Nonparametric Inference for Location Parameters via Fréchet functions

Vic Patrangenaru
Kouadio D. Yao and Ruite Guo

SMRLO'16 02/16/2016
Be'er Sheva, ISRAEL

Thanks

Thanks Jean Patrin for his tremendous help with logistics for my summer meetings in Europe and Israel!
Thanks Vladimir and Tedy Kiss and the for their help during my stay in Israel!
Thanks Ilia Frenkel for his invitation and help with scheduling my talk!
Thanks our NSA-MSP-H98230-15-1-0135 and NSF-DMS-1106935 for their support!

DASS-1

- Spaces of phylogenetic trees were first examples of sample spaces that are not manifolds were given in Billera et al.(2001). My original contact with DASS was during a 2002 IMS meeting in Banff, when S. Holmes presented some of her results on phylogenetic trees.
- During a 2007 visit in Göttingen, I learned from Stephan Huckemann that Fréchet sample means on *mean of a cone* might have a different behavior, if the population mean is its vertex, when compared with the case, when they don't.
- Another note was a in an 2010 e-mail conversation with J. S. Marron, organizer of the *Analysis of Object Data (AOD)* program at SAMSI in 2010 - 2011, where he made the point there is an interesting behavior of Fréchet means on sample spaces that have singular points, such as spaces of trees, compared with the case of sample means on manifolds.
- Thirdly, during a *crucial discussion* with Ezra Miller at SAMSI, 2010, I realized that AOD = DASS

DASS-2

- A working group of *Data Analysis on Sample Spaces with a Manifold Stratification*, was created at SAMSI in 2010, and an early version of the CLT for Fréchet sample means on Spiders was proved by Hotz, Huckemann et al.(2010), including a *stickiness* property
- The very inclusive subject of DASS, advanced during the MBI Workshop of *Statistics, Geometry, and Combinatorics on Stratified Spaces Arising from Biological Problems*, and in a Meeting of the ISNPS (Chalkidiki, 2012), and picked up some steam in the *Statistics and Geometry in Bioimaging: Manifolds and Stratified Spaces* conference, in Denmark, in 2013.
- some real data driven computation of means on phylogenetic trees are due Sqwerer et al (2014), Patrangenaru et al (2014)
- additional TDA work in DASS is due to Marron et al.(2014)

So ... What is a Stratified Space?

- A **stratified space (space with a manifold stratification (Hasler Whitney))** is a metric space M that admits a *filtration* $\emptyset = F_{-1} \subseteq F_0 \subseteq F_1 \ldots \subseteq F_n \subset \cdots = M = \cup_{i=0}^{\infty} F_i$, by closed subspaces, such that for each $i = 1, \ldots, F_i \backslash F_{i-1}$ is empty or is an i-dimensional manifold, called the $i - th$ stratum.
- The *regular part* of M is the highest dimensional stratum. At each regular point the stratified space has a tangent space
- The dimension of the stratified space is m if $M = F_m \neq F_{m-1}$, otherwise $dimM = \infty$. All the strata of dimension lower than m are *singular:* at each of their points, the stratified space does not have a tangent point.

Figure: A 2D stratified space - T_4, space of trees with four leafs

Manifolds and Sample Spaces with a Manifold Structure

Manifold stratified space is defined in terms of a **manifold** stratification.

- A manifold is a metric space M is locally homeomorphic to an Euclidean space, via homemorphisms forming an *atlas*; the transition map $k \circ k^{-1}$ of two such homeomorphisms h, k is smooth.
- At each of its points, a manifold has a *tangent space*, made of vectors tangent at that point, to differentiable paths on the manifold.
- Manifolds arise as non-Euclidean sample spaces: *spaces of axes in 3D, spaces of planar direct similarity shapes of k-ads (Kendall shape spaces), spaces of projective shapes of spatial configurations of k-ads in general position, etc*
- All the sample spaces listed above are **compact**.

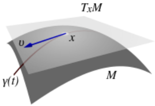


Figure: A tangent space at a point on a 2D manifold

Why Should Statisticians Care about Stratified Spaces?

Data **does not live on numerical spaces**. The sample spaces have in fact a complicated nonlinear structure. Simple example include

- tree structured data including phylogenetic trees
- astronomy and cosmology data: galaxies, stars and planetary orbits data
- Spatial Statistics : temperatures, snow and other functions measured across the planet
- vector fields of wind velocities on the Earth surface
- Geology : paleomagnetic data, plate tectonics, volcanos
- morphometric data
- protein and DNA structures
- medical imaging outputs, including : angiography data (such as brain vessels structure), CT, MRI
- satellite or aerial imaging
- digital camera imaging data
- internet data, may be (Show me the data)

Means on Manifolds

- Cartan (1929) considers for the first time the barycenter of a finite set of points on a Hadamard manifold. Such a barycenter will be called a Cartan (sample) mean. Note that the notion of manifold was introduced much later by de Rham (1946).
- Fréchet (1946) extends the notion of Cartan sample mean (without referring to Cartan, most likely being unaware of Cartan's work in Riemannian Geometry) to what today is called Fréchet mean set of a random object on a complete separable metric space. This is a more general setting than manifolds .
- Fréchet (1946) suggests to analyze, for example the shape of a (random) contour or the shape of eggs in an basket.
- Kobayashi and Nomizu (1969) mention Cartan's work on barycenters on CAT(0) spaces
- Karcher (1977) considered barycenters on finite sets of points on arbitrary manifolds

