A Lower Bound for Estimating Fréchet Means In Memory of Laurent Cavalier († 2014)

CIRM Workshop: Geometric Sciences in Action: from Geometric Statistics to Shape Analysis at Luminy (May 27 - 31, 2024)

Stephan F. Huckemann
joint work with Benjamin Eltzner and **Shayan Hundrieser**

University of Göttingen,
Felix Bernstein Institute for Mathematical Statistics in the Biosciences

May 27, 2024

supported by the

Niedersachsen Vorab of the Volkswagen Foundation

and DFG SFB 1456, DFG HU 1575-7, DFG GK 2088

Fréchet (1948) Means

Ingredients:

- metric space $(\mathfrak{X}, d)$, $\rho : \mathfrak{X} \times \mathfrak{X} \to \mathbb{R}$ continuous,
- Borel m'ble random variables $X_1, \ldots, X_n \overset{i.i.d.}{\sim} X \sim \mathbb{P}$ on $\mathfrak{X}$

$\rightsquigarrow$ Fréchet ρ-mean sets

$$\begin{aligned} \text{population} \quad \mathbb{M}(\mathbb{P}) &:= \operatorname{argmin}_{x \in \mathfrak{X}} \mathbb{E}[\rho(x, X)]\,, \\ \text{sample} \quad \mathbb{M}_n &:= \operatorname{argmin}_{x \in \mathfrak{X}} \frac{1}{n} \sum_{j=1}^{n} \rho(x, X_i)\,. \end{aligned}$$

Fréchet (1948) Means

Ingredients:

- metric space $(\mathfrak{X}, d)$, $\rho : \mathfrak{X} \times \mathfrak{X} \to \mathbb{R}$ continuous,
- Borel m'ble random variables $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} X \sim \mathbb{P}$ on $\mathfrak{X}$

$\rightsquigarrow$ Fréchet ρ-mean sets

$$\begin{aligned} \text{population} \quad \mathbb{M}(\mathbb{P}) &:= \operatorname{argmin}_{x \in \mathfrak{X}} \mathbb{E}[\rho(x, X)]\,, \\ \text{sample} \quad \mathbb{M}_n &:= \operatorname{argmin}_{x \in \mathfrak{X}} \frac{1}{n} \sum_{j=1}^{n} \rho(x, X_i)\,. \end{aligned}$$

Existence for $\rho = d^2$:

- $\mathbb{M}_n \neq \emptyset$ a.s. if $(\mathfrak{X}, d)$ complete,
- $\mathbb{M}(\mathbb{P}) \neq \emptyset$ if additionally, $\exists x \in \mathfrak{X}$ with $\mathbb{E}[d(x, X)] < \infty$.

Fréchet (1948) Means

Ingredients:

- metric space $(\mathfrak{X}, d)$, $\rho : \mathfrak{X} \times \mathfrak{X} \to \mathbb{R}$ continuous,
- Borel m'ble random variables $X_1, \ldots, X_n \overset{i.i.d.}{\sim} X \sim \mathbb{P}$ on $\mathfrak{X}$

$\rightsquigarrow$ Fréchet ρ-mean sets

$$\text{population} \quad \mathbb{M}(\mathbb{P}) := \operatorname{argmin}_{x \in \mathfrak{X}} \mathbb{E}[\rho(x, X)]\,,$$

$$\text{sample} \quad \mathbb{M}_n := \operatorname{argmin}_{x \in \mathfrak{X}} \frac{1}{n} \sum_{j=1}^{n} \rho(x, X_i)\,.$$

Existence for $\rho = d^2$:

- $\mathbb{M}_n \neq \emptyset$ a.s. if $(\mathfrak{X}, d)$ complete,
- $\mathbb{M}(\mathbb{P}) \neq \emptyset$ if additionally, $\exists x \in \mathfrak{X}$ with $\mathbb{E}[d(x, X)] < \infty$.

Uniqueness:

- $|\,\mathbb{M}(\mathbb{P})| = 1$ if additionally $\operatorname{supp}(X)$ in ball, smaller than a goedesic half ball (Karcher, 1977; Kendall, 1990; Le, 2001; Groisser, 2005; Afsari, 2011).
- $|\,\mathbb{M}_n\,| = 1$ a.s. if additionally $\mathfrak{X}$ is a Riemannain manifold and X is absolutely continuous w.r.t. Riemannian volume (Arnaudon and Miclo, 2014).

Strong Law of Large Numbers

Kuratowski (1948) convergence of sets $C_n \subseteq \mathfrak{X}$, $n \in \mathbb{N}$:

$$\begin{aligned}
\underset{n\to\infty}{Ls}\, C_n &:= \{x \in Q : \liminf_{n\to\infty} d(x, C_n) = 0\} = \bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} C_k} \\
\underset{n\to\infty}{Li}\, C_n &:= \{x \in Q : \limsup_{n\to\infty} d(x, C_n) = 0\}
\end{aligned}$$

superior (or outer) limit and inner limit, $Li_{n\to\infty} C_n \subseteq Ls_{n\to\infty} C_n$

Strong Law of Large Numbers

Kuratowski (1948) convergence of sets $C_n \subseteq \mathfrak{X}$, $n \in \mathbb{N}$:

$$\begin{aligned} \underset{n\to\infty}{Ls}\, C_n &:= \{x \in Q : \liminf_{n\to\infty} d(x, C_n) = 0\} = \bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} C_k} \\ \underset{n\to\infty}{Li}\, C_n &:= \{x \in Q : \limsup_{n\to\infty} d(x, C_n) = 0\} \end{aligned}$$

superior (or outer) limit and inner limit, $Li_{n\to\infty} C_n \subseteq Ls_{n\to\infty} C_n$

Theorem (Ziezold (1977); Bhattacharya and Patrangenaru (2003))

If $\exists \in \mathfrak{X}$ *with* $\mathbb{E}[d(x, X)] < \infty$, $\mathfrak{X}$ *separable, then*

(ZC) $Ls_{n\to\infty} \mathbb{M}_n \subseteq \mathbb{M}(\mathbb{P})$ *a.s.,*

(BPC) *if* $\mathfrak{X}$ *is Heine-Borel and* $\mathbb{M}(\mathbb{P}) \neq \emptyset$ *then* $\mathbb{M}(\mathbb{P}) = Li_{n\to\infty} \mathbb{M}_n$ *a.s.*

Strong Law of Large Numbers

Kuratowski (1948) convergence of sets $C_n \subseteq \mathfrak{X}$, $n \in \mathbb{N}$:

$$\underset{n\to\infty}{Ls}\, C_n := \{x \in Q : \liminf_{n\to\infty} d(x, C_n) = 0\} = \bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} C_k}$$

$$\underset{n\to\infty}{Li}\, C_n := \{x \in Q : \limsup_{n\to\infty} d(x, C_n) = 0\}$$

superior (or outer) limit and inner limit, $Li_{n\to\infty} C_n \subseteq Ls_{n\to\infty} C_n$

Theorem (Ziezold (1977); Bhattacharya and Patrangenaru (2003))

If $\exists \in \mathfrak{X}$ with $\mathbb{E}[d(x, X)] < \infty$, $\mathfrak{X}$ separable, then

(ZC) $Ls_{n\to\infty} \mathbb{M}_n \subseteq \mathbb{M}(\mathbb{P})$ *a.s.,*

(BPC) *if $\mathfrak{X}$ is Heine-Borel and $\mathbb{M}(\mathbb{P}) \neq \emptyset$ then $\mathbb{M}(\mathbb{P}) = Li_{n\to\infty} \mathbb{M}_n$ a.s.*

Extensions to generalized Fréchet means by H. (2011); e.g.

$$\text{argmin}_{\gamma\in\Gamma} \frac{1}{n}\sum_{j=1}^{n} d(\gamma, X_j)^2 \to \text{argmin}_{\gamma\in\Gamma} \mathbb{E}[d(\gamma, X)^2]$$

Strong Law of Large Numbers

Kuratowski (1948) convergence of sets $C_n \subseteq \mathfrak{X}$, $n \in \mathbb{N}$:

$$\underset{n\to\infty}{Ls}\, C_n := \{x \in Q : \liminf_{n\to\infty} d(x, C_n) = 0\} = \bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} C_k}$$

$$\underset{n\to\infty}{Li}\, C_n := \{x \in Q : \limsup_{n\to\infty} d(x, C_n) = 0\}$$

superior (or outer) limit and inner limit, $Li_{n\to\infty} C_n \subseteq Ls_{n\to\infty} C_n$

