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

BY BENJAMIN ELTZNER
Felix-Bernstein-Institut für Mathematische Statistik in den Biowissenschaften, Georg-August-Universität Göttingen

AND

BY STEPHAN F. HUCKEMANN
Felix-Bernstein-Institut für Mathematische Statistik in den Biowissenschaften, 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 (Bhattacharya and Patrangenaru (2005)). 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 \( n^\alpha \) with exponents \( \alpha < 1/2 \) with a non-normal 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

MSC 2010 subject classifications: Primary 62G20, 62H11; secondary 53C22, 58K45
Keywords and phrases: Fréchet means, asymptotic consistency and normality, asymptotics on manifolds, lower asymptotic rate, high dimension low sample size
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, p. 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 not realistic. We propose to consider instead a Taylor expansion of the population Fréchet function, combined with more involved Donsker 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.

Fig 1: Left: Directions (blue) of 76 hatching turtles from (Mardia and Jupp, 2000, p. 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 with smeariness (black dashed). Applying the classical Euclidean CLT to bootstrapped tangent space images gives classical asymptotically constant behavior (orange) corresponding to non-smeariness (black solid).

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); Huckemann (2011a); Bhattacharya and Patrangenaru (2013); Ellingson, Patrangenaru and Ruymgaart (2013); Patrangenaru and Ellingson (2015); Bhattacharya and Lin (2017)), rests on a Taylor expansion

\[
\sqrt{n} \text{grad}|_{x=x_0} F_n(x) = \sqrt{n} \text{grad}|_{x=0} F_n(x) + \text{Hess}|_{x=x} F_n(x) \sqrt{n} x_0
\]

(with suitable \( x \) between 0 and \( x_0 \)) and a generalized strong law (\( n \to \infty \) and \( x_0 \to 0 \))

\[
\text{Hess}|_{x=x} F_n(x) \overset{p}{\to} \text{Hess}|_{x=0} F(x).
\]

Here, \( X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \) is a sample on a smooth manifold \( M \),

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

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 \( \phi \) denotes a local smooth chart which maps the population Fréchet mean to 0.

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 \( \phi \) of any sample Fréchet mean, i.e. a minimizer of the sample Fréchet function \( F_n \), the l.h.s. of Equation (1.1) vanishes. The r.h.s. 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 (2013) 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

\[
\sqrt{n} x_n \overset{D}{\to} N(0, \Sigma), \quad \Sigma
\]

with suitable covariance matrix \( \Sigma \), if the Hessian on the r.h.s 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

\[
nF_n(x) = \sum_{x-\pi \leq X_j} (X_j - x)^2 + \sum_{X_j < x-\pi} (X_j + 2\pi - x)^2
= \sum_{j=1}^n (X_j - x)^2 + 4\pi \sum_{X_j < x-\pi} (X_j - x + \pi)
\]

so that the l.h.s. 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 r.h.s., 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 McKilliam, Quinn and Clarkson (2012); Hotz and Huckemann (2015), 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 \( n^\alpha \) with exponents \( \alpha \) 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 Donsker Theory (e.g. from van der Vaart (2000)) 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

\[
\sqrt{n}x_n|x_n|^\kappa \overset{D}{\to} \mathcal{N}(0, \Sigma),
\]

where \( x_n|x_n|^\kappa \) is defined componentwise. Then, \( x_n \) scales with \( n^\alpha \), \( \alpha = \frac{1}{2(\kappa+1)} \), 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, e.g. geodesic principal components (Huckemann and Ziezold (2006); Huckemann, Hotz and Munk (2010)) 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 \( \mu \),
(ii) a local manifold structure near \( \mu \), sufficiently smooth,
(iii) an a.s. Lipschitz condition and an a.s. differentiable distance between \( \mu \) and data, and
(iv) a population Fréchet function, sufficiently smooth at \( \mu \).