Data Analysis on Manifolds - History Note - I

- Ziezold reconsiders *Fréchet mean sets* of random points X on complete separable metric spaces (M, ρ), as minimizers of $x \to E(\rho^2(x, X)$, proving their *consistency*
- Fisher et al.(1996) estimate the mean axis for an arbitrary distribution on $\mathbb{R}P^d$.
- Patrangenaru (1998) defines *extrinsic and intrinsic means on manifolds* and derives a CLT for extrinsic means on manifolds, using a Fisher et al.(1996) type of sample estimator
- Hendricks and Landsman (1996, 1998) derive a CLT for extrinsic means on manifolds, using a different estimator, whose covariance matrix obtained from the Weingarten map Weingarten , and derive the first two sample test for extrinsic means on manifolds
- Bhattacharya and Patrangenaru (2003, 2005) derive a general asymptotic theory for Fréchet means on manifolds, as well as its nonparametric bootstrap corollaries.
- Bandulasiri et al.(2009), Patrangenaru et. al(2010), Crane and Patrangenaru (2011) use nonparametric bootstrap methods in 3D shape analysis.

Data Analysis on Manifolds - History Note - II

- Bhattacharya et. al. (2012) show that extrinsic data analysis on manifolds is faster than intrinsic analysis, thus rending intrinsic data analysis in general obsolete.
- Bhattacharya and Bhattacharya (2012) extend the work of Hendricks and Landsman (1998) to two sample tests for means extrinsic means of independent populations on arbitrary manifolds
- Osborne et al (2013) used Cholesky decompositions to derive to sample tests for mean DTI outputs on the symmetric space of positive semidefinite 3×3 matrices.
- Guo at al. (2014) considered empirical methods for nonparametric Bootstrap test for equality of extrinsic Mean reflection shapes with application to biological shape change
- Patrangenaru et al (2014) use nonparametric bootstrap to derive two sample tests for mean change on Lie groups, and apply their results to 3D projective shape analysis

Example of Stratified Space of Phylogenetic Trees with *p* Leaves

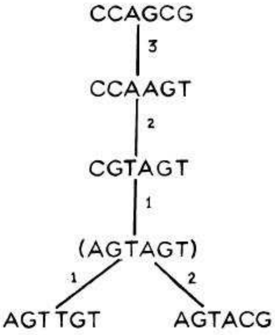
A phylogenetic tree with *p* leaves is an equivalence class based on a certain equivalence, of a DNA-based connected directed graph of species with no loops, having an unobserved *root* (common ancestor) and *p* observed *leaves* (current observed species of a certain family of living creatures).

```
Human    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
Gorilla  MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
Rabbit   MVHLSSEEKSAVTALWGKVNVEEVGGEALGRLLVVYPWTQRFFESFGDLSSANAVMNNPK
Cow      M..LTAEEKAAVTAFWGKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLSTADAVMNNPK
Goat     M..LTAEEKAAVTGFWGKVKVDEVGAEALGRLLVVYPWTQRFFEHFGDLSSADAVMNNAK
Mouse    MVHLTDAEKAAVSCLWGKVNSDEVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGNAK
Chicken  MVHWTAEEKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFFASFGNLSSPTAILGNPM
Carp     MVEWTDAERSAIIGLWGKLNPDELGPQALARCLIVYPWTQRYFASFGNLSSPAAIMGNPK

         61                                                          120
Human    VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
Gorilla  VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFKLLGNVLVCVLAHHFG
Rabbit   VKAHGKKVLAAFSEGLSHLDNLKGTFAKLSELHCDKLHVDPENFRLLGNVLVIVLSHHFG
Cow      VKAHGKKVLDSFSNGMKHLDDLKGTFAALSELHCDKLHVDPENFKLLGNVLVVVLARNFG
Goat     VKAHGKKVLDSFSNGMKHLDDLKGTFAQLSELHCDKLHVDPENFKLLGNVLVVVLARHHG
Mouse    VKAHGKKVITAFNDGLNHLDSLKGTFASLSELHCDKLHVDPENFRLLGNMIVIVLGHHLG
Chicken  VRAHGKKVLTSFGDAVKNLDNIKNTFSQLSELHCDKLHVDPENFRLLGDILIIVLAAHFS
Carp     VAAHGRTVMGGLERAIKNMDNIKATYAPLSVMHSEKLHVDPDNFRLLADCITVCAAMKFG

          121                        148
Human    .KEFTPPVQAAYQKVVAGVANALAHKYH
Gorilla  .K..........................
Rabbit   .KEFTPQVQAAYQKVVAGVANALAHKYH
Cow      .KEFTPVLQADFQKVVAGVANALAHRYH
Goat     .SEFTPLLQAEFQKVVAGVANALAHRYH
Mouse    .KDFTPAAQAAFQKVVAGVATALAHKYH
Chicken  .KDFTPECQAAWQKLVRVVAHALARKYH
Carp     PSGFSPNVQEAWQKFLSVVVSALCRQYH
```

Example of a Phylogenetic Tree Building



Tree Spaces

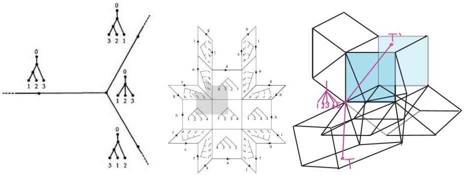


Figure: Tree spaces T_3, T_4, T_5.

