Extrinsic vs Intrinsic Means on the Circle

Thomas Hotz

Institute for Mathematics
Technische Universität Ilmenau

Geometric Science of Information
Paris, 28 August 2013
Circular data

- given: i.i.d. random elements $X, X_1, \ldots, X_n$ on

\[ T^1 = \mathbb{R} / 2\pi \mathbb{Z}, \]

i.e. abstract set of equivalence classes

- representation:

\[ p : T^1 \rightarrow [-\pi, \pi), \]

s.t. $x = [p(x)]$
Circular data

- given: i.i.d. random elements $X, X_1, \ldots, X_n$ on

$$\mathbb{T}^1 = \mathbb{R} / 2\pi\mathbb{Z},$$

i.e. abstract set of equivalence classes

- representation:

$$\rho : \mathbb{T}^1 \to [-\pi, \pi),$$

s.t. $x = [\rho(x)]$

aim: define mean!
Circular data

- given: i.i.d. random elements $X, X_1, \ldots, X_n$ on $T^1 = \mathbb{R} / 2\pi \mathbb{Z}$,

  i.e. abstract set of equivalence classes

- representation:

  $p : T^1 \rightarrow [-\pi, \pi)$,

  s.t. $x = [p(x)]$

aim: define mean!

problem: $T^1$ not a vector space!
Fréchet mean

given a metric $d$ on $\mathbb{T}^1$, define

- the set of *population Fréchet means*

$$\mu_d = \arg\min_{\nu \in \mathbb{T}^1} \mathbb{E} d(X, \nu)^2$$

- the set of *empirical Fréchet means*

$$\hat{\mu}_{d,n} = \arg\min_{\nu \in \mathbb{T}^1} \sum_{j=1}^{n} d(X_j, \nu)^2$$
Fréchet mean

given a metric $d$ on $\mathbb{T}^1$, define

- the set of *population Fréchet means*

\[
\mu_d = \arg\min_{\nu \in \mathbb{T}^1} \mathbb{E} \, d(X, \nu)^2
\]

- the set of *empirical Fréchet means*

\[
\hat{\mu}_{d,n} = \arg\min_{\nu \in \mathbb{T}^1} \sum_{j=1}^{n} d(X_j, \nu)^2
\]

**needed:** metric on $\mathbb{T}^1$
Extrinsic and intrinsic metric

extrinsic metric:

- embed $\mathbb{T}^1$ as unit circle in $\mathbb{C}$:
  \[ \zeta : \mathbb{T}^1 \to \mathbb{C}, \, x \mapsto \exp(ip(x)) \]
- use Euclidean distance on $\mathbb{C}$:
  \[ d_E(x, y) = |\zeta(x) - \zeta(y)|, \]
  i.e. chordal distance
- obtain extrinsic means
  \[ \varepsilon = \mu_{d_E} \text{ and } \hat{\varepsilon}_n = \hat{\mu}_{d_E,n} \]
Extrinsic and intrinsic metric

**Extrinsic metric:**

- **Embed** $\mathbb{T}^1$ as unit circle in $\mathbb{C}$:
  \[ \zeta : \mathbb{T}^1 \rightarrow \mathbb{C}, x \mapsto \exp(ip(x)) \]

- **Use Euclidean distance on $\mathbb{C}$:**
  \[ d_E(x, y) = |\zeta(x) - \zeta(y)|, \]
  i.e. chordal distance

- **Obtain extrinsic means**
  \[ \varepsilon = \mu_{d_E} \text{ and } \hat{\varepsilon}_n = \hat{\mu}_{d_E,n} \]

**Intrinsic metric:**

- **Use quotient metric**
  \[ d_I(x, y) = |p(x - y)| = \min_{a \in [x], b \in [y]} |a - b|, \]
  i.e. arc-length