Theorem (Ziezold (1977); Bhattacharya and Patrangenaru (2003))

If $\exists \in \mathfrak{X}$ with $\mathbb{E}[d(x, X)] < \infty$, $\mathfrak{X}$ separable, then

(ZC) $Ls_{n\to\infty} \mathbb{M}_n \subseteq \mathbb{M}(\mathbb{P})$ *a.s.,*

(BPC) *if $\mathfrak{X}$ is Heine-Borel and $\mathbb{M}(\mathbb{P}) \neq \emptyset$ then $\mathbb{M}(\mathbb{P}) = Li_{n\to\infty} \mathbb{M}_n$ a.s.*

Extensions to generalized Fréchet means by H. (2011); e.g.

$$\operatorname{argmin}_{\gamma\in\Gamma} \frac{1}{n}\sum_{j=1}^{n} d(\gamma, X_j)^2 \to \operatorname{argmin}_{\gamma\in\Gamma} \mathbb{E}[d(\gamma, X)^2]$$

Recent extensions by Schötz (2022); Wiechers et al. (2023); Evans and Jaffe (2024): (ZC) inclusion can be strict.

On Complete CAT(0) (Hadamard) Spaces

1. $|\mathbb{M}(\mathbb{P})| = 1 = |\mathbb{M}_n|$ a.s. (Sturm, 2003)

On Complete CAT(0) (Hadamard) Spaces

1. $|\mathbb{M}(\mathbb{P})| = 1 = |\mathbb{M}_n|$ a.s. (Sturm, 2003)
2. $\exists$ involved CLT by Mattingly et al. (2023),
   - see Ezra's talk (Thu afternoon)

On Complete CAT(0) (Hadamard) Spaces

1. $|\mathbb{M}(\mathbb{P})| = 1 = |\mathbb{M}_n|$ a.s. (Sturm, 2003)
2. $\exists$ involved CLT by Mattingly et al. (2023),
   - see Ezra's talk (Thu afternoon)
3. $\exists$ phenomenon of stickiness,
   - is a curse for statistics, e.g. of phylogenetic trees (e.g. Billera et al. (2001), Barden et al. (2018)),

On Complete CAT(0) (Hadamard) Spaces

1. $|\mathbb{M}(\mathbb{P})| = 1 = |\mathbb{M}_n|$ a.s. (Sturm, 2003)
2. $\exists$ involved CLT by Mattingly et al. (2023),
   - see Ezra's talk (Thu afternoon)
3. $\exists$ phenomenon of stickiness,
   - is a curse for statistics, e.g. of phylogenetic trees (e.g. Billera et al. (2001), Barden et al. (2018)),
   - that can be alleviated in wald space (Lueg et al., 2024)?
   - see Mahshid's poster,

On Complete CAT(0) (Hadamard) Spaces

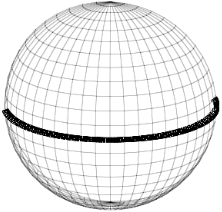
1. $|\mathbb{M}(\mathbb{P})| = 1 = |\mathbb{M}_n|$ a.s. (Sturm, 2003)
2. $\exists$ involved CLT by Mattingly et al. (2023),
   - see Ezra's talk (Thu afternoon)
3. $\exists$ phenomenon of stickiness,
   - is a curse for statistics, e.g. of phylogenetic trees (e.g. Billera et al. (2001), Barden et al. (2018)),
   - that can be alleviated in wald space (Lueg et al., 2024)?
   - see Mahshid's poster,
   - that comes in different flavors (Lammers et al., 2023).
   - Can it be turned into a blessing?
   - See Lars' poster,

Nonuniqueness

Consider $x, y \in \mathfrak{X} = \mathbb{S}^2$ with

- $d_E(x, y) = \|x - y\|$ = extrinsic metric
- $d_I(x, y) = \arccos y^T x$ = intrinsic (spherical) metric
- $d_R(x, y) = \|x - yy^T x\|$ = residual quasi metric

and $X \sim$ uniform on equator:

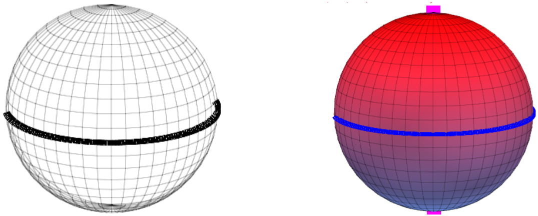


Nonuniqueness

Consider $x, y \in \mathfrak{X} = \mathbb{S}^2$ with

- $d_E(x, y) = \|x - y\|$ = extrinsic metric
- $d_I(x, y) = \arccos y^T x$ = intrinsic (spherical) metric
- $d_R(x, y) = \|x - yy^T x\|$ = residual quasi metric

and $X \sim$ uniform on equator:

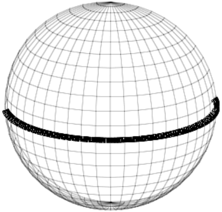


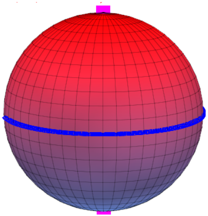
Nonuniqueness

Consider $x, y \in \mathfrak{X} = \mathbb{S}^2$ with

- $d_E(x, y) = \|x - y\|$ = extrinsic metric
- $d_I(x, y) = \arccos y^T x$ = intrinsic (spherical) metric
- $d_R(x, y) = \|x - yy^T x\|$ = residual quasi metric

and $X \sim$ uniform on equator:





How, and how good is the estimation of random variables “near” X?

Honesty

Definition

A Fréchet ρ-mean $\mathbb{M}$ is honest if $\mathbb{M}(\delta_x) = \{x\}$ for all $x \in \mathfrak{X}$.

Honesty

Definition

A Fréchet ρ-mean $\mathbb{M}$ is honest if $\mathbb{M}(\delta_x) = \{x\}$ for all $x \in \mathfrak{X}$.

Lemma 1 (Instability of Nonuniqueness)

For all honest $\mathbb{M}$ on $\mathfrak{X}$, all $t \in (0, 1]$ and all $x \in \mathbb{M}(\mathbb{P})$ we have

$$\mathbb{M}(\mathbb{P}_{x,t}) = \{x\}$$

with the perturbed measure $\mathbb{P}_{x,t} := (1 - t)\,\mathbb{P} + t\delta_x$.

Honesty

Definition

A Fréchet ρ-mean $\mathbb{M}$ is honest if $\mathbb{M}(\delta_x) = \{x\}$ for all $x \in \mathfrak{X}$.

Lemma 1 (Instability of Nonuniqueness)

For all honest $\mathbb{M}$ on $\mathfrak{X}$, all $t \in (0, 1]$ and all $x \in \mathbb{M}(\mathbb{P})$ we have

$$\mathbb{M}(\mathbb{P}_{x,t}) = \{x\}$$

with the perturbed measure $\mathbb{P}_{x,t} := (1 - t)\,\mathbb{P} + t\delta_x$.

Fréchet d^2-means based on a metric d are honest, however, not if based on the residual quasi metric.

