

A SMEARY CENTRAL LIMIT THEOREM FOR MANIFOLDS WITH APPLICATION TO HIGH-DIMENSIONAL SPHERES¹

BY BENJAMIN ELTZNER AND STEPHAN F. HUCKEMANN

*Felix-Bernstein-Institut für Mathematische Statistik in den Biowissenschaften and
Georg-August-Universität Göttingen*

The (CLT) central limit theorems for generalized Fréchet means (data descriptors assuming values in manifolds, such as intrinsic means, geodesics, etc.) on manifolds from the literature are only valid if a certain empirical process of Hessians of the Fréchet function converges suitably, as in the proof of the prototypical BP-CLT [*Ann. Statist.* **33** (2005) 1225–1259]. This is not valid in many realistic scenarios and we provide for a new very general CLT. In particular, this includes scenarios where, in a suitable chart, the sample mean fluctuates asymptotically at a scale n^α with exponents $\alpha < 1/2$ with a nonnormal distribution. As the BP-CLT yields only fluctuations that are, rescaled with $n^{1/2}$, asymptotically normal, just as the classical CLT for random vectors, these lower rates, somewhat loosely called smeariness, had to date been observed only on the circle. We make the concept of smeariness on manifolds precise, give an example for two-smeariness on spheres of arbitrary dimension, and show that smeariness, although “almost never” occurring, may have serious statistical implications on a continuum of sample scenarios nearby. In fact, this effect increases with dimension, striking in particular in high dimension low sample size scenarios.

1. Introduction. The classical central limit theorem (CLT) for i.i.d. random vectors with second moments states, in particular, that the multiple nV of the variance of the fluctuation of sample means \bar{X}_n around the population mean $\mathbb{E}[X]$ with sample size n , is asymptotically constant. Under specific conditions, the BP-CLT by Bhattacharya and Patrangenaru (2005) for intrinsic means on manifolds extends this result to images in a local chart of Fréchet sample means and Fréchet population means. If data are sufficiently dispersed, however, as in the “turtles” data set from Mardia and Jupp ((2000), page 9), bootstrapping even rather high sample sizes seems to render the BP-CLT not applicable; cf. Figure 1.

The reason is that classical arguments employed by Bhattacharya and Patrangenaru control the underlying empirical process of Hessians of sample Fréchet functions, only under strong constraining conditions, which as it turns out, are often

Received January 2018; revised August 2018.

¹Supported by DFG 755 B8, DFG GRK 2088 and the Niedersachsen Vorab of the Volkswagen Foundation.

MSC2010 subject classifications. Primary 62G20, 62H11; secondary 53C22, 58K45.

Key words and phrases. Fréchet means, asymptotic consistency and normality, asymptotics on manifolds, lower asymptotic rate, high dimension low sample size.

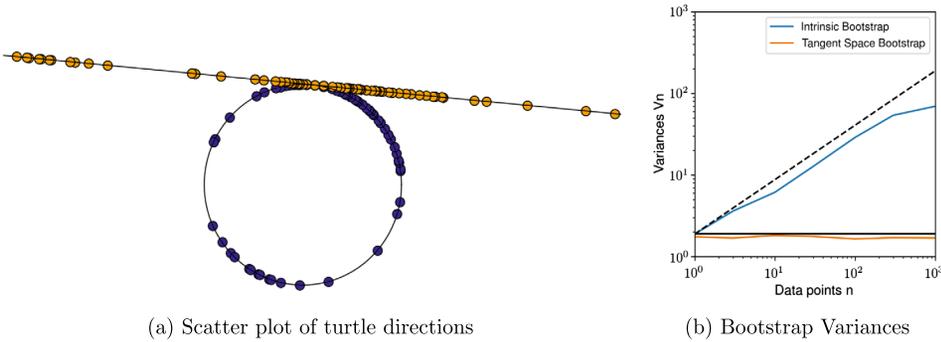


FIG. 1. Left: Directions (blue) of 76 female turtles after laying eggs from *Mardia and Jupp* (2000), page 9, and their images (orange) in the tangent space of their intrinsic mean. Right: variance (blue) of the fluctuation of intrinsic bootstrap sample means times sample size (vertical) over varying bootstrap sample sizes (horizontal), starting off close to 2-smeariness (black dashed). Applying the classical Euclidean CLT to bootstrapped tangent space images gives classical asymptotically constant behavior (orange) corresponding to nonsmeariness (black solid).

not realistic. We propose to consider instead a Taylor expansion of the population Fréchet function, combined with more involved empirical process theory, to develop a new line of argument, which is applicable under very mild conditions. This general approach to asymptotic theory is of mathematical interest in itself, and in retrospect, quite natural. In passing, we also get rid of another constraining condition by allowing singularity of the Hessian which precisely opens the CLT to *smeary* scenarios as in Figure 1.

Precisely verifying a specific type of smeariness on spheres is a challenging task, and by tackling it, we pave the way for the exploration of other types of smeariness on spheres, as well as on other standard data and descriptor manifolds. Notably, our smeary CLT also holds for suitable general M-estimators.

The BP-CLT. The celebrated central limit theorem (CLT) for intrinsic sample means on manifolds by *Bhattacharya and Patrangenaru* (2005), and many subsequent generalizations (e.g., *Bhattacharya and Bhattacharya* (2008), *Bhattacharya and Lin* (2017), *Bhattacharya and Patrangenaru* (2014), *Ellingson, Patrangenaru and Ruymgaart* (2013), *Huckemann* (2011a), *Patrangenaru and Ellingson* (2016)), rests on a *Taylor expansion*

$$(1.1) \quad \sqrt{n} \operatorname{grad} |_{x=x_0} F_n(x) = \sqrt{n} \operatorname{grad} |_{x=0} F_n(x) + \operatorname{Hess} |_{x=\tilde{x}} F_n(x) \sqrt{n} x_0$$

(with suitable \tilde{x} between 0 and x_0) and a generalized *strong law* ($n \rightarrow \infty$ and $x_0 \rightarrow 0$)

$$(1.2) \quad \operatorname{Hess} |_{x=\tilde{x}} F_n(x) \xrightarrow{\mathbb{P}} \operatorname{Hess} |_{x=0} F(x).$$

Here, $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X$ is a sample on a smooth manifold M ,

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n d(X_j, \phi(x))^2, \quad F(x) = \mathbb{E}[d(X, \phi(x))^2]$$

are the *sample* and *population Fréchet functions* and d is a smooth distance on M . Further it is assumed that F has a unique minimizer, called the *population Fréchet mean* and ϕ denotes a local smooth chart which maps 0 to the population Fréchet mean.

Note that the Taylor expansion (1.1) and all others in this paper are formulated in charts and involve only classical multivariate, not covariant, derivatives. We show below in Lemma 3.2 that smeariness in CLTs is invariant under chart changes.

For the preimage $x_0 = x_n$ under ϕ of any *sample Fréchet mean*, that is, a minimizer of the sample Fréchet function F_n , the left-hand side of equation (1.1) vanishes. The right-hand side is guaranteed to be well defined; however, only if the line segment between 0 and x_0 carries no sample points. This can be ensured for fixed sample size n , if x_0 is sufficiently close to 0 because due to [Le and Barden \(2014\)](#) the cut locus of the population Fréchet mean carries no mass. If X has a density near the cut locus and $x_0 = x_n$ is random, this is no longer clear. Even if (1.1) were well defined, it is not clear under which circumstances (1.2) also holds for random $x_0 = x_n$.

If both (1.1) and (1.2) hold, since the properly rescaled sum of i.i.d. random variables $\sqrt{n} \text{grad}|_{x=0} F_n(x)$ converges to a Gaussian, this strain of argument then gives the BP-CLT,

$$(1.3) \quad \sqrt{n}x_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma),$$

with suitable covariance matrix Σ , if the Hessian on the right-hand side of equation (1.2) is invertible.

