Abstracts

Contents

Updated: 24 October 2024

John Aston
University of Cambridge, UK

Using Geometry in Non-Parametric Statistics.

In non-parametric regression, the rates of estimation critically depend on the dimension, usually known as the curse of dimensionality. However, it has been long known that incorporating structure into the regression (such as sparsity) can improve on general rates. However, sparsity and most related concepts are linear in the data, while many patterns in regressions are non-linear in nature, and crucially depend on all the covariates. In this talk, we will consider how more general structure can be incorporated in non-parametric regression through the use of symmetries. General notions of symmetry corresponding to algebraic group structures can exhibit the similar dimension reduction phenomena, even in non-linear settings, and even in the case where the symmetries need to be estimated as part of the regression. We will also show that, by considering lattice structures, efficient computational estimation schemes to determine such symmetries are possible.

[Joint work with Louis Christie]

Back to Contents Page →

Benjamin Eltzner
Max Planck Institute for Mathematics in the Sciences, Germany

Testing for Uniqueness of Estimators

When considering quantities describing a random population of systems, one usually assumes uniqueness of such descriptors. However, in many models this assumption is non-trivial. We consider the general class of m-estimation descriptors defined as minimizers of a loss function. This class includes the MLE, the Frechet mean, and many geometric descriptors like principal geodesics. When estimating these descriptors from a finite sample of measured values, there are often local minima of the sample loss function. These local minima may stem from multiple global minima of the underlying population loss function. We present a hypothesis test to systematically determine for a given sample whether the underlying population loss function may have multiple global minima.

The test is applicable to many types of estimation problems and we discuss several typical applications.

Back to Contents Page →

Charles Fefferman
Princeton University, USA

Joint Distinguished Talk of Statistics and Mathematics
Fitting Smooth Functions to Data

Suppose we are given data points (χ_1, y_1),…, $(\chi_N, y_N) \in \mathbb{R}^n \times \mathbb{R}^D$. We want to compute a function $F(\chi_i)= y_i$ for i = 1,…, N, and we want our F to be nearly as smooth as possible. To measure smoothness we work in a Banach space of continuous $\mathbb{R}^D$ valued functions on $\mathbb{R}^n$. For some Banach spaces, but not for others, we describe algorithms to compute such an F. Variants of our algorithms allow us to look for F that agree with the data up to a given tolerance, and also to make a wise choice of data points to be discarded as outlilners. Our algorithms are theoretically optimal but not practical.

Back to Contents Page →

Fernando Galaz Garcia
Durham University, UK

Metric geometry of spaces of persistence diagrams

Persistence diagrams are central objects in topological data analysis. They are pictorial representations of persistence homology modules and describe topological features of a data set at different scales. In this talk, I will discuss the geometry of spaces of persistence diagrams and connections with the theory of Alexandrov spaces, which are metric generalizations of complete Riemannian manifolds with sectional curvature bounded below. In particular, I will discuss how one can assign to a metric pair (X, A) a one-parameter family of pointed metric spaces of (generalized) persistence diagrams Dp(X, A) with points in (X, A) via a family of functors Dp with p $\in$ [1, ∞]. These spaces are equipped with the p- Wasserstein distance when p $\geqslant$ 1 and the bottleneck distance when p = ∞. The functors Dp preserve natural metric properties of the space X, including non-negative curvature in the triangle comparison sense when p = 2. When p = ∞, the functor D∞ is continuous with respect to a suitable notion of Gromov–Hausdorff distance of metric pairs. When (X, A) = (R2, Δ), where Δ is the diagonal of R2, one recovers previously known properties of the usual spaces of persistence diagrams. I will also discuss some connections of these results with optimal partial trans-port.

This is joint work with Mauricio Che, Luis Guijarro, Ingrid Membrillo Solis, and Motiejus Valiunas.

