 4 we investigate the asymptotics of circular Fréchet sample means and prove a consistency result for bootstrap Fréchet sample means under FSS. For the tests to follow, we also verify that all moments of properly centered and scaled Fréchet sample means converge to the respective moments of the limit distribution. On Riemannian manifolds, Section 5 explores the quantile based test by Bhattacharya and Patrangenaru (2005) and an implementation of the bootstrap test, which was also suggested by Bhattacharya and Patrangenaru (2005). Bootstrap consistency is then established on the circle and the torus. This is followed by simulations in

Section 6 and the application to wind direction data in Section 7. We conclude in Section 8 with an outlook and list open problems that arise from our findings. Here, we also explain why we expect that FSS affects *extrinsic* means as well and leave details for future research.

All proofs, some more simulations and extended data analysis are deferred to the supplement (Hundrieser, Eltzner and Huckemann, 2024) with key ideas sketched only in the main text. The code for every simulation and analysis is available on `https://github.com/hundrieser/FSS`. In particular, we provide a package "`FSS`" with an implementation of the asymptotic quantile and bootstrap based test as well as the test for the presence of finite sample smeariness.

2. Finite sample smeariness

Throughout this section, we assume that (M, d) is a complete metric space such that every bounded, closed set is compact.

Remark 2.1 (Existence and Uniqueness of Fréchet Means). Indeed, for random variables $X_1, \ldots, X_n \sim X$ on such spaces, due to the triangle inequality, $\mathbb{E}[d^2(p, X)] < \infty$ for some $p \in M$ guarantees that the Fréchet population and sample mean sets $\mathbb{M}(X)$ and $\mathbb{M}_n(X_1, \ldots, X_n)$ are not empty. On the circle, under very general conditions detailed in Hotz and Huckemann (2015), these sets are unique single points. Also, e.g. (M, d) being a Hadamard space, cf. Sturm (2003), or a finite dimensional manifold with X sufficiently concentrated, cf. Afsari (2011), the mean sets are guaranteed to be unique single points. Moreover, sample means of random variables featuring a density with respect to Riemannian measure on a Riemannian manifold are a.s. unique (Arnaudon and Miclo, 2014). In fact, nonuniqueness is a rather exceptional and unstable property in the following sense: If $\mathbb{M}(X)$ is nonunique, then for $\epsilon \in (0, 1]$ arbitrarily small and $\mu \in \mathbb{M}(X)$ the perturbed probability measure $\mathbb{P}_{\mu,\epsilon} := \epsilon\delta_\mu + (1-\epsilon)\mathbb{P}$ has, as is easy to see, the unique Fréchet population mean μ.

2.1. Definition of finite sample smeariness

We first generalize the definition of the variance modulation (1.5) from Pennec (2019) without requiring Fréchet population or sample means to be unique. We only require existence of mean sets, which, by Remark 2.1, is guaranteed by our assumption on (M, d) that it is complete and that every bounded closed subset is compact.

Definition 2.2 (Variance modulation). Let X be a random variable on (M, d) such that $\mathbb{E}[d^2(p, X)] < \infty$ for some $p \in M$ with Fréchet population mean set $\mathbb{M}(X)$. Then, with i.i.d. samples $X_1, \ldots, X_n \sim X$, $n \in \mathbb{N}$ and Fréchet sample mean sets, $\mathbb{M}_n(X_1, \ldots, X_n)$, the *variance modulation* is defined as

$$\mathfrak{m}_n := \frac{n\mathbb{E}\left[d_H^2(\mathbb{M}(X), \mathbb{M}_n(X_1, \ldots, X_n))\right]}{\min_{p \in M} \mathbb{E}[d^2(p, X)]},$$

with the *Hausdorff distance* d_H induced by the metric.[3]

Next, we formalize FSS in terms of the variance modulation.

Definition 2.3 (Finite sample smeariness). Let X be a random variable on (M, d) such that $\mathbb{E}[d^2(p, X)] < \infty$ for some $p \in M$. Then, X is called *finite sample smeary* (FSS) if

$$1 < \sup_{n\in\mathbb{N}} \mathfrak{m}_n < \infty.$$

We speak of *Type I* FSS if $\lim_{n\to\infty} \mathfrak{m}_n > 1$ and of *Type II* FSS if $\lim_{n\to\infty} \mathfrak{m}_n = 1$.

Remark 2.4 (Relation with smeariness). Recalling the definition of smeariness from (D) in the introduction, $\sup_{n\in\mathbb{N}} \mathfrak{m}_n = \infty$, note that the properties of smeariness and finite smeariness are strictly disjoint.

The distinction into two types of FSS highlights two different phenomena. Under FSS of Type I for increasing sample size the effect of finite sample smeariness perseveres, whereas under FSS of Type II it will disappear. Nonetheless, both phenomena affect inference tasks involving the Fréchet sample mean in the finite sample regime. In Section 3 we exhaustively characterize the two types of FSS on the circle and torus, and relate the concept to whether the underlying distribution is concentrated within a sufficiently small set.

Notably, if there exists a random variable X on a metric space (M, d) with a nonunique Fréchet population mean, Hundrieser, Eltzner and Huckemann (2024, Corollary 2.6) have shown that for every $K > 1$ there exist "nearby" random variables X_K, $K > 0$, with unique Fréchet population means such that

$$\sup_{n\in\mathbb{N}} \mathfrak{m}_n \geq K.$$

The variance modulation $\mathfrak{m}_n$ quantifies the scale of finite sample smeariness for the sample size $n \in \mathbb{N}$ and showcases a potential loss in reliability of naive quantile based testing methodology. Indeed, in Section 5 we theoretically underpin this insight for the circle and torus (Proposition 5.2) and additionally confirm it in various simulations in Section 6. Based on these insights, the variance modulation $\mathfrak{m}_n$ can be used as a diagnostic tool to assess the reliability of quantile based inference. For practical contexts it is therefore only relevant to assess whether the variance modulation is considerably larger than one. To this end, we construct in the following a suitable statistical test.

2.2. Test for presence of finite sample smeariness and its heuristic

For a random variable X on (M, d) and sample size $n \in \mathbb{N}$ we seek to test the hypothesis

$$H_{0,n} \colon \mathfrak{m}_n \leq 1 \quad \text{vs.} \quad H_{1,n} \colon \mathfrak{m}_n > 1.$$

[3]Which is defined as $d_H(A, B) = \max(\sup_{p\in A} \tilde{d}(p, B), \sup_{p\in B} \tilde{d}(p, A))$ where $\tilde{d}(p, A) \coloneqq \inf_{x\in A} d(p, x)$ for $A, B \subseteq M$.

Note that rejection of the null hypothesis implies the presence of smeariness or finite sample smeariness. As discussed above, since smeariness has only been confirmed for distributions from parametric families at the boundary of the regime with nonunique Fréchet population means, the rejection of $H_{0,n}$ in practice indicates the presence of finite sample smeariness. Notably, if X is FSS of Type II, it follows that $\lim_{n\to\infty} \mathfrak{m}_n = 1$, and therefore with increasing sample size the hypothesis $H_{0,n}$ becomes increasingly challenging to discern from $H_{1,n}$. This is not problematic for practical considerations since we are only interested if the variance modulation $\mathfrak{m}_n$ is considerably large to assess the unreliability of quantile based inference. Hence, if $H_{0,n}$ is not rejected, even though $H_{0,n}$ might actually be false, the variance modulation $\mathfrak{m}_n$ is likely not too large and potential effects of finite sample smeariness can be ignored.

To test the null hypothesis $H_{0,n}$, for a given sample $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} X$, we utilize the empirical Fréchet variance

$$\widehat{V}_n := \frac{1}{n}\sum_{j=1}^{n} d^2(\widehat{\mu}_n, X_j) = F_n(\widehat{\mu}_n)\,, \tag{2.1}$$

and design a bootstrap procedure to estimate the numerator of the variance modulation. To this end, let $B > 0$ be a large integer, μ_b^* be a Fréchet sample mean based on an n-out-of-n bootstrap sample of $X_1, \dots, X_n$ for $b = 1, \dots, B$ and μ^* the Fréchet sample mean of the μ_b^* $(b = 1, \dots, B)$, yielding

$$\widehat{V}^*_{n,B} := \frac{1}{B}\sum_{b=1}^{B} d^2(\widehat{\mu}^*, \mu_b^*) \quad \text{and} \quad \widehat{W}^*_{n,B} := \frac{1}{B}\sum_{b=1}^{B} d^4(\widehat{\mu}^*, \mu_b^*)^4\,. \tag{2.2}$$

With these we formulate our statistical test.

Test 2.5 (for presence of finite sample smeariness)**.** *For random variables $X_1, \dots, X_n$ on (M, d) with sample mean $\widehat{\mu}_n$, bootstrap means $\mu_1^*, \dots, \mu_B^*$, and their mean $\widehat{\mu}^*$, as above and the notation from (2.1) and (2.2), define the* empirical (bootstrap-based) variance modulation*,*

$$\widehat{\mathfrak{m}}_n^* := \frac{n\widehat{V}^*_{n,B}}{\widehat{V}_n}\,. \tag{2.3}$$

With the standard normal $(1-\alpha)$-quantile $\phi_{1-\alpha}$ for $\alpha \in (0, 1)$ and again the notation from (2.1) and (2.2), reject the hypothesis $H_{0,n}$ at nominal level α if

$$\widehat{\mathfrak{m}}_n^* - 1 > h_{n,1-\alpha} \textit{ where } h_{n,1-\alpha} = \frac{n\phi_{1-\alpha}}{\sqrt{B}}\,\frac{\sqrt{\widehat{W}^*_{n,B} - (\widehat{V}^*_{n,B})^2}}{\widehat{V}_n}\,.$$

Remark 2.6 (Computation). The computational complexity of Test 2.5 is of order $\mathcal{O}(B \cdot \mathcal{C}(n))$, where B denotes the number of bootstrap samples B and $\mathcal{C}(n)$ denotes the computational effort in computing the Fréchet sample mean for a sample of size n. For instance, $\mathcal{C}(n) = \mathcal{O}(n\log(n))$ on the circle with the algorithm from Hotz and Huckemann (2015, Remark 3.8).

To illustrate the performance of the test we conduct in Section 6 multiple simulations, showcasing that the test keeps the nominal level if the variance modulation is equal or close to one, and rejects if it is much larger than one.

In the following we provide a heuristic derivation why we believe this test to be asymptotically of nominal level α. For this argument we rely on results by Dubey and Müller (2019) which assume the metric space (M, d) to be bounded and the Fréchet population and sample mean for the random variable X is (almost surely) unique. Upon defining the quantities,

$$V := \mathbb{E}[d^2(\mu, X)], \quad W := \mathbb{E}[d^4(\mu, X)], \quad \widehat{W}_n := \frac{1}{n}\sum_{j=1}^{n} d^4(\widehat{\mu}_n, X_j)\,,$$

where the 4-th moment exist due to boundedness of (M, d), Dubey and Müller (2019) show under additional mild assumptions[4] that

$$\sqrt{n}(\widehat{V}_n - V) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, W - V^2\right). \tag{2.4}$$

Herein, $\widehat{W}_n - \widehat{V}_n^2$ is an asymptotically unbiased estimator for $W - V^2$, that also satisfies $\sqrt{n}$ asymptotic normality (Dubey and Müller, 2019).

Moreover, since the bootstrap samples are all independent, bootstrap sample means $\mu_1^*, \ldots, \mu_B^*$ are i.i.d. (conditionally on $X_1, \ldots, X_n$). Hence, under the assumptions of Dubey and Müller (2019), with the Fréchet sample mean μ^* of the μ_b^* $(b = 1, \ldots, B)$, defining

$$V_n^* := \mathbb{E}[d^2(\mu^*, \mu_1^*)|X_1, \ldots, X_n], \quad W_n^* := \mathbb{E}[d^4(\mu^*, \mu_1^*)|X_1, \ldots, X_n],$$
$$\widehat{W}_{n,B}^* := \frac{1}{B}\sum_{b=1}^{B} d^4(\widehat{\mu}^*, \mu_b^*)^4\,,$$

it follows for $B \to \infty$ (conditionally on $X_1 \ldots, X_n$) that

$$\sqrt{B}\left(\widehat{V}_{n,B}^* - V_n^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, W_n^* - (V_n^*)^2\right). \tag{2.5}$$

Here, the limiting variance can also be consistently estimated by $\widehat{W}_{n,B}^* - (\widehat{V}_{n,B}^*)^2$.

For the bootstrap Test 2.5 to perform reliably, we require that the following approximate equality is met for $n \to \infty$,

$$n\mathbb{E}[d^2(\hat{\mu}_n, \mu_1^*)|X_1, \ldots, X_n] = n\mathbb{E}[d^2(\mu, \hat{\mu}_n)] + o_{\mathbb{P}^*}(1), \tag{2.6}$$

where $\mathbb{P}^*$ denotes the outer probability measure, see van der Vaart and Wellner (1996). This means that the distribution of the bootstrap sample mean consistently estimates (conditional on $X_1, \ldots, X_n$) the underlying distribution of $\hat{\mu}_n$. For the circle and the torus we show that condition (2.6) holds under mild assumptions (Propositions 4.6 and 4.7). Then, plugging $\hat{\mu}_n$ into the (conditional)

[4] These conditions encompass (almost sure) uniqueness of Fréchet sample and population mean, a separability condition on the population Fréchet function near the Fréchet mean and a growth bound on the metric entropy for (M, d).