Lower Bound

Theorem 1

Let the Fréchet ρ-mean $\mathbb{M}$ be honest,

- $|\mathbb{M}(\mathbb{P})| > 1$,
- $n \in \mathbb{N}$ *fixed, whereas* $x \in \mathbb{M}(\mathbb{P})$, $0 < t \leq 1$ *arbitrary*
- $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathbb{P}_{x,t}$,
- $\hat{\mu}_n$ *any estimator of* $\mathbb{M}$, *e.g.* $\hat{\mu}_n = \mathbb{M}_n$,

then for all $\eta > 0$,

$$\sup_{\substack{x \in \mathbb{M}(\mathbb{P}) \\ 0 < t \leq 1}} \mathbb{P}_{x,t}^{\otimes n} \left\{ d(\hat{\mu}_n, \mathbb{M}(\mathbb{P}_{x,t})) \geq \frac{\operatorname{diam} \mathbb{M}(\mathbb{P}) - \eta}{2} \right\} \geq \frac{1}{2} \,.$$

Proof

Apply argument chain of Tsybakov (2009, p. 79 ff.):

1. $y, z \in \mathbb{M}(\mathbb{P}), d(y, z) \geq \operatorname{diam}(\mathbb{M}(\mathbb{P}) - \eta,$

Proof

Apply argument chain of Tsybakov (2009, p. 79 ff.):

1. $y, z \in \mathbb{M}(\mathbb{P}), d(y, z) \geq \mathrm{diam}(\mathbb{M}(\mathbb{P}) - \eta$,
2. two hypotheses $\mathbb{P}_{y,t}, \mathbb{P}_{z,t}$,

Proof

Apply argument chain of Tsybakov (2009, p. 79 ff.):

1. $y, z \in \mathbb{M}(\mathbb{P}), d(y, z) \geq \text{diam}(\mathbb{M}(\mathbb{P}) - \eta,$
2. two hypotheses $\mathbb{P}_{y,t}, \mathbb{P}_{z,t},$
3. $d(\mathbb{M}(\mathbb{P}_{y,t}), \mathbb{M}(\mathbb{P}_{z,t})) \overset{\text{honesty}}{\geq} \text{diam}(\mathbb{M}(\mathbb{P})) - \eta =: 2s,$

Proof

Apply argument chain of Tsybakov (2009, p. 79 ff.):

1. $y, z \in \mathbb{M}(\mathbb{P}), d(y,z) \geq \text{diam}(\mathbb{M}(\mathbb{P}) - \eta$,
2. two hypotheses $\mathbb{P}_{y,t}, \mathbb{P}_{z,t}$,
3. $d(\mathbb{M}(\mathbb{P}_{y,t}), \mathbb{M}(\mathbb{P}_{z,t})) \overset{\text{honesty}}{\geq} \text{diam}(\mathbb{M}(\mathbb{P})) - \eta =: 2s$,
4. $\max_{x \in \{y,z\}} \mathbb{P}_{x,t}^{\otimes n}\{d(\hat{\mu}_n, \mathbb{M}(\mathbb{P}_{x,t})) \geq s\}$
   $\overset{(2.9),\ p.80}{\underset{\Delta\text{ ineq}}{\geq}} \quad \inf_{\psi_n} \max_{x \in \{y,z\}} \mathbb{P}_{x,t}^{\otimes n}\{\psi_n \neq x\} =: p_{e,1}$
   over all tests $\psi_n : \mathfrak{X}^n \to \{y,z\}$;
   a candidate for ψ_n is the test to reject if distance to $\hat{\mu}_n$ is $\geq s$, and to do a coin flip (reject less often) if both are $\geq s$ (both cannot be $< s$),

Proof

Apply argument chain of Tsybakov (2009, p. 79 ff.):

1. $y, z \in \mathbb{M}(\mathbb{P}), d(y, z) \geq \operatorname{diam}(\mathbb{M}(\mathbb{P}) - \eta,$
2. two hypotheses $\mathbb{P}_{y,t}, \mathbb{P}_{z,t}$,
3. $d(\mathbb{M}(\mathbb{P}_{y,t}), \mathbb{M}(\mathbb{P}_{z,t})) \overset{\text{honesty}}{\geq} \operatorname{diam}(\mathbb{M}(\mathbb{P})) - \eta =: 2s,$
4. $\max_{x \in \{y,z\}} \mathbb{P}_{x,t}^{\otimes n}\{d(\hat{\mu}_n, \mathbb{M}(\mathbb{P}_{x,t})) \geq s\}$
   $\overset{(2.9),\, p.80 \atop \Delta \text{ ineq}}{\geq} \quad \inf_{\psi_n} \max_{x \in \{y,z\}} \mathbb{P}_{x,t}^{\otimes n}\{\psi_n \neq x\} =: p_{e,1}$
   over all tests $\psi_n : \mathfrak{X}^n \to \{y, z\}$;
   a candidate for ψ_n is the test to reject if distance to $\hat{\mu}_n$ is $\geq s$, and to do a coin flip (reject less often) if both are $\geq s$ (both cannot be $< s$),
5. $TV(\mathbb{P}_{y,t}^{\otimes n}, \mathbb{P}_{z,t}^{\otimes n}) \overset{\text{next slide}}{\leq} \sqrt{2}\sqrt{1 - (1 - t)^n} := \alpha,$

Proof

Apply argument chain of Tsybakov (2009, p. 79 ff.):

1. $y, z \in \mathbb{M}(\mathbb{P}), d(y,z) \geq \mathrm{diam}(\mathbb{M}(\mathbb{P}) - \eta,$
2. two hypotheses $\mathbb{P}_{y,t}, \mathbb{P}_{z,t}$,
3. $d(\mathbb{M}(\mathbb{P}_{y,t}), \mathbb{M}(\mathbb{P}_{z,t})) \overset{\text{honesty}}{\geq} \mathrm{diam}(\mathbb{M}(\mathbb{P})) - \eta =: 2s$,
4. $\max_{x\in\{y,z\}} \mathbb{P}_{x,t}^{\otimes n}\{d(\hat{\mu}_n, \mathbb{M}(\mathbb{P}_{x,t})) \geq s\}$
   $\overset{(2.9),\, p.80 \atop \Delta \text{ ineq}}{\geq} \quad \inf_{\psi_n} \max_{x\in\{y,z\}} \mathbb{P}_{x,t}^{\otimes n}\{\psi_n \neq x\} =: p_{e,1}$
   over all tests $\psi_n : \mathfrak{X}^n \to \{y, z\}$;
   a candidate for ψ_n is the test to reject if distance to $\hat{\mu}_n$ is $\geq s$, and to do a coin flip (reject less often) if both are $\geq s$ (both cannot be $< s$),
5. $TV(\mathbb{P}_{y,t}^{\otimes n}, \mathbb{P}_{z,t}^{\otimes n}) \overset{\text{next slide}}{\leq} \sqrt{2}\sqrt{1-(1-t)^n} := \alpha$,
6. $\overset{\text{p. 90,Thm 2.2 (i)}}{\Rightarrow} \quad p_{e,1} \geq \frac{1-\alpha}{2}$,

Proof

Apply argument chain of Tsybakov (2009, p. 79 ff.):

1. $y, z \in \mathbb{M}(\mathbb{P}), d(y, z) \geq \operatorname{diam}(\mathbb{M}(\mathbb{P}) - \eta$,
2. two hypotheses $\mathbb{P}_{y,t}, \mathbb{P}_{z,t}$,
3. $d(\mathbb{M}(\mathbb{P}_{y,t}), \mathbb{M}(\mathbb{P}_{z,t})) \overset{\text{honesty}}{\geq} \operatorname{diam}(\mathbb{M}(\mathbb{P})) - \eta =: 2s$,
4. $\max_{x \in \{y,z\}} \mathbb{P}_{x,t}^{\otimes n}\{d(\hat{\mu}_n, \mathbb{M}(\mathbb{P}_{x,t})) \geq s\}$
   $\overset{\substack{(2.9),\, p.80 \\ \Delta \text{ ineq}}}{\geq} \inf_{\psi_n} \max_{x \in \{y,z\}} \mathbb{P}_{x,t}^{\otimes n}\{\psi_n \neq x\} =: p_{e,1}$
   over all tests $\psi_n : \mathfrak{X}^n \to \{y, z\}$;
   a candidate for ψ_n is the test to reject if distance to $\hat{\mu}_n$ is $\geq s$, and to do a coin flip (reject less often) if both are $\geq s$ (both cannot be $< s$),
5. $TV(\mathbb{P}_{y,t}^{\otimes n}, \mathbb{P}_{z,t}^{\otimes n}) \overset{\text{next slide}}{\leq} \sqrt{2}\sqrt{1 - (1-t)^n} := \alpha$,
6. $\overset{\text{p. 90,Thm 2.2 (i)}}{\Rightarrow} p_{e,1} \geq \frac{1-\alpha}{2}$,
7. $t \to 0 \Rightarrow \alpha \to 0 \Rightarrow$ Theorem 1 (max above $\leq$ sup of theorem).

Step 5: The Missing Inequality

Tsybakov (2009, p. 83, 86), two prob. distributions P, Q:

- squared Hellinger distance
  $$H(P,Q)^2 := \int \left(\sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}} \right)^2 d\nu \text{ , } \nu = P + Q,$$
- total variation distance $TV(P,Q) := \sup_{A \text{ m'ble}} |P(A) - Q(A)|$,
- $H(P^{\otimes n}, Q^{\otimes n})^2 = 2\left(1 - \left(1 - \frac{H(P,Q)^2}{2}\right)^n\right)$,
- Le Cam's inequalities $H^2/2 \leq TV \leq H$.