Beyond the BP-CLT. Recently in [Hotz and Huckemann \(\(2015\), Example 1\)](#), an example on the circle with log coordinates $x \in [-\pi, \pi)$ has been provided, with population Fréchet mean at $x = 0$ and a local density f near the antipodal $-\pi$. For $x > 0$ sufficiently small, the rescaled sample Fréchet function takes the value

$$\begin{aligned} nF_n(x) &= \sum_{x-\pi \leq X_j} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2 \\ &= \sum_{j=1}^n (X_j - x)^2 + 4\pi \sum_{X_j < x-\pi} (X_j - x + \pi) \end{aligned}$$

so that the left-hand side of equation (1.2) is only a.s. well defined with value $\text{Hess}|_x F_n(x) = 2$ a.s. (as in the Euclidean case). The right-hand side, however, assumes the value $\text{Hess}|_{x=0} F(x) = 2 - 4\pi f(-\pi)$. Hence, in case of $f(-\pi) \neq 0$, the convergence (1.2) is no longer valid, making the above strain of argument no

longer viable. Still, as shown in Hotz and Huckemann (2015), McKilliam, Quinn and Clarkson (2012), as long as $2\pi f(-\pi) < 1$, the BP-CLT (1.3) remains valid.

Further, in Hotz and Huckemann (2015) it was shown that $1 = 2\pi f(-\pi)$ is possible, so that the BP-CLT (1.3), which universally holds for Euclidean spaces under square integrability, fails for such 2D vectors confined to a circle, by giving examples in which the fluctuations may asymptotically scale with n^α with exponents α strictly lower than one-half.

This new phenomenon has, somewhat loosely, been called *smeariness*; it can only manifest in a non-Euclidean geometry. Examples beyond the circle were not known to date.

A general CLT. Making the concept of smeariness on manifolds precise, using empirical process theory (e.g., from van der Vaart (1998)) and avoiding the sample Taylor expansion (1.1) as well as the not generally valid convergence condition (1.2), we provide for a general CLT on manifolds that requires no assumptions other than a unique population mean and a sufficiently well behaved distance. With the *degree of smeariness* $\kappa \geq 0$, our general CLT takes the form

$$(1.4) \quad \sqrt{n}x_n|x_n|^\kappa \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma),$$

where $x_n|x_n|^\kappa$ is defined componentwise. Then x_n scales with n^α , $\alpha = \frac{1}{2(\kappa+1)}$, and $\kappa = 0$ corresponds to the usual CLT valid on Euclidean spaces, and to the BP-CLT (1.3).

We phrase our general CLT in terms of sufficiently well behaved generalized Fréchet means, for example, geodesic principal components (Huckemann, Hotz and Munk (2010), Huckemann and Ziezold (2006)) or principal nested spheres (Jung, Dryden and Marron (2012), Jung, Foskey and Marron (2011)). While we discuss some intricacies in Remark 2.8, their details are beyond the scope of this paper and left for future research. In general, generalized Fréchet means are random object descriptors (e.g., Marron and Alonso (2014)) that take values in a manifold, or more generally, in a stratified space, and for our general CLT we require only:

- (i) a law of large numbers for a unique generalized Fréchet mean μ ,
- (ii) a local manifold structure near μ , sufficiently smooth,
- (iii) an a.s. Lipschitz condition and an a.s. differentiable distance between μ and data, and
- (iv) a population Fréchet function, sufficiently smooth at μ .

Further, we give an example for two-smeariness on spheres of arbitrary dimension, and show that smeariness, although “almost never” occurring, may have serious statistical implications on a continuum of sample scenarios nearby. Remarkably, this effect increases with dimension, striking in particular in high dimension low sample size scenarios.

2. A general central limit theorem. In a typical scenario of non-Euclidean statistics, a two-sample test is applied to two groups of manifold-valued data or more generally to data on a manifold-stratified space. Such a test can be based on certain data descriptors such as intrinsic means (e.g., [Bhattacharya and Patrangenaru \(2005\)](#), [Munk et al. \(2008\)](#), [Patrangenaru and Ellingson \(2016\)](#)), best approximating geodesics (e.g., [Huckemann \(2011b\)](#)), best approximating subspaces within a given family of subspaces and entire flags thereof (cf. [Huckemann and Eltzner \(2018\)](#)), and asymptotic confidence regions can be constructed from a suitable CLT for such descriptors. In this section, we first introduce the setting of generalized Fréchet means along with standard assumptions, we then recollect and expand some empirical process theory from [van der Vaart \(1998\)](#) and state and prove our general CLT.

2.1. *Generalized Fréchet means and assumptions.* Fréchet functions and Fréchet means have been first introduced by [Fréchet \(1948\)](#) for squared metrics $\tilde{\rho} : Q \times Q \rightarrow [0, \infty)$ on a topological space Q and later extended to squared quasimetrics by [Ziezold \(1977\)](#). Generalized Fréchet means as follows have been introduced by [Huckemann \(2011b\)](#). A simple setting is given when $P = Q$ is a Riemannian manifold and $\tilde{\rho} = d^2$ is the squared geodesic intrinsic distance. Then a generalized Fréchet mean is a minimizer with respect to squared distance, often called a *barycenter*.

NOTATION 2.1. Let P and Q be separable topological spaces, Q is called the *data space* and P is called the *descriptor space*, linked by a continuous map $\tilde{\rho} : P \times Q \rightarrow [0, \infty)$ reflecting distance between a data descriptor $p \in P$ and a datum $q \in Q$. Further, with a silently underlying probability space $(\Omega, \mathfrak{A}, \mathbb{P})$, let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X$ be random elements on Q , that is, they are Borel-measurable mappings $\Omega \rightarrow Q$. They give rise to *generalized population* and *generalized sample Fréchet functions*,

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

respectively, and their *generalized population* and *generalized sample Fréchet means*

$$\tilde{E} = \left\{ p \in P : \tilde{F}(p) = \inf_{p \in P} \tilde{F}(p) \right\}, \quad \tilde{E}_n = \left\{ p \in P : \tilde{F}_n(p) = \inf_{p \in P} \tilde{F}_n(p) \right\},$$

respectively. Here, the former set is empty if the expected value is never finite.

All of the theory developed in this paper concerns only distributions with unique population minimizer $\mu \in \{\mu\} = \tilde{E}$; cf. Assumption 2.2. With Assumption 2.3 further down, P is a manifold locally near μ , so that convergence in probability in the following assumption is well defined.

ASSUMPTION 2.2 (Unique mean with law of large numbers). In fact, we assume that \tilde{E} is not empty but contains a single descriptor $\mu \in P$ and that for every measurable selection $\mu_n \in \tilde{E}$,

$$\mu_n \xrightarrow{\mathbb{P}} \mu.$$

ASSUMPTION 2.3 (Local manifold structure). With $2 \leq r \in \mathbb{N}$, assume that there is a neighborhood \tilde{U} of μ that is an m -dimensional Riemannian manifold, $m \in \mathbb{N}$, such that with a neighborhood U of the origin in \mathbb{R}^m the exponential map $\exp_\mu : U \rightarrow \tilde{U}$, $\exp_\mu(0) = \mu$, is a C^r -diffeomorphism, and we set for $p = \exp_\mu(x)$, $p' = \exp_\mu(x') \in \tilde{U}$ and $q \in Q$,

$$\begin{aligned} \rho &: (x, q) \mapsto \tilde{\rho}(\exp_\mu(x), q), \\ F &: x \mapsto \tilde{F}(\exp_\mu(x)), \quad F_n : x \mapsto \tilde{F}_n(\exp_\mu(x)). \end{aligned}$$

It will be convenient to extend F_n to all of \mathbb{R}^m via $F_n(x) = F_n(0)$ for $x \in \mathbb{R}^m \setminus U$.

ASSUMPTION 2.4 (Almost surely locally Lipschitz and differentiable at mean). Further assume that:

- (i) the gradient $\dot{\rho}_0(X) := \text{grad}_x \rho(x, X)|_{x=0}$ exists almost surely;
- (ii) there is a measurable function $\dot{\rho} : Q \rightarrow \mathbb{R}$ satisfying $\mathbb{E}[\dot{\rho}(X)^2] < \infty$ for all $x \in U$ and that the following Lipschitz condition:

$$|\rho(x_1, X) - \rho(x_2, X)| \leq \dot{\rho}(X) \|x_1 - x_2\| \quad \text{a.s.}$$

holds for all $x_1, x_2 \in U$.