Fréchet population function of μ_1^*, it follows under (2.6) and the validity of $H_{0,n}$ that

$$\begin{aligned} V_n^* &\leq \mathbb{E}[d^2(\hat{\mu}_n, \mu_1^*)|X_1, \ldots, X_n] \\ &= \mathbb{E}[d^2(\mu, \hat{\mu}_n)] + o_{\mathbb{P}^*}(n^{-1}) \\ &\leq \frac{\mathbb{E}[d^2(\mu, X)]}{n} + o_{\mathbb{P}^*}(n^{-1}) = \frac{V}{n} + o_{\mathbb{P}^*}(n^{-1}). \end{aligned} \tag{2.7}$$

Moreover, recalling (2.4), it follows that V can be estimated by $\widehat{V}_n$ up to a stochastic error of order $\mathcal{O}_P\big(\sqrt{(W - V^2)/n}\big)$. Altogether, under $H_{0,n}$ and assuming (2.6), by (2.5) with B large enough and imposing that $\sqrt{B}/n = o(1)$ it follows for every $\alpha \in (0,1)$ that

$$\mathbb{P}^* \left(\mathbb{P} \left(\widehat{V}_{n,B}^* > \frac{\widehat{V}_n}{n} + \phi_{1-\alpha} \sqrt{\frac{\widehat{W}_{n,B}^* - (\widehat{V}_{n,B}^*)^2}{B}} \,\middle|\, X_1, \ldots, X_n \right) \leq \alpha \right) \to 1. \tag{2.8}$$

Overall, this suggests that in the large sample regime and if B is moderately large, the test for finite sample smeariness attains the correct level. Formalizing this statement would amount to confirming (2.5) uniformly in n for $B \to \infty$. Moreover, to assess the performance under the alternative, one would additionally need to understand the sharpness of inequality (2.7). These issues are left for future research.

3. Finite sample smeariness on circles and tori

To analyze the concept of finite sample smeariness in greater detail, we establish in this section a complete description for the circle and the torus. To this end, we parametrize the circle as the space $\mathbb{S}^1 = [-\pi, \pi)$ with $-\pi$ and π identified and the usual arc length distance

$$d_{\mathbb{S}^1}(x, y) = \min\left\{|x - y|, 2\pi - |x - y|\right\}.$$

This parametrization of $\mathbb{S}^1$ enables us to interpret elements of $\mathbb{S}^1$ also as elements in $\mathbb{R}$. The following result provides an exhaustive characterization of the behavior of the variance modulation.

Theorem 3.1. *For arbitrary $n \geq 2$ suppose that X is a random variable on $\mathbb{S}^1$ with unique Fréchet mean μ and support $J \subseteq \mathbb{S}^1$. Then $\mathfrak{m}_n > 1$ for all $2 \leq n \in \mathbb{N}$ under either of the two following conditions*

(i) J is not contained in a closed half circle,
(ii) J contains two antipodal points, both are attained by X with positive probability.

Moreover, $\mathfrak{m}_n = 1$ for all $n \in \mathbb{N}$ under either of the two following conditions

(iii) J is strictly contained in a closed half circle,

(iv) J is contained in a closed half circle, at most one end point is attained by X with positive probability.

Finally, suppose that the distribution of X has a continuous density f near the antipode $\overline{\mu}$ of μ.

(v) If $f(\overline{\mu}) = 0$ then $\lim_{n\to\infty} \mathfrak{m}_n = 1$,
(vi) if $0 < f(\overline{\mu}) < (2\pi)^{-1}$ then $\lim_{n\to\infty} \mathfrak{m}_n = (1 - f(\overline{\mu})2\pi)^{-2} > 1$.

The proof is provided in Section A of the supplement (Hundrieser, Eltzner and Huckemann, 2024). It is based on our observation that the variance modulation is strictly larger than one if the Fréchet sample mean differs with positive probability from a Euclidean average in a chart at the Fréchet population mean. In particular, under conditions (i) and (ii), this difference occurs, whereas under conditions (iii) and (iv) the two quantities coincide.

As a consequence of this result, we derive the following characterizations for both types of FSS on the circle.

Corollary 3.2. *Let X be a random variable on $\mathbb{S}^1$ whose distribution has a continuous density f near the antipode $\overline{\mu}$ of the Fréchet population mean μ.*

(i) Then, X is FSS of Type I if and only if $0 < f(\overline{\mu}) < 1/(2\pi)$.
(ii) Further, X is FSS of Type II if and only if $f(\overline{\mu}) = 0$ and condition (i) or (ii) of Theorem 3.1 is fulfilled.

Remark 3.3. In the above case of a continuous density f near the antipode $\overline{\mu}$ of the Fréchet mean, $f(\overline{\mu}) > (2\pi)^{-1}$ is not possible and $f(\overline{\mu}) = (2\pi)^{-1}$ leads to smeariness, as shown by Hotz and Huckemann (2015, Theorems 1(ii) and 3(ii)).

Simple cases of FSS on the circle are illustrated in the following.

Example 3.4 (Von Mises mixtures)**.** *On $\mathbb{S}^1$ we consider von Mises mixtures with antipodal modes with parameters $\kappa, \lambda \geq 0, \beta \in [0, 1]$ and density with respect to arc length measure*

$$d\mathbb{P}^{\kappa,\beta,\lambda}_{vMm}(x) := \beta\,\frac{\exp\left(\kappa\cos(x)\right)}{2\pi I_0(\kappa)}\,dx + (1-\beta)\frac{\exp\left(\lambda\cos(x+\pi)\right)}{2\pi I_0(\lambda)}\,dx \tag{3.1}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of order 0, e.g., Mardia and Jupp (2000, p. 36). By symmetry, the von Mises mixture $d\mathbb{P}^{\kappa,\beta,\lambda}_{vMm}$ attains either a unique mean at 0 or π, or nonunique means at $\{-t, t\}$ for some $t \in (0, \pi)$. Furthermore, we define for $r \geq 0$ the function cutting out and mirroring a disk of radius r about $-\pi$:

$$\zeta^r : \mathbb{S}^1 \to \mathbb{S}^1, \quad p \mapsto \begin{cases} p & \text{if } p \in [-\pi + r, \pi - r) \\ p + \pi & \text{if } p \in [-\pi, -\pi + r) \\ p - \pi & \text{if } p \in [\pi - r, \pi) \end{cases} \tag{3.2}$$

For the von Mises mixture $\mathbb{P}^{\kappa,\beta,\lambda}_{vMm}$ we then denote the push-forward measure under ζ^r by

$$\mathbb{P}^{\kappa,\beta,\lambda,r}_{vMm} := \zeta^r_*\,\mathbb{P}^{\kappa,\beta,\lambda}_{vMm}\,, \tag{3.3}$$

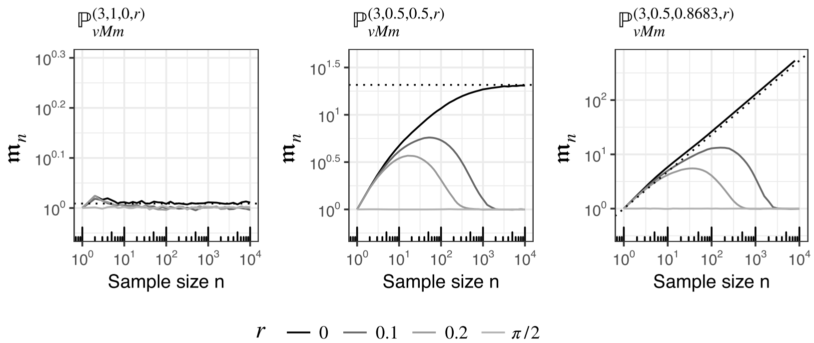

FIG 1. *Log-log plots of variance modulation curves $n \mapsto \mathfrak{m}_n$, for varying von Mises mixture (vMm) distributions defined in (3.1) in black ($r = 0$) and vMm distributions with a disk cut out of radius $r = 0.1, 0.2, \pi/2$ (in fading gray) around the antipode of the mean, as defined in (3.3) based on 100 000 simulation runs for each sample size. In the left and middle display, the vMm distributions (black) feature Type I FSS, with the dotted line giving the asymptotic scaled variance, the two vMm distributions with an interval cut out (gray curves) of radius $0 < r < \pi/2$ feature Type II FSS. In the right display, the vMm distribution (black) features smeariness, with theoretical scaled variance (dotted), the vMm distributions with an interval cut out (gray curves) of radius $0 < r < \pi/2$ feature Type II FSS. Only the vMm distributions with an interval cut out of radius $r = \pi/2$ (light gray) features no FSS at all.*

which preserves all mass except for that in the disk which is mirrored. Recall that by Theorem 3.1, all the distributions $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda,r}$ with unique mean $\mu = 0$ and $r < \pi/2 - \varepsilon$ for some $\varepsilon > 0$ feature FSS (if they are not smeary themselves), which is of Type I if $r = 0$ and Type II if $r > 0$. For the parameters $(\kappa, \beta, \lambda) \in \{(3, 1, 0), (3, 0.5, 0.5), (3, 0.5, 0.8683)\}$ (the last configuration leads to smeariness of order two, cf. Hotz and Huckemann (2015), as the density attains the local maximum $1/(2\pi)$ at $-\pi$ with negative second derivative) and varying choices for the parameter $r \in \{0, 0.1, 0.2, \pi\}$ the respective variance modulation curves are depicted in Figure 1. Notably, in case of Type I FSS, the curve $\mathfrak{m}_n$ rises from 1 to $(1 - 2\pi f(\overline{\mu}))^{-2}$ whereas under Type II FSS, the curve first rises from 1 and eventually drops to 1.

Depending on the probability distribution near $\overline{\mu}$, also more complicated versions of increase and decrease may occur, as Example 3.5 and Figure 2 teach: every pair of bumps of the density near the antipode may result in a bump of $\mathfrak{m}_n$. As before, however, it starts at 1 and eventually settles at $(1 - 2\pi f(\overline{\mu}))^{-2}$, producing Type I FSS if $f(\overline{\mu}) > 0$ and Type II FSS else.

Example 3.5 (Relating antipodal density to variance modulation)**.** *To investigate the relationship between the variance modulation of intrinsic sample means $n \mapsto \mathfrak{m}_n$ and the density $f^{(t,w)}$ of X near the antipode of the intrinsic population mean $\mu = 0$ we consider a family of distributions for which the density near the antipode $\overline{\mu}$ is piecewise constant. Let $l \in \mathbb{N}$, $w = (w_1, \dots, w_l) \in [0, 1]^l$, $t = (t_1, \dots, t_l) \in [0, \pi)^l$ with $t_0 \coloneqq 0 < t_1 < \dots < t_l < \pi$ and define the*

TABLE 2
Selected values for the parameters l, t, w and their resulting type of FSS.

	(a)	(b)	(c)	(d)	(e)
l	1	2	4	2	4
t	1.5	(0.8, 2)	(0.1, 0.2, 0.5, 2)	(0.8, 2)	(0.1, 0.2, 0.5, 2)
w	0.5	(0.5, 1)	(0.5, 0.8, 0.0, 1)	(0.0, 1)	(0.0, 0.85, 0.0, 1)
Type of FSS	I	I	I	II	II

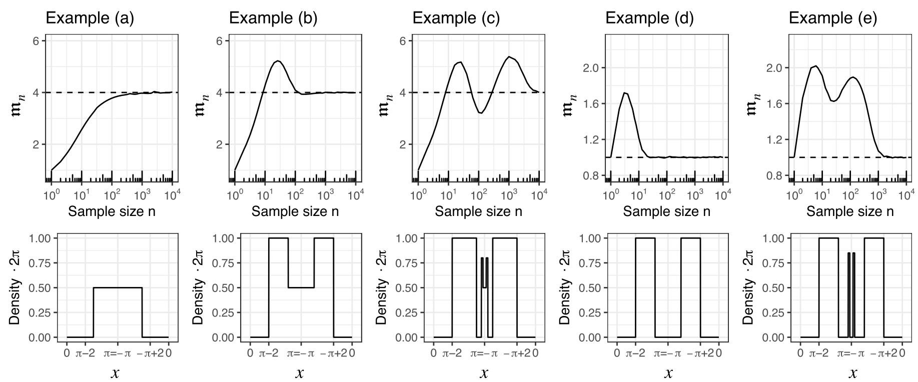

FIG 2. *Top: Variance modulation curves as in Figure 1 (here the vertical is not in log-scale), for each distribution in Table 2 of Example 3.5 based on 100 000 simulations for each sample size. Bottom: Density part of the corresponding distributions (that comprise a δ-measure at the origin). In examples (a)-(c) we have Type I FSS, whereas for (d), (e) we have Type II FSS.*

distribution $\mathbb{P}_U^{(t,w)}$ by

$$d\mathbb{P}_U^{(t,w)}(x) := k \cdot d\delta_0(x) + \frac{1}{2\pi} \cdot f^{(t,w)}(x)dx \qquad \textit{with}$$

$$f^{(t,w)}(x) := \sum_{i=1}^{l} w_i \mathbb{1}_{[-\pi+t_{i-1}, -\pi+t_i) \cup (\pi-t_i, \pi-t_{i-1}]}(x)$$

where $k = k(t,w) > 0$ is a normalization constant to ensure that $\mathbb{P}_U^{(t,w)}$ is a probability measure. In Table 2 we list cases (a) – (e) of parameter choices considered.