Step 5: The Missing Inequality

Tsybakov (2009, p. 83, 86), two prob. distributions P, Q:

- squared Hellinger distance
  $H(P,Q)^2 := \int \left(\sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}} \right)^2 d\nu \text{ , } \nu = P + Q,$
- total variation distance $TV(P,Q) := \sup_{A \text{ m'ble}} |P(A) - Q(A)|,$
- $H(P^{\otimes n}, Q^{\otimes n})^2 = 2\left(1 - \left(1 - \frac{H(P,Q)^2}{2}\right)^n\right),$
- Le Cam's inequalities $H^2/2 \leq TV \leq H$.

Application:

- $H(\mathbb{P}_{y,t}, \mathbb{P}_{z,t})^2/2 \leq TV(\mathbb{P}_{y,t}, \mathbb{P}_{z,t}) = t,$

Step 5: The Missing Inequality

Tsybakov (2009, p. 83, 86), two prob. distributions P, Q:

- squared Hellinger distance
  $H(P,Q)^2 := \int \left(\sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}} \right)^2 d\nu \, , \nu = P + Q,$
- total variation distance $TV(P,Q) := \sup_{A \text{ m'ble}} |P(A) - Q(A)|,$
- $H(P^{\otimes n}, Q^{\otimes n})^2 = 2 \left(1 - \left(1 - \frac{H(P,Q)^2}{2} \right)^n \right),$
- Le Cam's inequalities $H^2/2 \leq TV \leq H$.

Application:

- $H(\mathbb{P}_{y,t}, \mathbb{P}_{z,t})^2/2 \leq TV(\mathbb{P}_{y,t}, \mathbb{P}_{z,t}) = t,$
- $\left(1 - \frac{H(\mathbb{P}_{y,t}, \mathbb{P}_{z,t})^2}{2} \right)^n \geq (1-t)^n,$

Step 5: The Missing Inequality

Tsybakov (2009, p. 83, 86), two prob. distributions P, Q:

- squared Hellinger distance
  $$H(P,Q)^2 := \int \left(\sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}} \right)^2 d\nu \, , \, \nu = P + Q,$$
- total variation distance $TV(P,Q) := \sup_{A \text{ m'ble}} |P(A) - Q(A)|$,
- $H(P^{\otimes n}, Q^{\otimes n})^2 = 2\left(1 - \left(1 - \frac{H(P,Q)^2}{2}\right)^n\right)$,
- Le Cam's inequalities $H^2/2 \leq TV \leq H$.

Application:

- $H(\mathbb{P}_{y,t}, \mathbb{P}_{z,t})^2/2 \leq TV(\mathbb{P}_{y,t}, \mathbb{P}_{z,t}) = t$,
- $\left(1 - \frac{H(\mathbb{P}_{y,t}, \mathbb{P}_{z,t})^2}{2}\right)^n \geq (1-t)^n$,
- $TV(\mathbb{P}_{y,t}^{\otimes n}, \mathbb{P}_{z,t}^{\otimes n})^2 \leq H(\mathbb{P}_{y,t}^{\otimes n}, \mathbb{P}_{z,t}^{\otimes n})^2 \leq 2\,(1 - (1-t)^n)$.

Remark 1

Given $\mathbb{M}(\mathbb{P}) \ni y \neq z \in \mathbb{M}(\mathbb{P})$, uniformly over

$$\{\mathbb{P}_{x,t} : x \in \{y, z\}, t \in (0, \epsilon)\}, \forall 0 < \epsilon \leq 1$$

no test $\psi_n \colon \mathfrak{X}^n \to \{y, z\}$ will perform better than a coin flip.

Corollary 1

Markov inequality $\frac{1}{2} \leq \mathbb{P}\{Y \geq s\} \leq \frac{\mathbb{E}[Y]}{s}$ and $\eta \to 0$ gives

$$\inf_{\hat{\mu}_n} \sup_{\substack{x \in \mathbb{M}(\mathbb{P}) \\ 0 < t \leq 1}} \mathbb{E}_{\mathbb{P}_{x,t}^{\otimes n}} [d(\hat{\mu}_n, \mathbb{M}(\mathbb{P}_{x,t}))] \geq \frac{\operatorname{diam} \mathbb{M}(\mathbb{P})}{4} .$$

Corollary 1

Markov inequality $\frac{1}{2} \leq \mathbb{P}\{Y \geq s\} \leq \frac{\mathbb{E}[Y]}{s}$ and $\eta \to 0$ gives

$$\inf_{\hat{\mu}_n} \sup_{\substack{x \in \mathbb{M}(\mathbb{P}) \\ 0 < t \leq 1}} \mathbb{E}_{\mathbb{P}_{x,t}^{\otimes n}} [d(\hat{\mu}_n, \mathbb{M}(\mathbb{P}_{x,t}))] \geq \frac{\operatorname{diam} \mathbb{M}(\mathbb{P})}{4} .$$

Example:

- $\mathbb{P} = \frac{\delta_N + \delta_S}{2}$ on $\mathbb{S}^2$ with poles N, S, resp.,
- d: spherical distance,
- $\mathbb{M}(\mathbb{P}) =$ equator, $\operatorname{diam} = \pi$,
- pick $x \in$ equator, zero meridian,
- as $t \to 0$ (by symmetry)
  - $\mathbb{M}_n$ of $\mathbb{P}_{x,t}^{\otimes n}$ tends to be on far side with 50 %,
  - in these cases $d(\mathbb{M}_n, x) \geq \pi/2$,
  - in the mean, $d(\mathbb{M}_n, x) \geq \pi/4$.

The Wasserstein World

For an invitation, see, e.g. Panaretos and Zemel (2020):

- $\mathcal{P}(\mathbb{R}^m) :=$ all probability measures on $\mathbb{R}^m$,

The Wasserstein World

For an invitation, see, e.g. Panaretos and Zemel (2020):

- $\mathcal{P}(\mathbb{R}^m) :=$ all probability measures on $\mathbb{R}^m$,
- $\mathcal{P}_2(\mathbb{R}^m) := \left\{\mu \in \mathcal{P}(\mathbb{R}^m) : \exists x \in \mathfrak{X} \text{ with } \mathbb{E}[d(x,X)^2] < \infty\right\}$

The Wasserstein World

For an invitation, see, e.g. Panaretos and Zemel (2020):

- $\mathcal{P}(\mathbb{R}^m) :=$ all probability measures on $\mathbb{R}^m$,
- $\mathcal{P}_2(\mathbb{R}^m) := \left\{\mu \in \mathcal{P}(\mathbb{R}^m) : \exists x \in \mathfrak{X} \text{ with } \mathbb{E}[d(x,X)^2] < \infty\right\}$
- $\mathcal{W}(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)} \int \|x-y\|^2 \, d\pi(x,y)\right)^{1/2}$,

$$\Pi(\mu,\nu) := \left\{\pi \in \mathcal{P}(\mathbb{R}^m \times \mathbb{R}^m) \middle| \begin{array}{c} \pi(A \times \mathbb{R}^m) = \mu(A)\ \forall A \in \mathcal{B}(\mathbb{R}^m) \\ \pi(\mathbb{R}^m \times B) = \nu(B)\ \forall B \in \mathcal{B}(\mathbb{R}^m) \end{array} \right\}$$

The Wasserstein World

For an invitation, see, e.g. Panaretos and Zemel (2020):

- $\mathcal{P}(\mathbb{R}^m) :=$ all probability measures on $\mathbb{R}^m$,
- $\mathcal{P}_2(\mathbb{R}^m) := \left\{\mu \in \mathcal{P}(\mathbb{R}^m) : \exists x \in \mathfrak{X} \text{ with } \mathbb{E}[d(x,X)^2] < \infty\right\}$
- $\mathcal{W}(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)} \int \|x-y\|^2\, d\pi(x,y)\right)^{1/2}$,

  $$\Pi(\mu,\nu) := \left\{\pi \in \mathcal{P}(\mathbb{R}^m \times \mathbb{R}^m) \,\middle|\, \begin{array}{c} \pi(A\times\mathbb{R}^m) = \mu(A)\ \forall A \in \mathcal{B}(\mathbb{R}^m) \\ \pi(\mathbb{R}^m \times B) = \nu(B)\ \forall B \in \mathcal{B}(\mathbb{R}^m) \end{array}\right\}$$