ASSUMPTION 2.5 (Smooth Fréchet function). With $2 \leq r \in \mathbb{N}$ and a nonvanishing tensor $T = (T_{j_1, \dots, j_r})_{1 \leq j_1 \leq \dots \leq j_r \leq m}$, assume that the Fréchet function admits the power series expansion

$$(2.1) \quad F(x) = F(0) + \sum_{1 \leq j_1 \leq \dots \leq j_r \leq m} x_{j_1} \dots x_{j_r} T_{j_1, \dots, j_r} + o(\|x\|^r).$$

The tensor in T in (2.1) can be very complicated. As is well known, for $r = 2$, every symmetric tensor is diagonalizable ($m(m + 1)/2$ parameters involved), which is, however, not true in general. For simplicity of argument, however, we assume that T is diagonalizable with nonzero diagonal elements so that Assumption 2.5 rewrites as follows. In this formulation, we can also drop our assumption that $r \in \mathbb{N}$.

ASSUMPTION 2.6. With $2 \leq r \in \mathbb{R}$, a rotation matrix $R \in SO(m)$ and $T_1, \dots, T_m \neq 0$ assume that the Fréchet function admits the power series expansion

$$(2.2) \quad F(x) = F(0) + \sum_{j=1}^m T_j |(Rx)_j|^r + o(\|x\|^r).$$

REMARK 2.7 (Typical scenarios). Let us briefly recall typical scenarios. In many applications, Q is:

(a) globally a complete smooth Riemannian manifold, for example, a sphere (cf. Mardia and Jupp (2000) for directional data), a real or complex projective space (cf. Kendall (1984), Mardia and Patrangenaru (2005) for certain shape spaces) or the space of positive definite matrices (cf. Dryden, Koloydenko and Zhou (2009) for diffusion tensors),

(b) a nonmanifold *shape space* which is a quotient of a Riemannian manifold under an isometric group action with varying dimensions of isotropy groups (e.g., Dryden and Mardia (1998), Kendall et al. (1999), for spaces of three- and higher-dimensional shapes),

(c) a general stratified space where all strata are manifolds with compatible Riemannian structures, for example, phylogenetic tree spaces (cf. Billera, Holmes and Vogtmann (2001), Moulton and Steel (2004), for varying geometries).

On these spaces:

(α) in most of the above applications, $P = Q$ and *intrinsic means* are considered where $\tilde{\rho}$ is the squared geodesic distance induced from the Riemannian structure.

(β) In other examples, $P = \Gamma$, the space of geodesics on Q is considered, in view of PCA-like dimension reduction methods (e.g., Fletcher and Joshi (2004), Huckemann, Hotz and Munk (2010), Huckemann and Ziezold (2006)), or

(γ) P is a family of subspaces of Q , or even a space of *nested subspheres* in Jung, Dryden and Marron (2012), Jung, Foskey and Marron (2011); more general families have been recently considered in generic dimension reduction methods, for example, Pennec (2018), Sommer (2016).

REMARK 2.8. Of the above assumptions, some are harder to prove in real examples than others.

(i) Of all above assumptions, uniqueness (first part of Assumption 2.2) seems most challenging to verify, as there are cases of nonuniqueness. For example, both north and south poles are intrinsic means of a uniform distribution along the equator of a two-sphere. For more examples, see Huckemann (2012). To date, only for intrinsic means on the circle the entire picture is known; cf. Hotz and Huckemann ((2015), page 182ff.). For complete Riemannian manifolds, uniqueness for intrinsic means has been shown if the support is sufficiently concentrated (cf. Afsari (2011), Groisser (2005), Karcher (1977), Kendall (1990), Le (2001)) and intrinsic sample means are unique a.s. if from a distribution absolutely continuous w.r.t. Riemannian measure; cf. Bhattacharya and Patrangenaru ((2003), Remark 2.6) (for the circle) and Arnaudon and Miclo ((2014), Theorem 2.1) (in general).

(ii) For the above typical scenarios, we anticipate that the other assumptions are often valid in concrete applications.

For instance, Assumption 2.3 is also true on nonmanifold shape spaces, because the mean is assumed on the highest dimensional manifold stratum intersecting with the support, due to the *manifold stability theorem* Huckemann ((2012), Corollary 1).

On arbitrary stratified spaces, it may only be valid if the mean is assumed in the top dimensional stratum, while means on lower dimensional strata may feature stickiness, as is typically the case in BHV spaces for phylogenetic trees; cf. Billera, Holmes and Vogtmann (2001), Barden, Le and Owen (2013, 2018), Hotz et al. (2013), Huckemann et al. (2015).

Our results are not applicable to current spaces of persistence diagrams (e.g., Turner et al. (2014)) because they do not provide local manifold structures.

(iii) Smeariness as described here is always connected to a vanishing Hessian of the Fréchet function. Varying the probability measure slightly can easily turn the Hessian positive definite, negative definite or indefinite, corresponding to a local minimum, maximum or saddle point. In particular, it seems that smeariness can be viewed as a boundary case between probability measures with unique means and measures with nonunique means.

(iv) Moreover, Assumption 2.4 is only slightly stronger than *uniform coercivity* (condition (2) in Huckemann ((2011b), page 1118)) which suffices for the strong law (second part of Assumption 2.2); cf. Huckemann ((2011b), Theorem A4) and Huckemann and Eltzner ((2018), Theorem 4.1), and this has been established for principal nested spheres in Huckemann and Eltzner ((2018), Theorem 3.8) and for geodesics with nested mean on Kendall's shape spaces in Huckemann and Eltzner ((2018), Theorem 3.9). In consequence of Lemma 4.4 below, we have that Assumption 2.4 holds for intrinsic means of distributions on spheres which feature a density near the antipodal point of the intrinsic population mean.

A more detailed analysis is beyond the scope of this paper and left for future research.

2.2. General CLT. For the following, fix a measurable selection $\mu_n \in \tilde{E}_n$. Due to $\mu_n \xrightarrow{\mathbb{P}} \mu$ from Assumption 2.2, we have $\mathbb{P}\{\mu_n \in \tilde{U}\} \rightarrow 1$, and in accordance with the convention in Assumption 2.3, setting

$$x_n := \begin{cases} \exp_{\mu}^{-1}(\mu_n) & \text{if } \mu_n \in \tilde{U}, \\ 0 & \text{else,} \end{cases}$$

note that

$$(2.3) \quad F_n(0) \geq F_n(x_n) = \tilde{F}_n(\mu_n) + o_p(1),$$

because $\mathbb{P}\{F_n(x_n) - \tilde{F}_n(\mu_n) > \epsilon\} = \mathbb{P}\{\mu_n \notin \tilde{U}\} \rightarrow 0$ for all $\epsilon > 0$.

The following is a direct consequence of [van der Vaart \(\(1998\), Lemma 5.52\)](#), replacing maxima with minima, where due to continuity of $\tilde{\rho}$, we have no need for outer measure and outer expectation, and due to our setup, no need for approximate minimizers.

LEMMA 2.9. *Assume that for fixed constants C and $\alpha > \beta$ for every n and for sufficiently small δ ,*

$$(2.4) \quad \sup_{\|x\| < \delta} |F(x) - F(0)| \geq C\delta^\alpha,$$

$$(2.5) \quad \mathbb{E} \left[n^{1/2} \sup_{\|x\| < \delta} |F_n(x) - F(x) - F_n(0) + F(0)| \right] \leq C\delta^\beta.$$

Then any a random sequence $\mathbb{R}^m \ni y_n \xrightarrow{\mathbb{P}} 0$ that satisfies $F_n(y_n) \leq F_n(0)$ also satisfies $n^{1/(2\alpha-2\beta)} y_n = \mathcal{O}_P(1)$.

As a first step, the following generalization of [van der Vaart \(\(1998\), Corollary 5.53\)](#), only treating the case $r = 2$) gives a bound for the scaling rate in the general CLT, so that also in case of $r \geq 2$, $n^{1/(2r-2)} x_n = \mathcal{O}_P(1)$.

COROLLARY 2.10. *Under Assumptions 2.2, 2.3 and 2.4, as well as Assumptions 2.5 or 2.6,*

$$n^{1/(2r-2)} x_n = \mathcal{O}_P(1).$$

PROOF. By Assumption 2.2 and definition, $x_n \xrightarrow{\mathbb{P}} 0$ with $F_n(x_n) \leq F_n(0)$; cf. (2.3). Hence, Lemma 2.9 yields the assertion, because for $\alpha = r$, (2.4) follows at once from (2.1) or from (2.2), and under Assumption 2.4, (2.5), for $\beta = 1$ follows word by word from the proof of [van der Vaart \(\(1998\), Corollary 5.53\)](#). \square

As the second step, the following theorem, which is a generalization and adaptation of [van der Vaart \(\(1998\), Theorem 5.23\)](#), shows that under Assumption 2.6 the above bound gives the exact scaling rate, including the explicit limiting distribution.

