# Smeariness: Central Limit Theorem and Examples

Benjamin Eltzner

Max-Planck-Institute for Multidisciplinary Sciences, Göttingen

October 17, 2024

## Contents



### Initial Remark

**Parts that go beyond the lecture presentation are printed in blue.**

## 1 Basic Concepts and Notation

We will use standard notation:

- A "silently underlying" probability space $(\Omega, \mathcal{A}(\Omega), \mathcal{P})$, where $\mathcal{A}(\Omega)$ is a $\sigma$-algebra and $\mathcal{P}$ is a probability measure.
- A measurable function $X : \Omega \to Q$ is called a random variable (RV) and $Q$ is the topological measurable data space.
- We make heavy use of the following notation: Let $T(X)$ be an equality, inequality or set inclusion involving $X$, then $\mathbb{P}(T(X)) := \mathcal{P}(\{\omega \in \Omega : T(X(\omega))\})$. In these statements, it is always silently required that the set $\{x \in Q : T(x)\}$ is measurable.
- Expected values for any function $f : Q \to V$ for any vector space $V$ are defined as

$$\mathbb{E}[f(X)] := \int f \circ X(\omega) d\mathcal{P}(\omega) \in V \,.$$

$\Omega$ is always assumed "sufficiently large", but what does that mean?

- Assume a RV $X : \Omega \to \mathbb{R}$, then $\Omega = [0, 1]$ suffices.
- Assume a RV $X : \Omega \to \mathbb{R}^2$, then $\Omega = [0, 1]$ suffices.
- Assume two RVs $X_1, X_2 : \Omega \to \mathbb{R}$, then $\Omega = [0, 1]$ suffices.
- Assume a sequence of RVs $X_1, X_2, \dots : \Omega \to \mathbb{R}$, then $\Omega = [0, 1]$ suffices.

Arguments $\omega \in \Omega$ are usually not written out. One usually writes some RVs as capital letters (but not all, e.g. $\widehat{\mu}_n$), which are otherwise reserved for matrices, which one continues to write as capital letters. Please ask, whenever notation is unclear.

In all asymptotic theory, there are three principal objects:

1. A sequence of RVs $X_1, X_2, \dots : \Omega \to Q$ on some measurable data space $Q$. For now, we consider so-called univariate RVs $X : \Omega \to Q = \mathbb{R}$.
2. A descriptor (often called parameter) $p$ which maps a RV $X$ to an element of a measurable parameter space $P$.
3. An estimator, which is a sequence $\widehat{p}_1, \widehat{p}_2, \dots$ of functions $\widehat{p}_n : Q^n \to P$. We always write the RV $\widehat{p}_n := \widehat{p}_n(X_1, \dots, X_n)$ for any $n \in \mathbb{N}$.

**Definition 1.1.** *Let $Q$ be a topological space called the data space and $P$ a separable complete metric space with continuous metric function $d : P \times P \to \mathbb{R}_{\geq 0}$ and topology derived from the metric called the parameter space. Let $\rho : P \times Q \to \mathbb{R}$ be a function which is continuous in $P$ and measurable in $Q$, $X$ a $Q$-valued random variable and $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} X$ a random sample. Define*

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

$$E := \left\{ p \in P : F(p) = \inf_{\widetilde{p} \in P} F(\widetilde{p}) \right\} \qquad E_n := \left\{ p \in P : F_n(p) = \inf_{\widetilde{p} \in P} F_n(\widetilde{p}) \right\} ,$$

*where $F$ and $F_n$ are called the population and sample loss functions and $E$ and $E_n$ are the sets of population and sample descriptors. The elements of $E_n$, are commonly called m-estimators. We call $(Q, P, \rho)$ the m-estimation model.*

Many descriptors of interest can be expressed as m-estimators:

**Examples 1.2.**

- *If $P = Q$ and $\rho(\mu, x) := d(\mu, x)^2$, the descriptor $\mu$ is the Fréchet mean.*
- *If $P = Q$ and $\rho(\mu, x) := d(\mu, x)^p$, the descriptor $\mu$ is called the $L^p$-mean.*

- *If $\rho(\theta, x)$ is the negative logarithm of a probability density function from a parametric family with parameter set $P = \Theta$, the descriptor $\theta$ is the maximum likelihood parameter.*
- *If $Q$ is a Riemannian manifold, $P$ is the set of geodesics on $Q$ and $\rho(\gamma, x) := \inf_{y \in \mathrm{Img}(\gamma)} d(y, x)^2$, the descriptor $\mu$ is called the principal geodesic.*

**Assumption 1.** *The sequence of RVs $X_1, X_2, \cdots : \Omega \to Q$ are identically distributed to a random variable $X : \Omega \to Q$. This means that for every set $A \in \mathcal{A}(Q)$, one has for all $j$*

$$\mathcal{P}(X_j^{-1}(A)) = \mathcal{P}(X^{-1}(A)) \,.$$

**Assumption 2.** *The sequence of RVs $X_1, X_2, \cdots : \Omega \to Q$ are independent. This means that for any sequence of sets $A_1, A_2, \cdots \in \mathcal{A}(Q)$, and any subset $\{n_1, \ldots, n_k\} \subset \mathbb{N}$ for any $k \in \mathbb{N}$ one has*

$$\mathcal{P}\left(\bigcap_{j=1}^{k} X_{n_j}^{-1}(A_j)\right) = \prod_{j=1}^{k} \mathcal{P}(X_{n_j}^{-1}(A_j)) \,.$$

## 2 Generalized Central Limit Theorem

**Assumption 3** (Weak consistency)**.** *The set of sample descriptors $E_n$ is weakly consistent, that means*

$$\forall \varepsilon > 0 \quad \forall \delta > 0 \quad \exists n > 0 \quad \forall k \geq n : \mathbb{P}\left(\max_{\widehat{\mu}_k \in E_k} \min_{\mu \in E} d(\widehat{\mu}_k, \mu) > \delta\right) < \varepsilon$$

**Assumption 4** (Almost Surely Locally Lipschitz II)**.** *There is a measurable function $\dot{\rho} : E \times Q \to \mathbb{R}$ and a $K < \infty$ such that $\forall \mu \in E : \mathbb{E}[\dot{\rho}(\mu, X)^2] < K$ and a $\delta > 0$ such that for all $p_1, p_2 \in B_\delta(\mu)$ the following Lipschitz condition holds*

$$|\rho(p_1, X) - \rho(p_2, X)| \leq \dot{\rho}(\mu, X) d(p_1, p_2) \quad \text{a.s.}$$

This assumption corresponds to the assumption that the second moment of $X$ exists in the case of the Euclidean mean $P = Q = \mathbb{R}^m$ and $\rho(p, x) := |p - x|^2$.