In all cases the population mean of $\mathbb{P}_U^{(t,w)}$ is unique and located at $\mu = 0$. Whenever the density at the antipode is strictly between zero and $1/2\pi$ we have FSS of Type I. Regardless of the type of FSS, every pair of bumps in the density near the antipode corresponds to a single bump in the variance modulation curve. Thus, in case of Type I FSS the rescaled Fréchet sample variance $n\mathbb{E}[d(\widehat{\mu}_n, \mu)^2]$ approaches asymptotically a value strictly above the Euclidean variance $\mathbb{E}[d(X,\mu)^2]$, in case of Type II FSS it approaches asymptotically the Euclidean variance.

Remark 3.6. As Example 3.5 and Figure 2 teach, the variance modulation $\mathfrak{m}_n$

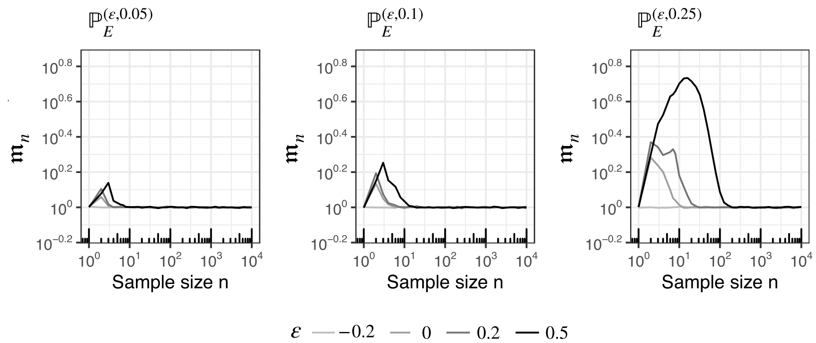

FIG 3. *Variance modulation curves, as in Figure 1, for the three types of distributions $\mathbb{P}_E^{(\varepsilon,w)}$ from Example 3.7 with weight w of point mass at $\pm(\pi/2+\varepsilon)$ based on 100 000 simulation runs for each sample size. Black: $\varepsilon = 0.5$, dark gray: $\varepsilon = 0.2$, gray: $\varepsilon = 0$, light gray: $\varepsilon = -0.2$. We have Type II FSS for all values of $\varepsilon \geq 0$ and no FSS (constant scaled variance) for $\varepsilon < 0$.*

may be non-monotonic, exhibiting different behaviors for different sample sizes.

The following example illustrates Theorem 3.1 with two point masses at or beyond the equator (case (*ii*)), or before (case (*i*)), with a possibly (in case (*iii*)) nonunique sample mean set, of which uniformly at random, a sample mean is selected.

Example 3.7. *Letting $\varepsilon \in [-\pi/2, \pi/2)$ and $w \in (0, 1/4]$, define the circular distribution $\mathbb{P}_E^{(\varepsilon,w)}$ by*

$$d\mathbb{P}_E^{(\varepsilon,w)}(x) \coloneqq (1-2w) \cdot \mathbb{1}_{[-1/2,1/2]}(x)dx + w\, d\delta_{\pi/2+\varepsilon}(x) + w\, d\delta_{-\pi/2-\varepsilon}(x)\,,$$

which assigns at least half the mass to $[-1/2, 1/2]$ and the rest is evenly distributed close to the equator at $\pi/2 + \varepsilon$ and $-\pi/2 - \varepsilon$.

- *(i) For $\varepsilon < 0$ these distributions are supported in $(-\pi/2, \pi/2)$ and thus by Theorem 3.1 (iii) feature no FSS.*
- *(ii) For $\varepsilon > 0$, in contrast by Theorem 3.1 (i), they always feature FSS, which is, by Corollary 3.2 (ii) of Type II.*
- *(iii) For $\varepsilon = 0$, by Theorem 3.1 (ii) and Corollary 3.2 (ii), they also feature FSS of Type II. Indeed, samples featuring only points at $\pm\pi/2$ and no others, occurring with positive probability, lead to nonunique sample means, namely at $\alpha \in (-\pi/2, \pi/2)$ and its counterpart $\tilde{\alpha} \coloneqq \mathbb{1}(\alpha > 0)\pi - \mathbb{1}(\alpha \leq 0)\pi - \alpha$, mirrored along the equator. In this setting, the Hausdorff distance between population mean set $\{0\}$ and sample mean set $\{\alpha, \tilde{\alpha}\}$ equals $|\tilde{\alpha}|$, corresponding to the situation where $\hat{\mu}_n = \tilde{\alpha}$ was selected as the sample mean.*

Each panel in Figure 3 illustrates the three cases, with larger effect on FSS the larger the weight w of each of the point masses.

While we defined FSS for arbitrary metric spaces, we have investigated it only for the circle. These results extend at once to the m-torus $\mathbb{T}^m = \left(\mathbb{S}^1\right)^m =$

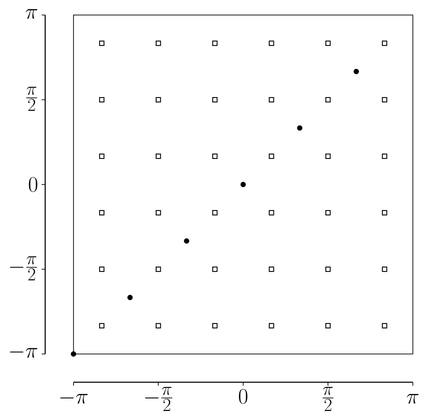

FIG 4. *Six sample points (black circles) on the diagonal of the two-torus* $\mathbb{T}^2$ *and their equally spaced* $6^2 = 36$ *Fréchet sample means (empty squares), cf. Remark 4.5.*

$\bigtimes_{i=1}^m [-\pi, \pi)$, $m \in \mathbb{N}$, equipped with the canonical product metric

$$d_{\mathbb{T}^m}(x, y) = \sqrt{\sum_{i=1}^m d(x^{(i)}, y^{(i)})^2}, \quad x = \begin{pmatrix} x^{(1)} \\ \vdots \\ x^{(m)} \end{pmatrix}, \; y = \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{pmatrix} \in \mathbb{T}^m \,.$$

Indeed, $\mu \in \mathbb{T}^m$ is a minimizer of the population Fréchet function (1.1) of a random variable X on $\mathbb{T}^m$ if and only if all of its coordinates $\mu^{(i)}$ are minimizers of the population Fréchet functions of the marginals $X^{(i)}$ on the i-th circle, $i \in \{1, \dots, m\}$.

Remark 3.8. In consequence, due to Hotz and Huckemann (2015, Corollary 3), for a sample $X_1, \dots, X_n$ on the m-torus $\mathbb{T}^m$ there are at most n^m minimizers of the sample Fréchet function (1.2). For instance, every one of the n^m grid points

$$\left(-\pi + \frac{2\pi(2j_1 - 1)}{2n}, \dots, -\pi + \frac{2\pi(2j_m - 1)}{2n}\right), \quad j_1, \dots, j_m \in \{1, \dots, n\}$$

is a minimizer of the sample Fréchet function if the sample is given by $X_j = \big(-\pi + \frac{2\pi(j-1)}{n}\big)(1, \dots, 1)^T \in \mathbb{T}^m$ for $j \in \{1, \dots, n\}$, cf. Figure 4. Indeed, taking the i-th component ($1 \leq i \leq m$) of the sample on $\mathbb{T}^m$ yields a sample on $\mathbb{S}^1$, given by $\{X_n^{(i)}, \dots, X_n^{(i)}\}$ with $X_j^{(i)} = \big(-\pi + \frac{2\pi(j-1)}{n}\big)$ for each $j \in \{1, \dots, n\}$. The corresponding sample Fréchet function for the projection has exactly n minimizers, all equally spaced and located at $-\pi + \frac{2\pi(2j_i - 1)}{2n}$ for $j_i \in \{1, \dots n\}$.

In particular, Theorem 3.1 yields the following behavior for the variance modulation of a random variable on $\mathbb{T}^m$.

Corollary 3.9. *For arbitrary* $n \geq 2$ *and* $m \in \mathbb{N}$ *let* $X = (X^{(1)}, \dots, X^{(m)})$ *be a random variable on the torus* $\mathbb{T}^m$ *with unique Fréchet population mean* μ*. For*

$i \in \{1, \ldots m\}$ let the support of $X^{(i)}$ be $J^{(i)} \subseteq \mathbb{S}^1$. Then $\mathfrak{m}_n > 1$ for all $n \in \mathbb{N}$ under either of the following two conditions

- *(i) there exists $i \in \{1, \ldots m\}$ such that $J^{(i)}$ is not contained in a closed half circle,*
- *(ii) there exists $i \in \{1, \ldots m\}$ such that $J^{(i)}$ contains two antipodal points, each of which is attained by $X^{(i)}$ with positive probability.*

Moreover, $\mathfrak{m}_n = 1$ for all $n \in \mathbb{N}$ under either of the following two conditions

- *(iii) for each $i \in \{1, \ldots, m\}$ the support $J^{(i)}$ is strictly contained in a closed half circle,*
- *(iv) for each $i \in \{1, \ldots, m\}$ the support $J^{(i)}$ is contained in a closed half circle and at most one of the end points is attained by $X^{(i)}$ with positive probability.*

Finally, suppose for each $i \in \{1, \ldots, m\}$ that the component $X^{(i)}$ has a continuous density $f^{(i)}$ near the antipode $\overline{\mu}^{(i)}$ of $\mu^{(i)}$.

- *(v) If $f^{(i)}(\overline{\mu}^{(i)}) = 0$ for all $i \in \{1, \ldots, m\}$, then $\lim_{n\to\infty} \mathfrak{m}_n = 1$,*
- *(vi) if $0 \le f^{(i)}(\overline{\mu}^{(i)}) < \frac{1}{2\pi}$ for all $i \in \{1, \ldots, m\}$ with $f^{(i)}(\overline{\mu}^{(i)}) > 0$ for at least one component then*

$$\lim_{n\to\infty} \mathfrak{m}_n = \left(\sum_{k=1}^{m} \frac{\mathbb{E}[d^2(X^{(k)}, \mu^{(i)})]}{(1 - f^{(i)}(\overline{\mu}^{(i)})2\pi)^2} \right) \cdot \left(\sum_{i=1}^{m} \mathbb{E}[d^2(X^{(i)}, \mu^{(i)})] \right)^{-1} > 1.$$

As a consequence, we obtain a similar characterization as in Corollary 3.2 for FSS on $\mathbb{T}^m$.

Corollary 3.10. *Let $X = (X^{(1)}, \ldots, X^{(m)})$ be a random variable on $\mathbb{T}^m$ and suppose for each $i \in \{1, \ldots, m\}$ there exists a continuous density near the antipode $\overline{\mu}^{(i)}$ of the respective coordinate's Fréchet population mean $\mu^{(i)}$.*

- *(i) Then, X is FSS of Type I if and only if $0 \le f^{(i)}(\overline{\mu}^{(i)}) < \frac{1}{2\pi}$ for all $i \in \{1, \ldots, m\}$ with $f^{(i)}(\overline{\mu}^{(i)}) > 0$ for at least one component.*
- *(ii) Further, X is FSS of Type II if and only if $f^{(i)}(\overline{\mu}^{(i)}) = 0$ for all $i \in \{1, \ldots, m\}$ and condition (i) or (ii) of Corollary 3.9 is fulfilled.*

4. Asymptotics of Fréchet means on the circle and the torus

In this section, we give with an exposition on the asymptotic behavior of Fréchet sample means on the circle. We then proceed with a proof on consistency of the bootstrap and afterwards show under the presence of FSS that moments of the Fréchet sample means and their bootstrapped versions converge to the respective moment of the limit distribution.

4.1. Central limit theorem and bootstrap consistency for circular Fréchet means

We begin with the central limit theorem for Fréchet sample means on the circle under the following assumption.

Assumption 4.1. Let X be a random element on $\mathbb{S}^1$

(i) with unique Fréchet population mean $\mu = 0$ and
(ii) that features a continuous density f on $(\pi - \delta, \pi) \cup [-\pi, -\pi + \delta)$ with respect to the arc length measure for some $\delta > 0$ such that

$$f(-\pi) = \lim_{x \searrow -\pi} f(x) = \lim_{x \nearrow \pi} f(x) < \frac{1}{2\pi}.$$

Theorem 4.2 (Central Limit Theorem on the Circle by McKilliam, Quinn and Clarkson (2012); Hotz and Huckemann (2015))**.** *Let X be a random element on $\mathbb{S}^1$ which fulfills Assumption 4.1. Consider $n \in \mathbb{N}$ i.i.d. random elements $X_1, \dots, X_n \sim X$ on $\mathbb{S}^1$ with a measurable selection $\widehat{\mu}_n$ of Fréchet sample means. Then, for $n \to \infty$, it follows,*

$$\sqrt{n}\, \widehat{\mu}_n \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\mathbb{E}[X^2]}{\big(1 - 2\pi f(-\pi)\big)^2}\right),$$

where $\mathbb{E}[X^2]$ denotes Euclidean variance.