- A tree with p leaves is a connected, simply connected graph , with a distinguished vertex, labeled o, called the *root*, and p vertices of degree 1, called leaves, that are labeled from 1 to p. In addition, we assume that with all interior edges have positive lengths. (An edge of a p-tree is called interior if it is not connected to a leaf.) See Billera et. al.(2001).
- Now consider a tree T, with interior edges $e_1, \ldots, e_r$ of lengths $l_1, \ldots, l_r$ respectively. If T is binary, then $r = p - 2$, otherwise $r < n - 2$. The vector $(l_1, \ldots, l_r)^T$ specifies a point in the positive open *orthant* $(0, \infty)^r$.

Eukaryotes mean tree example

A first data driven example of data driven intrinsic mean computation on the space of phylogenetic trees T_4 is was given in Ellingson et. al.(2014).

	Taxon \ Character	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	00001::Perkinsus_marinus::AF497479	-	-	-	-	-	-	A	A	C	C	U	G	G	U	-	U	G	A	-	U	C	C	U	-	G
2	00009::Perkinsus_atlanticus::AF140295	-	-	-	-	-	-	-	A	C	C	U	G	G	U	-	U	G	A	-	U	C	C	U	-	G
3	00010::Perkinsus_atlanticus::AF509333	-	-	-	-	-	-	-	-	C	C	U	G	G	U	-	U	G	A	-	U	C	C	U	-	G
4	00011::Perkinsus_marinus::X75762	-	-	-	-	-	-	A	A	C	C	U	G	G	U	-	U	G	A	-	U	C	C	U	-	G
5	00012::Perkinsus_marinus::AF126013	-	-	-	-	-	-	A	A	C	C	U	G	G	U	-	U	G	A	-	U	C	C	U	-	G
6	00013::Perkinsus_marinus::AF497479	-	-	-	-	-	-	A	A	C	C	U	G	G	U	-	U	G	A	-	U	C	C	U	-	G
7	00014::Perkinsus_sp::L07375	-	-	-	-	-	-	-	-	-	-	U	G	G	U	-	U	G	A	-	U	C	C	U	-	G
8	00015::Perkinsus_sp.__isolate_G117_::AF042707	-	-	-	-	-	-	A	A	C	C	U	G	G	U	-	U	G	A	-	U	C	C	U	-	G
9	00016::Perkinsus_sp.__isolate_H49_::AF042708	-	-	-	-	-	-	A	A	C	C	U	G	G	U	-	U	G	A	-	U	C	C	U	-	G
10	00017::Perkinsus_sp._CCA2001::AF252288	-	-	-	-	-	-	-	A	C	C	U	G	G	U	-	U	G	A	-	U	C	C	U	-	G

	Taxon \ Character	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	00001::Perkinsus_marinus::AF497479	C	C	A	G	U	A	G	U	C	A	-	U	A	U	G	C	U	U	G	U	C	U	C	A	A
2	00009::Perkinsus_atlanticus::AF140295	C	C	A	G	U	A	G	U	C	A	-	U	A	U	G	C	U	U	G	U	C	U	C	A	A
3	00010::Perkinsus_atlanticus::AF509333	C	C	A	G	U	A	G	U	C	A	-	U	A	U	G	C	U	U	G	U	C	U	C	A	A
4	00011::Perkinsus_marinus::X75762	C	C	A	G	U	A	G	U	C	A	-	U	A	U	G	C	U	U	G	U	C	U	C	A	A
5	00012::Perkinsus_marinus::AF126013	C	C	A	G	U	A	G	U	C	A	-	U	A	U	G	C	U	U	G	U	C	U	C	A	A
6	00013::Perkinsus_marinus::AF497479	C	C	A	G	U	A	G	U	C	A	-	U	A	U	G	C	U	U	G	U	C	U	C	A	A
7	00014::Perkinsus_sp::L07375	C	C	A	G	U	A	G	U	C	A	-	U	A	U	G	C	U	U	G	U	C	U	C	A	A
8	00015::Perkinsus_sp.__isolate_G117_::AF042707	C	C	A	G	U	A	G	U	C	A	-	U	A	U	G	C	U	U	G	U	C	U	C	A	A
9	00016::Perkinsus_sp.__isolate_H49_::AF042708	C	C	A	G	U	A	G	U	C	A	-	U	A	U	G	C	U	U	G	U	C	U	C	A	A
10	00017::Perkinsus_sp._CCA2001::AF252288	C	C	A	G	U	A	G	U	C	A	-	U	A	U	G	C	U	U	G	U	C	U	C	A	A

Figure: DNA sequences Eukaryotes species

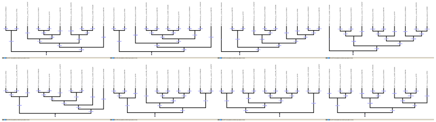


Figure: Yule speciation trees for Perkinsus DNA data

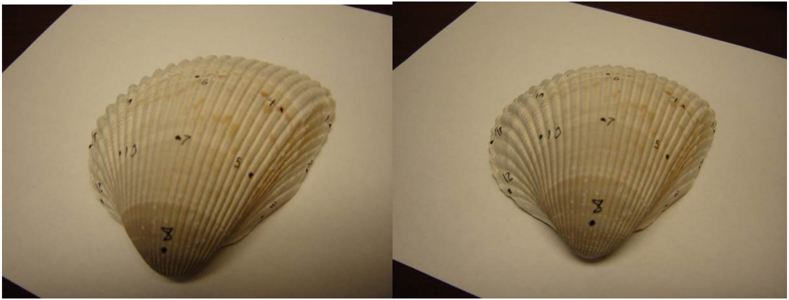
Acrosterigma Magnum Clamshell

- One of the largest species from the Cardiidae bivalve family
- Large numbers washed ashore St George Island during Deepwater Horizon oil spill in May 2010
- Bilateral Symmetry implies that a shape data analysis can be performed based on any one of the two shells found
- one such live specimen is pictured in the figure below