While consistency can be shown under only weak assumptions on the population descriptor set, we assume a unique population descriptor for the CLT.

**Assumption 5.** *The population descriptor is unique $E = \{\mu\}$.*

In all of the following let $\mathcal{F}$ denote some class of functions from the data space $Q$ to $\mathbb{R}$.

**Definition 2.1.** *For a class of functions $\mathcal{F}$ from $Q$ to $\mathbb{R}$, an envelope function $C_\mathcal{F} : Q \to \mathbb{R}$ is any function, such that for any $f \in \mathcal{F}$ and $q \in Q$, $|f(q)| \leq C_\mathcal{F}(q)$.*

Here, we consider the function class $\mathcal{F} := \Big\{\rho(x,\cdot) - \rho(\mu,\cdot), x \in B_\delta(\mu)\Big\}$, which has the envelope functions $C_{\mathcal{F}} := 2\delta\dot{\rho}$ due to the local Lipschitz condition in Assumption 4.

The most important notion for the following is the bracketing entropy of a function class.

**Definition 2.2.** *For a class of functions $\mathcal{F}$ from $Q$ to $\mathbb{R}$, a norm $\|\,\|$ on $\mathcal{F}$ and $\varepsilon > 0$, consider two functions $f_1, f_2 \in \mathcal{F}$ such that $\|f_1 - f_2\| < \varepsilon$ and call*

$$[f_1, f_2](\varepsilon, \mathcal{F}, \|\,\|) := \{f \in \mathcal{F} : \forall_0 q \in Q \quad f_1(q) \le f(q) \le f_2(q)\}$$

*an $\varepsilon$ bracket. Then the bracketing number $N_{[]}(\varepsilon, \mathcal{F}, \|\,\|)$ is defined as the minimal number of $\varepsilon$ brackets needed such that their union is $\mathcal{F}$. If $\mathcal{F}$ has an envelope function $C_{\mathcal{F}}$, the bracketing entropy is defined as*

$$J_{[]}(\eta, \mathcal{F}, \|\,\|) := \int_0^\eta \sqrt{1 + \log N_{[]}(\varepsilon\|C_{\mathcal{F}}\|, \mathcal{F}, \|\,\|)}\, d\varepsilon$$

The bracketing entropy for a function class is difficult to determine. However, there is a similar concept for general metric spaces, namely the covering entropy, which is usually much simpler to determine.

**Definition 2.3.** *For a totally bounded metric space $(P, d)$ and a size $\varepsilon$ the covering number $N(\varepsilon, P, d)$ is defined as the minimal number of $\varepsilon$ balls needed to cover $P$.*

**Definition 2.4.** *The covering entropy is defined as*

$$J(\eta, P, d) := \int_0^\eta \sqrt{1 + \log N(\varepsilon, P, d)}\, d\varepsilon$$

For the next theorem we introduce the empirical process and two relevant norms.

**Definition 2.5.** *Let $Q$ be the data space and $\mathcal{P}_X = \mathcal{P} \circ X^{-1}$ the distribution measure in $Q$ corresponding to the random variable $X$.*

*a) For an i.i.d. sample $X_1, \ldots, X_n$ from $\mathcal{P}_X$ let $\mathcal{P}_n := \frac{1}{n}\sum_{j=1}^n \delta_{X_j}$ be the empirical measure.*

*b) For any measurable function $f : Q \to \mathbb{R}$ let $\mathbb{G}_n(f) := \sqrt{n}\big(\mathcal{P}_n f - \mathcal{P}_X f\big)$ be the empirical process.*

*c) For any measurable function $f : Q \to \mathbb{R}$ let $\|f\|_{\mathcal{P}_X, r} := \left(\int |f|^r d\mathcal{P}_X\right)^{\frac{1}{r}}$ be the $L_r(\mathcal{P}_X)$ norm. For any class $\mathcal{F}$ of measurable functions let $\|\mathbb{G}_n\|_{\mathcal{F}} := \sup_{f\in\mathcal{F}} |\mathbb{G}_n(f)|$.*

*d) For any $\mathbb{R}$-valued RV $Z$ The outer expectation is defined as*

$$\mathbb{E}^*[Z] := \inf\left\{\mathbb{E}[Y] : Y \ge Z, Y : \Omega \to \overline{R} \text{ measurable and } \mathbb{E}[Y] \text{ exists}\right\}$$

**Theorem 2.6** (van der Vaart and Wellner (1996), Theorem 2.14.2). *Under Assumptions 1 and 2 for any class $\mathcal{F}$ of measurable functions with measurable envelope function $C_{\mathcal{F}}$,*

$$\mathbb{E}^* \left[ \|\mathbb{G}_n\|_{\mathcal{F}} \right] \lesssim J_{[]}(1, \mathcal{F}, L_2(\mathcal{P}_X)) \|C_{\mathcal{F}}\|_{\mathcal{P}_X,2} \,.$$

The local Lipschitz condition in Assumption 4 connects the bracketing number of a function class indexed by a totally bounded metric space to the covering number of this metric space. Consider the function class $\mathcal{F} := \left\{ \rho(x, \cdot) - \rho(\mu, \cdot), x \in B_\delta(\mu) \right\}$, which has the envelope functions $2\delta\dot{\rho}$, then the connection is as follows.

**Theorem 2.7** (van der Vaart and Wellner (1996), Theorem 2.7.11). *Under Assumptions 1, 2, 3, and 4, we have for any norm $\| \, \|$ on $\mathcal{F}$*

$$N_{[]}(2\varepsilon\|\dot{\rho}\|, \mathcal{F}, \| \, \|) \leq N(\varepsilon, B_\delta(\mu), d) \,.$$

Now, we can introduce a bound on the covering entropy to get a bound on the empirical process.

**Assumption 6** (Entropy bound). *For the class $\mathcal{F}$ the following entropy bound holds*

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

This entropy bound is fairly general as discussed in Dubey and Müller (2019). The following Lemma shows how it translates to a bound on the empirical process.