Remark 4.3. The case $f(-\pi) = \frac{1}{2\pi}$ where the density f is $(k+1)$-times continuously differentiable near $-\pi$ for $k \in \mathbb{N}$, such that the first k derivatives vanish at $-\pi$ whereas the derivative of order $k+1$ at $-\pi$ does not, has been investigated by Hotz and Huckemann (2015). In this setting, the convergence rate for Fréchet sample means turns out to be of order $n^{\frac{-1}{2(k+1)}}$ which is strictly slower than the standard $n^{-1/2}$-rate. The phenomenon of a slower convergence rate is known as *smeariness*. Moreover, it is not possible that the antipode of the Fréchet population mean exhibits a point mass or fulfills $f(-\pi) > \frac{1}{2\pi}$, cf. Hotz and Huckemann (2015, Theorem 1).

The limit distribution of the Fréchet sample mean depends on the behavior of the density near the antipode of the Fréchet population mean. In particular, in case the density near the antipode of the population mean does not vanish, approximating the distribution of $\sqrt{n}\widehat{\mu}_n$ by a centered Gaussian with an estimated variance $\frac{1}{n}\sum_{i=1}^n d(\hat{\mu}_n, X_i)$ is not suited. Unfortunately, estimating the density at the antipode poses a considerable statistical challenge, since the population mean additionally needs to be estimated. Further, available density estimation methods typically impose certain regularity assumptions at the density, and deteriorate in performance with decreasing regularity (Tsybakov, 2009).

An alternative approach to imitate the law of scaled Fréchet sample means without imposing regularity assumptions on the antipodal density is by means of bootstrap methods. In the following, we show that the naive n-out-of-n bootstrap is indeed (asymptotically) consistent in approximating the distribution of

Fréchet sample means. For this purpose, we denote by

$$\mathrm{BL}_1(\mathbb{R}) := \left\{ f\colon \mathbb{R} \to [-1,1] \;\middle|\; |f(x)-f(y)| \le |x-y| \text{ for all } x,y \in \mathbb{R} \right\}$$

the space of bounded Lipschitz functions on $\mathbb{R}$ which are bounded by one and have Lipschitz modulus at most one. Further, we denote by $\xrightarrow{\mathbb{P}^*}$ the convergence in outer probability, cf. van der Vaart and Wellner (1996).

Theorem 4.4 (Consistency of Bootstrap)**.** *Let X be a random element on $\mathbb{S}^1$ which fulfills Assumption 4.1. Consider $n \in \mathbb{N}$ i.i.d. random elements $X_1, \dots, X_n \sim X$ on $\mathbb{S}^1$ with a measurable selection $\widehat{\mu}_n$ of Fréchet sample means. Further, consider a bootstrap sample $X_1^*, \dots, X_n^* \overset{\text{i.i.d.}}{\sim} \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ with measurable selection $\widehat{\mu}_n^*$ of its Fréchet sample mean. Then, for $n \to \infty$, it follows,*

$$\sup_{h \in \mathrm{BL}_1(\mathbb{R})} \left| \mathbb{E}\left[h\left(\sqrt{n}\left(\widehat{\mu}_n^* - \widehat{\mu}_n\right)\right) \middle| X_1, \dots, X_n \right] - \mathbb{E}\left[h\left(\sqrt{n}\,\widehat{\mu}_n\right)\right] \right| \xrightarrow{\mathbb{P}^*} 0.$$

The proof is stated in Section B of the supplement (Hundrieser, Eltzner and Huckemann, 2024) and relies on a careful analysis of the empirical and population Fréchet function. Along the way, we also employ consistency of the bootstrap for the Euclidean sample mean (van der Vaart, 2000, Theorem 23.4) as well as utilize empirical process theory to suitably bound certain errors.

Remark 4.5. These results extend at once to the m-torus $\mathbb{T}^m = \left(\mathbb{S}^1\right)^m$ equipped with the canonical product metric as the Fréchet mean on the m-torus is given by the vector of circular Fréchet means for each component of the torus.

4.2. Moment convergence of Fréchet means

In order to employ the theory on the asymptotics of Fréchet sample means and bootstrap version for the formulation of Hotelling tests it is necessary to estimate the variance of the limit distribution. The following result guarantees that the naive plug-in estimator for the covariance is indeed consistent. In fact, we prove that all moments of scaled Fréchet sample means and bootstrap variants converge to the corresponding moment of their limit distribution.

Proposition 4.6. *Let X be a random element on $\mathbb{S}^1$ that meets Assumption 4.1. Consider $n \in \mathbb{N}$ i.i.d. random elements $X_1, \dots, X_n \sim X$ on $\mathbb{S}^1$ with a measurable selection $\widehat{\mu}_n$ of Fréchet sample means. Then for all $p \ge 1$ we have*

$$\sup_{n \in \mathbb{N}} \mathbb{E}\left[\left| \sqrt{n} d(\hat{\mu}_n, \mu) \right|^p \right] = \sup_{n \in \mathbb{N}} \mathbb{E}\left[\left| \sqrt{n} \hat{\mu}_n \right|^p \right] < \infty.$$

Further, for $Z \sim \mathcal{N}\left(0, \mathbb{E}[X^2]/(1-2\pi f(-\pi))^2\right)$ we have for every $p \ge 1$ as n tends to infinity that

$$\mathbb{E}\left[\left(\sqrt{n} d(\hat{\mu}_n, \mu) \right)^p \right] = \mathbb{E}\left[\left(\sqrt{n} \hat{\mu}_n \right)^p \right] \to \mathbb{E}[Z^p].$$

The proof is stated in Section C of the Supplement (Hundrieser, Eltzner and Huckemann, 2024) and uses a moment convergence result for M-estimators relying on the theory by Nishiyama (2010). Explicit bounds for $\mathbb{E}\left[|\sqrt{n}d(\hat{\mu}_n,\mu)|^p\right]$ for Fréchet means on general metric spaces were also established by Schötz (2019b), Ahidar-Coutrix, Le Gouic and Paris (2019), and Le Gouic et al. (2022).

For the circular setting, moments of bootstrap based Fréchet sample means are also consistent.

Proposition 4.7. *Let X be a random element on $\mathbb{S}^1$ that meets Assumption 4.1. Consider $n \in \mathbb{N}$ i.i.d. random elements $X_1, \ldots, X_n \sim X$ on $\mathbb{S}^1$ with a measurable selection $\widehat{\mu}_n$ of Fréchet sample means. Further, consider a bootstrap sample $X_1^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ with measurable selection $\widehat{\mu}_n^*$ of its Fréchet sample mean. Then,*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[\left|\sqrt{n}d(\widehat{\mu}_n^*, \widehat{\mu}_n)\right|^p\right] \leq \sup_{n\in\mathbb{N}} \mathbb{E}\left[\left|\sqrt{n}(\widehat{\mu}_n^* - \widehat{\mu}_n)\right|^p\right] < \infty$$

for all $p \geq 1$ where the expectation is taken with respect to $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X$ and bootstrap based $X_1^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} n^{-1}\sum_{i=1}^n \delta_{X_i}$. Further, for a random variable $Z \sim \mathcal{N}\big(0, \mathbb{E}[X^2]/(1-2\pi f(-\pi))^2\big)$ we have for every $p \geq 1$ as n tends to infinity,

$$\mathbb{E}\left[\left(\sqrt{n}(\widehat{\mu}_n^* - \widehat{\mu}_n)\right)^p \,\middle|\, X_1, \ldots, X_n\right] \xrightarrow{\mathbb{P}^*} \mathbb{E}[Z^p].$$

The proof is deferred to Section C of the Supplement (Hundrieser, Eltzner and Huckemann, 2024) and relies on a general result for conditional moment convergence of bootstrap M-estimators by Kato (2011).

5. One- and two-sample tests for Fréchet means under finite sample smeariness

We begin with a brief review of the celebrated central limit theorem (CLT) by Bhattacharya and Patrangenaru (2005) for a k-dimensional Riemannian manifold M and corresponding tests proposed therein. The CLT states that, under suitable conditions (further clarified in Bhattacharya and Lin (2017); Eltzner and Huckemann (2019); Eltzner et al. (2021)), in a local chart $\phi : U \to \mathbb{R}^k$, $U \subset M$, the fluctuation $\phi(\widehat{\mu}_n) - \phi(\mu)$ of the Fréchet sample mean $\widehat{\mu}_n$ about the Fréchet population mean μ, rescaled with the square root of sample size $\sqrt{n}$, is asymptotically for $n \to \infty$ Gaussian with zero mean and covariance given by

$$\Sigma = H^{-1} C H^{-1}\,,$$

cf. (1.3). Here, C is the population covariance of the gradient of the Fréchet function F from (1.1) at μ in the local chart and H is twice the expected value of the Hessian of the Fréchet function at μ in the local chart. One of the above mentioned conditions is that H be positive definite. In our language, this means that $\widehat{\mu}_n$ is nonsmeary.

For mutually independent samples $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} X$ and $Y_1,\dots,Y_n \overset{\text{i.i.d.}}{\sim} Y$ with population Fréchet means $\mu^{(X)}$ and $\mu^{(Y)}$, respectively, and $\mu_0 \in M$, consider the hypotheses

$$\begin{array}{lll} H_0^1: & \mu^{(X)} = \mu_0 & \text{for the one-sample test,} \\ H_0^2: & \mu^{(X)} = \mu^{(Y)} & \text{for the two-sample test.} \end{array}$$

5.1. Quantile based tests

With a local chart $\phi: U \to \mathbb{R}^k$ where $U \subset M$ contains μ (for H_0^1), or $\mu^{(X)}$ and $\mu^{(Y)}$ (for H_0^2), respectively, Bhattacharya and Patrangenaru (2005) thus suggest considering the following test statistics,

$$\begin{aligned} T^1 &= \quad n\left(\phi(\widehat{\mu}_n^{(X)}) - \phi(\mu_0)\right)^T \widehat{H}_X \widehat{C}_X^{-1} \widehat{H}_X \left(\phi(\widehat{\mu}_n^{(X)}) - \phi(\mu_0)\right), \\ T^2 &= (n+m)\left(\phi(\widehat{\mu}_n^{(X)}) - \phi(\widehat{\mu}_m^{(Y)})\right)^T \widehat{H}_{X,Y} \widehat{C}_{X,Y}^{-1} \widehat{H}_{X,Y} \left(\phi(\widehat{\mu}_n^{(X)}) - \phi(\widehat{\mu}_m^{(Y)})\right) \end{aligned}$$

for the one- and two-sample test, and use χ_k^2 as their asymptotic approximation under the respective null hypothesis. Here $\widehat{H}_X, \widehat{C}_X$ or $\widehat{H}_{X,Y}, \widehat{C}_{X,Y}$ are the suitable plugin estimators of H and C based on the first sample, or the pooled sample, respectively, cf. also Bhattacharya and Bhattacharya (2012, Section 5.4.1).

Test 5.1 (Bhattacharya and Patrangenaru (2005); Bhattacharya and Bhattacharya (2012))**.** *For $0 < \alpha < 1$ and the $1-\alpha$ quantile $\chi^2_{k,1-\alpha}$ of the χ_k^2 distribution, reject at level α*

(i) H_0^1, if $T^1 \geq \chi^2_{k,1-\alpha}$,
(ii) H_0^2, if $T^2 \geq \chi^2_{k,1-\alpha}$.

Asymptotically, while both tests are invariant under diffeomorphisms due to the delta method (Fang and Santos, 2019) and thus independent of the chart chosen, in general, they do not keep the nominal level.

Proposition 5.2. *Let X and Y be random elements on $\mathbb{S}^1$ with unique Fréchet population means $\mu^{(X)}$ and $\mu^{(Y)}$, respectively, such that $\mathbb{P}(X \neq \mu^{(X)}) > 0$ and $\mathbb{P}(Y \neq \mu^{(Y)}) > 0$. Further, assume that X (resp. Y) have continuous densities $f^{(X)}$ (resp. $f^{(Y)}$) near the antipode of $\mu^{(X)}$ (resp. $\mu^{(Y)}$) with $f^{(X)}(\bar{\mu}^{(X)}) < \frac{1}{2\pi}$ (resp. $f^{(Y)}(\bar{\mu}^{(Y)}) < \frac{1}{2\pi}$). Let $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} X$ and $Y_1,\dots,Y_m \overset{\text{i.i.d.}}{\sim} Y$ be mutually independent. Then, for $0 < \alpha < 1$ it follows*

(i) under $\mu^{(X)} = \mu_0$ and $f^{(X)}(\bar{\mu}^{(X)}) > 0$ that

$$\lim_{n\to\infty} \mathbb{P}\left(T^1 \geq \chi^2_{1,1-\alpha}\right) > \alpha,$$

(ii) under $\mu^{(X)} = \mu^{(Y)}$ and $f^{(X)}(\bar{\mu}^{(X)}) + f^{(Y)}(\bar{\mu}^{(Y)}) > 0$ that

$$\lim_{n,m\to\infty} \mathbb{P}\left(T^2 \geq \chi^2_{1,1-\alpha}\right) > \alpha \quad \textit{where} \quad \frac{m}{n+m} \to \delta \in (0,1).$$

In particular, under Type I FSS of X or Y the Tests 5.1 have asymptotically a level strictly higher than the nominal level.