THEOREM 2.11 (General CLT for generalized Fréchet means). *Under Assumptions 2.2, 2.3, 2.4 and 2.6, we have*

$$n^{1/2} ((Rx_n)_1 | (Rx_n)_1|^{r-2}, \dots, (Rx_n)_m | (Rx_n)_m|^{r-2})^T \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{1}{r^2} T^{-1} \text{Cov}[\text{grad}|_{x=0} \rho(x, X)] T^{-1} \right)$$

with $T = \text{diag}(T_1, \dots, T_m)$. In particular for every coordinate $j = 1, \dots, m$,

$$n^{\frac{1}{2r-2}} (R^T x_n)_j \xrightarrow{\mathcal{D}} \mathcal{H}_j,$$

where $(\mathcal{H}_1, \dots, \mathcal{H}_m)$ is a random vector such that $(\mathcal{H}_1|\mathcal{H}_1|^{r-2}, \dots, \mathcal{H}_m|\mathcal{H}_m|^{r-2})$ has the above multivariate Gaussian limiting distribution.

PROOF. For $z \in U$ and $2(r - 1) = 1/s$, let us abbreviate

$$\begin{aligned} \tau_n(z, X) &:= n^s(\rho(zn^{-s}, X) - \rho(0, X)) - z^T \dot{\rho}_0(X), \\ G_n &:= n^{1/2} \left(\frac{1}{n} \sum_{j=1}^n \dot{\rho}_0(X_j) - \mathbb{E}[\dot{\rho}_0(X)] \right), \end{aligned}$$

where we set $\rho(zn^{-s}, X) = \rho(0, X)$ if $zn^{-s} \notin U$. Then, due to Assumptions 2.4 and 2.6, and $1/2 + s - sr = 0$,

$$\begin{aligned} &n^{1/2} \left(\frac{1}{n} \sum_{j=1}^n (\tau_n(z, X_j)) - \mathbb{E}[\tau_n(z, X)] \right) \\ &= n^{1/2+s} (F_n(zn^{-s}) - F_n(0) - F(zn^{-s}) + F(0)) - z^T G_n \\ &= n^{1/2+s} (F_n(zn^{-s}) - F_n(0)) - \sum_{j=1}^m T_j |(Rz)_j|^r - z^T G_n + o(\|z\|^r) \end{aligned}$$

is a sequence of stochastic processes, indexed in $z \in U$, with zero expectation and variance converging to zero. By argument from the proof of van der Vaart ((1998), Lemma 19.31), due to Assumption 2.4, z can be replaced with any random sequence $z_n = O_p(1)$ (cf. also the proof of van der Vaart (1998), Lemma 5.23) for $r = 2$, yielding

$$(2.6) \quad n^{1/2+s} (F_n(z_n n^{-s}) - F_n(0)) = \sum_{j=1}^m T_j |(Rz_n)_j|^r + z_n^T G_n + o_p(1).$$

By Corollary 2.10, $z_n = x_n n^s$ is a valid choice in equation (2.6). Comparison with any other $z_n = O_p(1)$, because μ_n is a minimizer for \tilde{F}_n and $F_n(x_n)$ deviates only up to $o_p(1)$ from $\tilde{F}_n(\mu_n)$, due to (2.3), reveals

$$n^{1/2+s} (F_n(x_n) - F_n(0)) \leq n^{1/2+s} (F_n(z_n n^{-s}) - F_n(0)) + o_p(1).$$

This asserts that $Rx_n n^s$ is a minimizer, up to $o_p(1)$, of the right-hand side of (2.6), that is, of

$$f : w \mapsto f(w) := \sum_{j=1}^m T_j |(w)_j|^r + w^T R G_n.$$

This function, however, has a unique minimizer, given on the component level ($j = 1, \dots, m$) by

$$r T_j \text{sign}((w_n)_j) |(w_n)_j|^{r-1} = -(R G_n)_j \quad \text{that is, } (w_n)_j |(w_n)_j|^{r-2} = -\frac{(R G_n)_j}{r T_j},$$

yielding

$$\sqrt{n}(Rx_n)_j |(Rx_n)_j|^{r-2} = -\frac{(RG_n)_j}{rT_j} + o_P(1).$$

Now the classical CLT gives the first assertion. The second also follows from the above display, since for $z = (Rx_n)_j$ and $H = -(RG_n)_j/rT_j$, the equation $\sqrt{n} \operatorname{sign}(z)|z|^{r-1} = H$ implies $\operatorname{sign}(z) = \operatorname{sign}(H)$, and hence

$$n^{\frac{1}{2r-2}}z = n^{\frac{1}{2r-2}} \operatorname{sign}(z)|z| = \operatorname{sign}(H)|H|^{\frac{1}{r-1}}. \quad \square$$

REMARK 2.12. The above arguments rely among others on the fact that due to Assumption 2.4, a specific convergence, different from (1.2), that can be easily verified for empirical processes indexed in a deterministic bounded variable, is also valid if the index varies randomly, bounded in probability. This can be weakened to the requirement, that the function class $\rho(x, \cdot)$ possesses the *Donsker* property; cf. van der Vaart ((1998), Chapter 19).

3. Smeariness. Recall from Huckemann (2015) that a sequence of random vectors X_n is *kth order smeary* if $n^{\frac{1}{2(k+1)}}X_n$ has a nontrivial limiting distribution as $n \rightarrow \infty$.

With this notion, the classical central limit theorem in particular asserts for random vectors with existing second moments that the fluctuation of sample means around the population mean is 0th order smeary, also called *nonsmeary*.

It has been shown in Hotz and Huckemann (2015) that the fluctuation of random directions on the circle of sample means around the population mean may feature smeariness of any given positive integer order. It has been unknown to date, however, whether the phenomenon of smeariness extends to higher dimensions, in particular, to positive curvature.

To this end, we now make the concept of smeariness on manifolds precise.

DEFINITION 3.1. Let $(\Omega, \mathfrak{A}, \mathbb{P})$ be a probability space, $X : \Omega \rightarrow \mathbb{R}^m$ a random vector and $k > -1$. Then a sequence of Borel measurable mappings $X_n : \Omega_n \rightarrow \mathbb{R}^m$ ($n \in \mathbb{N}$) with $\Omega_n \in \mathfrak{A}$, $\mathbb{P}(\Omega_n) \rightarrow 1$ ($n \rightarrow \infty$) is *k-smeariness with limiting distribution* of X if

$$\mathbb{P}\{n^{\frac{1}{2(k+1)}}X_n \in B | \Omega_n\} \rightarrow \mathbb{P}\{X \in B\} \quad \text{as } n \rightarrow \infty \text{ for all Borel sets } B \subset \mathbb{R}^m.$$

In this case, we write $n^{\frac{1}{2(k+1)}}X_n \xrightarrow{\mathcal{D}} X$.

Note that $-1 < k$ -smeariness implies that $\mathbb{P}\{X_n \in B | \Omega_n\} \rightarrow 1_{0 \in B}$ for all Borel $B \subset \mathbb{R}^m$. As usual, we abbreviate this with $X_n \xrightarrow{\mathbb{P}} 0$.

LEMMA 3.2. Let $X_n : \Omega_n \rightarrow \mathbb{R}^m$ be Borel measurable with $\mathbb{P}(\Omega_n) \rightarrow 1$ and $X_n \xrightarrow{\mathbb{P}} 0$, consider a continuously differentiable local bijection $\Phi : U \rightarrow V$ preserving the origin $0 \in U, V$ open $\subset \mathbb{R}^m$, set $Y_n = \Phi(X_n) : \Omega_n \cap \{X_n \in U\} \rightarrow \mathbb{R}^m$ and let $k > -1$. Then

$$X_n \text{ is } k\text{-smeary} \iff Y_n \text{ is } k\text{-smeary}.$$

In particular, if X has the limiting distribution of $n^{\frac{1}{2(k+1)}} X_n$, then $D\Phi(0)X$ has the limiting distribution of $n^{\frac{1}{2(k+1)}} Y_n$. Here, $D\Phi(x)$ denotes the differential of Φ at $x \in U$ and $\det(D\Phi(0)) \neq 0$ due to invertibility of Φ .