**Lemma 2.8.** *Under Assumptions 1, 2, 3, 4, and 5 the following bound holds for the empirical process*

$$\begin{aligned} \mathbb{E}^* \left[ n^{1/2} \sup_{x \in B_\delta(\mu)} \left| F_n(x) - F(x) - F_n(\mu) + F(\mu) \right| \right] &= \mathbb{E}^* \left[ \sup_{x \in B_\delta(\mu)} \|\mathbb{G}_n(\rho(x, \cdot)) - \mathbb{G}_n(\rho(\mu, \cdot))\| \right] \\ &= \mathbb{E}^* \left[ \|\mathbb{G}_n\|_{\mathcal{F}} \right] \lesssim 2\delta \|\dot{\rho}\|_{\mathcal{P}_X,2} \int_0^1 \sqrt{1 + \log N(\varepsilon\delta, B_\delta(\mu), d)} \, d\varepsilon \,. \end{aligned}$$

*Under Assumption 6 this implies* $\lim_{\delta \to 0} \mathbb{E}^* \left[ n^{1/2} \sup_{x \in B_\delta(\mu)} \left| F_n(x) - F(x) - F_n(\mu) + F(\mu) \right| \right] = o_P(1)$.

*Proof.*

$$
\begin{aligned}
\mathbb{E}^*\left[\|\mathbb{G}_n\|_{\mathcal{F}^i}\right] &\lesssim 2\delta\|\dot\rho^i\|_{\mathcal{P}_X,2} J_{[]}(1,\mathcal{F}^i, L_2(\mathcal{P}_X)) \\
&= 2\delta\|\dot\rho^i\|_{\mathcal{P}_X,2} \int_0^1 \sqrt{1+\log N_{[]}(2\varepsilon\delta\|\dot\rho^i\|, \mathcal{F}_i, L_2(\mathcal{P}_X))}\, d\varepsilon \\
&\le 2\delta\|\dot\rho^i\|_{\mathcal{P}_X,2} \int_0^1 \sqrt{1+\log N(\varepsilon\delta, B_\delta(\mu^i), d_i)}\, d\varepsilon\,.
\end{aligned}
$$

The limit follows from Assumption 6. □

Based on these preliminary results, we can state a limited CLT for m-estimators on general metric spaces.

**Lemma 2.9** (van der Vaart (2000) Lemma 5.52)**.** *Under Assumptions 1, 2, 3, 4, and 5 and assuming that for fixed constants $C$ and $\alpha > \beta$ for every $n$ and for every sufficiently small $\delta > 0$*

$$
\sup_{x\in B_\delta(\mu)\setminus B_{\delta/2}(\mu)} |F(x) - F(\mu)| \ge C\delta^\alpha\,, \tag{1}
$$

$$
\mathbb{E}^*\left[n^{1/2} \sup_{x\in B_\delta(\mu)} \left|F_n(x) - F(x) - F_n(\mu) + F(\mu)\right|\right] \le C\delta^\beta\,, \tag{2}
$$

*then any a random sequence $P \ni y_n \xrightarrow{\mathcal{P}} \mu$ that satisfies $F_n(y_n) \le F_n(\mu) + o_P(1)$ also satisfies the asymptotic rate $n^{1/(2\alpha-2\beta)} d(y_n,\mu) = \mathcal{O}_P(1)$.*

In consequence of this lemma, one can get an asymptotic convergence result of the sample descriptor to the population descriptor. However, while a convergence rate can be shown, the limiting distribution cannot be easily determined, since a normal distribution can only be defined on a linear space and a generic metric space cannot be canonically mapped to a linear space. To achieve a fully fledged central limit theorem, we therefore restrict attention to the case that $P$ is a Riemannian manifold.

**Notation 2.10.** *In the case that $P$ is a $p$-dimensional Riemannian manifold and Assumption 5 holds, we denote for any $x \in T_\mu P$, $q \in Q$*

$$
\begin{aligned}
&\tau : (x,q) \mapsto \rho(\exp_\mu(x), q)\,, \\
&G : x \mapsto F(\exp_\mu(x))\,, \qquad\qquad G_n : x \mapsto F_n(\exp_\mu(x))\,.
\end{aligned}
$$

**Assumption 4a** (Almost Surely Locally Lipschitz II on manifolds)**.** *Let Assumption 5 hold and use Notation 2.10.*

*(i) The function $\tau$ is differentiable at $x = 0$ for $\mathcal{P}_X$-almost every $q \in Q$ and its derivative vector is denoted as $\dot{\tau}_0(q)$.*

*(ii) There is a measurable function $\dot{\tau} : Q \to \mathbb{R}$ satisfying $\mathbb{E}[\dot{\tau}(X)^2] < \infty$ and a $\delta > 0$ such that for all $x_1, x_2 \in B_\delta(0) \subset T_\mu P$ the following Lipschitz condition holds*

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

First, we note that we can forego Assumption 6 due to the following.

**Lemma 2.11.** *Under Assumptions 1, 2, 4a, and 5 and using notation 2.10 we have for small enough $\delta > 0$, with a dimension dependent constant $K_p$,*

$$\int_0^1 \sqrt{1 + \log N(\varepsilon\delta, B_\delta(0), \|\,\|)}\, d\varepsilon \leq \int_0^1 \sqrt{1 + \log K_p - p\log(\varepsilon)}\, d\varepsilon$$

*for any norm $\|\,\|$ on $\mathcal{F} := \Big\{\tau(y, \cdot) - \tau(0, \cdot), y \in B_\delta(0)\Big\}$*

**Lemma 2.12.** *Under Assumptions 1, 2, 4a, and 5 and using notation 2.10 we have, with a dimension dependent constant $K_p$,*

$$J_{[]}(\eta, \mathcal{F}, \|\,\|) \leq \int_0^\eta \sqrt{1 + \log K_p - p\log(\varepsilon)}\, d\varepsilon$$

*for any norm $\|\,\|$ on $\mathcal{F} := \Big\{\tau(y, \cdot) - \tau(0, \cdot), y \in B_\delta(0)\Big\}$*

*Proof.* Note that $2\delta\|\dot{\tau}\|$ is an envelope function of $\mathcal{F}$, $\|\,\|$ is a norm and $\dot{\tau}^i$ takes the role of the Lipschitz "constant" in Theorem 2.7. This yields

$$J_{[]}(\eta, \mathcal{F}, \|\,\|) = \int_0^\eta \sqrt{1 + \log N_{[]}(2\varepsilon\delta\|\dot{\tau}\|, \mathcal{F}, \|\,\|)}\, d\varepsilon \leq \int_0^\eta \sqrt{1 + \log N(\varepsilon\delta, B_\delta(0), \|\,\|)}\, d\varepsilon\,.$$

Next, we see that

$$N(\kappa, B_\delta(0), \|\,\|) = K_p\left(\frac{\delta}{\kappa}\right)^p = K_p\left(\frac{\delta}{\kappa}\right)^p$$

with $K_p$ being a dimension dependent constant and thus

$$J_{[]}(\eta, \mathcal{F}, \|\,\|) \leq \int_0^\eta \sqrt{1 + \log K_p - p\log(\varepsilon)}\, d\varepsilon\,.$$

□

Now, we can prove a bound on the empirical process.