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 (2015)), best approximating geodesics (e.g. Huckemann (2011b)), best approximating subspaces within a given family of subspaces and entire flags thereof (cf. Huckemann and Eltzner (2017)), 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 Donsker Theory from van der Vaart (2000) 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 \( \hat{\rho} : Q \times Q \to [0, \infty) \) 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 \( \hat{\rho} = d^2 \) 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 $\tilde{\rho} : P \times Q \to [0, \infty)$ reflecting distance between a data descriptor $p \in P$
and a datum $q \in Q$. Further, with a silently underlying probability space
$(\Omega, \mathcal{A}, \mathbb{P})$, let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} X$ be random elements on $Q$, i.e. they are Borel-
measurable mappings $\Omega \to Q$. 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 $\mu$, 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 $\mu_n \in \tilde{E},$

\[ \mu_n \xrightarrow{p} \mu. \]

**Assumption 2.3 (Local Manifold Structure).** With $2 \leq r \in \mathbb{N}$ assume
that there is a neighborhood $\bar{U}$ of $\mu$ 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 $\exp_\mu : U \to \bar{U}, \exp_\mu(0) = \mu$, is a $C^r$-diffeomorphism, and
we set for $p = \exp_\mu(x), p' = \exp_\mu(x') \in \bar{U}$ and $q \in Q$,

\[ \rho : (x, q) \mapsto \tilde{\rho}(\exp_\mu(x), q), \]

\[ F : x \mapsto \tilde{F}(\exp_\mu(x)), \quad F_n : x \mapsto \tilde{F}_n(\exp_\mu(x)). \]

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 \( \dot{\rho} : \mathcal{Q} \to \mathbb{R} \) satisfying \( \mathbb{E}[\dot{\rho}(X)^2] < \infty \) for all \( x \in U \) and that the following Lipschitz condition

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

holds for all \( x_1, x_2 \in U \).

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

\[ F(x) = F(0) + \sum_{1 \leq j_1 \leq \ldots \leq j_r \leq m} x_{j_1} \ldots x_{j_r} T_{j_1, \ldots, j_r} + o(\|x\|^r). \]

(2.1)

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 non-zero 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 \( T_1, \ldots, T_m \neq 0 \) assume that the Fréchet function admits the power series expansion

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

(2.2)

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

(a) globally a complete smooth Riemannian manifold, e.g. 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, Kolodyeneko and Zhou (2009) for diffusion tensors),
(b) a non-manifold 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, e.g. 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 $\bar{\rho}$ 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 and Ziezold (2006); Huckemann, Hotz and Munk (2010)), 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, e.g. Sommer (2016); Pennec (2017).

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 non-uniqueness. For example, both north and south pole 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, p. 182 ff.). For complete Riemannian manifolds, uniqueness for intrinsic means has been shown if the support is sufficiently concentrated (cf. Karcher (1977); Kendall (1990); Le (2001); Groisser (2005); Assari (2011)) 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 non-manifold 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 as-
sumed 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); Hotz et al. (2013); Huckemann et al. (2015); Barden, Le and Owen (2018).

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 non-unique means.

(iv) Moreover, Assumption 2.4 is only slightly stronger than uniform coercivity (condition (2) in Huckemann (2011b, p. 1118)) which suffices for the strong law (second part of Assumption 2.2), cf. Huckemann (2011b, Theorem A4) and Huckemann and Eltzner (2017, Theorem 4.1), and this has been established for principal nested spheres in Huckemann and Eltzner (2017, Theorem 3.8) and for geodesics with nested mean on Kendall’s shape spaces in Huckemann and Eltzner (2017, 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 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 \( \mu_n \xrightarrow{P} \mu \) from Assumption 2.2, we have \( \mathbb{P}\{\mu_n \in \tilde{U}\} \to 1 \), and in accordance with the convention in Assumption 2.3, setting

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

note that

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

because \( \mathbb{P}\{F_n(x_n) - \tilde{F}_n(\mu_n) > \epsilon\} = \mathbb{P}\{\mu_n \not\in \tilde{U}\} \to 0 \) for all \( \epsilon > 0 \).
The following is a direct consequence of van der Vaart (2000, 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 $\delta$

\begin{equation}
\sup_{\|x\| < \delta} |F(x) - F(0)| \leq C\delta^\alpha,
\end{equation}

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

Then, any a random sequence $\mathbb{R}^m \ni y_n \Rightarrow 0$ 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 (2000, 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 $r \geq 2$, $\sqrt{n}x_n = o_P(1)$.

**Corollary 2.10.** Under Assumptions 2.2, 2.3 and 2.4, as well as Assumption 2.5 or 2.6, $n^{1/(2r-2)}x_n = \mathcal{O}_P(1)$.