- $(\mathfrak{X}, d) := (\mathcal{P}_2(\mathbb{R}^m), \mathcal{W}) =$ the 2-Wasserstein space over $\mathbb{R}^m$,

The Wasserstein World

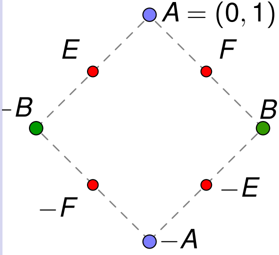
For an invitation, see, e.g. Panaretos and Zemel (2020):

- $\mathcal{P}(\mathbb{R}^m) :=$ all probability measures on $\mathbb{R}^m$,
- $\mathcal{P}_2(\mathbb{R}^m) := \left\{\mu \in \mathcal{P}(\mathbb{R}^m) : \exists x \in \mathfrak{X} \text{ with } \mathbb{E}[d(x,X)^2] < \infty\right\}$
- $\mathcal{W}(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)} \int \|x-y\|^2\, d\pi(x,y)\right)^{1/2}$,

$$\Pi(\mu,\nu) := \left\{\pi \in \mathcal{P}(\mathbb{R}^m \times \mathbb{R}^m) \,\middle|\, \begin{array}{c} \pi(A\times\mathbb{R}^m) = \mu(A)\ \forall A \in \mathcal{B}(\mathbb{R}^m) \\ \pi(\mathbb{R}^m \times B) = \nu(B)\ \forall B \in \mathcal{B}(\mathbb{R}^m) \end{array}\right\}$$

- $(\mathfrak{X}, d) := (\mathcal{P}_2(\mathbb{R}^m), \mathcal{W}) =$ the 2-Wasserstein space over $\mathbb{R}^m$,
- it's metric, hence its Fréchet means are honest.

Nonuniqueness in the Wasserstein World



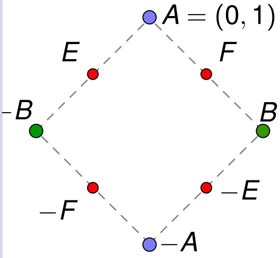
In $\mathbb{R}^2$, let

$$\mu := \frac{1}{2}\left(\delta_A + \delta_{-A}\right),$$
$$\nu := \frac{1}{2}\left(\delta_B + \delta_{-B}\right),$$
$$\mathbb{P} := \frac{1}{2}\left(\delta_\mu + \delta_\nu\right),$$

yielding,

$$\mathbb{M}(\mathbb{P}) = \left\{ \frac{\alpha}{2}(\delta_E + \delta_{-E}) + \frac{1-\alpha}{2}(\delta_F + \delta_{-F}) : 0 \le \alpha \le 1 \right\}.$$

Nonuniqueness in the Wasserstein World



In $\mathbb{R}^2$, let
$\mu := \frac{1}{2}\left(\delta_A + \delta_{-A}\right)$,
$\nu := \frac{1}{2}\left(\delta_B + \delta_{-B}\right)$,
$\mathbb{P} := \frac{1}{2}\left(\delta_\mu + \delta_\nu\right)$,

yielding,

$$\mathbb{M}(\mathbb{P}) = \Big\{ \tfrac{\alpha}{2}(\delta_E + \delta_{-E}) + \tfrac{1-\alpha}{2}(\delta_F + \delta_{-F}) : 0 \le \alpha \le 1 \Big\}.$$

By Corollary 1, for $\mathbb{M}_n$ of $\mu_1, \dots, \mu_n \overset{i.i.d.}{\sim} \mathbb{P}_{\mu,t} = (1-t)\,\mathbb{P} + t\delta_\mu$, $0 < t \le 1$ (by symmetry):

$$\sup_{t>0} \mathbb{E}_{\mathbb{P}_{\mu,t}^{\otimes n}}[\mathcal{W}(\mathbb{M}_n, \mu)] \ge \frac{\operatorname{diam} \mathbb{M}(\mathbb{P})}{4} = \frac{1}{4}\,,$$

as $\|F - E\| = 1 = \|F + E\|$.

Parametric Rates ($1/n$)

Theorem (Le Gouic et al. (2022), Cor. 1.2)

Let/assume:

- $0 < \alpha \leq \beta$, $\Phi_{\alpha,\beta} := \Big\{\phi \in \mathcal{C}^1(\mathbb{R}^m \to \mathbb{R}) :$
  $\phi(y) \geq \phi(x) + \nabla\phi(x)^T(y - x) + \frac{\alpha}{2}\|x - y\|^2$

Parametric Rates ($1/n$)

Theorem (Le Gouic et al. (2022), Cor. 1.2)

Let/assume:

- $0 < \alpha \leq \beta$, $\Phi_{\alpha,\beta} := \Big\{ \phi \in \mathcal{C}^1(\mathbb{R}^m \to \mathbb{R}) :$
  $\phi(y) \geq \phi(x) + \nabla\phi(x)^T(y-x) + \frac{\alpha}{2}\|x-y\|^2$
  $\phi(y) \leq \phi(x) + \nabla\phi(x)^T(y-x) + \frac{\beta}{2}\|x-y\|^2$ *for all* $x, y \in \mathbb{R}^m \Big\}$,

Parametric Rates ($1/n$)

Theorem (Le Gouic et al. (2022), Cor. 1.2)

Let/assume:

- $0 < \alpha \leq \beta$, $\Phi_{\alpha,\beta} := \Big\{\phi \in \mathcal{C}^1(\mathbb{R}^m \to \mathbb{R}) :$
  $\phi(y) \geq \phi(x) + \nabla\phi(x)^T(y - x) + \frac{\alpha}{2}\|x - y\|^2$
  $\phi(y) \leq \phi(x) + \nabla\phi(x)^T(y - x) + \frac{\beta}{2}\|x - y\|^2$ *for all* $x, y \in \mathbb{R}^m \Big\}$,
- $\mathbb{M}(\mathbb{P}) = \{\tau\}$.
- $\mathrm{supp}(\mathbb{P}) \subseteq \{(\nabla\phi)_\sharp\tau : \phi \in \Phi_{\alpha,\beta}\}$,

Parametric Rates ($1/n$)

Theorem (Le Gouic et al. (2022), Cor. 1.2)

Let/assume:

- $0 < \alpha \leq \beta$, $\Phi_{\alpha,\beta} := \Big\{\phi \in \mathcal{C}^1(\mathbb{R}^m \to \mathbb{R}) :$
  $\phi(y) \geq \phi(x) + \nabla\phi(x)^T(y-x) + \frac{\alpha}{2}\|x-y\|^2$
  $\phi(y) \leq \phi(x) + \nabla\phi(x)^T(y-x) + \frac{\beta}{2}\|x-y\|^2$ *for all* $x, y \in \mathbb{R}^m \Big\}$,
- $\mathbb{M}(\mathbb{P}) = \{\tau\}$.
- $\operatorname{supp}(\mathbb{P}) \subseteq \{(\nabla\phi)_\sharp\tau : \phi \in \Phi_{\alpha,\beta}\}$,

Then, if $\beta < \alpha + 1$,

$$\mathbb{E}_{\mathbb{P}^{\otimes n}}[\mathcal{W}(\mathbb{M}_n, \tau)^2] \leq \frac{1}{n}\,\frac{4\,\mathbb{E}_{P\sim\mathbb{P}}[\mathcal{W}(P,\tau)^2]}{(\beta-\alpha-1)^2}\,.$$

Parametric Rates ($1/n$)

Theorem (Le Gouic et al. (2022), Cor. 1.2)

Let/assume:

- $0 < \alpha \leq \beta$, $\Phi_{\alpha,\beta} := \Big\{ \phi \in \mathcal{C}^1(\mathbb{R}^m \to \mathbb{R}) :$
  $\phi(y) \geq \phi(x) + \nabla\phi(x)^T(y - x) + \frac{\alpha}{2}\|x - y\|^2$
  $\phi(y) \leq \phi(x) + \nabla\phi(x)^T(y - x) + \frac{\beta}{2}\|x - y\|^2$ *for all* $x, y \in \mathbb{R}^m \Big\}$,
- $\mathbb{M}(\mathbb{P}) = \{\tau\}$.
- $\mathrm{supp}(\mathbb{P}) \subseteq \{(\nabla\phi)_\sharp\tau : \phi \in \Phi_{\alpha,\beta}\}$,