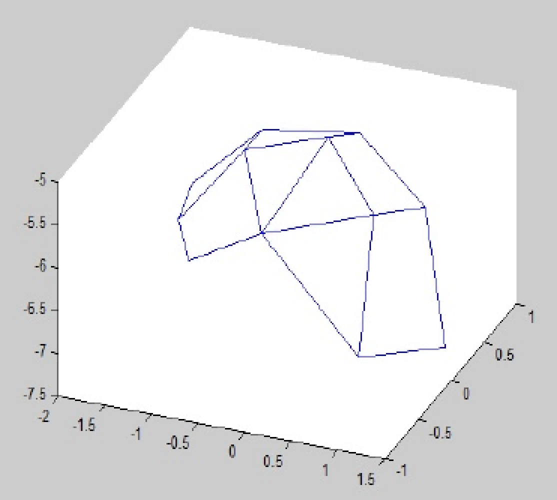
Compare Mean Reflection Shape Change

- Two samples: large shells and small shells
- Select landmarks consistently throughout the two samples
- Obtain Euclidean 3D similarity reconstructions using
- Compute Schoenberg (extrinsic) sample means
- Compare Schoenberg (extrinsic) population means
- The methodology is nonparametric nonpivotal bootstrap
- For details on the 3D Scoenberg means computations see Bandulasiri et al.(2009).

Two pictures of shell with landmarks



Reconstruction obtained from landmark correspondences



Schoenberg sample mean computations for the 3D data

Similarly, $\hat{C}_{small} = \frac{1}{7}\sum_{j=\{10,11,12,14,17,18,21\}} U_j^T U_j$.
The eigenvalues of $\hat{C}_{small}$ satisfie $\lambda_p > \lambda_{p+1}$. So the Schoenberg mean reflection shape $\hat{\mu}_E$ exists and $\mu_E = [u]_R$ where u^Tcan be taken as

$$V = (v_1 v_2 v_3)$$

whose columns are orthogonal eigenvectors of $\hat{C}_{small}$ corresponding to the largest eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ of $\hat{C}_{small}$, with

$$v_j^T v_j = \lambda_j + \frac{1}{3}(\lambda_4 + \ldots + \lambda_k), \quad \forall j = 1, 2, 3.$$

The Schoenberg sample mean $[\hat{\xi}_{large}]_R$ reflection shape of the small group is

$$[\hat{\xi}_{small}]_R = \{A\hat{u}_{small} : A \in O(3)\}$$

Schoenberg sample mean computations for the 3D data-2

Schoenberg sample mean reflection shapes of large and small shells

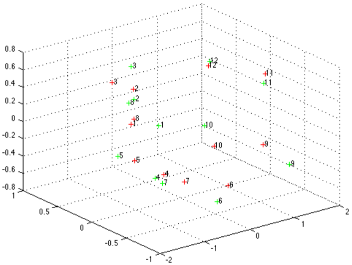


Figure: Icons of extrinsic mean shape for large shells sample(red) and small shells sample(blue)

Schoenberg sample mean computations for the 3D data

Icons of bootstrap distributions of Schoenberg sample mean reflection shapes of large and small shells

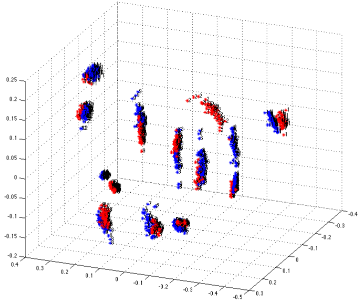


Figure: Distributions of bootstrapped extrinsic mean shape for large shells sample(red) and small shells sample(blue).

3D Mean Projective Shape Change Analysis for Lily Data

- In this example we will test the existence of mean 3D projective shape change to differentiate between two Lily flowers.
- The methodology was developed in Patrangenaru et al. (2014)
- The MATLAB codes were borrowed from that work
- Flowers belonging to the genus ***Lilum*** have three petals and three petal-like sepals. It may be difficult to distinguish the lily petals from the sepals, all six are referred to as ***tepals***.
- We will use pairs of pictures of two flowers for our study. We will recover the 3D projective shape of a spatial k-ad (in our case $k = 13$) from the pairs of images, which will allow us to test for mean 3D projective shape change detection. See Patrangenaru et.al. (2010).
- We used digital images of two flowers: 11 pairs of pictures for one lily and 8 pairs of pictures for the other one.
- We slected 13 landmarks anatomic landmaks, 5 of which will be used to construct a projective frame.