- **Obtain intrinsic means**
  \[ \eta = \mu_{d_I} \text{ and } \hat{\eta}_n = \hat{\mu}_{d_I,n} \]
Extrinsic and intrinsic metric

**Extrinsic metric:**

- embed $\mathbf{T}^1$ as unit circle in $\mathbf{C}$:
  \[ \zeta : \mathbf{T}^1 \to \mathbf{C}, \, x \mapsto \exp(ip(x)) \]
- use Euclidean distance on $\mathbf{C}$:
  \[ d_E(x, y) = |\zeta(x) - \zeta(y)|, \]
  i.e. chordal distance
- obtain extrinsic means
  \[ \varepsilon = \mu_{d_E} \text{ and } \hat{\varepsilon}_n = \hat{\mu}_{d_E,n} \]

**Intrinsic metric:**

- use quotient metric
  \[ d_I(x, y) = |p(x - y)| = \min_{a \in [x], b \in [y]} |a - b|, \]
  i.e. arc-length
- obtain intrinsic means
  \[ \eta = \mu_{d_I} \text{ and } \hat{\eta}_n = \hat{\mu}_{d_I,n} \]

Topologies agree: compact space $\Rightarrow$ existence of means
Extrinsic and intrinsic metric

**extrinsic metric:**
- embed $\mathbb{T}^1$ as unit circle in $\mathbb{C}$:
  \[ \zeta : \mathbb{T}^1 \to \mathbb{C}, x \mapsto \exp(ip(x)) \]
- use Euclidean distance on $\mathbb{C}$:
  \[ d_E(x, y) = |\zeta(x) - \zeta(y)|, \]
  i.e. chordal distance
- obtain extrinsic means
  \[ \varepsilon = \mu_{d_E} \text{ and } \hat{\varepsilon}_n = \hat{\mu}_{d_E,n} \]

**intrinsic metric:**
- use quotient metric
  \[ d_I(x, y) = |p(x - y)| = \min_{a \in [x], b \in [y]} |a - b|, \]
  i.e. arc-length
- obtain intrinsic means
  \[ \eta = \mu_{d_I} \text{ and } \hat{\eta}_n = \hat{\mu}_{d_I,n} \]

Topologies agree: compact space $\Rightarrow$ existence of means

Which mean is “better”?
Uniqueness

extrinsic:

- for $Z = \zeta(X)$:
  
  $$\varepsilon = \text{Arg} \ E \ Z$$

- if $EZ \neq 0$, $\varepsilon = T^1$ otherwise
- “classic” circular mean
- thus $\varepsilon$ unique iff $EZ \neq 0$
Uniqueness

extrinsic:

- for $Z = \zeta(X)$:
  \[ \varepsilon = \text{Arg } E Z \]

  if $EZ \neq 0$, $\varepsilon = T^1$ otherwise

- “classic” circular mean

- thus $\varepsilon$ unique iff $EZ \neq 0$

intrinsic:

Theorem (H. & Huckemann (2011))

If $p(X)$ has Lebesgue density $f$ which is $< \frac{1}{2\pi}$ on open arcs $S_1, \ldots S_k$ then $X$ has at most $k$ intrinsic means, at most one opposite each arc.
Uniqueness

extrinsic:

- for $Z = \zeta(X)$:
  
  $$\varepsilon = \text{Arg } E Z$$

- if $EZ \neq 0$, $\varepsilon = T^1$ otherwise

- "classic" circular mean

- thus $\varepsilon$ unique iff $EZ \neq 0$

intrinsic:

Theorem (H. & Huckemann (2011))

If $p(X)$ has Lebesgue density $f$ which is $< \frac{1}{2\pi}$ on open arcs $S_1, \ldots S_k$ then $X$ has at most $k$ intrinsic means, at most one opposite each arc.

neither uniqueness implies the other, but
Uniqueness

extrinsic:

- for $Z = \zeta(X)$:
  