The proof is provided in Section D of the supplement (Hundrieser, Eltzner and Huckemann, 2024). The gist of the proof relies on the fact that under an exponential chart ϕ (which contains the population means) the estimators for $HC^{-1}H$ for the one-sample case,

$$\widehat{H}_X \widehat{C}_X^{-1} \widehat{H}_X = \left(\tfrac{1}{n} \sum_{j=1}^n d^2(\widehat{\mu}_n^{(X)}, X_j)\right)^{-1},$$

and the two-sample case,

$$\widehat{H}_{X,Y} \widehat{C}_{X,Y}^{-1} \widehat{H}_{X,Y} = \left(\tfrac{n+m}{n} \tfrac{1}{n} \sum_{j=1}^n d^2(\widehat{\mu}_n^{(X)}, X_j) + \tfrac{n+m}{m} \tfrac{1}{m} \sum_{j=1}^m d^2(\widehat{\mu}_m^{(Y)}, Y_j)\right)^{-1},$$

do not account for a large (asymptotic) variance modulation. This leads to a systematic underestimation of the asymptotic variance of the Fréchet sample mean and implies the test to reject too often under the null.

Remark 5.3. On the circle, in case of Type II FSS, Tests 5.1 also have a true level higher than their nominal level in the range where the variance modulation $\mathfrak{m}_n$ is strictly larger than 1. This deviation is especially substantial if $\mathfrak{m}_n$ is large, as the simulations in Section 6 show.

5.2. Bootstrap based tests

Tests based on a bootstrap principle have also been proposed by Bhattacharya and Patrangenaru (2005); Bhattacharya and Bhattacharya (2012) where $\Sigma = H^{-1}CH^{-1}$ is estimated by bootstrapping from the samples. As they have not provided details, in the following, we employ the bootstrap one- and two-sample tests by Eltzner and Huckemann (2017), for a given level $0 < \alpha < 1$.

Resample $B \in \mathbb{N}$ times n-out-of-n from the sample $X_1, \dots, X_n$ and let $\widehat{\mu}_{n,b}^{(X),*}$ be the corresponding Fréchet sample means for $b = 1, \dots, B$. Mapping these sample means under ϕ to a Euclidean space yields the covariance estimate $\Sigma_B^{(X),*}$. From another round of resampling obtain new $\widehat{\mu}_{n,b}^{(X),*}$ set

$$T_b^* = \left(\phi(\widehat{\mu}_{n,b}^{(X),*}) - \phi(\widehat{\mu}_n^{(X)})\right)^T (\Sigma_B^{(X),*})^{-1} \left(\phi(\widehat{\mu}_{n,b}^{(X),*}) - \phi(\widehat{\mu}_n^{(X)})\right)$$

for $b \in \{1, \dots, B\}$ and determine $e_{1-\alpha}^*$ such that

$$\frac{\sharp\{b \in \{1, \dots, B\} : T_b^* \le e_{1-\alpha}^*\} - 1}{B} \le 1 - \alpha \le \frac{\sharp\{b \in \{1, \dots, B\} : T_b^* \le e_{1-\alpha}^*\}}{B}.$$

Indeed, under bootstrap consistency, the distribution of the T_b^*, $b \in \{1, \dots, B\}$, mimics the distribution of the unavailable oracle statistic under the null

$$T_{orc}^{1,*} = \left(\phi(\widehat{\mu}_n^{(X)}) - \phi(\mu^{(X)})\right)^T (\Sigma_B^{(X),*})^{-1} \left(\phi(\widehat{\mu}_n^{(X)}) - \phi(\mu^{(X)})\right).$$

Further, define for the one-sample null hypothesis the test statistic

$$T^{1,*} = \big(\phi(\widehat{\mu}_n^{(X)}) - \phi(\mu_0)\big)^T (\Sigma_B^{(X),*})^{-1} \big(\phi(\widehat{\mu}_n^{(X)}) - \phi(\mu_0)\big)\,.$$

Similarly, using m-out-of-m sampling with replacement from $Y_1, \dots, Y_m$, obtain a bootstrap covariance estimate $\Sigma_B^{(Y),*}$. Now, set $A_B = \Sigma_B^{(X),*} + \Sigma_B^{(Y),*}$. This choice of A_B, and not using the pooled variance, proves to be more robust, cf. Huckemann and Eltzner (2020). Then, from another round of sampling obtain $\mu_b^{(X),*}$ and $\widehat{\mu}_{n,b}^{(Y),*}$, for $b \in \{1, \dots, B\}$ set $d_b^{(X),*} = \phi(\widehat{\mu}_{n,b}^{(X),*}) - \phi(\widehat{\mu}_n^{(X)})$, $d_b^{(Y),*} = \phi(\widehat{\mu}_{n,b}^{(Y),*}) - \phi(\widehat{\mu}_m^{(Y)})$ and define

$$T_b^* = \big(d_b^{(X),*} - d_b^{(Y),*}\big)^T A_B^{-1} \big(d_b^{(X),*} - d_b^{(Y),*}\big).$$

Then, determine $f_{1-\alpha}^*$ such that

$$\frac{\sharp\{b \in \{1, \dots, B\} : T_b^* \le f_{1-\alpha}^*\} - 1}{B} \le 1 - \alpha \le \frac{\sharp\{b \in \{1, \dots, B\} : T_b^* \le f_{1-\alpha}^*\}}{B}\,.$$

Further, define for the two-sample hypothesis the test statistic

$$T^{2,*} = \big(\phi(\widehat{\mu}_n^{(X)}) - \phi(\widehat{\mu}_m^{(Y)})\big)^T A_B^{-1} \big(\phi(\widehat{\mu}_n^{(X)}) - \phi(\widehat{\mu}_m^{(Y)})\big)\,.$$

Tests 5.4 (Bootstrap Based)**.** *With the notation above, for* $0 < \alpha < 1$*, reject at level* α

(i) H_0^1*, if* $T^{1,*} \ge e_{1-\alpha}^*$*,*
(ii) H_0^2*, if* $T^{2,*} \ge f_{1-\alpha}^*$*.*

Asymptotically for $n, B \to \infty$, the bootstrap based test are also independent of the chart. Indeed, by the delta method for the bootstrap (Fang and Santos, 2019) the limit law of a diffeomorphism applied on bootstrap quantities is given by the derivative evaluated at the conditional limit. Hence, invoking the delta method for empirical Fréchet means, we find by Theorem 4.4 for every chart that the conditional law of bootstrap Fréchet sample means asymptotically coincides with distribution of empirical Fréchet sample means.

Moreover, for circular data the bootstrap tests attain asymptotically, in contrast to the quantile based Tests 5.1, under both types of FSS the correct level.

Proposition 5.5. *Assume the same setting as in Proposition 5.2. Then, for* $0 < \alpha < 1$*, it follows as* $n, m, B \to \infty$ *with* $m/(n+m) \to \delta \in (0,1)$ *that*

(i) under $\mu^{(X)} = \mu_0$ *that* $\mathbb{P}\left(T^{1,*} \ge e_{1-\alpha}^* \middle| X_1, \dots, X_n\right) \xrightarrow{\mathbb{P}^*} \alpha$*,*

(ii) and under $\mu^{(X)} \ne \mu_0$ *that* $\mathbb{P}\left(T^{1,*} \ge e_{1-\alpha}^* \middle| X_1, \dots, X_n\right) \xrightarrow{\mathbb{P}^*} 1$*.*

(iii) Under $\mu^{(X)} = \mu^{(Y)}$ *it follows that* $\mathbb{P}\left(T^{2,*} \ge f_{1-\alpha}^* \middle| X_1, \dots, X_n\right) \xrightarrow{\mathbb{P}^*} \alpha$*,*

(iv) and under $\mu^{(X)} \ne \mu^{(Y)}$ *that* $\mathbb{P}\left(T^{2,*} \ge f_{1-\alpha}^* \middle| X_1, \dots, X_n\right) \xrightarrow{\mathbb{P}^*} 1$*.*

The proof is stated in Section D of the supplement (Hundrieser, Eltzner and Huckemann, 2024). It relies on the consistency of the bootstrap for the circular Fréchet sample mean (Theorem 4.4) in conjunction with moment convergence of bootstrap sample means (Proposition 4.7).

In computational terms, the bootstrap-based test presents the challenge of computing B times a test statistic similar to that of the quantile-based test. Therefore, the computational complexity of the bootstrap-based test scales by an order of magnitude B compared to the quantile-based test. Hence, unless the data is affected by FSS, the quantile based test should be considered. Based on the findings on the nominal level of the two tests for the equality of Fréchet population means, we provide two guidelines (Algorithms 1 and 2) for practitioners to decide which test should be considered. The two guidelines and the asymptotic quantile and bootstrap based test are implemented in our package "`FSS`" on `https://github.com/hundrieser/FSS`.

Algorithm 1 Guideline for one-sample test

```
Given X_1,...,X_n ∈ S^1, μ_0 ∈ S^1, α ∈ (0,1), select B ≥ 1000
if X_1,...,X_n is contained in an open half circle then
    Perform quantile based test (Test 5.1(i)) for X_1,...,X_n and μ_0 with α
else
    Perform Test 2.5 for the presence of FSS on X_1,...,X_n with level α
    if Test 2.5 does not reject then
        Perform quantile based test (Test 5.1(i)) for X_1,...,X_n and μ_0 with level α
    else
        Perform bootstrap based test (Test 5.4(i)) for X_1,...,X_n and μ_0 with level α
    end if
end if
```

Algorithm 2 Guideline for two-sample test

```
Given X_1,...,X_n ∈ S^1, Y_1,...,Y_m ∈ S^1, α ∈ (0,1), select B ≥ 1000
if X_1,...,X_n and Y_1,...,Y_m are each contained in an open half circle then
    Perform quantile based test (Test 5.1(ii)) for X_1,...,X_n and Y_1,...,Y_m with α
else
    Perform Test 2.5 for the presence of FSS on X_1,...,X_n and on Y_1,...,Y_m with level α
    if Test 2.5 does not reject for either sample then
        Perform quantile based test (Test 5.1(ii)) for X_1,...,X_n and Y_1,...,Y_m with level α
    else
        Perform bootstrap based test (Test 5.4(ii)) for X_1,...,X_n and Y_1,...,Y_m with level α
    end if
end if
```

Remark 5.6. Simulations in Section 6 depicted in Figures 6 and 7 show that Tests 5.4 keep the nominal level $\alpha = 0.05$ fairly well, in particular for Type I FSS. Upon (very) close inspection, for Type II FSS, the Tests 5.4 may be slightly too conservative, for Type I FSS too liberal. Investigating this effect and correcting for it is left for future research and beyond the scope of this work.

Remark 5.7 (Testing between toroidal data). Similarly, under uniqueness of means, the bootstrap based test is also asymptotically consistent on tori if the

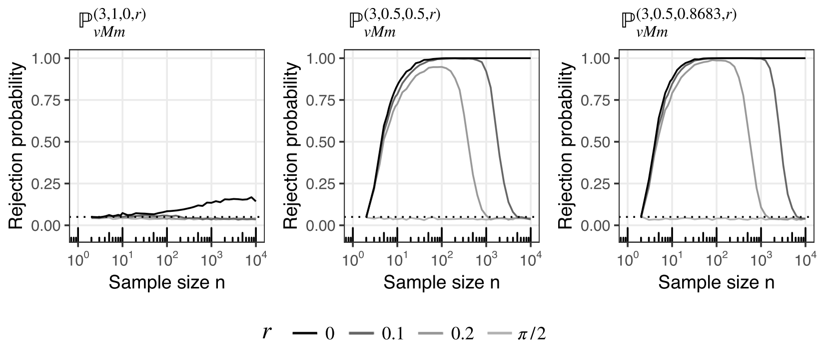

FIG 5. *Empirical rejection probability curves for Test 2.5 with nominal level $\alpha = 0.05$ (dotted horizontal) for the vMm distribution from Figure 1 based on 5000 simulation runs for each sample size and $B = 1000$. For the vMm distribution with a disk cut out of size $r = \pi/2$ about the antipode of its mean, which feature no FSS, all the curves (light gray) keep the level. For Type II FSS (gray: $r = 0.2$, dark gray: $r = 0.1$) rejection probabilities are almost one within the regime of a large variance modulation; beyond that regime they drop back to the nominal level (see Figure 1). Also for the vMm distribution consisting of only one vM distribution (left panel, black) with almost no FSS visible, rejection probabilities increase visibly with sample size. For Type I FSS (middle, black) and smeariness (right, black), the rejection probabilities remain one also for higher sample sizes.*

covariance of X (one-sample) or the sum of covariances of X and Y (two-samples), respectively, is non-singular and each marginal distribution has a density near the antipode that fulfills the assumptions of Proposition 5.2. This follows from the bootstrap consistency of the Fréchet sample mean on tori (Theorem 4.4 and Remark 4.5) and conditional convergence in outer probability of the covariance estimators $\Sigma_B^{(X),*}$ and A_B for $\phi(\widehat{\mu}_n^{(X)})$ and $\phi(\widehat{\mu}_n^{(X)}) - \phi(\widehat{\mu}_m^{(Y)})$, respectively, as $n, B \to \infty$ to the corresponding population covariance quantities.

6. Simulations

In this section we conduct various of simulations to assess the nominal level of the different test procedures as well as their power. We start with the test for the presence of finite sample smeariness (Test 2.5) and continue with a comparison of the quantile (Test 5.1) with the bootstrap test (Test 5.4).

6.1. *Simulating the test for the presence of FSS*

In Figure 5 we display the rejection probability of Test 2.5 for various distributions $\mathbb{P}_{vMm}^{(3,\beta,\lambda,r)}$ on the circle, for which the assumptions of Dubey and Müller (2019) are fulfilled. For $r = 0$ the distributions are defined as in (3.1) and feature Type I FSS, while for $r > 0$ they are defined as in (3.3) and exhibit Type II FSS. Indeed, for $r = \pi/2$, the null hypothesis of absence of FSS is true and Test 2.5 keeps the level $\alpha = 0.05$. For Type II FSS, rejection probabilities

TABLE 3
Parameters of $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda,r}$ with modulation $\mathfrak{m}_n$ of distributions considered for simulations in Figures 6 and 7 with Tests 5.1 and 5.4 and sample sizes $n \in \{30, 100, 300\}$.