First Sample Data

We collected a set of $n_1 = 11$ pairs of pictures for the first flower and it is show in the figures below;

Figure: Lily flower 1

Second Sample Data

For the second flower we have a set of $n_2 = 8$ pairs of pictures;

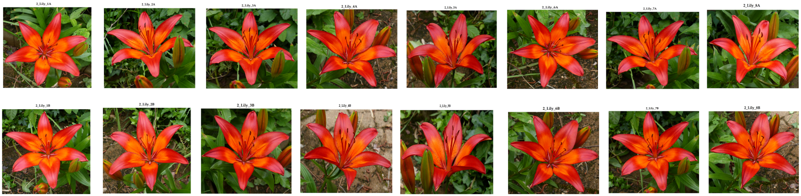


Figure: Lily flower 2

Landmarks-I

Throughout the first sample (size $n_1 = 11$) we kept the same labeling of landmarks and the same configuration. The tepals landmarks were labeled 1 through 6;

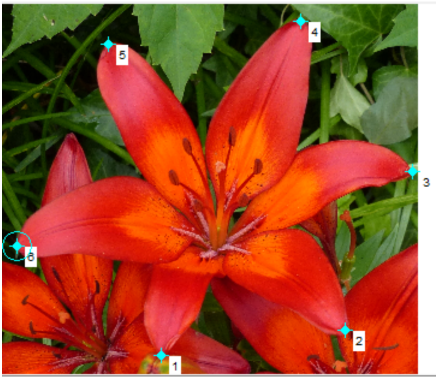


Figure: Landmark Placements for Tepals

Landmarks-I

The six ***stamens*** (male part of the flower) ,were labeled 7 through 12 starting with the stamen that is closely related to tepal 1 and continuing in the same fashion. The landmarks were placed at the tip of the ***anther*** of each of the six stamens and in the center of the ***stigma*** for the ***carpel*** (the female part).

Figure: Landmark Placements for Stamens and Carpel

Landmarks-II

For the second sets of data, we kept a constant configuration for all the landmarks, using the same orientation for all $n_2 = 8$ pairs, as illustrated below

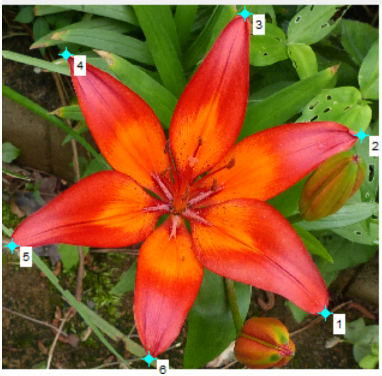


Figure: Landmark Placements for Tepals

Landmarks-II

Landmarks for Stamens and Carpel :

Figure: Landmark Placements for Stamens and Carpel of second flower

Projective Frame and Projective Coordinates

The data analysis was carried out on the space of 3D projective shape of k-ads ($k = 13$), a sample space homeomorphic to $(\mathbb{R}P^3)^8$. Here $\mathbb{R}P^3$ is the 3D real projective space.

- For all of our 3D reconstructed configurations, the first 5 (our 13) landmarks were selected as our projective frame.
- To each projective point we associated its projective coordinate with respect to the projective frame (see Mardia and Patrangenaru (2005),Patrangenaru et al.(2010, 2014))
- The projective shape of the 3D k-ad, is then determined by the 8 projective coordinates of the remaining landmarks of the reconstructed configuration.

Two sample test for change in the VW means

- The Veronese-Whitney (VW) mean is the extrinsic mean associated with the VW embedding (see Patrangenaru et al.(2010, 2014) for details).
- We tested for the VW mean change, since $(\mathbb{R}P^3)^8$ has a Lie group structure (Crane and Patrangenaru(2010)).
- Two types of VW mean changes were considered: one for cross validation, and the other for comparing the WV mean shapes of the two flowers.
- Suppose Q_1 and Q_2 are independent r.o.'s, the hypothesis for there mean change are

$$H_0 : \mu_{1,13}^{-1} \odot \mu_{2,13} = 1_8$$

- Let φ be the Log chart on this Lie group, $\varphi(1_8) = 0_8$, compute the **bootstrap** distribution.

$$D_* = \varphi(\bar{X}_{11,j_{13}}^{-1} \otimes \bar{X}_{8,j_{13}})$$

- Construct the $100(1-\alpha)\%$ nonparametric bootstrap confidence region by D^*,
- For simplicity one may use simultaneous confidence intervals.
- We expect to fail to reject the null, for both tests.

Results for the Two sample Test

In this slide and the next two, we have eight figures showing the various simultaneaous nonparametric bootstrap confidence regions for all remaining 8 reconstructed landmarks.

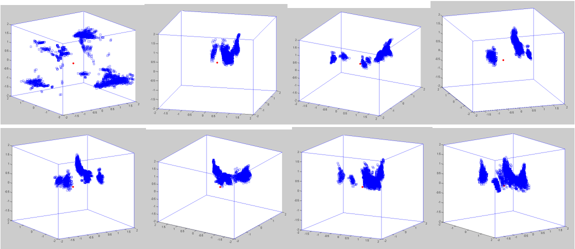


Figure: Eight bootstrap projective shape marginals for lily data

Conclusion: there is significant mean VW projective shape change between the two flowers

Simultaneaous Nonparametric Bootstrap Confidence intervals

Highlighted in blue are the intervals not containg $0 \in \mathbb{R}$.

simultaneous confidence intervals for Lily's landmarks 6 to 9				
	LM6	LM7	LM8	LM9
x	$(0.61, 1.64)$	$(0.32, 0.56)$	$(-0.43, 0.82)$	$(0.05, 0.88)$
y	$(-0.91, 0.10)$	$(-0.20, 0.35)$	$(-0.25, 0.58)$	$(-0.36, 0.46)$
z	$(-1.5, 1.2)$	$(0.18, 0.64)$	$(0.29, 0.83)$	$(0.21, 0.88)$

Simultaneous confidence intervals for Lily's landmarks 10 to 13				
	LM10	LM11	LM12	LM13
x	$(0.060, 0.82)$	$(0.50, 0.84)$	$(0.42, 0.65)$	$(0.47, 0.87)$
y	$(-0.34, 0.16)$	$(-0.05, 0.25)$	$(-0.08, 0.19)$	$(-0.07, 0.45)$
z	$(0.20, 0.80)$	$(0.059, 0.62)$	$(0.076, 0.57)$	$(-0.14, 0.50)$

Data for Cross-Validation

- For this particular portion we used a sample consisting of the same flower. We do so in part to validate the choice of the landmarks and to make sure the code can recognize the same flower displayed through many pictures.
- This sample was divided into two portion of $n_1 = 5$ data points and $n_2 = 6$ data points.