**Theorem 2.13.** *Under Assumptions 1, 2, 4a, and 5 and using notation 2.10 there is a constant $C$ such that*

$$\mathbb{E}\left[\sup_{\|y\|<\delta} |\mathbb{G}_n(\tau(y,X)-\tau(0,X))|\right] \leq C\delta$$

*Proof.* The claim follows by combining Theorem 2.6 with the envelope function $2\delta\|\dot{\tau}\|$ and Lemma 2.11 since

$$J_{[]}(1,\mathcal{F}_i,L_2(\mathcal{P}_X)) \leq \int\limits_0^1 \sqrt{1+\log K_p - p\log(\varepsilon)}\, d\varepsilon$$

is independent of $\delta$ and therefore only contributes a constant. $\square$

This theorem implies that the second inequality (2) in Lemma 2.9 holds with $\beta = 1$.

**Assumption 7** (Smooth Fréchet Function)**.** *Using notation 2.10, a rotation matrix $R \in SO(p)$ and $T_1,\ldots,T_p \in \mathbb{R}^{>0}$, assume that the Fréchet function admits the power series expansion*

$$G(x) = G(0) + \sum_{j=1}^{p} T_j|(Rx)_j|^r + o(\|x\|^r)\,. \tag{3}$$

Note that for $r = 2$ Equation (3) encompasses every possible covariance whereas for $r > 2$ this represents a restriction to easily tractable tensors.

**Notation 2.14.** *We denote $\widehat{\nu}_n := \log_\mu(\widehat{\mu}_n)$.*

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 $r \geq 2$, $\sqrt{n}\widehat{\nu}_n = o_p(1)$.

**Corollary 2.15.** *Under Assumptions 1, 2, 3, 4a, 5, and 7*

$$n^{1/(2r-2)}\widehat{\nu}_n = \mathcal{O}_P(1)\,.$$

*Proof.* By Assumptions 3 and 5, $\widehat{\nu}_n \xrightarrow{\mathcal{P}} 0$ with $G_n(\widehat{\nu}_n) \leq G_n(0)$. Hence, Lemma 2.9 yields the assertion, because for $\alpha = r$, (1) follows at once from (3), and under Assumption 4a, (2), for $\beta = 1$ follows word by word from the proof of van der Vaart (2000, Corollary 5.53). $\square$

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 7 the above bound gives the exact scaling rate, including the explicit limiting distribution.

**Theorem 2.16** (General CLT for Generalized Fréchet Means, see Eltzner and Huckemann (2019) Theorem 2.11)**.** *Under Assumptions 1, 2, 3, 4a, 5, and 7 and using Notation 2.10 we have*

$$n^{1/2}\left((R\widehat{\nu}_n)_1|(R\widehat{\nu}_n)_1|^{r-2},\dots,(R\widehat{\nu}_n)_p|(R\widehat{\nu}_n)_p|^{r-2}\right)^T \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{r^2}T^{-1}\mathrm{Cov}[\nabla_x\tau(x,X)|_{x=0}]T^{-1}\right)$$

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

$$n^{\frac{1}{2r-2}}(R\widehat{\nu}_n)_j \xrightarrow{\mathcal{D}} \mathcal{H}_j$$

*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.*

*Sketch of Proof.* We can follow the arguments in the proof of van der Vaart (2000, Lemma 5.23) to show that $R\widehat{\nu}_n n^{\frac{1}{2r-2}}$ is a minimizer of the function

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

where $\Gamma_n := n^{1/2}\left(\frac{1}{n}\sum_{j=1}^n \dot\tau_0(X_j) - \mathbb{E}\left[\dot\tau_0(X)\right]\right)$. This function has a unique minimizer

$$rT_j\,\mathrm{sign}((w_n)_j)|(w_n)_j|^{r-1} = -(R\Gamma_n)_j \quad \text{ i.e. } \quad (w_n)_j|(w_n)_j|^{r-2} = -\frac{(R\Gamma_n)_j}{rT_j}\,,$$

yielding

$$\sqrt{n}(R\widehat{\nu}_n)_j|(R\widehat{\nu}_n)_j|^{r-2} = -\frac{(R\Gamma_n)_j}{rT_j} + o_P(1)\,.$$

Now the classical CLT for $\dot\tau_0(X_j)$ gives the first assertion. The second assertion follows with a simple calculation. □

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

$$\tau_n(z,X) := n^s(\tau(zn^{-s},X) - \tau(0,X)) - z^T\dot\tau_0(X)$$
$$\Gamma_n := n^{1/2}\left(\frac{1}{n}\sum_{j=1}^n \dot\tau_0(X_j) - \mathbb{E}\left[\dot\tau_0(X)\right]\right)\,,$$

where we set $\tau(zn^{-s},X) = \tau(0,X)$ if $zn^{-s} \notin U$. Then, due to Assumptions 4a (i) and 7, and $1/2 + s - sr = 0$,

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

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 (2000, Lemma 19.31), due to Assumption 4a, $z$ 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,

$$n^{1/2+s}\Big(G_n(z_n n^{-s}) - G_n(0)\Big) = \sum_{j=1}^{m} T_j |(Rz_n)_j|^r + z_n^T \Gamma_n + o_P(1)\,. \tag{4}$$

By Corollary 2.15, $z_n = \widehat{\nu}_n n^s$ is a valid choice in equation (4). Comparison with any other $z_n = O_P(1)$, using that $\widehat{\nu}_n$ is a minimizer for $G_n$, reveals,

$$n^{1/2+s}\Big(G_n(\widehat{\nu}_n) - G_n(0)\Big) \le n^{1/2+s}\Big(G_n(z_n n^{-s}) - G_n(0)\Big)\,.$$

This asserts that $R\widehat{\nu}_n n^s$ is a minimizer, of the right hand side of (4), i.e. of

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

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

$$rT_j \operatorname{sign}((w_n)_j)|(w_n)_j|^{r-1} = -(R\Gamma_n)_j \quad \text{i.e.} \quad (w_n)_j|(w_n)_j|^{r-2} = -\frac{(R\Gamma_n)_j}{rT_j}\,,$$

yielding

$$\sqrt{n}(R\widehat{\nu}_n)_j|(R\widehat{\nu}_n)_j|^{r-2} = -\frac{(R\Gamma_n)_j}{rT_j} + o_P(1)\,.$$

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

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

□

If one assumes $r = 2$ in Assumption 7, one gets a generalized version of the classical CLT. However, the theorem presented here goes beyond the classical CLT and encompasses also certain cases of slower rates of convergence. This leads to the concept of smeariness, which we will discuss in a later lecture.