κ	3		3		3	
β	1		1/2		1/2	
λ	0		0		1/2	
r	0	0.1	0	0.1	0	0.1
$\mathfrak{m}_{30}$	1.0	1.0	3.7	2.7	7.6	5.5
$\mathfrak{m}_{100}$	1.0	1.0	4.0	2.1	11.8	5.4
$\mathfrak{m}_{300}$	1.0	1.0	4.1	1.4	15.8	3.2

are almost one within the regime where the variance modulation $\mathfrak{m}_n$ is well above one (compare with Figure 1). In the large sample regime the variance modulation is close to one again, and the test rejects with probability close to the nominal level. For the von Mises mixture (vMm) distribution consisting of only one von Mises distribution ($\beta = 1$, left panel, black) with almost no FSS visible in Figure 1, rejection probabilities increase visibly with sample size. As expected, the power of Test 2.5 increases with proximity to a smeary distribution ($\beta = 0.5, \lambda = 0.8683, r = 0$, right panel, black), and specifically in case of Type II FSS also with smaller hole size and proximity to a distribution featuring Type I FSS.

6.2. Simulating the tests for equality of Fréchet means

To assess the performance of the quantile-based and bootstrap-based tests under the presence of FSS (i.e., in case of a large variance modulation $\mathfrak{m}_n$) we consider i.i.d. samples of some nonaltered (Type I FSS) von Mises mixtures, introduced in (3.1) and denoted by $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda}$, and altered (Type II FSS) von Mises mixtures, introduced in (3.3) and denoted by $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda,r}$. By Theorem 3.1 all the $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda,r}$ with unique mean $\mu = 0$ and $r < \pi/2 - \varepsilon$ for some $\varepsilon > 0$ are FSS if they are not smeary themselves. For the following simulations we considered parameters as described in Table 3. All of them give unique means at $\mu = 0$, none of which is smeary. Only the nearby $\mathbb{P}_{vMm}^{3,0.5,\lambda,0}$ is smeary for $\lambda \approx 0.8683$, (following the notation of Eltzner and Huckemann (2019), the order of smeariness is equal to 2, according to Theorem 3 (*ii*) in Hotz and Huckemann (2015)).

In Figures 6 and 7 we compare the power functions at nominal level $\alpha = 0.05$ of the quantile tests (gray lines, Test 5.1) to the power functions of the bootstrap tests (black lines, Test 5.4). For the one-sample tests (Figure 6) we consider simulations of $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda,r}$ and rotated null hypotheses $\mu_0 \in [-\pi, \pi)$, for the two sample-tests (Figure 7) we consider two simulations of $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda,r}$ that are rotated with respect to one another by the angle $p \in [-\pi, \pi)$. Random variables of von Mises distributions were drawn using the R-package *circular* (Lund et al., 2017).

With increasing variance modulation (see Table 3) we see that the quantile tests (gray) become more and more liberal while the bootstrap tests (black) maintain the correct level. In particular, the quantile based tests perform poorly in the presence of considerable Type I FSS, in consequence of a large variance

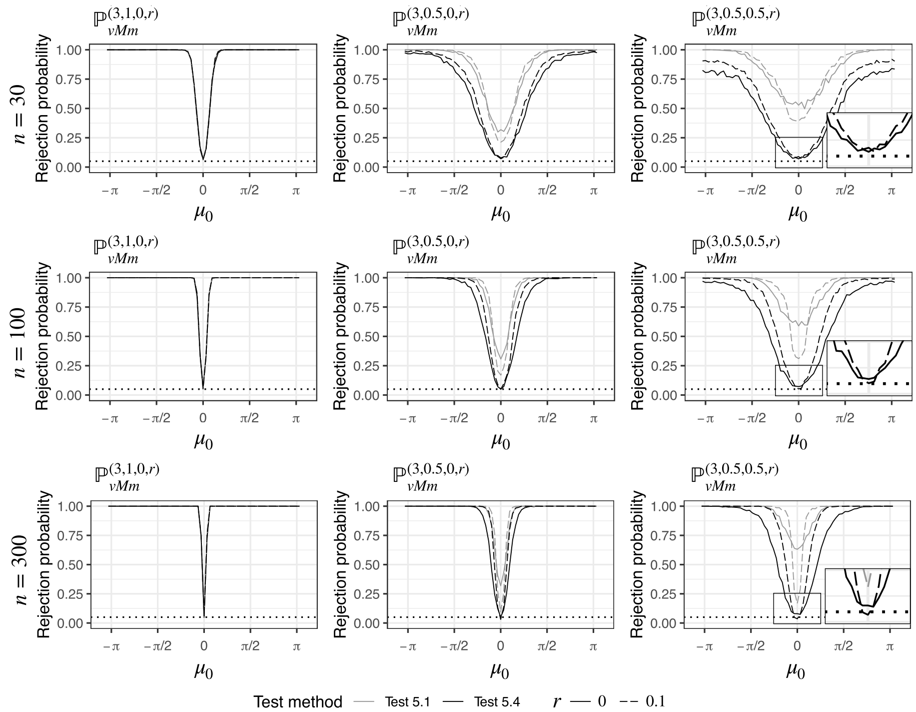

Fig 6. *Rejection probabilities for $H_0^1\colon \mu^{(X)} = \mu_0$ for $\mu^{(X)} = 0$ and varying $\mu_0 \in [-\pi, \pi)$ under the quantile based tests (Test 5.1, gray) and the bootstrap based tests (Test 5.4, black) for $B = 1000$ at nominal level 5% (dotted horizontal). For each μ_0 in total 1000 simulation runs with one sample of size $n = 30$ (top row), $n = 100$ (middle row), and $n = 300$ (bottom row), were performed. Solid lines represent samples from mixed von Mises distributions $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda,r}$, i.e., $r = 0$ (Type I FSS). Dashed lines represent samples from $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda,r}$ where all elements closer to $-\pi$ than $r = 0.1$ were mirrored (Type II FSS). Table 3 reports parameters and variance modulations.*

Table 4

Parameters of $\mathbb{P}_{BvMm}^{\kappa,\gamma,r}$ with variance modulation $\mathfrak{m}_n$ of distributions considered for simulations in Figure 8 on performance of Tests 5.1 and 5.4 and sample sizes $n \in \{30, 100, 300\}$.

κ	3		3		3	
γ	1		1.25		1.5	
r	0	0.1	0	0.1	0	0.1
$\mathfrak{m}_{30}$	1.1	1.1	1.7	1.6	26.1	22.7
$\mathfrak{m}_{100}$	1.1	1.0	1.2	1.1	47.6	32.0
$\mathfrak{m}_{300}$	1.1	1.0	1.2	1.0	32.3	11.7

modulation. Upon very close inspection, under Type II FSS (dashed lines, $r > 0$) we see that the bootstrap tests may be slightly too conservative. Conversely, in case of Type I FSS (solid lines, $r = 0$) the bootstrap tests may be slightly too liberal. Both effects may be due to a systematic bias, cf. Remark 5.6.

To further analyze the performance of the quantile and bootstrap based tests

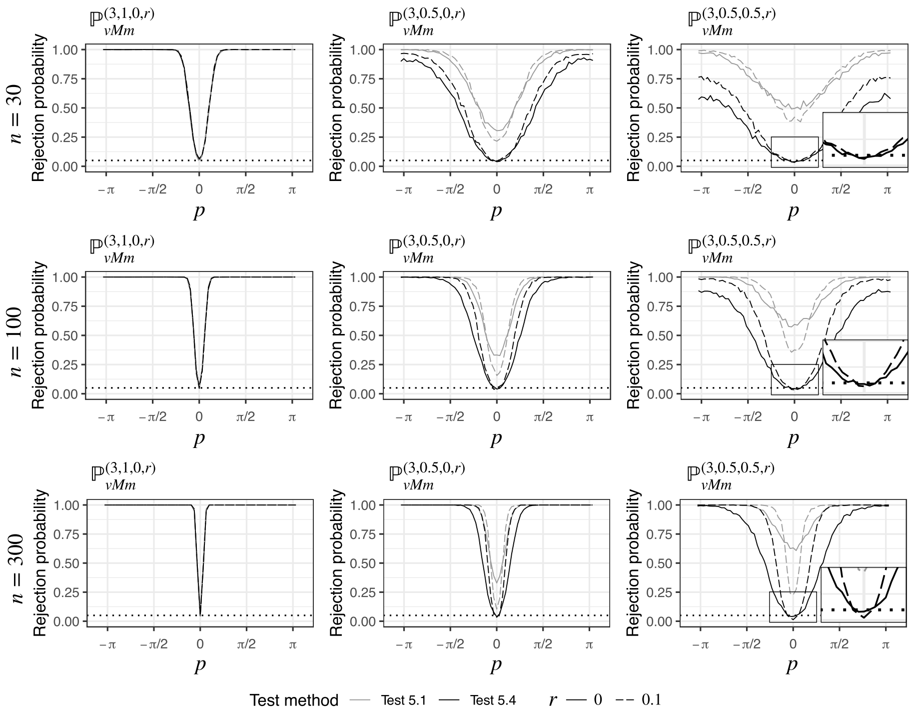

FIG 7. *Rejection probabilities for $H_0^2\colon \mu^{(X)} = \mu^{(Y)}$ with $\mu^{(X)} = 0$ and $\mu^{(Y)} = p \in [-\pi, \pi)$ under quantile based tests (Test 5.1, gray) and bootstrap based tests (Test 5.4, black) for $B = 1000$ at nominal level 5% (dotted horizontal). For each p in total 1000 simulation runs with two samples from the same distribution of size $n = 30$ (top row), $n = 100$ (middle row), and $n = 300$ (bottom row), but where the latter sample is rotated by p, were performed. Solid lines represent samples from mixed von Mises distributions $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda,0}$, (Type I FSS). Dashed lines represent samples from $\mathbb{P}_{vMm}^{\kappa,\beta,\lambda,r}$ where all elements closer to $-\pi$ than $r = 0.1$ were mirrored (Type II FSS).*

in the regime close to nonunique Fréchet means we conducted another set of simulations with samples taken from a suitable mixture of von Mises distributions with *equal* concentration parameter and nearly antipodal modes. More precisely, we took bimodal von Mises mixtures with symmetric modes around the origin parametrized by $\kappa \geq 0, \gamma \in [0, \pi/2]$ and with density under the arc length measure

$$d\mathbb{P}_{BvMm}^{\kappa,\gamma}(x) := \frac{1}{2}\,\frac{\exp\left(\kappa\cos(x-\gamma)\right)}{2\pi I_0(\kappa)}\,dx + \frac{1}{2}\,\frac{\exp\left(\kappa\cos(x+\gamma)\right)}{2\pi I_0(\kappa)}\,dx,$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of order 0, e.g., Mardia and Jupp (2000, p. 36). By symmetry, for $\kappa > 0$ and $\gamma = \pi/2$ the set of Fréchet means is nonunique and given by $\{0, -\pi\}$, whereas for $\gamma < \pi/2 \approx 1.57$ the mean is unique and located at the origin. Hence, with increasing γ the respective law approaches the nonuniqueness regime. Further, for $r \geq 0$ we

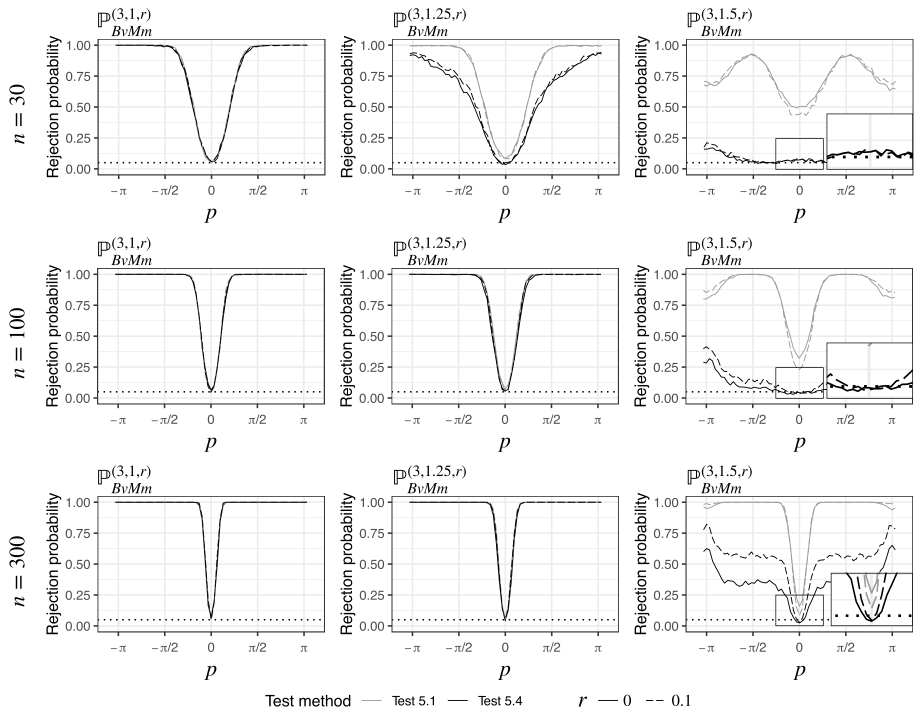

FIG 8. *Rejection probabilities for* $H_0^2\colon \mu^{(X)} = \mu^{(Y)}$ *with* $\mu^{(X)} = 0$ *and* $\mu^{(Y)} = p \in [-\pi, \pi)$ *under the quantile based tests (Test 5.1, gray) and the bootstrap based tests (Test 5.4, black) for* $B = 1000$ *at nominal level 5% (dotted horizontal). For each p in total 1000 simulation runs with two samples from the same distribution of size* $n = 30$ *(top row),* $n = 100$ *(middle row), and* $n = 300$ *(bottom row), but where the latter sample is rotated by p, were performed. Solid lines represent samples from bimodal von Mises mixtures* $\mathbb{P}^{\kappa,\gamma}_{BvMm}$*, i.e.,* $r = 0$ *(Type I FSS). Dashed lines represent samples from* $\mathbb{P}^{\kappa,\gamma,r}_{BvMm}$ *where all elements closer to* $-\pi$ *than* $r = 0.1$ *were mirrored (Type II FSS). Table 4 reports parameters and variance modulations.*

recall the function ζ^r, introduced in (3.2), which mirrors the disk of radius r at $-\pi$ to the origin and define

$$\mathbb{P}^{\kappa,\gamma,r}_{BvMm} \coloneqq \zeta^r_* \, \mathbb{P}^{\kappa,\gamma}_{BvMm}.$$

For $r > 0$ and $\kappa \geq 0, \gamma \in [0, \pi/2]$ the Fréchet mean of $\mathbb{P}^{\kappa,\gamma,r}_{BvMm}$ is unique and coincides with the origin.

