Mini-Workshop: Asymptotic Statistics on Stratified Spaces

Organised by
Aasa Feragen, Copenhagen
Stephan Huckemann, Göttingen
J.S. Marron, Chapel Hill
Ezra Miller, Durham

28 September – 4 October 2014

Abstract. Statistical analysis of non-Euclidean data such as data on manifolds is an active and established topic of research, for instance, in the statistical analysis of shape. However, many types of data naturally reside in metric spaces which are not smooth manifolds as a whole, rather they are unions of manifold strata of varying dimensions. These spaces form a key general family of geometric spaces for data analysis. Statistics in stratified spaces has recently found great interest in applications and mathematical theory building. While the fundamental theory is still in its beginnings, as a centerpiece the derivation and investigation of statistics and their asymptotics has materialized. Only a few basic results are known, but it is clear that the geometric constraints imposed by stratified spaces lead to unexpected asymptotic behavior of standard statistical properties, such as “stickiness” of means, see [4]. It is the scope of the proposed workshop to better understand fundamental relations between asymptotic behavior of statistical descriptors and global as well as local geometric and topological structures. This investigation calls for an intense collaboration of the fields involved: statistics & stochastics; geometry & topology; combinatorics, algorithms & numerics.

This workshop sought to bring together world-leading scientists and high-potential early career researchers working in this field to collaborate on a focused set of fundamental questions.

Mathematics Subject Classification (2010): 05C88, 51M09, 52C45, 57N80, 58J65, 60F05, 60J60, 62G20, 62H35.
Introduction by the Organisers

The mini-workshop *Asymptotic Statistics on Stratified Spaces* organized by Aasa Feragen, Stephan Huckemann, J.S. Marron and Ezra Miller, had 17 participants, 10 of whom were junior participants.

Many statistical problems of high current interest deal with the analysis of data sampled from spaces that are naturally non-Euclidean. For example, phylogenetic trees in evolutionary biology as well as anatomical tree-like structures such as the lung airway system or blood vessels can be viewed as sampled from stratified spaces with discontinuous non-positive curvatures (cf. [3, 11]). Stratified spaces appear naturally when data have varying topological structure but continuous interpolation between different topological structures is meaningful, as is the case with anatomical tree- or graph shapes or phylogenetic trees. Stratified spaces also appear naturally as quotients under group actions in shape analysis, for instance when data distances are to be invariant under symmetries such as translation, rotation, scaling (cf. [6]).

When data reside in stratified spaces, fundamental statistical concepts, even simple ones such as “mean” or “principal component”, have no canonical generalization. The non-differentiability (due to the stratification) combined with non-positivity of curvature at lower-dimensional strata results in particular statistical phenomena such as “stickiness of intrinsic means”, cf. [4] (intrinsic mean may stick to lower dimensional strata under arbitrary small perturbations). Smooth curvature may also exhibit unexpected limiting behaviors; already in directional statistics on the circle, mass distributed around the antipodal of an intrinsic mean can cause “smeariness” of intrinsic means, in some sense the effect opposite to stickiness, cf. [5] (suitable mass near the antipodal may reduce asymptotic rates arbitrarily).

Not only does geometry have an impact on defining statistical properties; it may also be non-trivial to compute distances and thus statistical descriptors in stratified spaces. Optimization near points of negative curvature faces combinatorical challenges of computation at lower-dimensional strata adjacent to vast numbers of higher-dimensional manifold strata, cf. [9,10]. While in the context of shape analysis curvature positive with respect to the top space ensures that intrinsic means of distributions not restricted to singular strata are likewise not singular, cf. [7], means may no longer be unique, but can be non-trivially set valued. In particular, positive curvature may lead to non-uniqueness of shortest paths and ambiguity which can lead to NP completeness, even for computing distances, cf. [2]. Understanding and solving computational problems often requires an interplay of geometry, stochastics, combinatorics and optimization, cf. [1,8–10].

To address these problems, the workshop brought together young and world-leading specialists in the fields of statistics, geometry, topology, combinatorics, and numerics currently grappling with related issues from their specific scientific viewpoint. Due to the interdisciplinary nature of our research, the workshop started out with five introductory lectures on current key problems:
Object oriented data analysis
- Dimension reduction on manifolds
- Geometric structure and statistical problems in tree spaces
- Diffusion in tree space
- Central limit theorem on $T_4$

The introductory lectures were held on the first day and the first half of the second day, with extremely engaged participation from the audience in the form of extended discussions of the presented problems. The introductory lectures were followed by response lectures from the remaining participants, who presented alternative viewpoints and partial solutions to the proposed problems.

The size and informal structure of the workshop resulted in a week of lively discussion and collaboration, leading to new insights and results, some of which were presented in the talks. This included the asymptotic confidence intervals on the three-spider presented by Thomas Hotz and Huiling Le, as well as very insightful discussions on topics such as diffusion on non-linear spaces, the embeddability of nonlinear distance functions in linear spaces, and a realization that kernel PCA can be viewed as slicing with varieties in higher dimensions thus potentially allowing for backward principal nested subspace analysis. We have no doubt that many of the new ideas generated at this workshop will spur new collaborations and papers in the near future.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-1049268, “US Junior Oberwolfach Fellows”.

References

Mini-Workshop: Asymptotic Statistics on Stratified Spaces

Table of Contents

J. S. Marron

*Introduction to Object Oriented Data Analysis* .......................... 2487

Sungkyu Jung (joint with Stephan Huckemann, J. S. Marron, and Thomas Hotz)

*Dimension reduction for directions and 2D shapes* ................. 2489

Aasa Feragen (joint with Sean Cleary, Megan Owen, and Daniel Vargas)

*On tree-space PCA* ......................................................... 2491

Tom M. W. Nye

*Construction of distributions on tree-space via diffusion processes* .... 2495

Huiling Le (joint with Dennis Barden and Megan Owen)

*Fréchet Means in the Space of Phylogenetic Trees* ................. 2497

Franz J. Király (joint with Duncan A.J. Blythe, Martin Kreuzer, Louis Theran, Ryota Tomioka)

*Algebraic Combinatorial Methods in Statistics and Machine Learning* ... 2499

Washington Mio (joint with Diego Diaz Martinez and Facundo Mémoli)

*On Geometry Underlying Borel Measures on Euclidean Space* ........ 2502

Peter W. Michor (joint with Thomas Hotz and Andreas Kriegl)

*Frölicher spaces as a setting for tree spaces and stratified spaces* .... 2505

Stefan Sommer

*Diffusion Processes and PCA on Manifolds* ............................. 2509

Megan Owen (joint with Sean Cleary, Aasa Feragen, Daniel Vargas)

*Multiple Principal Components Analysis in Tree Space* .............. 2514

Thomas Hotz (joint with Huiling Le)

*Asymptotic confidence sets for the Fréchet mean on the 3-spider* .... 2516

Stephan F. Huckemann (joint with Benjamin Eltzner)

*Stickiness and Smeariness* ............................................... 2520

Ezra Miller (joint with Stephan F. Huckemann, Jonathan C. Mattingly, and James Nolen)

*Topological definition of stickiness for means in arbitrary metric spaces* . 2521

Sean Skwerer (joint with J. S. Marron and Scott Provan)

*Optimization in Phylogenetic Treespace* ............................... 2522