Data for Cross-Validation

The sample is shown below;

Figure: Lily same Flower sample 1

Data for Cross-Validation

Figure: Lily same Flower sample 2

Confidence Region for Cross validation

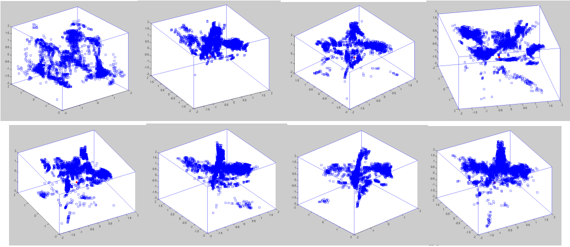


Figure: Eight bootstrap projective shape marginals for cross validation

One can show that, as expected, there is no mean VW projective shape change, based on the two samples with sample sizes respectively $n_1 = 5$ and $n_2 = 6$.

Fréchet Means and Fréchet Antimeans

Oller and Corcuera (1995) found for the von Mises distribution on S^d having two mean values, while in fact there is only one intrinsic (and extrinsic) mean.

Note that S^d us a compact manifold. One can easily see that the mean value in the sense of Oller and Corcuera (1995), is actually a *maximizer* of the expected mean chord square distance on S^d.

Note that the computations considered in the two last examples above were also on compact manifolds. We are led to consider the following

DEFINITION (Patrangenaru and Ellingson (2015), Patrangenaru et. al.(2016)). Consider a probability measure Q on a compact Hausdorf metric space (M, ρ), for which the Fréchet function is bounded. The set of all points of maximum (minimum) of this function is called the *Fréchet antimean set (Fréchet mean set)* . In case the Fréchet antimean set only point only, that point is the *Fréchet antimean (Fréchet mean)*. Given $X_1, \dots, X_n$ i.i.d.r.v.'s from Q, their Cartan sample antimean set (Cartan sample mean set) are the Cartan antimean set (Cartan mean set) of the empirical distribution $\hat{Q}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$.

Extrinsic Mean and Extrinsic Sample Mean

Definition (Extrinsic Mean)

The set of all minimizers of this function is called the *extrinsic mean set*. If Fréchet function has unique minimizer, then this point is the *extrinsic mean*, and is labeled $\mu_j(Q)$, or simply μ_E, when j is known.

$$\mu_E = \arg\min_p F(p) = \arg\min_p \mathbb{E}\left[\rho^2(p,x)\right] = \arg\min_p \mathbb{E}\|j(x) - j(p)\|^2$$

Definition (Extrinsic Sample Mean)

Given $X_1, \ldots, X_n$ i.i.d.r.v.'s from Q, if Fréchet function has unique minimizer, then this point is the *extrinsic sample mean* and its notation is $\bar{X}_E$.

$$\hat{\mu}_E = \arg\min_p F(p) = \arg\min_p \mathbb{E}\left[\rho^2(p,X)\right] = \arg\min_p \frac{1}{n}\sum_{i=1}^{n} \|j(x_i) - j(p)\|^2$$

Extrinsic AntiMean and Extrinsic Sample AntiMean

Definition (Extrinsic AntiMean)

The set of all maximizers of this function is called the *extrinsic antimean set*. If Fréchet function has unique maximizer, then this point is the *extrinsic antimean*, and is labeled $\alpha\mu_j(Q)$, or simply $\alpha\mu_E$, when j is known.

$$\alpha\mu_E = \arg\max_p F(p) = \arg\max_p \mathbb{E}\left[\rho^2(p,x)\right] = \arg\max_p \mathbb{E}\|j(x) - j(p)\|^2$$

Definition (Extrinsic Sample Mean)

Given $X_1, \dots, X_n$ i.i.d.r.v.'s from Q, if Fréchet function has unique maximizer, then this point is the *extrinsic sample antimean* and its notation is $a\bar{X}_E$.

$$\hat{\mu}_E = \arg\max_p F(p) = \arg\max_p \mathbb{E}\left[\rho^2(p,X)\right] = \arg\max_p \frac{1}{n}\sum_{i=1}^{n} \|j(x_i) - j(p)\|^2$$

Extrinsic Antimeans

Definition

The set of maximizers of the Fréchet function, is called the *extrinsic antimean set*. In case the extrinsic antimean set has one point only, that point is called **extrinsic antimean** of X, and is labeled $\alpha\mu_{j,E}(Q)$, or simply $\alpha\mu_E$, when j is known.

Extrinsic Antimeans

Theorem

Let $Q = P_X$ be a probability measure associated with the random object X on a compact metric space (M, ρ). So we have $F(p) = E(\rho^2(p, X))$ is finite on M.