Until now, the CLT on manifolds is formulated only in the exponential chart at the population descriptor. While this is typically the most useful chart, it is desirable to formulate a result which is independent of the specific chart.

**Theorem 2.17** (see Eltzner and Huckemann (2019) Lemma 3.2)**.** *Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence random variables in* $\mathbb{R}^m$ *and* $X_n \xrightarrow{\mathcal{P}} 0$*. Consider a local diffeomorphism* $\Phi : U \to V$ *preserving the origin* $0 \in U, V$ *open* $\subset \mathbb{R}^m$*, define the random variables* $Y_n = \Phi(X_n)$ *and let* $s \in (0, 1/2]$*. Then*

$$n^s X_n = \mathcal{O}_P(1) \quad \Leftrightarrow \quad n^s Y_n = \mathcal{O}_P(1)\,.$$

*In particular, if the random variable* $X$ *satisfies* $n^s X_n \xrightarrow{\mathcal{D}} X$*, then* $n^s Y_n \xrightarrow{\mathcal{D}} Y := D\Phi(0)X$*. Here* $D\Phi(x)$ *denotes the differential of* $\Phi$ *at* $x \in U$ *and* $\det\big(D\Phi(0)\big) \neq 0$ *due to invertibility of* $\Phi$*.*

*Proof.* In all of the following, $B \subset \mathbb{R}^m$ denotes any Borel set. The implication "$\Rightarrow$" is a direct consequence of a Taylor expansion and the continuity theorem with a suitable point $\widetilde{X}_n \xrightarrow{\mathcal{P}} 0$ between the origin and $X_n$ as follows

$$\begin{aligned}\mathbb{P}\left(\{n^s Y_n \in B\} \cap \{X_n \in U\}\right) &= \mathbb{P}\left(\left\{n^s D\Phi(\widetilde{X}_n)X_n \in B\right\} \cap \{X_n \in U\}\right)\\ &\to \mathbb{P}\left(D\Phi(0)X \in B\right)\end{aligned}$$

because $\mathbb{P}(X_n \in U) \to 1$ due to $X_n \xrightarrow{\mathcal{P}} 0$. Similarly, the implication "$\Leftarrow$" follows.

Denote by $Y$ the random variable satisfying $n^s Y_n \xrightarrow{\mathcal{D}} Y$. Then

$$\begin{aligned}\mathbb{P}\left(n^s X_n \in B\right) = \mathbb{P}&\left(\left\{n^s D\Phi(\widetilde{X}_n)^{-1} Y_n \in B\right\} \cap \{X_n \in U\}\right)\\ &+ \mathbb{P}\left(\{n^s X_n \in B\} \cap \{X_n \notin U\}\right)\\ \to \mathbb{P}&\left(D\Phi(0)^{-1} Y \in B\right)\,,\end{aligned}$$

again due to the hypothesis $X_n \xrightarrow{\mathcal{P}} 0$. □

To achieve a CLT in any chart, one applies Theorem 2.17 to the sequence of random variables $(X_n)_{n\in\mathbb{N}} := (\widehat{\nu}_n)_{n\in\mathbb{N}}$. The asymptotic distribution remains a normal distribution, albeit with a transformed covariance matrix.

# 3 Positively Curved Spaces and Smeariness

## 3.1 Smeariness

On spaces of positive curvature, asymptotics of the Fréchet mean can be compromised in a different way.

**Definition 3.1** (Smeariness of Random Variables)**.** *Assume that there is* $\zeta > 0$ *such that for every* $x \in B_\zeta(0) \setminus \{0\} \subset T_\mu P$ *one has* $G(x) > G(0)$*. Suppose that for fixed constants* $C_X > 0$

*and $2 < \kappa \in \mathbb{R}$ and a linear subspace $\mathcal{V} \subseteq T_\mu P$ we have for every sufficiently small $\delta > 0$*

$$\sup_{x \in \mathcal{V},\, \delta/2 \le \|x\| < \delta} |G(x) - G(0)| \ge C_X \delta^\kappa \,.$$

*Then we say that the descriptor $\mu$ is (directionally) smeary on the linear subspace $\mathcal{V}$ and that the model $(Q, P, \rho)$ admits (directional) smeariness. If $\mathcal{V} = T_\mu P$, we simply say that the descriptor $\mu$ is smeary.*

**Remark 3.2.** *If Assumption 7 holds with $r > 2$, the generalized CLT, Theorem 2.16, leads to an asymptotic rate $n^{\frac{1}{2r-2}} \widehat{\nu}_n = \mathcal{O}_P(1)$ of the sample descriptor. In this case, we say the descriptor is smeary with rate $\frac{1}{2r-2}$.*

In spaces of non-positive curvature, the Féchet mean is never smeary, as shown by Afsari (2009). However, examples of random variables with smeary Fréchet means were found on the circle $S^1$, the torus $S^1 \times \cdots \times S^1$, Spheres $S^m$ and real projective spaces $\mathbb{R}P^m$.

## 3.2 Origins of Smeariness

**Definition 3.3.** *Consider a point in a Riemannian manifold $p \in Q$. The cut locus $\mathrm{Cut}(p)$ of $p$ is the closure of the set of all points $q \in Q$ such that there is more than one shortest geodesic from $p$ to $q$.*

In all the following, consider the Fréchet mean, which means $P = Q$ and $\rho = d_Q^2$ being the squared geodesic distance on $Q$.