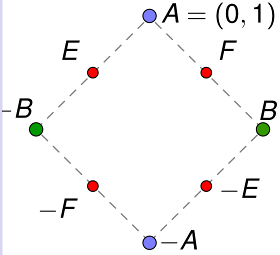
Then, if $\beta < \alpha + 1$,

$$\mathbb{E}_{\mathbb{P}^{\otimes n}}[\mathcal{W}(\mathbb{M}_n, \tau)^2] \leq \frac{1}{n} \frac{4\, \mathbb{E}_{P\sim\mathbb{P}}[\mathcal{W}(P, \tau)^2]}{(\beta - \alpha - 1)^2}.$$

Question

Is $\beta < \alpha + 1$ *sharp for parametric rates?*

Partial Answer



Let

$\mu := \frac{1}{2}(\delta_A + \delta_{-A})$,
$\nu := \frac{1}{2}(\delta_B + \delta_{-B})$,
$\tau := \frac{1}{2}(\delta_E + \delta_{-E})$,

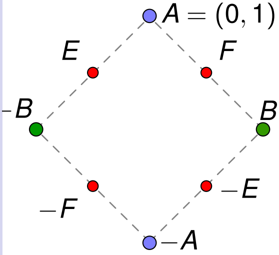
Theorem 2

Let $0 < \alpha < \beta$. *Then*

(i) $\mu = (\nabla\phi)_\sharp\tau$ *with* $\phi \in \Phi_{\alpha,\beta}$
$\Rightarrow \beta > \alpha + 2$

(ii) $\forall \delta > 0\ \exists \beta \leq \alpha + 2 + \delta$ *such that*
$\forall \zeta \in \{\mu, \nu, \tau\}$
$\zeta = (\nabla\phi)_\sharp\tau$ *for some* $\phi \in \Phi_{\alpha,\beta}$.

Partial Answer



Let
$\mu := \frac{1}{2}\left(\delta_A + \delta_{-A}\right),$
$\nu := \frac{1}{2}\left(\delta_B + \delta_{-B}\right),$
$\tau := \frac{1}{2}\left(\delta_E + \delta_{-E}\right),$

Theorem 2

Let $0 < \alpha < \beta$. Then

(i) $\mu = (\nabla\phi)_\sharp\tau$ *with* $\phi \in \Phi_{\alpha,\beta}$
$\Rightarrow \beta > \alpha + 2$

(ii) $\forall\delta > 0\ \exists\beta \leq \alpha + 2 + \delta$ *such that*
$\forall\zeta \in \{\mu, \nu, \tau\}$
$\zeta = (\nabla\phi)_\sharp\tau$ *for some* $\phi \in \Phi_{\alpha,\beta}$.

Hence $\beta \leq \alpha + 2$ is a realistic upper bound.

Conjecture

Under Le Gouic et al. (2022) hypotheses, with C indep. of $\mathbb{P}$,
$\mathbb{E}_{\mathbb{P}^{\otimes n}}[\mathcal{W}(\mathbb{M}_n, \tau)^2] \leq \frac{1}{n}\frac{C\,\mathbb{E}_{P\sim\mathbb{P}}[\mathcal{W}(P,\tau)^2]}{(\beta-\alpha-2)^2}$ *for all* $0 < \alpha < \beta < \alpha + 2$.

The Shape World

E.g. Gower (1975), H. (2010, 2012), Dryden and Mardia (2016), $2 \leq m < k$:

- landmark configurations $z = (z_1, \ldots, z_k) \in \mathbb{R}^{m \times k}$,

The Shape World

E.g. Gower (1975), H. (2010, 2012), Dryden and Mardia (2016), $2 \leq m < k$:

- landmark configurations $z = (z_1, \ldots, z_k) \in \mathbb{R}^{m \times k}$,
- removing translation and scaling $z \mapsto x \in \mathbb{S}^{m \times (k-1)-1}$,

The Shape World

E.g. Gower (1975), H. (2010, 2012), Dryden and Mardia (2016), $2 \leq m < k$:

- landmark configurations $z = (z_1, \ldots, z_k) \in \mathbb{R}^{m \times k}$,
- removing translation and scaling $z \mapsto x \in \mathbb{S}^{m \times (k-1)-1}$,
- removing rotation:
  $\Sigma_m^k := \{[x] : x \in \mathbb{S}^{m \times (k-1)-1}\}, \quad [x] := \{gx : g \in SO(m)\}$,

The Shape World

E.g. Gower (1975), H. (2010, 2012), Dryden and Mardia (2016), $2 \leq m < k$:

- landmark configurations $z = (z_1, \dots, z_k) \in \mathbb{R}^{m \times k}$,
- removing translation and scaling $z \mapsto x \in \mathbb{S}^{m\times(k-1)-1}$,
- removing rotation:
  $\Sigma_m^k := \{[x] : x \in \mathbb{S}^{m\times(k-1)-1}\}, \quad [x] := \{gx : g \in SO(m)\}$,
- Procrustes metric (quotient of residual quasi metric)
  $d_P([x],[y]) := \min_{g\in SO(m)} \|gx - yy^T gx\|$,

The Shape World

E.g. Gower (1975), H. (2010, 2012), Dryden and Mardia (2016), $2 \leq m < k$:

- landmark configurations $z = (z_1, \dots, z_k) \in \mathbb{R}^{m \times k}$,
- removing translation and scaling $z \mapsto x \in \mathbb{S}^{m \times (k-1)-1}$,
- removing rotation:
  $\Sigma_m^k := \{[x] : x \in \mathbb{S}^{m \times (k-1)-1}\}, \quad [x] := \{gx : g \in SO(m)\}$,
- Procrustes metric (quotient of residual quasi metric)
  $d_P([x],[y]) := \min_{g \in SO(m)} \|gx - yy^T gx\|$,
- for $m \geq 3$, Σ_m^k is a stratified space with open and dense top-dimensional manifold stratum Σ^*,

The Shape World

E.g. Gower (1975), H. (2010, 2012), Dryden and Mardia (2016), $2 \leq m < k$:

- landmark configurations $z = (z_1, \ldots, z_k) \in \mathbb{R}^{m \times k}$,
- removing translation and scaling $z \mapsto x \in \mathbb{S}^{m \times (k-1)-1}$,
- removing rotation:
  $\Sigma_m^k := \{[x] : x \in \mathbb{S}^{m \times (k-1)-1}\}, \quad [x] := \{gx : g \in SO(m)\}$,
- Procrustes metric (quotient of residual quasi metric)
  $d_P([x],[y]) := \min_{g \in SO(m)} \|gx - yy^T gx\|$,
- for $m \geq 3$, Σ_m^k is a stratified space with open and dense top-dimensional manifold stratum Σ^*,
- $\forall \xi \in \mathbb{S}^{m \times (k-1)-1}$ with $[\xi] \in \Sigma^* \ \exists \ \Sigma_{[\xi]}^{**} \subseteq \Sigma^*$ open and a smooth lift $\psi_\xi : \Sigma_{[\xi]}^{**} \to \mathbb{S}^{m \times (k-1)-1}$ in optimal position to ξ, i.e.
  $d_P([x],[\xi]) = \|\psi_\xi([x]) - \xi\xi^T \psi_\xi([x])\|$,

The Shape World

E.g. Gower (1975), H. (2010, 2012), Dryden and Mardia (2016), $2 \leq m < k$:

- landmark configurations $z = (z_1, \dots, z_k) \in \mathbb{R}^{m \times k}$,
- removing translation and scaling $z \mapsto x \in \mathbb{S}^{m \times (k-1)-1}$,
- removing rotation:
  $\Sigma_m^k := \{[x] : x \in \mathbb{S}^{m \times (k-1)-1}\}, \quad [x] := \{gx : g \in SO(m)\}$,
- Procrustes metric (quotient of residual quasi metric)
  $d_P([x],[y]) := \min_{g \in SO(m)} \|gx - yy^T gx\|$,
- for $m \geq 3$, Σ_m^k is a stratified space with open and dense top-dimensional manifold stratum Σ^*,
- $\forall \xi \in \mathbb{S}^{m \times (k-1)-1}$ with $[\xi] \in \Sigma^* \ \exists \ \Sigma_{[\xi]}^{**} \subseteq \Sigma^*$ open and a smooth lift $\psi_\xi : \Sigma_{[\xi]}^{**} \to \mathbb{S}^{m \times (k-1)-1}$ in optimal position to ξ, i.e. $d_P([x],[\xi]) = \|\psi_\xi([x]) - \xi\xi^T \psi_\xi([x])\|$,
- chart at $\xi \in \mathbb{S}^{m \times (k-1)-1}$ for $\mathrm{tr}(x^T\xi) > 0$: $\phi_\xi : x \mapsto \frac{x - \mathrm{tr}(x^T\xi)\xi}{\mathrm{tr}(x^T\xi)}$.