PROOF. The implication “ \Rightarrow ” is a direct consequence of a Taylor expansion and the continuity theorem with a suitable point $\tilde{X}_n \xrightarrow{\mathbb{P}} 0$ between the origin and X_n as follows:

$$\begin{aligned} &\mathbb{P}\{n^{\frac{1}{2(k+1)}} Y_n \in B \mid \Omega_n \cap \{X_n \in U\}\} \\ &= \mathbb{P}\{n^{\frac{1}{2(k+1)}} D\Phi(\tilde{X}_n)X_n \in B \mid \Omega_n \cap \{X_n \in U\}\} \\ &\rightarrow \mathbb{P}\{D\Phi(0)X \in B\} \end{aligned}$$

because $\mathbb{P}\{X_n \in U\} \rightarrow 1$ due to $X_n \xrightarrow{\mathbb{P}} 0$.

Similarly, the implication “ \Leftarrow ” follows. Suppose that Y has the limiting distribution of $n^{\frac{1}{2(k+1)}} Y_n$. Then

$$\begin{aligned} &\mathbb{P}\{n^{\frac{1}{2(k+1)}} X_n \in B \mid \Omega_n\} \\ &= \mathbb{P}\{n^{\frac{1}{2(k+1)}} D\Phi(\tilde{X}_n)^{-1}Y_n \in B \mid \Omega_n \cap \{X_n \in U\}\} \\ &\quad + \mathbb{P}\{n^{\frac{1}{2(k+1)}} X_n \in B \mid \Omega_n \cap \{X_n \notin U\}\} \\ &\rightarrow \mathbb{P}\{D\Phi(0)^{-1}Y \in B\}, \end{aligned}$$

again due to the hypothesis $X_n \xrightarrow{\mathbb{P}} 0$. \square

In consequence of Lemma 3.2, we have the following general definition.

DEFINITION 3.3. A sequence $\mu_n \xrightarrow{\mathbb{P}} \mu$ of random variables on an m -dimensional manifold M is k -smeary if in one—and hence in every—continuously differentiable chart $\phi^{-1} : \tilde{U} \rightarrow \mathbb{R}^m$ around $\mu \in \tilde{U} \subset M$ the sequence of vectors $\phi^{-1}(\mu_n) - \phi^{-1}(\mu) : \{\mu_n \in \tilde{U}\} \rightarrow \mathbb{R}^m$ is k -smeary.

REMARK 3.4. In particular, the order of smeariness is independent of the chart chosen.

4. An example of two-smeariness on spheres.

4.1. *Setup.* Consider a random variable X distributed on the m -dimensional unit sphere \mathbb{S}^m ($m \geq 2$) that is uniformly distributed on the lower half-sphere $\mathbb{L}^m = \{q \in \mathbb{S}^m : q_2 \leq 0\}$ with total mass $0 < \alpha < 1$ and assuming the *north pole* $\mu = (0, 1, 0, \dots, 0)^T$ with probability $1 - \alpha$. Then we have the *Fréchet function*

$$\tilde{F} : \mathbb{S}^m \rightarrow [0, \infty), \quad p \mapsto \int_{\mathbb{S}^m} d(p, q)^2 d\mathbb{P}^X(q)$$

involving the *squared spherical distance* $d(p, q)^2 = \arccos\langle p, q \rangle^2$ based on the standard inner product $\langle \cdot, \cdot \rangle$ of \mathbb{R}^{m+1} . Every minimizer $p^* \in \mathbb{S}^m$ of F is called an *intrinsic Fréchet population mean* of X .

With the volume of \mathbb{S}^m given by

$$v_m = \text{vol}(\mathbb{S}^m) = \frac{2\pi^{\frac{m+1}{2}}}{\Gamma(\frac{m+1}{2})}$$

define

$$\gamma_m = \frac{v_{m+1}}{2v_m} = \frac{\sqrt{\pi} \Gamma(\frac{m+1}{2})}{2 \Gamma(\frac{m+2}{2})}.$$

Moreover, we have the *exponential chart* centered at $\mu \in \mathbb{S}^m$ with inverse

$$\exp_{\mu}^{-1}(p) = (e_1, e_3, \dots, e_{m+1})^T (p - \langle p, \mu \rangle \mu) \frac{\arccos\langle p, \mu \rangle}{\|p - \langle p, \mu \rangle \mu\|} = x \in \mathbb{R}^m,$$

where e_1, \dots, e_{m+1} are the standard unit column vectors in \mathbb{R}^{m+1} . Note that \exp_{μ}^{-1} has continuous derivatives of any order in $\tilde{U} = \mathbb{S}^m \setminus \{-\mu\}$ and recall that $e_2 = \mu$.

4.2. *Derivatives of the Fréchet function.*

LEMMA 4.1. *With the above notation, the function $F = \tilde{F} \circ \exp_{\mu}$ has derivatives of any order for $x \in \exp_{\mu}^{-1}(\tilde{U})$ with $\|x\| < \pi/2$. For $\alpha = 1/(1 + \gamma_m)$, the north pole μ gives the unique intrinsic Fréchet mean with $\text{Hess}|_{x=0} F \circ \exp_{\mu}(x) = 0$. Moreover, for any choice of $0 < \alpha < 1$,*

$$\begin{aligned} \partial_i \partial_k \partial_l |_{x=0} F &= 0, \\ \partial_i \partial_k \partial_l \partial_s |_{x=0} F &= c_m \delta_{i,k,l,s} \end{aligned}$$

for every $1 \leq i, k, l, s \leq m$ with the constant $c_m = \frac{2\gamma_m}{1+\gamma_m} \frac{m-1}{m+2} > 0$.

PROOF. For convenience, we choose polar coordinates $\theta_1, \dots, \theta_{m-1} \in [-\pi/2, \pi/2]$ and $\phi \in [-\pi, \pi)$ in the nonstandard way

$$q = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_{m-1} \\ q_m \\ q_{m+1} \end{pmatrix} = \begin{pmatrix} -\left(\prod_{j=1}^{m-1} \cos \theta_j\right) \cos \phi \\ -\left(\prod_{j=1}^{m-1} \cos \theta_j\right) \sin \phi \\ \vdots \\ -\cos \theta_1 \cos \theta_2 \sin \theta_3 \\ -\cos \theta_1 \sin \theta_2 \\ \sin \theta_1 \end{pmatrix},$$

such that the north pole μ has coordinates $(0, \dots, 0, -\pi/2)$. In fact, we have chosen these coordinates so that w.l.o.g. we may assume that the arbitrary but fixed point $p \in \mathbb{S}^m$ has coordinates $(0, 0, \dots, 0, -\pi/2 + \delta)$ with suitable $\delta \in [0, \pi]$. Setting $\Theta = [-\pi/2, \pi/2]$, with the function

$$g : \Theta^{m-1} \rightarrow [0, 1], \quad \theta = (\theta_1, \dots, \theta_{m-1}) \mapsto \prod_{j=1}^{m-1} \cos^{m-j} \theta_j$$

we have the spherical volume element $g(\theta) d\theta d\phi$. Additionally, defining

$$h(\theta) = \prod_{j=1}^{m-1} \cos \theta_j,$$

we have that

$$\tilde{F}(p) = \tilde{F}(\mu) + \frac{2\alpha}{v_m} (C_+(\delta) - C_-(\delta)) + \delta^2(1 - \alpha) =: G(\delta)$$

with the two ‘‘crescent’’ integrals

$$\begin{aligned} C_+(\delta) &= \int_{\Theta^{m-1}} d\theta g(\theta) \int_{-\delta}^0 d(\mu, q)^2 d\phi \\ &= \int_{\Theta^{m-1}} d\theta g(\theta) \int_0^\delta (\arccos(h(\theta) \sin \phi))^2 d\phi, \\ C_-(\delta) &= \int_{\Theta^{m-1}} d\theta g(\theta) \int_{\pi-\delta}^\pi d(\mu, q)^2 d\phi \\ &= \int_{\Theta^{m-1}} d\theta g(\theta) \int_0^\delta (\arccos(-h(\theta) \sin \phi))^2 d\phi \end{aligned}$$

(cf. Figure 2), because the spherical measure of \mathbb{L}^m is $v_m/2$.