**Notation 3.4.** *Let $\widetilde{P} = P \setminus \mathrm{Cut}(\mu)$ and $d_{\widetilde{P}}$ the geodesic distance on $\widetilde{P}$. Then we write*

$$\widetilde{\tau} : (x, q) \mapsto d_{\widetilde{P}}^2(\exp_\mu(x), q)\,, \qquad\qquad \widetilde{G} : x \mapsto \mathbb{E}[\widetilde{\tau}(x, X)]\,.$$

*Here, $d_{\widetilde{P}}^2(q, p) \ge d_P^2(q, p)$ is given by the infimum over the length of all curves in $\widetilde{P}$ connecting $q$ and $p$.*

**Definition 3.5** (Cut Locus Smeariness and Geometrical Smeariness Eltzner (2022))**.** *Assume that $Q$ admits smeariness and $X$ is a random variable with smeary Fréchet mean $\mu \in Q$ on the linear subspace $\mathcal{V} \subseteq T_\mu Q$. This means that Definition 3.1 is satisfied. The restriction of the Hessian matrix to $\mathcal{V}$ is denoted by $\mathrm{Hess}_\mathcal{V}$.*

*(i) If for every linear subspace neighborhood $U \subset \mathcal{V}$ of 0, there is an $x \in U$ such that $G(x) \neq \widetilde{G}(x)$ and $\mathrm{Hess}_\mathcal{V}(G - \widetilde{G})(0) < 0$, then the mean of $X$ is called cut locus smeary on the linear subspace $\mathcal{V}$ and we say that $Q$ admits cut locus smeariness.*

*(ii) If there is a linear subspace neighborhood $U \subset \mathcal{V}$ of 0, such that for every $x \in U$ one has $G(x) = \widetilde{G}(x)$ or if $\operatorname{Hess}_{\mathcal{V}}(G - \widetilde{G})(0) \geq 0$, the mean of $X$ is called geometrically smeary on the linear subspace $\mathcal{V}$ and we say that $Q$ admits geometrical smeariness.*

To motivate the term *cut locus smeariness*, first note that smeariness always implies that $\operatorname{Hess}_{\mathcal{V}} F(0) = 0$. On Euclidean space, the Hessian matrix of the Fréchet function is always positive definite. In reference to this, smeariness depends on a negative contribution to the Hessian that leads to a vanishing Hessian overall. The terms *cut locus smeariness* and *geometrical smeariness* point to the origin of this negative contribution to the Hessian. Note that one can write

$$0 = \operatorname{Hess}_{\mathcal{V}} G(0) = \underbrace{\operatorname{Hess}_{\mathcal{V}} \widetilde{G}(0)}_{>0} + \underbrace{\operatorname{Hess}_{\mathcal{V}}(G - \widetilde{G})(0)}_{<0}$$

to illustrate that cut locus smeariness crucially hinges on $G \neq \widetilde{G}$. Since the only difference between these two functions is that in $\widetilde{G}$ the geodesics crossing $\operatorname{Cut}(\mu)$ are excluded, the negative term $\operatorname{Hess}(G - \widetilde{G})(0)$ can be understood as the contribution of the cut locus.

Cut locus smeariness requires a cut locus of codimension 1 with respect to the smeary subspace. See Hotz et al. (2024) for a comprehensive treatment.

However, smeariness can also occur in more general settings. Eltzner and Huckemann (2019) present and example of smeariness on $S^m$ for $m \geq 2$. Moreover, Eltzner (2022) shows that no density is required at the cut locus for geometric smeariness of the mean, but the support of the measure must extend to more than a half sphere on $S^m$ with $m \geq 2$.

For cut locus smeariness, we consider the example of the Fréchet mean on the circle first treated by Hotz and Huckemann (2015).

**Theorem 3.6.** *Let $X$ be a random variable on the circle parametrized by $(-\pi, \pi]$ with continuous density function $f_X$ and with Fréchet mean $\mu = 0$. Then the following holds:*

*(i) $f_X(\pi) \leq \frac{1}{2\pi}$.*

*(ii) If $f_X(\pi) < \frac{1}{2\pi}$ then Assumption 7 holds with $r = 2$ and thus a standard CLT holds.*

*(iii) If there are constants $s > 0$, $d > 0$, and $C > 0$, such that for every $0 \leq y < d$ it holds $f_X(\pi - y) = f_X(-\pi + y) = \frac{1}{2\pi} - Cy^s$, then 7 holds with $r = s + 2$ and thus a the mean is smeary with rate $\frac{1}{2s+2}$.*

*Proof.* Consider without loss of generality an $x \in [0, \pi]$. Then, we can write the Fréchet loss function as

$$
\begin{aligned}
F(x)-F(0) &= \int_{-\pi+x}^{\pi}(x-y)^2 f_X(y)dy + \int_{-\pi}^{-\pi+x}(x-2\pi-y)^2 f_X(y)dy - \int_{-\pi}^{\pi} y^2 f_X(y)dy\\
&= \int_{-\pi+x}^{\pi}(-2xy+x^2) f_X(y)dy + \int_{-\pi}^{-\pi+x}(-4\pi x-2xy+4\pi y+4\pi^2+x^2) f_X(y)dy\\
&= x^2 - 2x\int_{-\pi}^{\pi} y f_X(y)dy + 4\pi\int_{-\pi}^{-\pi+x}(y-x+\pi) f_X(y)dy\\
&= x^2 + 4\pi\int_{-x}^{0} z f_X(z+x-\pi)dz
\end{aligned}
$$

Using this formula, we can now tackle the various cases.

(i) Assume that $f_X(\pi) > \frac{1}{2\pi}$, then there is some $\varepsilon > 0$ such that for all $x \in (-\pi, -\pi-\varepsilon]$ one has $f_X(x) > \frac{1}{2\pi}$. This leads to

$$
F(x)-F(0) < x^2 + 4\pi\int_{-x}^{0} z\frac{1}{2\pi}dz = 0
$$

for $x \in (0,\varepsilon]$, which is in contradiction to the assumption that the Fréchet mean is $\mu = 0$.

(ii) If $f_X(\pi) < \frac{1}{2\pi}$, then there is some $\varepsilon > 0$ such that for all $x \in (-\pi, -\pi-\varepsilon]$ one has $f_X(x) \leq \frac{\rho}{2\pi} < \frac{1}{2\pi}$. This leads to

$$
F(x)-F(0) \geq x^2 + 4\pi\int_{-x}^{0} z\frac{\rho}{2\pi}dz = (1-\rho)x^2
$$

for $x \in (0,\varepsilon]$. In fact, for $x \to 0$, we get $\rho \to 2\pi f_X(\pi) < 1$. Thus one gets a standard CLT.

(iii) Plugging in $f_X(-\pi+y) = \frac{1}{2\pi} + Cy^s$ we get

$$
\begin{aligned}
F(x)-F(0) = x^2 + 4\pi\int_{-x}^{0} z\left(\frac{1}{2\pi} - C(z+x)^s\right)dz &= -C4\pi\int_0^x (z-x)z^s dz\\
= -4\pi C\left(\frac{1}{s+2} - \frac{1}{s+1}\right)x^{s+2} &= \frac{4\pi C}{(s+2)(s+1)}x^{s+2}
\end{aligned}
$$

which leads to a smeary Fréchet mean and proves the claim.

□

For geometrical smeariness, consider the example on $S^m$ for $m \geq 2$ treated by Eltzner and Huckemann (2019).