A CLT for Procrustes Means

Theorem 3

Let $Z_1, \ldots, Z_n, Z \overset{i.i.d.}{\sim} \mathbb{P} \in \mathcal{P}(\Sigma_m^k)$ *be random shapes, that a.s. assume* $\Sigma_{[\xi]}^{**}$, $\mathbb{M}(\mathbb{P}) = \{[\xi]\}$, $\hat{\mu}_n \in \mathbb{M}_n$ *m'ble selection. Then*

$$\sqrt{n}\,\mathrm{vec}\left(\phi_\xi \circ \psi_\xi(\hat{\mu}_n)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\, H^- D H^-\right)$$

with

$$D = \mathrm{cov}[\mathrm{vec}(\psi_\xi(Z))\mathrm{vec}(\psi_\xi(Z))^T] = \sum_{j=1}^{N} \lambda_j v_j v_j^T\,,$$

$$H^- = \frac{1}{2}\sum_{j=2}^{N+1} (\lambda_1 - \lambda_j)^{-1} v_j v_j^T$$

where $N = m(k-1)$, $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_N$, $\mathrm{vec}(\xi) = v_1$, $(v_1, \ldots, v_N) \in SO(N)$.

A CLT for Procrustes Means

Theorem 3

Let $Z_1, \ldots, Z_n, Z \overset{i.i.d.}{\sim} \mathbb{P} \in \mathcal{P}(\Sigma_m^k)$ *be random shapes, that a.s. assume* $\Sigma^{**}_{[\xi]}$, $\mathbb{M}(\mathbb{P}) = \{[\xi]\}$, $\hat{\mu}_n \in \mathbb{M}_n$ *m'ble selection. Then*

$$\sqrt{n}\,\mathrm{vec}\left(\phi_\xi \circ \psi_\xi(\hat{\mu}_n)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ H^- D H^-\right)$$

with

$$D = \mathrm{cov}[\mathrm{vec}(\psi_\xi(Z))\mathrm{vec}(\psi_\xi(Z))^T] = \sum_{j=1}^{N} \lambda_j v_j v_j^T\,,$$

$$H^- = \frac{1}{2}\sum_{j=2}^{N+1}(\lambda_1 - \lambda_j)^{-1} v_j v_j^T$$

where $N = m(k-1)$, $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_N$, $\mathrm{vec}(\xi) = v_1$, $(v_1, \ldots, v_N) \in SO(N)$.

Remark

The $v_2, \ldots, v_N$ *span the tangent space of* $\mathbb{S}^{N-1}$ *at* $\mathrm{vec}(\xi)$.
If $\lambda_1 - \lambda_2 \to 0$ *we loose uniqueness,* H^- *explodes, i.e. the constants deteriorate.*

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,
- Alleviated for diffusion means (Eltzner et al., 2023).

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,
- Alleviated for diffusion means (Eltzner et al., 2023).
- Only rather well understood (McKilliam et al. (2012), Hotz and H. (2015), Hundrieser et al. (2020)) on the circle $\mathfrak{X} = \mathbb{S}^1$, $d =$ arc length, with new example (for all $r > 0$, $\mu = 0$):

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,
- Alleviated for diffusion means (Eltzner et al., 2023).
- Only rather well understood (McKilliam et al. (2012), Hotz and H. (2015), Hundrieser et al. (2020)) on the circle $\mathfrak{X} = \mathbb{S}^1$, $d =$ arc length, with new example (for all $r > 0$, $\mu = 0$):
  - $\hat{\mu}_n \xrightarrow{a.s.} 0$

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,
- Alleviated for diffusion means (Eltzner et al., 2023).
- Only rather well understood (McKilliam et al. (2012), Hotz and H. (2015), Hundrieser et al. (2020)) on the circle $\mathfrak{X} = \mathbb{S}^1$, $d =$ arc length, with new example (for all $r > 0$, $\mu = 0$):
  - $\hat{\mu}_n \xrightarrow{a.s.} 0$
  - $\left(log\sqrt{n}\right)^{1/r} \hat{\mu}_n \xrightarrow{\mathcal{D}} \frac{1}{2}(\delta_{-1} + \delta_1)$

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,
- Alleviated for diffusion means (Eltzner et al., 2023).
- Only rather well understood (McKilliam et al. (2012), Hotz and H. (2015), Hundrieser et al. (2020)) on the circle $\mathfrak{X} = \mathbb{S}^1$, $d =$ arc length, with new example (for all $r > 0$, $\mu = 0$):
  - $\hat{\mu}_n \xrightarrow{a.s.} 0$
  - $\left(log\sqrt{n}\right)^{1/r} \hat{\mu}_n \xrightarrow{\mathcal{D}} \frac{1}{2}(\delta_{-1} + \delta_1)$
  - $r \left(\log \sqrt{n}\right)^{(1+r)/r} \left(\hat{\mu}_n - \frac{\mathrm{sign}(\hat{\mu}_n)}{\left(\log \sqrt{n}\right)^{1/r}}\right) \xrightarrow{\mathcal{D}} \mathrm{sign}(Z) \cdot \log |Z|$

    with $Z \sim \mathcal{N}(0, \sigma^2)$.

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,
- Alleviated for diffusion means (Eltzner et al., 2023).
- Only rather well understood (McKilliam et al. (2012), Hotz and H. (2015), Hundrieser et al. (2020)) on the circle $\mathfrak{X} = \mathbb{S}^1$, $d =$ arc length, with new example (for all $r > 0$, $\mu = 0$):
  - $\hat{\mu}_n \xrightarrow{a.s.} 0$
  - $\left(log\sqrt{n}\right)^{1/r} \hat{\mu}_n \xrightarrow{\mathcal{D}} \frac{1}{2}(\delta_{-1} + \delta_1)$
  - $r\left(\log\sqrt{n}\right)^{(1+r)/r}\left(\hat{\mu}_n - \frac{\mathrm{sign}(\hat{\mu}_n)}{\left(\log\sqrt{n}\right)^{1/r}}\right) \xrightarrow{\mathcal{D}} \mathrm{sign}(Z)\cdot\log|Z|$

    with $Z \sim \mathcal{N}(0, \sigma^2)$.
- Kroshnin (2021, Theorem 4.6.1) has a logarithmic rate Wasserstein example.

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,
- Alleviated for diffusion means (Eltzner et al., 2023).
- Only rather well understood (McKilliam et al. (2012), Hotz and H. (2015), Hundrieser et al. (2020)) on the circle $\mathfrak{X} = \mathbb{S}^1$, $d =$ arc length, with new example (for all $r > 0$, $\mu = 0$):
  - $\hat{\mu}_n \xrightarrow{a.s.} 0$
  - $\left(log\sqrt{n}\right)^{1/r} \hat{\mu}_n \xrightarrow{\mathcal{D}} \frac{1}{2}(\delta_{-1} + \delta_1)$
  - $r \left(\log \sqrt{n}\right)^{(1+r)/r} \left(\hat{\mu}_n - \frac{\mathrm{sign}(\hat{\mu}_n)}{\left(\log \sqrt{n}\right)^{1/r}}\right) \xrightarrow{\mathcal{D}} \mathrm{sign}(Z) \cdot \log |Z|$

    with $Z \sim \mathcal{N}(0, \sigma^2)$.
- Kroshnin (2021, Theorem 4.6.1) has a logarithmic rate Wasserstein example.
- In addition to deteriorating constants (near nonuniqeness) $\exists$ finite sample (stickiness/smeariness) phenomena.