(a) *Then, given any $\varepsilon > 0$, there exist a P-null set N and $n(\omega) < \infty$ $\forall\, \omega \in N^c$ such that the Fréchet (sample) antimean set of $\hat{Q}_n = \hat{Q}_{n,\omega}$ is contained in the ε-neighborhood of the Fréchet antimean set of Q for all $n \geq n(\omega)$.*

(b) *If the Fréchet antimean of Q exists then every measurable choice from the Fréchet (sample) antimean set of $\hat{Q}_n$ is a strongly consistent estimator of the Fréchet antimean of Q.*

Non-focal condition

Definition (αj-nonfocal)

(a) A point $y \in \mathbb{R}^N$ for which there is a unique point $p \in \mathcal{M}$ satisfying the equality,

$$\sup_{x \in \mathcal{M}} \|y - j(x)\|_0 = d_0(y, j(p)) \tag{1}$$

is called αj-nonfocal. A point which is not αj-nonfocal is said to be αj-focal.

(b) If y is an αj-nonfocal point, its projection on $j(\mathcal{M})$ is the unique point $z = P_{F,j}(y) \in j(\mathcal{M})$ with $\sup_{x \in \mathcal{M}} \|y - j(x)\|_0 = d_0(y, j(p))$.

Non-focal condition

Definition

A probability distribution Q on $\mathcal{M}$ is said to be αj-nonfocal if the mean μ of $j(Q)$ is αj-nonfocal.

Theorem

Let μ be the mean vector of $j(Q)$ in $\mathbb{R}^N$. Then the following hold true:

(i) *The extrinsic antimean set is the set of all points $x \in \mathcal{M}$ such that $\sup_{p \in \mathcal{M}} \|\mu - j(p)\|_0 = d_0(\mu, j(x))$.*

(ii) *If $\alpha\mu_{j,E}(Q)$ exists, then μ is αj-nonfocal and $\alpha\mu_{j,E}(Q) = j^{-1}(P_{F,j}(\mu))$.*

Extrinsic Sample Antimean

Definition

Let $x_1, ..., x_n$ be random observations from a distribution Q on a compact metric space $(\mathcal{M}, \rho)$, then their extrinsic sample antimean set, is the set of maximizers of the Fréchet function $\hat{\mathcal{F}}_n$ associated with the empirical distribution $\hat{Q}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$, which is given by

$$\hat{\mathcal{F}}_n(p) = \frac{1}{n}\sum_{i=1}^{n}\|j(x_i) - j(p)\|_0^2 \tag{2}$$

If $\hat{Q}_n$ has an extrinsic antimean, its extrinsic antimean is called extrinsic sample antimean, and it is denoted $a\bar{X}_{j,E}$.

Extrinsic Sample Antimean

Theorem

Assume Q is an αj-nonfocal probability measure on the manifold $\mathcal{M}$ and $X = \{X_1, ..., X_n\}$ are i.i.d r.o.'s from Q. Then,

(a) *If $\overline{j(X)}$ is an αj-nonfocal, then the extrinsic sample antimean is given by $a\bar{X}_{j,E} = j^{-1}(P_{F,j}(\overline{j(X)}))$.*

(b) *The set $(\alpha F)^c$ of αj-nonfocal points is a generic subset of $\mathbb{R}^N$, and if $\alpha\mu_{j,E}(Q)$ exists, then the extrinsic sample antimean $a\bar{X}_{j,E}$ is a consistent estimator of $\alpha\mu_{j,E}(Q)$.*

Central Limit Theorem for Sample Antimean

Theorem

$n^{1/2}(P_{F,j}(\overline{j(X)}) - P_{F,j}(\mu))$ converges weakly to $N_k(0_k, \alpha\Sigma_\mu)$, where $\overline{j(X)} = \frac{1}{n}\sum_{i=1}^{n} j(X_i)$ and

$$\begin{aligned}\alpha\Sigma_\mu = &[\sum_{a=1}^{d} d_\mu P_{F,j}(e_b)\cdot e_a(P_{F,j}(\mu))e_a(P_{F,j}(\mu))]_{b=1,\dots,k} \\ &\times\Sigma[\sum_{a=1}^{d} d_\mu P_{F,j}(e_b)\cdot e_a(P_{F,j}(\mu))e_a(P_{F,j}(\mu))]^{T}_{b=1,\dots,k}\end{aligned} \tag{3}$$

Here Σ is the covariance matrix of $j(X_1)$ w.r.t the canonical basis $e_1, e_2, \dots, e_k$.

References I

A. Bandulasiri, R.N. Bhattacharya and V. Patrangenaru (2009). Nonparametric Inference for Extrinsic Means on Size-and-(Reflection)-Shape Manifolds with Applications in Medical Imaging. *Journal of Multivariate Analysis*. **100** 1867-1882.

Bojan Basrak (2010). *Limit theorems for the inductive mean on metric trees* J. Appl. Prob. **47**, 1136–1149.

Rabi N. Bhattacharya, Marius Buibas, Ian L. Dryden, Leif A. Ellingson, David Groisser, Harrie Hendriks, Stephan Huckemann, Huiling Le, Xiuwen Liu, James S.Marron, Daniel E. Osborne, Vic Patrângenaru, Armin Schwartzman, Hilary W. Thompson, and Andrew T. A.Wood. (2013). Extrinsic data analysis on sample spaces with a manifold stratification. In *Advances in Mathematics, Invited Contributions at the Seventh Congress of Romanian Mathematicians, Brasov, 2011*, Publishing House of the Romanian Academy (Editors: Lucian Beznea, Vasile

References II

Brîzanescu, Marius Iosifescu, Gabriela Marinoschi, Radu Purice and Dan Timotin), pp. 148–156.