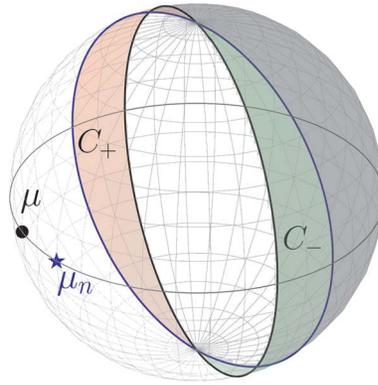


FIG. 2. Depicting the two crescents $C_+ = C_+(\delta)$ and $C_- = C_-(\delta)$ for $\delta = \arccos\langle \mu, \mu_n \rangle$ on \mathbb{S}^m for $m = 2$ with north pole μ and nearby sample Fréchet mean μ_n .

Since for $a \in [0, 1]$,

$$\begin{aligned} & (\arccos(a))^2 - (\arccos(-a))^2 \\ &= (\arccos(a) + \arccos(-a))(\arccos(a) - \arccos(-a)) \\ &= 2\pi \left(\frac{\pi}{2} - \arccos(a) \right) \\ &= -2\pi \arcsin(a), \end{aligned}$$

which has arbitrary derivatives if $-1 < a < 1$, we have that

$$\begin{aligned} \tilde{F} \circ \exp_\mu(x) &= G(\delta) \\ (4.1) \quad &= G(0) - \frac{4\pi\alpha}{v_m} \int_{\Theta^{m-1}} d\theta g(\theta) \int_0^\delta \arcsin(h(\theta) \sin \phi) d\phi \\ &\quad + \delta^2(1 - \alpha) \end{aligned}$$

for every $x \in \exp_\mu^{-1}(\tilde{U})$ with $\|x\| = \delta$, yielding the first assertion of the Lemma.

For the second assertion, we use the Taylor expansion

$$(4.2) \quad \arcsin(h(\theta) \sin \phi) = \phi h(\theta) + \frac{\phi^3}{6}(h(\theta)^3 - h(\theta)) + O(\phi^5)$$

and compute for $k = 0, 1, \dots$,

$$\begin{aligned} & \int_{\Theta^{m-1}} g(\theta) h(\theta)^k d\theta \\ (4.3) \quad &= \int_{\Theta^{m-1}} \prod_{j=1}^{m-1} (\cos^{m-j+k} \theta_j d\theta_j) \end{aligned}$$

$$\begin{aligned}
 &= \int_{\Theta^{m+k-1}} \prod_{j=1}^{m+k-1} (\cos^{m+k-j} \theta_j \, d\theta_j) / \int_{\Theta^k} \prod_{j=1}^k (\cos^j \theta_j \, d\theta_j) \\
 &= \frac{v_{m+k}}{v_{k+1}}
 \end{aligned}$$

to obtain, in conjunction with (4.1),

$$G(\delta) = G(0) + \delta^2 \left(1 - \alpha \left(1 + \frac{v_{m+1}}{2v_m} \right) \right) + \frac{\delta^4}{24} \frac{\alpha v_{m+1}}{v_m} \frac{m-1}{m+2} + \dots,$$

which yields that for any choice of $\alpha \in [0, 1]$ we have $G'(0) = 0 = G'''(0)$, as well as $G''(0) \geq 0$ for $1 \geq \alpha(1 + \gamma_m)$ with equality for $\alpha = 1/(1 + \gamma_m)$. Since $G''''(0) = \frac{\alpha v_{m+1}}{v_m} \frac{m-1}{m+2} = c_m > 0$ for all $\alpha \in (0, 1)$, this guarantees a local minimum for $\alpha = 1/(1 + \gamma_m)$.

In order to see that μ gives the global minimum in case of $\alpha = 1/(1 + \frac{v_{m+1}}{2v_m})$ we consider the derivatives

$$\begin{aligned}
 G'(\delta) &= -\frac{4\pi\alpha}{v_m} \int_{\Theta^{m-1}} g(\theta) \arcsin(h(\theta) \sin \delta) \, d\theta + 2\delta(1 - \alpha), \\
 G''(\delta) &= -\frac{4\pi\alpha}{v_m} \int_{\Theta^{m-1}} g(\theta)h(\theta) \frac{\cos \delta}{\sqrt{1 - h(\theta)^2 \sin^2 \delta}} \, d\theta + 2(1 - \alpha) \\
 (4.4) \quad &\geq -\frac{4\pi\alpha}{v_m} \int_{\Theta^{m-1}} g(\theta)h(\theta) \, d\theta + 2(1 - \alpha) \\
 &= 2 - \alpha \left(2 + \frac{v_{m+1}}{v_m} \right) = 0,
 \end{aligned}$$

where the inequality is strict for $\delta \neq 0, \pi$, that is, $p \neq \pm\mu$, due to $0 < h(\theta) < 1$ for all $\theta \in (-\pi/2, \pi/2)^{m-1}$. Hence we infer that $G'(\delta)$ is strictly increasing in δ from $G'(0) = 0$, yielding that there is no stationary point for F other than $p = \mu$. \square

REMARK 4.2.

(i) Note that the result of [Bhattacharya and Lin \(\(2017\), Proposition 3.1\)](#) is not applicable to our setup as they have shown that on an arbitrary dimensional sphere the Fréchet function is twice differentiable, if the random direction has a density that is twice differentiable w.r.t. spherical measure. For the theorem to follow, we require fourth derivatives.

(ii) Note that $O(\phi^5)$ in the Taylor expansion (4.2) stands for

$$\sum_{j=2}^{\infty} \phi^{2j+1} \sum_{r=0}^j a_{2r+1, 2j+1} h(\theta)^{2r+1}$$

with suitable coefficients $a_{2r+1,2j+1} \in \mathbb{R}$. Moreover, due to (4.3) we have

$$\begin{aligned} \frac{1}{v_m} \int_{\Theta^{m-1}} g(\theta)h(\theta)^{2r+1} d\theta &= \frac{1}{v_{2r+2}} \frac{v_{m+2r+1}}{v_m} \\ &= \begin{cases} \frac{1}{v_2} \frac{v_{m+1}}{v_m} & \text{for } r = 0 \\ \frac{1}{v_2} \frac{v_{m+1}}{v_m} \prod_{k=1}^r \frac{2k}{m+2k-1} & \text{for } r > 0 \end{cases} \\ &= O\left(\frac{1}{\sqrt{m}}\right), \end{aligned}$$

due to Stirling’s formula $\Gamma(z) = \sqrt{\frac{2\pi}{z}} \left(\frac{z}{e}\right)^z (1 + O(\frac{1}{z}))$. In consequence (cf. (4.4)), $G^{(k)}(0) = 0$ for odd $k \in \mathbb{N}$ and $G^{(k)}(0) = O(\frac{1}{\sqrt{m}})$ for even $4 \leq k \in \mathbb{N}$, as $m \rightarrow \infty$.

4.3. *A two-smearly central limit theorem.* For a sample X_1, \dots, X_n on \mathbb{S}^m , recall the empirical Fréchet function

$$\tilde{F}_n : \mathbb{S}^m \rightarrow [0, \infty), \quad p \mapsto \frac{1}{n} \sum_{j=1}^n d(p, X_j)^2,$$

where every minimizer $\mu_n \in \mathbb{S}^m$ of \tilde{F}_n is called an *intrinsic Fréchet sample mean* or short just a *sample mean*.

THEOREM 4.3. *Let X_1, \dots, X_n be a sample from X as introduced in the setup Section 4.1 with $\alpha = 1/(1 + \gamma_m)$. Then every measurable selection of sample means*

$$\mu_n \in \operatorname{argmin}_{p \in \mathbb{S}^m} \tilde{F}_n(p)$$

is two-smearly. More precisely, with the exponential chart \exp_μ at the north pole, there is a full rank $m \times m$ matrix Σ such that

$$\sqrt{n}(\exp_\mu^{-1}(\mu_n))^3 \rightarrow \mathcal{N}(0, \Sigma),$$

where the third power is taken componentwise.

PROOF. From Lemma 4.1, we infer that μ is the unique intrinsic Fréchet mean, and hence by the strong law of [Bhattacharya and Patrangenaru \(\(2003\), Theorem 2.3\)](#) we have that $\mu_n \rightarrow \mu$ almost surely yielding that Assumption 2.2 holds. Since \mathbb{S}^m is an analytic Riemannian manifold also, Assumption 2.3 holds for arbitrary $r \in \mathbb{N}$. With the exponential chart $\exp_\mu^{-1} : \tilde{U} \rightarrow \mathbb{R}^m$, we have $\exp_\mu^{-1}(\mu) = 0$ and we set $\exp_\mu^{-1}(\mu_n) = Z_n$ on $\{\mu_n \neq -\mu\}$ with $\mathbb{P}\{\mu_n \neq -\mu\} \rightarrow 1$ and $Z_n \xrightarrow{\text{a.s.}} 0$.