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,
- Alleviated for diffusion means (Eltzner et al., 2023).
- Only rather well understood (McKilliam et al. (2012), Hotz and H. (2015), Hundrieser et al. (2020)) on the circle $\mathfrak{X} = \mathbb{S}^1$, $d =$ arc length, with new example (for all $r > 0$, $\mu = 0$):
  - $\hat{\mu}_n \xrightarrow{a.s.} 0$
  - $\left(log\sqrt{n}\right)^{1/r} \hat{\mu}_n \xrightarrow{\mathcal{D}} \frac{1}{2}(\delta_{-1} + \delta_1)$
  - $r\left(\log\sqrt{n}\right)^{(1+r)/r}\left(\hat{\mu}_n - \frac{\mathrm{sign}(\hat{\mu}_n)}{\left(\log\sqrt{n}\right)^{1/r}}\right) \xrightarrow{\mathcal{D}} \mathrm{sign}(Z)\cdot\log|Z|$

    with $Z \sim \mathcal{N}(0, \sigma^2)$.
- Kroshnin (2021, Theorem 4.6.1) has a logarithmic rate Wasserstein example.
- In addition to deteriorating constants (near nonuniqeness)
  $\exists$ finite sample (stickiness/smeariness) phenomena.
- Smeariness begets finite sample smeariness, detect by bootstrapping variance modulation (Pennec, 2019):

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

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,
- Alleviated for diffusion means (Eltzner et al., 2023).
- Only rather well understood (McKilliam et al. (2012), Hotz and H. (2015), Hundrieser et al. (2020)) on the circle $\mathfrak{X} = \mathbb{S}^1$, $d =$ arc length, with new example (for all $r > 0$, $\mu = 0$):
  - $\hat{\mu}_n \xrightarrow{a.s.} 0$
  - $\left(log\sqrt{n}\right)^{1/r} \hat{\mu}_n \xrightarrow{\mathcal{D}} \frac{1}{2}(\delta_{-1} + \delta_1)$
  - $r\left(\log\sqrt{n}\right)^{(1+r)/r}\left(\hat{\mu}_n - \frac{\mathrm{sign}(\hat{\mu}_n)}{\left(\log\sqrt{n}\right)^{1/r}}\right) \xrightarrow{\mathcal{D}} \mathrm{sign}(Z)\cdot\log|Z|$

  with $Z \sim \mathcal{N}(0, \sigma^2)$.
- Kroshnin (2021, Theorem 4.6.1) has a logarithmic rate Wasserstein example.
- In addition to deteriorating constants (near nonuniqeness) $\exists$ finite sample (stickiness/smeariness) phenomena.
- Smeariness begets finite sample smeariness, detect by bootstrapping variance modulation (Pennec, 2019):

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

- That the bootstrap is valid has only been shown on the circle (Hundrieser et al., 2020) ...

Outlook

- Research on nonparametric rates: stickiness (faster), smeariness (slower) is still in its beginnings, cf. also Eltzner (2022) for geometrical vs. topological smeariness and Eltzner (2020) how to test for nonuniqueness,
- Alleviated for diffusion means (Eltzner et al., 2023).
- Only rather well understood (McKilliam et al. (2012), Hotz and H. (2015), Hundrieser et al. (2020)) on the circle $\mathfrak{X} = \mathbb{S}^1$, $d =$ arc length, with new example (for all $r > 0$, $\mu = 0$):
  - $\hat{\mu}_n \xrightarrow{a.s.} 0$
  - $\left(log\sqrt{n}\right)^{1/r} \hat{\mu}_n \xrightarrow{\mathcal{D}} \frac{1}{2}(\delta_{-1} + \delta_1)$
  - $r\left(\log\sqrt{n}\right)^{(1+r)/r}\left(\hat{\mu}_n - \frac{\operatorname{sign}(\hat{\mu}_n)}{\left(\log\sqrt{n}\right)^{1/r}}\right) \xrightarrow{\mathcal{D}} \operatorname{sign}(Z)\cdot\log|Z|$

    with $Z \sim \mathcal{N}(0, \sigma^2)$.
- Kroshnin (2021, Theorem 4.6.1) has a logarithmic rate Wasserstein example.
- In addition to deteriorating constants (near nonuniqeness)
  $\exists$ finite sample (stickiness/smeariness) phenomena.
- Smeariness begets finite sample smeariness, detect by bootstrapping variance modulation (Pennec, 2019):
  $$\mathfrak{m}_n := \frac{n\,\mathbb{E}[d^2(\hat{\mu}_n, \mu)]}{\mathbb{E}[d^2(X, \mu)]}$$
- That the bootstrap is valid has only been shown on the circle (Hundrieser et al., 2020) ...
- All of the nonreferenced and new results are from Hundrieser et al. (2024).

References

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

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

Barden, D., H. Le, and M. Owen (2018). Limiting behaviour of fréchet means in the space of phylogenetic trees. *Annals of the Institute of Statistical Mathematics 70*(1), 99–129.

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

Billera, L., S. Holmes, and K. Vogtmann (2001). Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics 27*(4), 733–767.

Dryden, I. L. and K. V. Mardia (2016). *Statistical Shape Analysis* (2nd ed.). Chichester: Wiley.

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

Eltzner, B. (2022). Geometrical smeariness–a new phenomenon of fréchet means. *Bernoulli 28*(1), 239–254.

Eltzner, B., P. E. Hansen, S. F. Huckemann, and S. Sommer (2023). Diffusion means in geometric spaces. *Bernoulli 29*(4), 3141–3170.

Evans, S. N. and A. Q. Jaffe (2024). Limit theorems for fréchet mean sets. *Bernoulli 30*(1), 419–447.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut de Henri Poincaré 10*(4), 215–310.

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika 40*, 33–51.

Groisser, D. (2005). On the convergence of some Procrustean averaging algorithms. *Stochastics: Internatl. J. Probab. Stochstic. Processes 77*(1), 51–60.

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

Huckemann, S. (2011). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *The Annals of Statistics 39*(2), 1098–1124.

Huckemann, S. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Annals of the Institute of Statistical Mathematics 64*(6), 1227–1259.

Huckemann, S., T. Hotz, and A. Munk (2010). Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). *Statistica Sinica 20*(1), 1–100.

Hundrieser, S., B. Eltzner, and S. F. Huckemann (2020). Finite sample smeariness of Fréchet means and application to climate. *arXiv preprint arXiv:2005.02321*.

Hundrieser, S., B. Eltzner, and S. F. Huckemann (2024). A lower bound for estimating fréchet means.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics XXX*, 509–541.

Kendall, W. S. (1990). Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence. *Proceedings of the London Mathematical Society 61*, 371–406.

Kroshnin, A. (2021). *Inside and around Wasserstein barycenters.* Ph. D. thesis, Université de Lyon; Kharkevich Institute.

Kuratowski, K. (1948). *Topologie I.*

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

Le, H. (2001). Locating Fréchet means with an application to shape spaces. *Advances of Applied Probability (SGSA) 33*(2),

References

Le, H. (2001). Locating Fréchet means with an application to shape spaces. *Advances of Applied Probability (SGSA) 33*(2), 324–338.

Le Gouic, T., Q. Paris, P. Rigollet, and A. J. Stromme (2022). Fast convergence of empirical barycenters in Alexandrov spaces and the Wasserstein space. *Journal of the European Mathematical Society 25*(6), 2229–2250.

Lueg, J., M. K. Garba, T. M. Nye, J. Lueg, and S. F. Huckemann (2024). Foundations of wald space for statistics of phylogenetic trees. *Journal of the London Mathematical Society 109*(5), e12893.

Mattingly, J. C., E. Miller, and D. Tran (2023). Central limit theorems for féchet means on stratified spaces. manuscript.

McKilliam, R. G., B. G. Quinn, and I. V. L. Clarkson (2012). Direction estimation by minimum squared arc length. *IEEE Transactions on Signal Processing 60*(5), 2115–2124.

Panaretos, V. and Y. Zemel (2020). *An invitation to statistics in Wasserstein space*. Cham: Springer.

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

Schötz, C. (2022). Strong laws of large numbers for generalizations of fréchet mean sets. *Statistics 56*(1), 34–52.

Sturm, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature, heat kernels and analysis on manifolds, graphs, and metric spaces. *Contemporary Mathematics 338*, 357–390.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer.

Wiechers, H., M. Zobel, M. Bennati, I. Tkach, B. Eltzner, S. Huckemann, and Y. Pokern (2023). Drift models on complex projective space for electron-nuclear double resonance. *arXiv preprint arXiv:2307.12414*.

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