R. N. Bhattacharya and V. Patrangenaru (2014). Rejoinder of Discussion paper “Statistics on Manifolds and Landmarks Based Image Analysis: A Nonparametric Theory with Applications." *Journal of Statistical Planning and Inference.* **145**, 42–48.

Bhattacharya, R.; Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Annals of statistics*,**31**, 1-29.

Bhattacharya, R.; Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Annals of statistics*,**33**, 1211-1245.

R. N. Bhattacharya, L. Ellingson, X. Liu and V. Patrangenaru and M. Crane (2012). Extrinsic Analysis on Manifolds is Computationally Faster than Intrinsic Analysis, with Applications to Quality Control by Machine Vision. *Applied Stochastic Models in Business and Industry.* **28**, 222–235.

References III

- Bhattacharya, R.N. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds-Part I,*Ann. Statist.* **31**, no. 1, 1-29.
- Bhattacharya, R.N. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds-Part II,*Ann. Statist.* **33**, 1211– 1245.
- Billera, Louis J.; Holmes, Susan P.; Vogtmann, Karen (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** , no. 4, 733–767
- Cartan, É. (1929) Groupes simples clos et ouverts et géométrie riemannienne. *Journal de Mathématiques pures et appliqués* **8** , 1–43 (in French).
- M. Crane and V. Patrangenaru. (2011). Random Change on a Lie Group and Mean Glaucomatous Projective Shape Change Detection From Stereo Pair Images. *Journal of Multivariate Analysis*. **102**, 225-237.

References IV

- de Rham, Georges (1946). Sur la théorie des formes différentielles harmoniques. *Ann. Univ. Grenoble. Sect. Sci. Math. Phys.* 22.
- L. Ellingson, V. Patrangenaru, H. Hendriks, P. S. Valentin (2014). CLT on Low Dimensional Stratified Spaces. *In Press at Proceedings of the First INSPS Conference.*
- N. I. Fisher, P. Hall, B. Y. Jing and A. T. A. Wood (1996). Properties of principal component methods for functional and longitudinal data analysis. J. Amer. Statist. Assoc. **91**, 1062–1070.
- Fréchet, Maurice (1948). Les élements aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré* **10**, 215–310.

References V

- Guo, R., Patrangenaru, V. anf Lester, L. (2014) Nonparametric Bootstrap test for Equality of Extrinsic Mean Reflection Shapes in Large vs Small Populations of Acrosterigma Magnum Shells, poster, *Geometric Topological and Graphical Model Methods in Statistics Fields Institute, Toronto, Canada, May 22-23, 2014*
- Harrie Hendriks and Z. Landsman, *Asymptotic behaviour of sample mean location for manifolds*, Statistics and Probability Letters. **26** (1996), 169–178.
- Hendriks, H. and Landsman, Z. (1998). Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. *J. Multivariate Anal.* **67** no. 2, 227-243.

- Harrie Hendriks and Z. Landsman, *Asymptotic behaviour of sample mean location for manifolds*, Statistics and Probability Letters. **26** (1996), 169–178.

References VI

- Hendriks, H. and Patrangenaru (2014). Mean location and sample mean location on graphs: asymptotics, tests, confidence regions. *in work*.
- Thomas Hotz, Stephan Huckemann, Huiling Le, James S. Marron, Jonathan C. Mattingly, Ezra Miller, James Nolen, Megan Owen, Vic Patrangenaru and Sean Skwerer (2013). Sticky Central Limit Theorems on Open Books. *Annals of Applied Probability*, **23**, 2238–2258.
- Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30** 509-541.
- Mardia, K. V.; Patrangenaru, V. (2005). Directions and projective shapes. *The Annals of Statistics*, **33**, 1666-1699.
- Oller, J. M.; Corcuera, J. M.(1995). Intrinsic analysis of statistical estimation. *Ann. Statist.* **23**, 1562==1581

References VII

- D. Osborne, V. Patrangenaru, L. Ellingson, D. Groisser and A. Schwartzman. (2013). Nonparametric Two-Sample Tests on Homogeneous Riemannian Manifolds, Cholesky Decompositions and Diffusion Tensor Image Analysis. *Journal of Multivariate Analysis*. **119**, 163-175.
- Paige R. L.; Patrangenaru, V.; and Qiu, M. (2013). 3D Projective Shapes of Leaves from Image Data. *Abstracts of the 29th European Meeting of Statisticians, Budapest July 21-25.* p.232.
- V. Patrangenaru and L. A. Ellingson (2015). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. Chapman&Hall/CRC. ISBN-13: 978-1439820506
- V. Patrangenaru, M. Qiu and M. Buibas(2014). Two Sample Tests for Mean 3D Projective Shapes from Digital Camera Images. *Methodology and Computing in Applied Probability.* **16**, 485–506.

References VIII

- Patrangenaru, V; Liu, X. and Sugathadasa, S.(2010). Nonparametric 3D Projective Shape Estimation from Pairs of 2D Images - I, In Memory of W.P. Dayawansa. *Journal of Multivariate Analysis*. **101**, 11-31.
- V. Patrangenaru, K.D.Yao and R. Guo (2016). Means and Antimeans. Proceedings of the 2nd International Conference of ISNPS, Cadiz, Spain, June 12-16, 2014. *to appear*
- Wang, H.; and Marron, J. S. (2007). Object oriented data analysis: Sets of trees. *The Annals of Statistics*, **35**, 1849-1873.
- Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transactions of Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes A.* 591–602.