Further, due to Lemma 4.4, the family of functions

$$\rho(z, X) = d(\exp_\mu(z), X)^2, \quad z \in U$$

has a.s. derivatives $\text{grad}_z \rho(z, X)$, which are bounded, and on a compact set, are square integrable, so that Assumption 2.4 holds.

Recalling the function $G(\delta)$ from the proof of Lemma 4.1 with its Taylor expansion, we have with $\delta = d(\exp_\mu(z), \mu) = \|z\|$ that

$$\mathbb{E}[g_z(X)] = G(\delta) = G(0) - \delta^4 \frac{c_m}{24} + \dots$$

and in consequence, Assumption 2.6 holds with $r = 4$. Thus, Theorem 2.11 is applicable.

In particular, for the covariance we have

$$\Sigma = \frac{36}{c_m^2} \text{Cov}[\text{grad}_z |_{z=0} d(\exp_\mu(z), X)^2],$$

which has full rank because in the exponential chart, rotational symmetry is preserved. This yields the assertion. \square

LEMMA 4.4. For $x \in \mathbb{S}^m$ and $z \in \mathbb{R}^m \setminus \{\exp_\mu^{-1}(-x)\}$, $\|z\| < \pi$,

$$\text{grad}_z(d(\exp_\mu(z), x)^2)$$

is well defined and has bounded directional limits as $z \rightarrow \exp_\mu^{-1}(-x)$ or $\|z\| \rightarrow \pi$.

PROOF. Recalling that $d(\exp_\mu(z), x)^2 = \text{acos}\langle x, \exp_\mu(z) \rangle^2$, we have

$$(4.5) \quad \text{grad}_z(d(\exp_\mu(z), x)^2) = -2 \frac{\text{grad}_z \langle x, \exp_\mu(z) \rangle}{\sqrt{1 - \langle x, \exp_\mu(z) \rangle^2}} \text{acos}\langle x, \exp_\mu(z) \rangle.$$

In case of $x \neq 0$, this is bounded for $\|z\| \rightarrow \pi$. As we now show boundedness of (4.5) also for $z \rightarrow \exp_\mu^{-1}(-x)$ for arbitrary $x \in \mathbb{S}^m$, also the boundedness in case of $x = 0$ and $\|z\| \rightarrow \pi$ follows at once.

To this end, let z be near $\exp_\mu^{-1}(-x)$ such that $z = \exp_\mu^{-1}(-x) + w$ with $w = (w_1, \dots, w_m) \in \mathbb{R}^m$ small. Then the asserted boundedness of (4.5) follows from

$$(4.6) \quad \begin{aligned} \langle x, \exp_\mu(z) \rangle &= \langle x, \exp_\mu(\exp_\mu^{-1}(-x) + w) \rangle \\ &= -1 + w^T B w + \mathcal{O}(\|w\|^3) \end{aligned}$$

with a symmetric matrix B , because then

$$\frac{\text{grad}_z \langle x, \exp_\mu(z) \rangle}{\sqrt{1 - \langle x, \exp_\mu(z) \rangle^2}} = \frac{2Bw + \mathcal{O}(w^2)}{\sqrt{2w^T B w + \mathcal{O}(w^3)}},$$

which is bounded for $w \rightarrow 0$ with any (possibly vanishing) symmetric B .

Finally, in order to see that the gradient w.r.t. w of the left-hand side of (4.6) vanishes at $w = 0$, w.l.o.g. assume that $\exp_\mu^{-1}(x) = (\theta, 0, \dots, 0)^T$ for some $\theta \in$

$[0, \pi]$, such that $x = (\sin \theta, \cos \theta, 0, \dots, 0)^T$ and $\exp_\mu^{-1}(-x) = (\pi - \theta, 0, \dots, 0)^T$. Moreover, verify that

$$\exp_\mu(z) = \left(\frac{-\pi + \theta + w_1}{\|z\|} \sin \|z\|, \cos \|z\|, *, \dots, * \right),$$

which, in conjunction with $\|z\| = \sqrt{(\pi - \theta)^2 - 2w_1(\pi - \theta) + \|w\|^2}$, and hence, $\text{grad}_w|_{w=0}\|z\| = -e_1$, yields that

$$\begin{aligned} & \text{grad}_w \langle x, \exp_\mu(\exp_\mu^{-1}(-x) + w) \rangle \\ &= \left((-\pi + \theta + w_1) \sin \theta \left(\frac{\cos \|z\|}{\|z\|} - \frac{\sin \|z\|}{\|z\|^2} \right) - \cos \theta \sin \|z\| \right) \text{grad}_w \|z\| \\ & \quad + \sin \theta \frac{\sin \|z\|}{\|z\|} e_1, \end{aligned}$$

giving

$$\begin{aligned} & \text{grad}_w|_{w=0} \langle x, \exp_\mu(\exp_\mu^{-1}(-x) + w) \rangle \\ &= (\sin \theta (\text{sinc}(\pi - \theta) - \text{sinc}(\pi - \theta) - \cos(\pi - \theta)) - \cos \theta \sin(\pi - \theta)) e_1 \\ &= -\sin \pi e_1 = 0, \end{aligned}$$

which proves the claim (4.6). \square

5. High dimension low sample size effects near smeariness. We illustrate the relevance of our result by simulations of the variance $V = \tilde{F}(\mu)$ (the Fréchet function at the point mass at the north pole μ) from the distributions introduced above in the setup Section 4.1, which are parametrized in $\alpha = \alpha_{\text{crit}} + \beta \in [0, 1]$, on \mathbb{S}^m , for dimensions $m = 2, 10$ and 100 . Here, the critical value $\alpha_{\text{crit}} = \frac{1}{1+\gamma_m}$ is 0.56, 0.72 and 0.89, respectively.

We consider sample sizes ranging from 300 to $3 \cdot 10^5$ data points. For every sample size, we draw 1000 samples, determine the spherical mean for each sample and then determine their empirical Fréchet function at μ , that is, the sum of squared distances of the means from the north pole. As we have a nonunique circular minimum of the Fréchet function for $\beta > 0$, we expect in this case that the variance V approaches a finite value, namely the squared radius of the circular mean set. For $\beta \leq 0$, we have a unique minimum, where for $\beta = 0$ we expect a slow decay of V with rate approaching $n^{-\frac{1}{3}}$, due to Theorem 4.3, and for $\beta < 0$ we expect the decay rate to approach n^{-1} .

The results of our simulation are displayed in Figure 3. The asymptotic rates are clearly in agreement with our considerations based on the asymptotic theory. Strikingly, however, for $\beta < 0$ very close to 0, the decay rate stays close to $n^{-\frac{1}{3}}$ until very large sample sizes and only then settles into the asymptotic rate of n^{-1} . This illustrates that the slow convergence to the mean is an issue, which does not

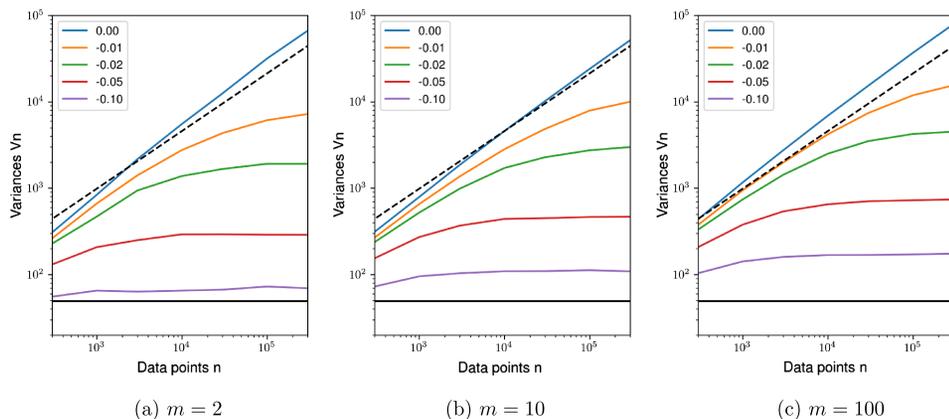


FIG. 3. Simulated variances V times n for different values of β for dimensions $m = 2, 10$ and 100 . Black lines $V \propto n^{-1}$ (solid) and $V \propto n^{-\frac{1}{3}}$ (dashed) for reference.

only plague the distribution with $\beta = 0$ but also sufficiently adjacent distributions for finite sample size.