**Proof.** By Assumption 2.2 and definition, $x_n \Rightarrow 0$ 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 (2000, Corollary 5.53).

As the second step, the following Theorem, which is a generalization and adaption of van der Vaart (2000, 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

\begin{equation}
\begin{aligned}
n^{1/2} \left( (Rx_n)_1 |(Rx_n)_1|^{r-2}, \ldots, (Rx_n)_m |(Rx_n)_m|^{r-2} \right)^T \\
\Rightarrow \mathcal{N} \left( 0, \frac{1}{r^2} T^{-1} \text{Cov}[\text{grad}|_x=0\rho(x,X)] T^{-1} \right)
\end{aligned}
\end{equation}
with $T = \text{diag}(T_1, \ldots, T_m)$. In particular for every coordinate $j = 1, \ldots, m$,

$$n^{r-2} \left( R^T x_n \right)_j \xrightarrow{D} H_j$$

where $(H_1, \ldots, H_m)$ is a random vector such that $(H_1|H_1|^{-2}, \ldots, H_m|H_m|^{-2})$ has the above multivariate Gaussian limiting distribution.

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

$$\tau_n(z, X) := n^s (\rho(zn^{-s}, X) - \rho(0, X)) - z^T \hat{\rho}_0(X)$$

and

$$G_n := n^{1/2} \left( \frac{1}{n} \sum_{j=1}^{n} \hat{\rho}_0(X_j) - \mathbb{E} [\hat{\rho}_0(X)] \right),$$

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 $1/2 + s - sr = 0$,

$$n^{1/2} \left( \frac{1}{n} \sum_{j=1}^{n} (\tau_n(z, X_j)) - \mathbb{E} [\tau_n(z, X)] \right)$$

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 (2000, 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 (2000, Lemma 5.23) for $r = 2$, yielding,

$$n^{1/2+s} \left( F_n(zn^{-s}) - F_n(0) \right) = \sum_{j=1}^{m} T_j |(Rz)_j|^r - z^T G_n + o_p(1).$$

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 $\mu_n$ 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,

$$n^{1/2+s} \left( F_n(x_n) - F_n(0) \right) \leq n^{1/2+s} \left( F_n(zn^{-s}) - F_n(0) \right) + o_p(1).$$
This asserts that \( R x_n n^s \) is a minimizer, up to \( o_P(1) \), of the right hand side of (2.6), i.e. 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 \((j = 1, \ldots, m)\) by
\[
 r T_j \text{sign}((w_n)_j)| (w_n)_j |^{r-1} = -(R G_n)_j \quad \text{i.e.} \quad (w_n)_j |(w_n)_j |^{r-2} = -\frac{(R G_n)_j}{r T_j},
\]
yielding
\[
 \sqrt{n} (R x_n)_j |(R x_n)_j |^{r-2} = -\frac{(R G_n)_j}{r T_j} + o_P(1). 
\]
Now the classical CLT gives the first assertion. The second also follows from the above display, since for \( z = (R x_n)_j \) and \( H = -(R G_n)_j/r T_j \), the equation
\[
 \sqrt{n} \text{sign}(z) |z|^{r-1} = H \text{ implies sign}(z) = \text{sign}(H) \text{ and hence}
\]
\[
 n^\frac{1}{r-2} z = n^\frac{1}{r-2} \text{sign}(z) |z| = \text{sign}(H) |H|^{1/r-1}.
\]

**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, are 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 (2000, Chapter 19).

3. Smeariness. Recall from Huckemann (2015) that a sequence of random vectors \( X_n \) is \( k \)-th order smeary if \( n^{\frac{1}{2k+1}} X_n \) has a non-trivial limiting distribution as \( n \to \infty \).

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 0-th order smeary, also called non-smeary.

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 \((Ω, ℱ, ℙ)\) be a probability space, \(X : Ω → ℜ^m\) a random vector and \(k > -1\). Then a sequence of Borel measurable mappings \(X_n : Ω_n → ℜ^m\) \((n ∈ ℤ)\) with \(Ω_n ∈ ℱ, ℙ(Ω_n) → 1\) \((n → ∞)\) is \(k\)-smeary with limiting distribution of \(X\) if

\[
ℙ \left\{ n^\frac{1}{2(k+1)} X_n ∈ B | Ω_n \right\} → ℙ \{ X ∈ B \} \text{ as } n → ∞ \text{ for all Borel sets } B ⊂ ℜ^m.
\]

In this case we write \(n^\frac{1}{2(k+1)} X_n \overset{P}{→} X\).