For our subsequent simulations we selected parameters described in Table 4 and computed the variance modulation for $n \in \{30, 100, 300\}$. Notably, if $\kappa > 0$ and $\gamma \in [0, \pi/2)$ the distribution $\mathbb{P}^{\kappa,\gamma,r}_{BvMm}$ is Type I FSS for $r = 0$, whereas for $r > 0$ it is Type II FSS.

If γ is close to $\pi/2$, i.e., if the underlying distribution is near a distribution with nonunique means, it is expected that $\mathfrak{m}_n$ attains fairly large values, indicating a high deviation of the quantile based test from its nominal level. Indeed, our simulations demonstrate that a large choice of γ results in the quantile test (gray) attaining a nominal level dramatically exceeding the correct level, whereas the bootstrap test (black) attains the correct level. Notably, for small sample sizes the bootstrap based test appears to be for $\gamma = 1.5$ slightly too liberal under the null which might be explained by the fact that the empirical sample means are likely to be located near the modes of the underlying distribution. Remarkably, for γ close to $\pi/2$ and $r = 0$, the distributions $\mathbb{P}^{\kappa,\gamma,r}_{BvMm}$ and shifted by p close to $\pm\pi$ are fairly close in TV-distance, but their Fréchet means are far apart. This explains the high power of the bootstrap test for p near $\pm\pi$.

7. Assessing significant change of wind direction

In application of our methods dealing with FSS, we analyze wind data from Basel and Göttingen (the city of the authors' institution) provided by meteoblue AG (2020). For our purpose, we consider daily Fréchet mean wind directions for the years 2000 to 2019 giving for each city 20 samples of two-dimensional circular data of size $n = 365$. The respective daily wind directions are illustrated in Figure 9. To assess a possible effect of climate change, we test for a significant change in wind direction.

For each of these 40 samples we computed the estimated variance modulation $\widehat{\mathfrak{m}}^*_n$ from (2.3) with $B = 1000$, cf. Table 5. Remarkably, for both cities nearly all years statistically indicate the presence of FSS (Test 2.5 rejects the presence of FSS at level $\alpha = 0.001$) except for Göttingen in the year 2017. To assess the stability of the test we also performed some sub-sampling procedures which are provided in the supplement Hundrieser, Eltzner and Huckemann (2024). Moreover, to assess the property of uniqueness of population means we additionally generated $R = 10\,000$ bootstrap sample means based on resampling $n = 365$ data points for each year and city and plotted the resulting histograms (see Figure 1 in the supplement, Hundrieser, Eltzner and Huckemann (2024)). For each year, with the exception of Basel 2004 and Göttingen 2007, 2010, 2013, we see that the distributions of bootstrap sample means admit a single cluster. This suggests that most of the wind data sets are not too close nearby distributions with nonunique means.

In total, we performed six series of tests for changes of wind directions in consecutive years were performed: for wind data from Basel and Göttingen (viewed as data on the two-torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$), as well as Basel and Göttingen individually (viewed as data on the circle $\mathbb{S}^1$) using the quantile test (Test 5.1) and the bootstrap test (Test 5.4). In consequence of detected FSS, we expect that the quantile based two-sample test (Test 5.1) will feature a considerably high error of the first kind as compared to the bootstrap based Test 5.4. For each series of 19 tests to compare daily wind directions of consecutive years from 2000 to 2019 at nominal level $\alpha = 0.05$ we performed a Benjamini-Hochberg correction, reported in Figure 10.

The first series of tests has been performed on $\mathbb{T}^2$ for Basel and Göttingen jointly (top rows of Figure 10). According to the quantile Test 5.1, a majority of years seem to be significantly different from the year following. Applying the

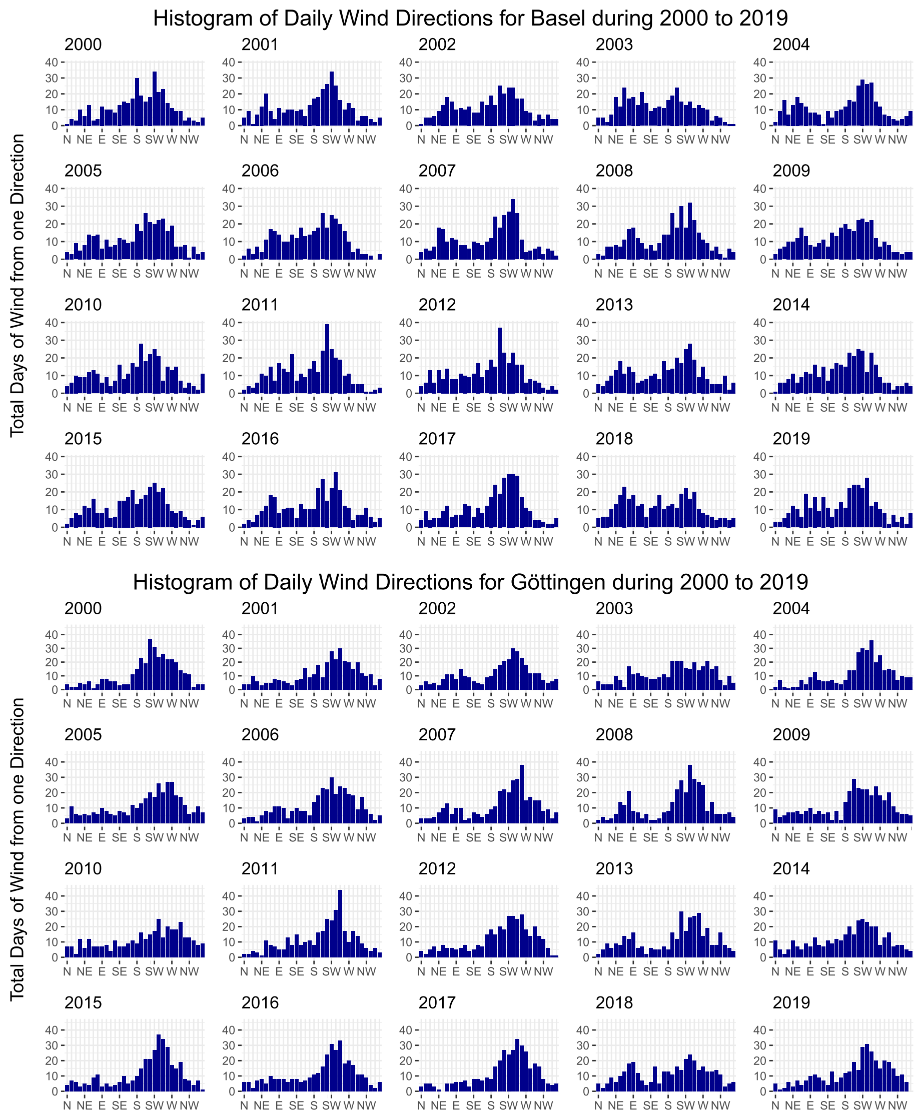

Fig 9. *Histograms of Fréchet means of daily wind direction data by meteoblue AG (2020) for Basel (top) and Göttingen (bottom) for the years* 2000 *to* 2019*. The x-axis is divided into 32 segments with labels N : "North", E : "East", S : "South", and W : "West", indicating the average direction of wind origin.*

bootstrap Test 5.4 for $B = 1000$ we see this "noise" disappearing, leaving only the changes in wind directions between 2005 and 2006, 2017 and 2018, as well as 2018 and 2019 as exceptional.

Looking only at Basel (now testing only on $\mathbb{S}^1$, middle rows of Figure 10), Test 5.1 seems to suggest that the years 2002 to 2004 and 2016 to 2019 are

TABLE 5

Empirical variance modulation $\widehat{\mathfrak{m}}_n^*$ for $n = 365$ from (2.3) for daily Fréchet mean wind directions in Basel and Göttingen for the years 2000 to 2019. All of these values indicate presence of FSS (Test 2.5 rejects absence of FSS with p-values upper bounded by 10^{-3}) except for Göttingen in the year 2017.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Basel	1.601	4.067	1.542	1.197	32.695	2.683	1.845	2.096	1.676	2.344
Göttingen	3.317	2.734	2.760	3.065	1.492	4.171	1.605	15.493	3.861	8.796
Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Basel	5.010	1.775	1.466	3.211	1.491	1.741	1.705	5.229	2.694	2.481
Göttingen	8.003	1.685	4.457	36.365	3.446	3.305	5.636	1.037	1.804	2.592

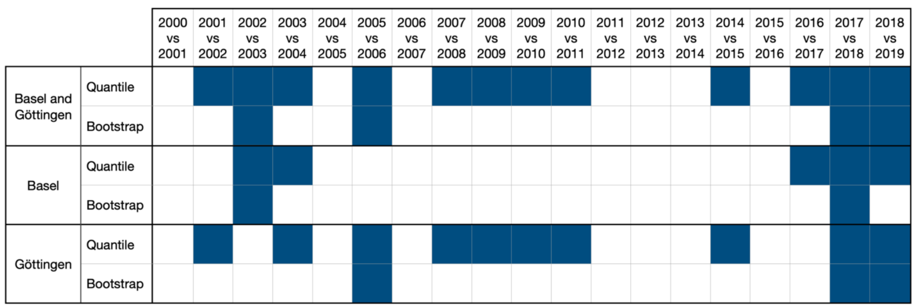

FIG 10. *Benjamini-Hochberg corrected test results (first, third, fifth row: Test 5.1; second, fourth, sixth row: Test 5.4 with $B = 1000$) of comparisons of daily Fréchet means of wind direction data for consecutive years 2000 to 2019 for Basel and Göttingen on $\mathbb{T}^2$ (top), Basel on $\mathbb{S}^1$ (middle), and Göttingen on $\mathbb{S}^1$ (bottom) for a significance level of 5%. Consecutive years with significant changes according to the test procedure are depicted in blue.*

exceptional. Test 5.4, however, clarifies: only the changes between 2002 and 2003 as well as between 2017 and 2018 are exceptional.

Comparison with Göttingen only (again testing on $\mathbb{S}^1$, bottom rows of Figure 10) shows that the "noise" from Test 5.1 is rooted in Göttingen. Again, Test 5.4 clarifies the picture: only the changes between 2005 and 2006, and 2017 to 2019 are exceptional. For both cities, the change from the year 2017 to 2018 seems most prominently exceptional. Notably, 2017 is also exceptional through absence of FSS, cf. Table 5.

These findings fall well into the climatological context of central European heat waves linked to exceptional wind constellations (Kornhuber et al. (2019) identify a *recurrent wave-7* wind pattern for the years 2003, 2006, 2015 and 2018). While the 2003 heat wave occurred in most of Europe, the 2018 heat wave manifested in a climatic dipole: hot and dry north of the Alps, comparably cool and moist across large parts of Mediterranean, cf. Buras, Rammig and Zang (2020).

In a debate quantifying climate change, its anthropogenic component and future costs, linking to changes of wind patterns (e.g. McInnes, Erwin and Bathols, 2011), our new inferential tools for cyclic data presented here, warrant a more detailed application in future work.