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 $\mathbb{L}^m = \{q \in S^m : q_2 \leq 0\}$ with total mass $0 < \alpha < 1$ and assuming the *north pole* $\mu = (0,1,0,\ldots,0)^T$ with probability $1-\alpha$.

With the volume of $S^m$ given by

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

define

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

**Lemma 3.7.** *With the above notation, the Freéchet function $G$ has derivatives of any order for $x \in \log_\mu(U)$ with $\|x\| < \pi/2$. For $\alpha = 1/(1+\gamma_m)$ the north pole $\mu$ gives the unique intrinsic Fréchet mean with $\mathrm{Hess}\,|_{x=0}G(x) = 0$. Moreover, for any choice of $0 < \alpha < 1$,*

$$\partial_i\partial_k\partial_l G(x)|_{x=0} = 0$$

$$\partial_i\partial_k\partial_l\partial_s G(x)|_{x=0} = c_m\delta_{i,k,l,s}$$

*for every $1 \le i,k,l,s \le m$ 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 non-standard 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 $\mu$ 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

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

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

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

with the two "crescent" integrals

$$C_+(\delta) = \int\limits_{\Theta^{m-1}} d\theta\, g(\theta) \int\limits_{-\delta}^{0} d\phi\, \widetilde{\rho}(\mu,q)^2 = \int\limits_{\Theta^{m-1}} d\theta\, g(\theta) \int\limits_{0}^{\delta} \Big(\arccos\big(h(\theta)\,\sin\phi\big)\Big)^2 d\phi$$

$$C_-(\delta) = \int\limits_{\Theta^{m-1}} d\theta\, g(\theta) \int\limits_{\pi-\delta}^{\pi} d\phi\, \widetilde{\rho}(\mu,q)^2 = \int\limits_{\Theta^{m-1}} d\theta\, g(\theta) \int\limits_{0}^{\delta} \Big(\arccos\big(-h(\theta)\,\sin\phi\big)\Big)^2 d\phi$$

cf. Figure 1, because the spherical measure of $\mathbb{L}^m$ is $v_m/2$.

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

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

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

$$\widetilde{F}\circ\exp_\mu(x) = G(\delta) = G(0) - \frac{4\pi\alpha}{v_m}\int\limits_{\Theta^{m-1}} d\theta\, g(\theta)\int\limits_0^\delta \arcsin\big(h(\theta)\,\sin\phi\big)\,d\phi + \delta^2(1-\alpha) \qquad (5)$$

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

For the second assertion we use the Taylor expansion

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

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

$$\begin{aligned}\int\limits_{\Theta^{m-1}} g(\theta)\,h(\theta)^k\,d\theta &= \int_{\Theta^{m-1}} \prod_{j=1}^{m-1}\left(\cos^{m-j+k}\theta_j\,d\theta_j\right)\\ &= \int_{\Theta^{m+k-1}} \prod_{j=1}^{m+k-1}\left(\cos^{m+k-j}\theta_j\,d\theta_j\right) \Big/ \int_{\Theta^k}\prod_{j=1}^{k}\left(\cos^{j}\theta_j\,d\theta_j\right)\\ &= \frac{v_{m+k}}{v_{k+1}}\,, \end{aligned} \qquad (7)$$

to obtain, in conjunction with (5),

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

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 $\alpha = 1/(1+\gamma_m)$.

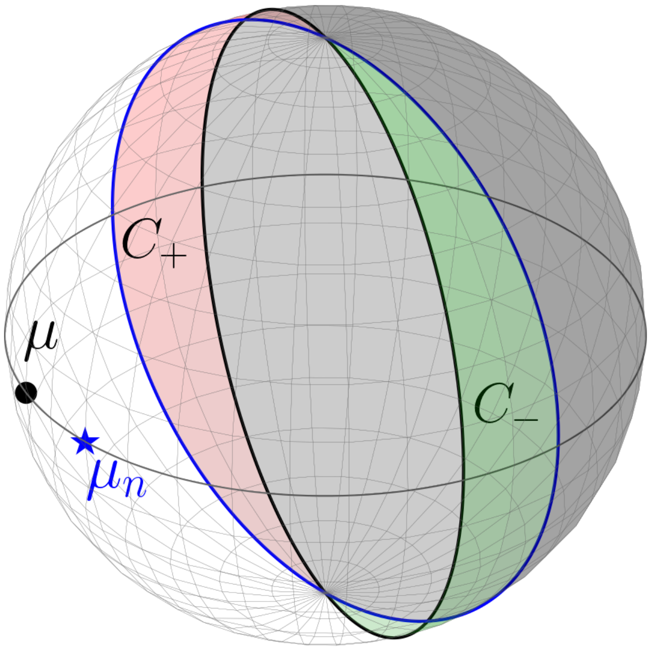



Figure 1: *Depicting the two crescents* $C_+ = C_+(\delta)$ *and* $C_- = C_-(\delta)$ *for* $\delta = \arccos\langle \mu, \mu_n \rangle$ *on* $\mathbb{S}^m$ *for* $m = 2$ *with north pole* $\mu$ *and nearby sample Fréchet mean* $\mu_n$ *.*

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

$$\begin{aligned} G'(\delta) &= -\frac{4\pi\alpha}{v_m} \int\limits_{\Theta^{m-1}} g(\theta) \arcsin\big(h(\theta) \sin\delta\big)\, d\theta + 2\delta(1-\alpha)\,, \\ G''(\delta) &= -\frac{4\pi\alpha}{v_m} \int\limits_{\Theta^{m-1}} g(\theta)\, h(\theta) \frac{\cos\delta}{\sqrt{1 - h(\theta)^2 \sin^2\delta}}\, d\theta + 2(1-\alpha) \\ &\geq -\frac{4\pi\alpha}{v_m} \int\limits_{\Theta^{m-1}} g(\theta)\, h(\theta)\, d\theta + 2(1-\alpha) = 2 - \alpha\left(2 + \frac{v_{m+1}}{v_m}\right) = 0\,, \end{aligned} \tag{8}$$

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 $p = \mu$. □

Since the leading non-trivial order of the Fréchet function is $|x|^4$, the asymptotic rate of the mean is $n^{-1/6}$.

### 3.3 Finite Sample Smeariness

We follow the treatment of Hundrieser et al. (2024).

Smeariness occurs only in finely tuned models. However, probability distributions which are "close" to distributions with a smeary mean, especially the empirical measures of samples from distributions with smeary mean, often exhibit finite sample smeariness of the mean.