Note that \(-1 < k\)-smearyness implies that \(ℙ \{ X_n ∈ B | Ω_n \} → 1\) for all Borel \(B ⊂ ℜ^m\). As usual, we abbreviate this with \(X_n \overset{P}{→} 0\).

Lemma 3.2. Let \(X_n : Ω_n → ℜ^m\) be Borel measurable with \(ℙ(Ω_n) → 1\) and \(X_n \overset{P}{→} 0\), consider a continuously differentiable local bijection \(Φ : U → V\) preserving the origin \(0 ∈ U, V\) open \(⊂ ℜ^m\), set \(Y_n = Φ(X_n) : Ω_n ∩ \{ X_n ∈ U \} → ℜ^m\) and let \(k > -1\). Then \(X_n\) is \(k\)-smeary \(⇔ Y_n\) is \(k\)-smeary.

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

Proof. The implication “⇒” is a direct consequence of a Taylor expansion and the continuity theorem with a suitable point \(X_n \overset{P}{→} 0\) between the origin and \(X_n\) as follows

\[
ℙ \left\{ n^\frac{1}{2(k+1)} Y_n ∈ B | Ω_n ∩ \{ X_n ∈ U \} \right\} = ℙ \left\{ n^\frac{1}{2(k+1)} DΦ(\bar{X}_n) X_n ∈ B | Ω_n ∩ \{ X_n ∈ U \} \right\} \rightarrow ℙ \{ DΦ(0) X ∈ B \}
\]

because \(ℙ \{ X_n ∈ U \} → 1\) due to \(X_n \overset{P}{→} 0\).

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

\[
ℙ \left\{ n^\frac{1}{2(k+1)} X_n ∈ B | Ω_n \right\} = ℙ \left\{ n^\frac{1}{2(k+1)} DΦ(\bar{X}_n)^{-1} Y_n ∈ B | Ω_n ∩ \{ X_n ∈ U \} \right\} \rightarrow ℙ \{ DΦ(0)^{-1} Y ∈ B \},
\]

again due to the hypothesis \(X_n \overset{P}{→} 0\). 

\(\square\)
In consequence of Lemma 3.2, we have the following general definition.

**Definition 3.3.** A sequence $\mu_n \xrightarrow{P} \mu$ of random variables on a $m$-dimensional manifold $M$ is $k$-smeary if in one – and hence in every – continuously differentiable chart $\phi^{-1} : \tilde{U} \to \mathbb{R}^m$ around $\mu \in \tilde{U} \subset M$ the sequence of vectors $\phi^{-1}(\mu_n) - \phi^{-1}(\mu) : \{\mu_n \in \tilde{U}\} \to \mathbb{R}^m$ 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 $S^m$ ($m \geq 2$) that is uniformly distributed on the lower half sphere $L^m = \{q \in S^m : q_2 \leq 0\}$ with total mass $0 < \alpha < 1$ and assuming the north pole $\mu = (0, 1, 0, \ldots, 0)^T$ with probability $1 - \alpha$. Then we have the Fréchet function

$$\tilde{F} : S^m \to [0, \infty), \ p \mapsto \int_{S^m} \rho(p, q) \ dP_X(q)$$

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

With the volume of $S^m$ given by

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

define

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

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

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

where $e_1, \ldots, e_{m+1}$ are the standard unit column vectors in $\mathbb{R}^{m+1}$. Note that $\exp^{-1}_\mu$ has continuous derivatives of any order in $\tilde{U} = S^m \setminus \{-\mu\}$ 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 $x \in \exp_\mu^{-1}(\bar{U})$ with $||x|| < \pi/2$. For $\alpha = 1/(1 + \gamma_m)$ the north pole $\mu$ gives the unique intrinsic Fréchet mean with $\text{Hess}|_{x=0}F \circ \exp_\mu(x) = 0$. Moreover, for any choice of $0 < \alpha < 1$,

\[
\partial_i \partial_k \partial_l \big|_{x=0} F = 0 \quad \partial_i \partial_h \partial_l \partial_s \big|_{x=0} F = c_m \delta_{i,k,l,s}
\]