Table 6

Empirical rejection probabilities of quantile Test 5.1 and bootstrap Test 5.4 for testing for equality of Fréchet means for significance level 5% based on split samples for daily Fréchet mean wind directions of Basel and Göttingen for the years 2000 to 2019. All values computed based on $R = 500$ repetitions.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Quantile Test	0.312	0.404	0.234	0.160	0.436	0.410	0.162	0.568	0.170	0.576
Bootstrap Test	0.042	0.028	0.024	0.016	0.030	0.044	0.040	0.024	0.006	0.030
Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Quantile Test	0.478	0.160	0.330	0.720	0.230	0.248	0.404	0.384	0.250	0.296
Bootstrap Test	0.032	0.040	0.034	0.026	0.016	0.022	0.036	0.048	0.020	0.042

To further emphasize the strong bias of the quantile test (Test 5.1) towards rejection and the consistency of the bootstrap test (Test 5.4) with $B = 1000$, we conducted additional analysis, splitting samples. For each year, we split the wind data sets for both cities, into two subsets of size 182 and 183 and tested for equality of Fréchet means at significance level 5% with both Tests 5.1 and 5.4. This procedure was repeated $R = 500$ times and the resulting empirical rejection probabilities are depicted in Table 6. In every year, the quantile based test rejects much more often compared to the bootstrap based test. The years for which the quantile test rejects least often (2003, 2006, 2008, and 2011) are also the years for which the empirical variance modulation of both cities is relatively small. These results suggest that the bootstrap test indeed keeps the level fairly well, even for wind data sets for Basel in 2004 and Göttingen in 2007, 2010, and 2013 which are affected by a distribution with nonunique means nearby.

8. Discussion and outlook

In this contribution, we have investigated two manifolds with codimension one cut loci, namely circles and tori and found FSS manifesting in two different types. For other spaces, an investigation of FSS is beyond the scope of this paper. We expect similar findings for other manifolds with codimension one cut loci, such as real projective spaces, modeling projective shapes, say, as in Mardia and Patrangenaru (2005); Hotz, Kelma and Kent (2016). We believe that this is, using the language of Eltzner (2022), a consequence of *topological smeariness*. On manifolds with higher codimension cut loci, in the language of Eltzner (2022), for instance on arbitrary spheres, there is the different phenomenon of *geometrical smeariness*. In the light of this, we conjecture that Type I FSS is present in all nondegenerate random variables on positively curved spaces. For such spaces, Afsari (2009) has shown that the Hessian of the Fréchet function is smaller than its Euclidean equivalent, which is twice the identity matrix. Moreover, for very small sample sizes (n approx. less than 10) of highly concentrated random variables, Pennec (2019) explicitly derived $\mathfrak{m}_n > 1$ but well below what is predicted asymptotically by the CLT. Indeed, on the sphere Eltzner, Hundrieser and Huckemann (2021) showed that every non-trivial random variable is FSS of Type I, see also section E of the supplement (Hundrieser, Eltzner and Huckemann, 2024). For more general manifolds it is also known that "smeariness begets FSS" as Tran, Eltzner and Huckemann (2021) constructed from every smeary random variable a random variable featuring FSS of Type I.

Moreover, we have shown that the proposed bootstrap tests are asymptotically consistent and demonstrated in simulations that they preserve the level fairly well for reasonable sample sizes. Inspired by recent findings of Zhilova (2020) on the non-asymptotic accuracy of the bootstrap for Euclidean sample means for the finite sample regime, we deem it possible to derive similar non-asymptotic results for the bootstrap Fréchet sample mean. This topic is beyond the scope of this work.

Notably, equipping a manifold M embedded in a Euclidean space with the *extrinsic* metric $d^E(p,q) := \|p-q\|$ for $p,q \in M$ the resulting Fréchet means have been called *extrinsic means* by Bhattacharya and Patrangenaru (2003) and *mean locations* by Hendriks and Landsman (1998); they encompass *mean directions* on circles and spheres, see Mardia and Jupp (2000). Thus, for an i.i.d. sample $X_1, \dots, X_n$ on M with extrinsic population and sample Fréchet mean μ^E and $\hat{\mu}_n^E$, respectively, we have the *extrinsic variance modulation*

$$\mathfrak{m}_n^E := \frac{n\mathbb{E}[\left\|\mu^E - \hat{\mu}_n^E\right\|^2]}{\mathbb{E}[\|\mu^E - X\|^2]}.$$

Although extrinsic means typically exhibit standard $n^{-1/2}$ convergence rates, as shown by Bhattacharya and Patrangenaru (2005), Schötz (2019a) demonstrated that also arbitrary slow (smeary) and fast (anti-smeary) convergence rates may occur. In particular, this hints towards existence of FSS also for extrinsic means.

Acknowledgments

S. Hundrieser acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2067/1- 390729940 and the RTG 2088. B. Eltzner and S. Huckemann gratefully acknowledge support by the DFG HU 1575/7, the DFG CRC 1456, the Niedersachsen Vorab of the Volkswagen Foundation and the Felix-Bernstein Institute for Mathematical Stochastics in the Biosciences. All authors also thank meteoblue AG for providing the wind data sets and Prof. Nishiyama for providing details to his work on moment convergence of M-estimators. The authors also acknowledge helpful comment and suggestions by two anonymous referees.

Supplementary Material

Auxiliary Results, Omitted Proofs for Sections 3–5, and Simulations (doi: 10.1214/24-EJS2276SUPP; .pdf). The supplementary material includes auxiliary results for the proofs for the characterization of FSS on the circle (Theorem 3.1), the bootstrap consistency of circular Fréchet sample means (Theorem 4.4) and their moment convergence (Propositions 4.6 and 4.7), as well as proofs on the performance of the quantile and bootstrap based tests (Propositions 5.2 and 5.5). We also present simulation results for FSS on $\mathbb{S}^2$ and detail additional analysis on the wind data.

References

Afsari, B. (2009). *Means and Averaging on Riemannian Manifolds*. University of Maryland. MR2718072

Afsari, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society* **139** 655–773. MR2736346

Ahidar-Coutrix, A., Le Gouic, T. and Paris, Q. (2019). Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendable geodesics. *Probability Theory and Related Fields*. https://doi.org/10.1007/s00440-019-00950-0. MR4095017

Arnaudon, M. and Miclo, L. (2014). Means in complete manifolds: uniqueness and approximation. *ESAIM: Probability and Statistics* **18** 185–206. MR3230874

Barden, D., Le, H. and Owen, M. (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electronic Journal of Probability* **18** 1–25. MR3035753

Barden, D., Le, H. and Owen, M. (2018). Limiting behaviour of Fréchet means in the space of phylogenetic trees. *Annals of the Institute of Statistical Mathematics* **70** 99–129. MR3742820

Bhattacharya, A. and Bhattacharya, R. (2012). *Nonparametric Inference on Manifolds*. Cambridge University Press. MR2934285

Bhattacharya, R. and Lin, L. (2017). Omnibus CLT for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proceedings of the American Mathematical Society* **145** 413–428. MR3565392

Bhattacharya, R. N. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds I. *The Annals of Statistics* **31** 1–29. MR1962498

Bhattacharya, R. N. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds II. *The Annals of Statistics* **33** 1225–1259. MR2195634

Billera, L. J., Holmes, S. P. and Vogtmann, K. (2001). Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics* **27** 733–767. MR1867931

Buras, A., Rammig, A. and Zang, C. S. (2020). Quantifying impacts of the drought 2018 on European ecosystems in comparison to 2003. *Biogeosciences* **17** 1655–1672. https://doi.org/10.5194/bg-17-1655-2020

Dubey, P. and Müller, H.-G. (2019). Fréchet analysis of variance for random objects. *Biometrika* **106** 803–821. MR4031200

Eltzner, B. (2020). Testing for uniqueness of estimators. *arXiv e-prints arXiv:2011.14762*.

Eltzner, B. (2022). Geometrical smeariness – A new phenomenon of Fréchet means. *Bernoulli* **28** 239–254. https://doi.org/10.3150/21-BEJ1340. MR4337704

Eltzner, B. and Huckemann, S. (2017). Bootstrapping descriptors for non-Euclidean data. In *Geometric Science of Information* (F. Nielsen and F. Barbaresco, eds.) 12–19. Springer International Publishing. MR3737707

Eltzner, B. and Huckemann, S. F. (2019). A smeary central limit theorem for manifolds with application to high dimensional spheres. *The Annals of Statistics* **47** 3360–3381. MR4025745

Eltzner, B., Hundrieser, S. and Huckemann, S. (2021). Finite sample smeariness on spheres. In *Geometric Science of Information* (F. Nielsen and F. Barbaresco, eds.) 12–19. Springer International Publishing, Cham. MR4424304

Eltzner, B., Galaz-Garcia, F., Huckemann, S. F. and Tuschmann, W. (2021). Stability of the cut locus and a central limit theorem for Fréchet means of Riemannian manifolds. *Proceedings of the American Mathematical Society* **149** 3947–3963. https://doi.org/10.1090/proc/15429. MR4291592

Fang, Z. and Santos, A. (2019). Inference on directionally differentiable functions. *The Review of Economic Studies* **86** 377–412. MR3936869

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut de Henri Poincaré* **10** 215–310. MR0027464

Hansen, P., Eltzner, B., Huckemann, S. F. and Sommer, S. (2021). Diffusion means in geometric spaces. *arXiv e-prints arXiv:2105.12061*. MR4632134

Hendriks, H. and Landsman, Z. (1998). Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. *Journal of Multivariate Analysis* **67** 227–243. MR1659156

Hotz, T. and Huckemann, S. (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics* **67** 177–193. MR3297863

Hotz, T., Kelma, F. and Kent, J. T. (2016). Manifolds of projective shapes. *arXiv e-prints arXiv:1602.04330*.

Hotz, T., Huckemann, S., Le, H., Marron, J. S., Mattingly, J., Miller, E., Nolen, J., Owen, M., Patrangenaru, V. and Skwerer, S. (2013). Sticky central limit theorems on open books. *Annals of Applied Probability* **23** 2238–2258. MR3127934

Huckemann, S. F. and Eltzner, B. (2020). Statistical methods generalizing principal component analysis to non-Euclidean spaces. In *Handbook of Variational Methods for Nonlinear Geometric Data* 317–388. Springer. MR4267074

Huckemann, S., Mattingly, J., Miller, E. and Nolen, J. (2015). Sticky central limit theorems at isolated hyperbolic planar singularities. *Electronic Journal of Probability* **20** 1–34. MR3371437

Hundrieser, S., Eltzner, B. and Huckemann, S. F. (2023). A lower bound for estimating Fréchet means. *arXiv e-prints arXiv:2402.12290*.

Hundrieser, S., Eltzner, B. and Huckemann, S. (2024). Supplementary material for "Finite sample smeariness of Fréchet means with application to

climate".
Kato, K. (2011). A note on moment convergence of bootstrap M-estimators. *Statistics & Decisions* **28** 51–61. MR2776694

Kornhuber, K., Osprey, S., Coumou, D., Petri, S., Petoukhov, V., Rahmstorf, S. and Gray, L. (2019). Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. *Environmental Research Letters* **14** 054002.

Lammers, L., Van, D. T., and Huckemann, S. F. (2023). Sticky flavors. *arXiv preprint arXiv:2311.08846*.

Le Gouic, T., Paris, Q., Rigollet, P. and Stromme, A. J. (2022). Fast convergence of empirical barycenters in Alexandrov spaces and the Wasserstein space. *Journal of the European Mathematical Society* **25** 2229–2250. MR4592868

Lund, U., Agostinelli, C. et al. (2017). Circular. `https://cran.r-project.org/web/packages/circular/index.html`. R package version 0.4-93.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Wiley, New York. MR1828667

Mardia, K. V. and Patrangenaru, V. (2005). Directions and projective shapes. *The Annals of Statistics* **33** 1666–1699. MR2166559

Mattingly, J. C., Miller, E. and Tran, D. (2023). Central limit theorems for Féchet means on stratified spaces. *arXiv preprint arXiv:2311.09455*.

McInnes, K. L., Erwin, T. A. and Bathols, J. M. (2011). Global climate model projected changes in 10 m wind speed and direction due to anthropogenic climate change. *Atmospheric Science Letters* **12** 325–333.

McKilliam, R. G., Quinn, B. G. and Clarkson, I. V. L. (2012). Direction estimation by minimum squared arc length. *IEEE Transactions on Signal Processing* **60** 2115–2124. MR2954196

meteoblue AG (2020). history+ platform. www.meteoblue.com/en/weather/archive/export. Accessed on 06/04/2020.

Nishiyama, Y. (2010). Moment convergence of M-estimators. *Statistica Neerlandica* **64** 505–507. MR2757870

Pennec, X. (2019). Curvature effects on the empirical mean in Riemannian and affine Manifolds: A non-asymptotic high concentration expansion in the small-sample regime. *arXiv e-prints arXiv:1906.07418*.

Pennec, X. (2020). Personal communication.

Schötz, C. (2019a). Arbitrary rates of convergence for projected and extrinsic means. *arXiv e-prints arXiv:1910.11223*.

Schötz, C. (2019b). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electronic Journal of Statistics* **13** 4280–4345. `https://doi.org/10.1214/19-EJS1618`. MR4023955

Sturm, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. *Contemporary Mathematics* **338** 357–390. `https://doi.org/10.1090/conm/338/06080`. MR2039961

Tran, D., Eltzner, B. and Huckemann, S. F. (2021). Smeariness begets finite sample smeariness. In *Geometric Science of Information* (F. Nielsen and F. Barbaresco, eds.) 29–36. Springer International Publishing, Cham.

MR4424306

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. Springer. MR2724359

VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. Cambridge Univ. Press. MR1652247

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer. MR1385671

ZHILOVA, M. (2020). Nonclassical Berry–Esseen inequalities and accuracy of the bootstrap. *The Annals of Statistics* **48** 1922–1939. https://doi.org/10.1214/18-AOS1802. MR4134780

ZIEZOLD, H. (1977). Expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the 7th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes* **A** 591–602. MR0501230