**Definition 3.8.** *Using the data variance $V := F(\mu) = \mathbb{E}[d^2(\mu, X)]$ and the sample mean variance $V_n := \mathbb{E}[d^2(\mu, \widehat{\mu}_n)]$ define the variance modulation as*

$$\mathfrak{m}_n := \frac{nV_n}{V}\,.$$

*We say that $\mu$ is*

(i) *Euclidean if $\mathfrak{m}_n = 1$ for all $n \in \mathbb{N}$,*

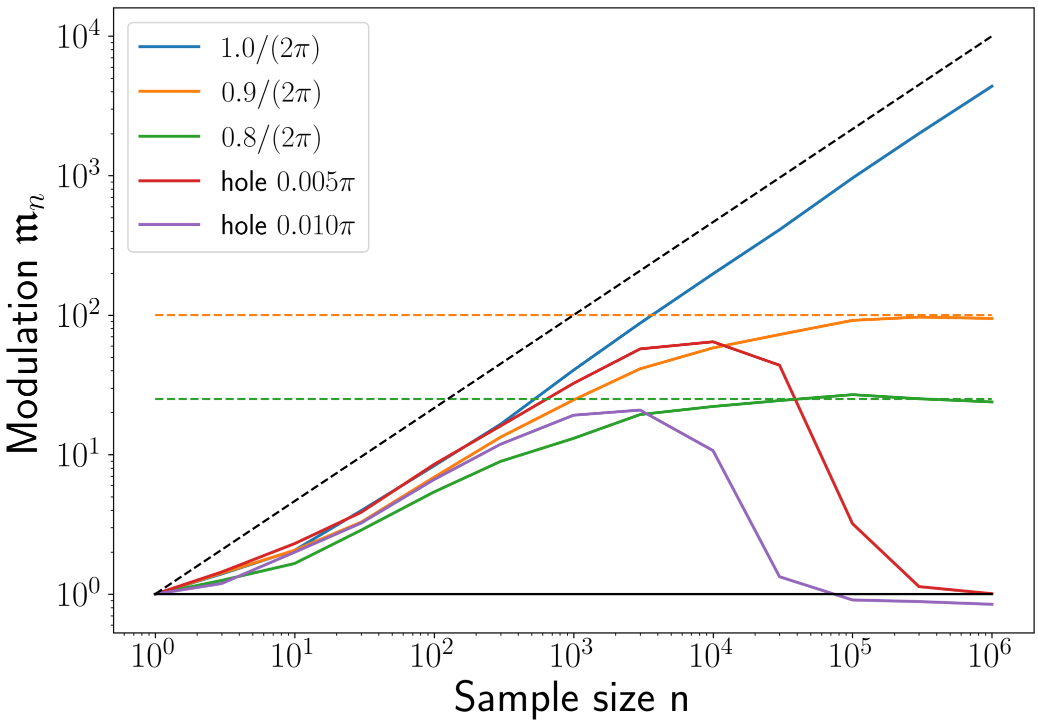
(ii) *finite sample (FS) smeary if $1 < \sup_{n\in\mathbb{N}} \mathfrak{m}_n < \infty$,*

  ($ii_1$) *Type I if $\lim_{n\to\infty} \mathfrak{m}_n > 1$,*

  ($ii_2$) *Type II if $\lim_{n\to\infty} \mathfrak{m}_n = 1$,*

(iii) *(asymptotically) smeary if $\sup_{n\in\mathbb{N}} \mathfrak{m}_n = \infty$.*

In Theorem 3.6 (ii) if $\rho > 0$, one has a Type I FS smeary mean. Samples from a distribution with smeary mean, have a type II FS smeary mean, if they have a FS smeary mean, because the density at the antipodal point of the mean is 0.




In Hundrieser et al. (2024), we propose a hypothesis test for finite sample smeariness and show that it asymptotically has the correct size using asymptotic theory for the Fréchet variance.

**Hypothesis Test 3.9** (Bootstrap test for Finite Sample Smeariness)**.**

*Setting: Fréchet mean setting with data space $Q$, descriptor space $P = Q$ and loss function $\rho = d^2$. I.i.d. sample $X_1, \ldots, X_{n_X} \in Q$ from unknown distribution such that Assumptions*

*1, 2, 3, 4a, 5, and 7 are satisfied. Let $\mu_b^*$ be a Fréchet sample mean based on an n-out-of-n bootstrap sample of $X_1, \dots, X_n$ for $b = 1, \dots, B$ and $\mu^*$ the Fréchet sample mean of the $\mu_b^*$ $(b = 1, \dots, B)$. We use notation*

$$\widehat{V}_n := \frac{1}{n} \sum_{j=1}^{n} d^2(\widehat{\mu}_n, X_j) = F_n(\widehat{\mu}_n)\,, \qquad \widehat{V}^*_{n,B} := \frac{1}{B} \sum_{b=1}^{B} d^2(\widehat{\mu}^*, \mu_b^*)\,,$$

$$\widehat{W}^*_{n,B} := \frac{1}{B} \sum_{b=1}^{B} d^4(\widehat{\mu}^*, \mu_b^*)^4\,.$$

*Null Hypothesis:* $H_{0,n} \colon \mathfrak{m}_n \leq 1$.

*Test Statistic:* *Using the notation introduced above, define*

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

*Rejection Criterion:* *Using the notation introduced above and the standard normal $(1-\alpha)$-quantile $\phi_{1-\alpha}$ for $\alpha \in (0,1)$, reject the hypothesis $H_{0,n}$ at nominal level $\alpha$ if*

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

# References


Afsari, B. (2009). *Means and averaging on Riemannian manifolds.* University of Maryland.

Dubey, P. and H.-G. Müller (2019). Fréchet analysis of variance for random objects. *Biometrika 106*(4), 803–821.

Eltzner, B. (2022). Geometrical Smeariness – A new Phenomenon of Fréchet Means. *Bernoulli 28*(1), 239–254.

Eltzner, B. and S. F. Huckemann (2019). A smeary central limit theorem for manifolds with application to high dimensional spheres. *Annals of Statistics 47*(6), 3360–3381.

Hotz, T. and S. F. Huckemann (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics 67*(1), 177–193.

Hotz, T., H. Le, and A. T. A. Wood (2024). Central limit theorem for intrinsic fréchet means in smooth compact riemannian manifolds. *Probability Theory and Related Fields 189*(3), 1219–1246.

Hundrieser, S., B. Eltzner, and S. Huckemann (2024). Finite sample smeariness of Fréchet means with application to climate. *Electronic Journal of Statistics 18*(2), 3274 – 3309.

van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.