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 $\theta_1, \ldots, \theta_{m-1} \in [-\pi/2, \pi/2]$ and $\phi \in [-\pi, \pi)$ in the non-standard way

\[
q = \begin{pmatrix}
g_1 \\
g_2 \\
\vdots \\
g_{m-1} \\
g_m \\
g_{m+1}
\end{pmatrix} = \begin{pmatrix}
-\left( \prod_{j=1}^{m-1} \cos \theta_j \right) \cos \phi \\
-\left( \prod_{j=1}^{m-1} \cos \theta_j \right) \sin \phi \\
\vdots \\
- \cos \theta_1 \cos \theta_2 \sin \theta_3 \\
- \cos \theta_1 \sin \theta_2 \\
\sin \theta_1
\end{pmatrix},
\]

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

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

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

\[ C_+ (\delta) = \int_{\Theta^{m-1}} d\theta \, g(\theta) \int_0^\delta d\phi \, \tilde{p}(\mu, q)^2 = \int d\theta \, g(\theta) \int_0^\delta \left( \arccos (h(\theta) \sin \phi) \right)^2 d\phi \]

\[ C_- (\delta) = \int_{\Theta^{m-1}} d\theta \, g(\theta) \int_{\pi-\delta}^{\pi} d\phi \, \tilde{p}(\mu, q)^2 = \int d\theta \, g(\theta) \int_0^\delta \left( \arccos (-h(\theta) \sin \phi) \right)^2 d\phi \]

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

Since for \( a \in [0, 1] \),

\[ (\arccos(a))^2 - (\arccos(-a))^2 = (\arccos(a) + \arccos(-a))(\arccos(a) - \arccos(-a)) \]

\[ = 2\pi \left( \frac{\pi}{2} - \arccos(a) \right) = -2\pi \arcsin(a) , \]

which has arbitrary derivatives if \(-1 < a < 1\), we have that

\[ (4.1) \]

\[ \tilde{F} \circ \exp^\mu (x) = G(\delta) = G(0) - \frac{4\pi \alpha}{v_m} \int_{\Theta^{m-1}} d\theta \, g(\theta) \int_0^\delta \arcsin \left( h(\theta) \sin \phi \right) d\phi + \delta^2 (1 - \alpha) \]

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

For the second assertion we use the Taylor expansion

\[ (4.2) \]

\[ \arcsin \left( h(\theta) \sin \phi \right) = \phi \, h(\theta) + \frac{\phi^3}{6} \left( h(\theta)^3 - h(\theta) \right) + O(\phi^5) \]

and compute for \( k = 0, 1, \ldots \),

\[ \int_{\Theta^{m-1}} g(\theta) \, h(\theta)^k \, d\theta = \int_{\Theta^{m-1}} \prod_{j=1}^{m-1} (\cos^{m-j+k} \theta_j \, d\theta_j) \]

\[ = \int_{\Theta^{m+k-1}} \prod_{j=1}^{m+k-1} (\cos^{m-j+k} \theta_j \, d\theta_j) / \int_{\Theta^k} \prod_{j=1}^k (\cos^j \theta_j \, d\theta_j) \]

\[ (4.3) \]

\[ = \frac{v_{m+k}}{v_{k+1}} , \]

to obtain, in conjunction with (4.1),

\[ G(\delta) = G(0) + \delta^2 \left( 1 - \alpha \left( 1 + \frac{v_{m+1}}{2v_m} \right) \right) + \frac{\delta^4}{24} \frac{\alpha v_{m+1}}{v_m} \frac{m - 1}{m + 2} + \ldots \]
Fig 2: Depicting the two crescents $C_+ = C_+(\delta)$ and $C_- = C_-(\delta)$ for $\delta = \arccos(\mu, \mu_n)$ on $S^m$ for $m = 2$ with north pole $\mu$ and nearby sample Fréchet mean $\mu_n$.

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 $G'''(0) = \frac{v_{m+1}}{v_m} \frac{m-1}{m+2} = c_m > 0$ for all $\alpha \in (0, 1)$, this guarantees a local minimum for $\alpha = 1 / (1 + \gamma_m)$.