Even more strikingly, Figure 3 shows that this phenomenon increases with dimension m . Indeed, due to Remark 4.2(ii), in the limit $m \rightarrow \infty$, all derivatives of G vanish with a uniform rate, so that we approach a situation of infinite smearing.

Acknowledgments. The authors thank Axel Munk for pointing to van der Vaart (1998) and John Kent for bringing the turtles data set in Figure 1 to our attention.

REFERENCES

- AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. [MR2736346](#)
- ARNAUDON, M. and MICLO, L. (2014). Means in complete manifolds: Uniqueness and approximation. *ESAIM Probab. Stat.* **18** 185–206. [MR3230874](#)
- BARDEN, D., LE, H. and OWEN, M. (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electron. J. Probab.* **18** 1–25. [MR3035753](#)
- BARDEN, D., LE, H. and OWEN, M. (2018). Limiting behaviour of Fréchet means in the space of phylogenetic trees. *Ann. Inst. Statist. Math.* **70** 99–129. [MR3742820](#)
- BHATTACHARYA, A. and BHATTACHARYA, R. (2008). Statistics on Riemannian manifolds: Asymptotic distribution and curvature. *Proc. Amer. Math. Soc.* **136** 2959–2967. [MR2399064](#)
- BHATTACHARYA, R. and LIN, L. (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proc. Amer. Math. Soc.* **145** 413–428. [MR3565392](#)
- BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. [MR1962498](#)
- BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. [MR2195634](#)
- BHATTACHARYA, R. and PATRANGENARU, V. (2014). Statistics on manifolds and landmarks based image analysis: A nonparametric theory with applications. *J. Statist. Plann. Inference* **145** 1–22. [MR3125343](#)

- BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. [MR1867931](#)
- DRYDEN, I. L., KOLOYDENKO, A. and ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123. [MR2750388](#)
- DRYDEN, I. L. and MARDIA, K. V. (1998). *Statistical Shape Analysis. Wiley Series in Probability and Statistics: Probability and Statistics.* Wiley, Chichester. [MR1646114](#)
- ELLINGSON, L., PATRANGENARU, V. and RUYMGAART, F. (2013). Nonparametric estimation of means on Hilbert manifolds and extrinsic analysis of mean shapes of contours. *J. Multivariate Anal.* **122** 317–333. [MR3189326](#)
- FLETCHER, P. T. and JOSHI, S. C. (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. *ECCV Workshops CVAMIA and MMBIA* 87–98.
- FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. [MR0027464](#)
- GROSSER, D. (2005). On the convergence of some Procrustean averaging algorithms. *Stochastics* **77** 31–60. [MR2138772](#)
- HOTZ, T. and HUCKEMANN, S. (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Ann. Inst. Statist. Math.* **67** 177–193. [MR3297863](#)
- HOTZ, T., HUCKEMANN, S., LE, H., MARRON, J. S., MATTINGLY, J. C., MILLER, E., NOLEN, J., OWEN, M., PATRANGENARU, V. et al. (2013). Sticky central limit theorems on open books. *Ann. Appl. Probab.* **23** 2238–2258. [MR3127934](#)
- HUCKEMANN, S. (2011a). Inference on 3D Procrustes means: Tree bole growth, rank deficient diffusion tensors and perturbation models. *Scand. J. Stat.* **38** 424–446. [MR2833839](#)
- HUCKEMANN, S. F. (2011b). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.* **39** 1098–1124. [MR2816349](#)
- HUCKEMANN, S. F. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Ann. Inst. Statist. Math.* **64** 1227–1259. [MR2981621](#)
- HUCKEMANN, S. (2015). (Semi-)intrinsic statistical analysis on non-Euclidean spaces. In *Advances in Complex Data Modeling and Computational Methods in Statistics.* Springer, Berlin.
- HUCKEMANN, S. F. and ELTZNER, B. (2018). Backward nested descriptors asymptotics with inference on stem cell differentiation. *Ann. Statist.* **46** 1994–2019. [MR3845008](#)
- HUCKEMANN, S., HOTZ, T. and MUNK, A. (2010) Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). *Statist. Sinica* **20** 1–100.
- HUCKEMANN, S. and ZIEZOLD, H. (2006). Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces. *Adv. in Appl. Probab.* **38** 299–319. [MR2264946](#)
- HUCKEMANN, S., MATTINGLY, J. C., MILLER, E. and NOLEN, J. (2015). Sticky central limit theorems at isolated hyperbolic planar singularities. *Electron. J. Probab.* **20** no. 78, 34. [MR3371437](#)
- JUNG, S., DRYDEN, I. L. and MARRON, J. S. (2012). Analysis of principal nested spheres. *Biometrika* **99** 551–568. [MR2966769](#)
- JUNG, S., FOSKEY, M. and MARRON, J. S. (2011). Principal arc analysis on direct product manifolds. *Ann. Appl. Stat.* **5** 578–603. [MR2810410](#)
- KARCHER, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30** 509–541. [MR0442975](#)
- KENDALL, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. Lond. Math. Soc.* **16** 81–121. [MR0737237](#)
- KENDALL, W. S. (1990). Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence. *Proc. Lond. Math. Soc.* (3) **61** 371–406. [MR1063050](#)
- KENDALL, D. G., BARDEN, D., CARNE, T. K. and LE, H. (1999). *Shape and Shape Theory. Wiley Series in Probability and Statistics.* Wiley, Chichester. [MR1891212](#)

- LE, H. (2001). Locating Fréchet means with application to shape spaces. *Adv. in Appl. Probab.* **33** 324–338. [MR1842295](#)
- LE, H. and BARDEN, D. (2014). On the measure of the cut locus of a Fréchet mean. *Bull. Lond. Math. Soc.* **46** 698–708. [MR3239607](#)
- MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics. Wiley Series in Probability and Statistics.* Wiley, Chichester. [MR1828667](#)
- MARDIA, K. V. and PATRANGENARU, V. (2005). Directions and projective shapes. *Ann. Statist.* **33** 1666–1699. [MR2166559](#)
- MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biom. J.* **56** 732–753. [MR3258083](#)
- MCKILLIAM, R. G., QUINN, B. G. and CLARKSON, I. V. L. (2012). Direction estimation by minimum squared arc length. *IEEE Trans. Signal Process.* **60** 2115–2124. [MR2954196](#)
- MOULTON, V. and STEEL, M. (2004). Peeling phylogenetic ‘oranges’. *Adv. in Appl. Math.* **33** 710–727. [MR2095862](#)
- MUNK, A., PAIGE, R., PANG, J., PATRANGENARU, V. and RUYMGAART, F. (2008). The one- and multi-sample problem for functional data with application to projective shape analysis. *J. Multivariate Anal.* **99** 815–833. [MR2405093](#)
- PATRANGENARU, V. and ELLINGSON, L. (2016). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis.* CRC Press, Boca Raton, FL. [MR3444169](#)
- PENNEC, X. (2018). Barycentric subspace analysis on manifolds. *Ann. Statist.* **46** 2711–2746. [MR3851753](#)
- SOMMER, S. (2016). Anisotropically weighted and nonholonomically constrained evolutions on manifolds. *Entropy* **18** Paper No. 425, 21. [MR3594070](#)
- TURNER, K., MILEYKO, Y., MUKHERJEE, S. and HARER, J. (2014). Fréchet means for distributions of persistence diagrams. *Discrete Comput. Geom.* **52** 44–70. [MR3231030](#)
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- ZIEZOLD, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians* 591–602. [MR0501230](#)

FELIX-BERNSTEIN-INSTITUT
FÜR MATHEMATISCHE STATISTIK
IN DEN BIOWISSENSCHAFTEN
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN
GOLDSCHMIDTSTRASSE 7
37077 GOETTINGEN
GERMANY
E-MAIL: benjamin.eltzner@mathematik.uni-goettingen.de
stephan.huckemann@mathematik.uni-goettingen.de