  $\varepsilon = \text{Arg} \ E \ Z$

  if $EZ \neq 0$, $\varepsilon = T^1$ otherwise

- “classic” circular mean

- thus $\varepsilon$ unique iff $EZ \neq 0$

intrinsic:

Theorem (H. & Huckemann (2011))

If $p(X)$ has Lebesgue density $f$ which is $< \frac{1}{2\pi}$ on open arcs $S_1, \ldots, S_k$ then $X$ has at most $k$ intrinsic means, at most one opposite each arc.

neither uniqueness implies the other, but

Proposition

If there is only one such arc, then both $\varepsilon$ and $\eta$ are unique.
Computational complexity

extrinsic:

\[ \hat{e}_n = \text{Arg } \bar{Z}_n \text{ for} \]

\[ \bar{Z}_n = \frac{1}{n} \sum_{j=1}^{n} \zeta(X_j) \]

thus \( \hat{e}_n \) computable in \( \mathcal{O}(n) \) time
Computational complexity

extrinsic:

- $\hat{e}_n = \text{Arg } \bar{Z}_n$ for

\[
\bar{Z}_n = \frac{1}{n} \sum_{j=1}^{n} \zeta(X_j)
\]

- thus $\hat{e}_n$ computable in $\mathcal{O}(n)$ time

intrinsic:

- H. & Huckemann (2011): intrinsic means $\hat{\eta}_n$ are vertices of regular $n$-gon
- can be computed in $\mathcal{O}(n)$ time after sorting, i.e. total $\mathcal{O}(n \log n)$ time

- McKilliam et al. (2012): computable in $\mathcal{O}(n)$ time using lattice algorithm
Robustness

breakdown point = 0 in both cases, but
Robustness

breakdown point = 0 in both cases, but

extrinsic:

Proposition

If $|EZ| > 2\delta$ then for any r.e. $\tilde{X}$ on $T^1$ with $d_{TV}(P^X, P^{\tilde{X}}) \leq \delta$ then

- $E \zeta(\tilde{X}) > \delta$,
- $\tilde{\epsilon} = \text{Arg } E \tilde{Z}$ is unique, and
- $|\epsilon - \tilde{\epsilon}| \leq \sin\left(\frac{2\delta}{|EZ|}\right)$. 
Robustness

breakdown point = 0 in both cases, but

extrinsic:

**Proposition**

*If \(|E Z| > 2\delta\) then for any r.e. \(\tilde{X}\) on \(T^1\) with \(d_{TV}(P^X, P^{\tilde{X}}) \leq \delta\) then*

- \(E \zeta(\tilde{X}) > \delta\),
- \(\tilde{\epsilon} = \text{Arg} E \tilde{Z}\) is unique, and
- \(|\epsilon - \tilde{\epsilon}| \leq \sin\left(\frac{2\delta}{|E Z|}\right)\).

intrinsic:

- uniqueness may be lost by arbitrarily small perturbations (in TV distance)
Asymptotics
Fréchet means fulfill strong law of large numbers for sets (Ziezold (1977), Bhattacharya & Patrangenaru (2003))
Asymptotics
Fréchet means fulfill strong law of large numbers for sets (Ziezold (1977), Bhattacharya & Patrangenaru (2003))

extrinsic:

Theorem (by $\delta$-method)

If $\varepsilon = [0]$ is unique, then

$$
\sqrt{n} \, p(\hat{\varepsilon}_n) \xrightarrow{d} \mathcal{N}
\left(0, \frac{\mathbb{E} \sin^2 p(X)}{\mathbb{E} Z^2}\right).
$$
Asymptotics

Fréchet means fulfill strong law of large numbers for sets (Ziezold (1977), Bhattacharya & Patrangenaru (2003))

extrinsic:

Theorem (by $\delta$-method)

If $\varepsilon = [0]$ is unique, then

$$\sqrt{n} \, p(\hat{\varepsilon}_n) \xrightarrow{D} \mathcal{N}\left(0, \frac{\mathbb{E} \sin^2 p(X)}{\mathbb{E} |Z|^2}\right).$$

intrinsic:

Theorem (H. & Huckemann (2011), McKilliam et al. (2012))

If $\eta = [0]$ is unique, and near $[\pi]$, $p(X)$ has cont. Lebesgue density $f$, then: if $f(-\pi) < \frac{1}{2\pi}$,

$$\sqrt{n} \, p(\hat{\eta}_n) \xrightarrow{D} \mathcal{N}\left(0, \frac{\mathbb{E} p(X)^2}{(1 - 2\pi f(-\pi))^2}\right)$$

while for $f(-\pi) = \frac{1}{2\pi}$, a CLT with slower rate holds if derivatives are continuous there.
Asymptotic relative efficiency

both means respect translations (rotations)
– which has smaller variance?
Asymptotic relative efficiency

both means respect translations (rotations) – which has smaller variance?

Proposition

Let $\mathcal{P}$ the class of distributions $p^X$ s.t. $p(X)$ has even Lebesgue density $f$, non-decreasing on $[-\pi, 0]$ with $f(0) > f(-\pi)$; then $\epsilon = \eta = [0]$ are both unique, and

$$\inf_{p^X \in \mathcal{P}} \lim_{n \to \infty} \frac{\text{Var} p(\hat{\eta}_n)}{\text{Var} p(\hat{\epsilon}_n)} = 0 \quad \text{while} \quad \inf_{p^X \in \mathcal{P}} \lim_{n \to \infty} \frac{\text{Var}(\hat{\epsilon}_n)}{\text{Var}(\hat{\eta}_n)} \geq \frac{1}{2\pi^2}.$$
Universal confidence sets

extrinsic:

**Proposition**

For $\alpha \in (0, 1)$ let $n \in \mathbb{N}$ be large enough such that $\alpha n > 1$, assume that $\mathbf{E} Z \neq 0$, i.e. unique $\varepsilon$. If $I_n = (\mathbf{A}rg \bar{Z}_n - \delta_n, \mathbf{A}rg \bar{Z}_n + \delta_n) \subset T^1$

with $\sin \delta_n = (|\bar{Z}_n| \sqrt{\alpha n})^{-1}$, then $P(\varepsilon \in I_n) > 1 - \alpha$ and the length of $I_n$, i.e. the Lebesgue measure of $p(I_n)$, tends to zero with the optimal $\sqrt{n}$-rate.
Universal confidence sets

extrinsic:

**Proposition**

For $\alpha \in (0, 1)$ let $n \in \mathbb{N}$ be large enough such that $\alpha n > 1$, assume that $E Z \neq 0$, i.e. unique $\epsilon$. If $I_n = (\text{Arg} \ Z_n - \delta_n, \text{Arg} \ Z_n + \delta_n) \subset T^1$ with $\sin \delta_n = (|Z_n| \sqrt{\alpha n})^{-1}$, then $P(\epsilon \in I_n) > 1 - \alpha$ and the length of $I_n$, i.e. the Lebesgue measure of $p(I_n)$, tends to zero with the optimal $\sqrt{n}$-rate.

intrinsic:

- $\sqrt{n}$-rate cannot be achieved universally
- such universal confidence sets appear impossible due to non-robustness
## Summary

<table>
<thead>
<tr>
<th>Extrinsic</th>
<th>vs</th>
<th>Intrinsic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+)</td>
<td>uniqueness</td>
<td>(−)</td>
</tr>
<tr>
<td>(+)</td>
<td>computational complexity</td>
<td>(−)</td>
</tr>
<tr>
<td>+</td>
<td>robustness</td>
<td>−</td>
</tr>
<tr>
<td>−</td>
<td>relative efficiency</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>universal conf. sets</td>
<td>−</td>
</tr>
</tbody>
</table>
Thank you for your attention!

Please see proceedings for references.