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

\[ G'(\delta) = -\frac{4\pi\alpha}{v_m} \int_{\Omega^m_{n-1}} g(\theta) \arcsin \left( h(\theta) \sin \delta \right) d\theta + 2\delta (1 - \alpha), \]

(4.4)

\[ G''(\delta) = -\frac{4\pi\alpha}{v_m} \int_{\Omega^m_{n-1}} g(\theta) h(\theta) \frac{\cos \delta}{\sqrt{1 - h(\theta)^2 \sin^2 \delta}} d\theta + 2(1 - \alpha) \]

\[ \geq -\frac{4\pi\alpha}{v_m} \int_{\Omega^m_{n-1}} g(\theta) h(\theta) d\theta + 2(1 - \alpha) = 2 - \alpha \left( 2 + \frac{v_{m+1}}{v_m} \right) = 0, \]

where the inequality is strict for $\delta \neq 0, \pi$, i.e. $p \neq \pm \mu$, 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 $\delta$ from $G'(0) = 0$, yielding that there is no stationary point for $F$ other than $p = \mu$.

**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
\[ \frac{1}{v_m} \int_{\Omega^{m-1}} g(\theta) h(\theta)^{2r+1} d\theta = \frac{1}{v_{2r+2}} \frac{v_m + 2r + 1}{v_m} \]
\[ = \left\{ \begin{array}{ll}
\frac{1}{v_2} \frac{v_m + 1}{v_m} \prod_{k=1}^{r} \frac{2k}{2k - 1} & \text{for } r = 0, \\
\frac{1}{v_2} \frac{v_m + 1}{v_m} \prod_{k=1}^{r} \frac{2k}{2k - 1} & \text{for } r > 0.
\end{array} \right\} = O \left( \frac{1}{\sqrt{m}} \right), \]
due to Stirling’s formula $\Gamma(z) = \sqrt{2\pi\frac{z}{e}} \left(1 + O\left(\frac{1}{z}\right)\right)$. In consequence, cf. (4.4), $G^{(k)}(0) = 0$ for odd $k \in \mathbb{N}$ and $G^{(k)}(0) = O\left(\frac{1}{\sqrt{m}}\right)$ for even $4 \leq k \in \mathbb{N}$, as $m \to \infty$.

4.3. A Two-Smeary Central Limit Theorem. For a sample $X_1, \ldots, X_n$ on $\mathbb{S}^m$ recall the empirical Fréchet function
\[ \tilde{F}_n : \mathbb{S}^m \to [0, \infty), \quad p \mapsto \frac{1}{n} \sum_{j=1}^{n} \tilde{\rho}(p, X_j)^2, \]

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 $X_1, \ldots, X_n$ 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
\[ \mu_n \in \arg\min_{p \in \mathbb{S}^m} \tilde{F}_n(p) \]
is two-smeary. More precisely, with the exponential chart $\exp_\mu$ at the north pole, there is a full rank $m \times m$ matrix $\Sigma$ such that
\[ \sqrt{n} \left( \exp_\mu^{-1}(\mu_n) \right)^3 \to \mathcal{N}(0, \Sigma) \]
where the third power is taken component-wise.
Proof. From Lemma 4.1 we infer that $\mu$ is the unique intrinsic Fréchet mean and hence by the strong law of Bhattacharya and Patrangenaru (2003, Theorem 2.3) we have that $\mu_n \to \mu$ almost surely yielding that Assumption 2.2 holds. Since $S^m$ is an analytic Riemannian manifold also Assumption 2.3 holds for arbitrary $r \in \mathbb{N}$. With the exponential chart $\exp_{\mu}^{-1} : \tilde{U} \to \mathbb{R}^m$ we have $\exp_{\mu}^{-1}(\mu) = 0$ and we set $\exp_{\mu}^{-1}(\mu_n) = Z_n$ on $\{\mu_n \neq -\mu\}$ with $\mathbb{P}\{\mu_n \neq -\mu\} \to 1$ and $Z_n \xrightarrow{a.s.} 0$.

Further, due to Lemma 4.4, the family of functions
$$\rho(z, X) = \tilde{\rho}(\exp_{\mu}(z), X)^2, \quad z \in U$$
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 $\delta = \tilde{\rho}(\exp_{\mu}(z), \mu) = \|z\|$ that
$$\mathbb{E}[g_z(X)] = G(\delta) = G(0) - \delta^4 \frac{c_m}{24} + \ldots$$
and in consequence, Assumption 2.6 holds with $r = 4$, Thus, Theorem 2.11 is applicable.

In particular, for the covariance we have
$$\Sigma = \frac{36}{c_m^2} \text{Cov}[\text{grad}_z \tilde{\rho}(\exp_{\mu}(z), X)^2],$$
which has full rank because in the exponential chart, rotational symmetry is preserved. This yields the assertion.

\begin{lemma}
For $x \in S^m$ and $z \in \mathbb{R}^m \setminus \{\exp_{\mu}^{-1}(-x)\}$, $\|z\| < \pi$,
$$\text{grad}_z \left( \tilde{\rho}(\exp_{\mu}(z), x) \right)$$
is well defined and has bounded directional limits as $z \to \exp_{\mu}^{-1}(-x)$ or $\|z\| \to \pi$.
\end{lemma}
\begin{proof}
Recalling that $\tilde{\rho}(\exp_{\mu}(z), x) = \text{acos} \langle x, \exp_{\mu}(z) \rangle^2$, we have
\begin{equation}
\text{grad}_z \left( \tilde{\rho}(\exp_{\mu}(z), x) \right) = -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.
\end{equation}
\end{proof}
In case of \( x \neq 0 \) this is bounded for \( \|z\| \to \pi \). As we now show boundedness of (4.5) also for \( z \to \exp^{-1}_\mu(-x) \) for arbitrary \( x \in \mathbb{S}^m \), also the boundedness in case of \( x = 0 \) and \( \|z\| \to \pi \) follows at once.

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

\[
\langle x, \exp_\mu(z) \rangle = \langle x, \exp_\mu(\exp^{-1}_\mu(-x) + w) \rangle = -1 + w^T B w + \mathcal{O}(\|w\|^3)
\]

with a symmetric matrix \( B \), because then

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

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

Finally, in order to see that the gradient w.r.t. \( w \) of the l.h.s. of (4.6) vanishes at \( w = 0 \), w.l.o.g. assume that \( \exp^{-1}_\mu(x) = (\theta, 0, \ldots, 0)^T \) for some \( \theta \in [0, \pi] \), such that \( x = (\sin \theta, \cos \theta, 0, \ldots, 0)^T \) and \( \exp^{-1}_\mu(-x) = (\pi - \theta, 0, \ldots, 0)^T \). Moreover, verify that

\[
\text{grad}_w |_{w=0} \langle x, \exp_\mu(\exp^{-1}_\mu(-x) + w) \rangle
\]

\[
= \left( (-\pi + \theta + w_1) \sin \theta \left( \frac{\cos \|z\|}{\|z\|} - \frac{\sin \|z\|}{\|z\|^2} \right) - \cos \theta \sin \|z\| \right) \text{grad}_w \|z\| + \sin \theta \left( \frac{\sin \|z\|}{\|z\|} \right) e_1,
\]

giving

\[
\text{grad}_w |_{w=0} \langle x, \exp_\mu(\exp^{-1}_\mu(-x) + w) \rangle
\]

\[
= \left( \sin \theta \left( \text{sinc}(\pi - \theta) - \text{sinc}(\pi - \theta) - \cos(\pi - \theta) \right) - \cos \theta \sin(\pi - \theta) \right) e_1
\]

\[
= - \sin \pi e_1 = 0,
\]

which proves the claim (4.6).
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 \( \mu \)) from the above in the setup Section 4.1 introduced distributions parametrized in \( \alpha = \alpha_{\text{crit}} + \beta \in [0, 1] \), on \( S^m \), 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 30 to \( 10^6 \) 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 \( \mu \), i.e. the sum of squared distances of the means from the north pole. As we have a non-unique 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 \( n^{-1} \).

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

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 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 \( m \to \infty \), all derivatives of \( G \) vanish with a uniform rate, so that we approach a situation
of infinite smeariness.

Acknowledgment. Both authors gratefully acknowledge DFG 755 B8, DFG HU 1575/4 and the Niedersachsen Vorab of the Volkswagen Foundation. We also thank Axel Munk for pointing to van der Vaart (2000) and John Kent for bringing the turtles data set in Fig. 1 to our attention.

References.


E-MAIL: benjamin.eltzner@mathematik.uni-goettingen.de  E-MAIL: stephan.huckemann@mathematik.uni-goettingen.de