Back to Contents Page →

Sungkyu Jung
Seoul National University, Korea

Huber means on Riemannian manifolds

In this talk, I will introduce Huber means on Riemannian manifolds, pro-viding a robust alternative to the Fr´echet mean by integrating elements of both L2 and L1 loss functions. The Huber means are designed to be highly resistant to outliers while maintaining efficiency, making it a valuable generalization of Huber's M-estimator for manifold-valued data. We comprehensively investigate the statistical and computational aspects of Huber means, demonstrating their utility in manifold-valued data analysis. Specifically, we establish nearly minimal conditions for ensuring the existence and uniqueness of the Huber mean and discuss regularity conditions for un-biasedness. The Huber means are consistent and enjoy the central limit theorem. Additionally, we propose a novel moment-based estimator for the limiting covariance matrix, which is used to construct a robust one-sample location test procedure and an approximate confidence region for location parameters. The Huber mean is shown to be highly robust and efficient in the presence of outliers or under heavy-tailed distribution. Specifically, it achieves a breakdown point of at least 0.5, the highest among all isometric equivariant estimators, and is more efficient than the Fréchet mean under heavy-tailed distribution.

This talk is based on a joint work with Jongmin Lee.

Back to Contents Page →

Huiling Le
The University of Nottingham, UK

The influence of the cut locus on the CLT for Frechet means

This talk presents a general result on the CLT for sample Frechet means on Riemannian manifolds when the support of a distribution meets the cut locus of its Frechet mean.

Back to Contents Page →

Steve Marron
University of North Carolina, USA

Object Oriented Data Analysis

The rapid change in computational capabilities has made Big Data a major modern statistical challenge. Less well understood is the rise of Complex Data as a perhaps greater challenge. Object Oriented Data Analysis (OODA) is a framework for addressing this, in particular providing a general approach to the definition, representation, visualization and analysis of Complex Data. The notion of OODA generally guides data analysis, through providing a useful terminology for interdisciplinary discussion of the many choices typically needed in modern complex data analyses. The main ideas are illustrated through several OODA contexts, including shapes, trees (in the sense of graph theory), covariance matrices and nonnegative curves as data objects.

Back to Contents Page →

Jonathan Mattingly
Duke University, USA

The random tangent field and central limit theorems on stratified spaces

I will discuss the part of a sequence of recent papers with Ezra Miller and Do Tran about central limit theorems on stratified spaces. I will emphasize the local analysis in the tangent space above the mean/barycenter.

Back to Contents Page →

Marina Meila
University of Washington, USA

Manifold Learning, Explanations and Eigenflows

This talk will expand Manifold Learning in two directions. First, we ask if it is possible, in the case of scientific data where quantitative prior knowledge is abundant, to explain a data manifold by new coordinates, chosen from a set of scientifically meaningful functions? Second, we ask how standard Manifold Learning tools and their applications can be recreated in the space of vector fields and flows on a manifold. Central to this approach is the order 1-Laplacian of a manifold, $\Delta 1$, whose eigendecomposition into gradient, harmonic, and curl, known as the Helmholtz-Hodge Decomposition, provides a basis for all vector fields on a manifold. We present an estimator for $\Delta 1$, and based on it we develop a variety of applications. Among them, visualization of the principal harmonic, gradient or curl flows on a manifold, smoothing and semi-supervised learning of vector fields, 1-Laplacian regularization. In topological data analysis, we describe the 1st-order analogue of spectral clustering, which amounts to prime manifold decomposition. Furthermore, from this decomposition a new algorithm for finding shortest independent loops follows. The algorithms are illustrated on a variety of real data sets.

Joint work with Yu-Chia Chen, Samson Koelle, Vlad Murad, Weicheng Wu, Hanyu Zhang and Ioannis Kevrekidis.

Back to Contents Page →

Washington Mio
Florida State University, USA

Probing the Shape of Metric and Networked Data Through Observables

We discuss an approach to statistical analysis of data in (compact) metric spaces X based on metric observables; that is, non-expansive scalar fields on X. This includes the notions of observable mean and covariance operators, stable with respect to the Wasserstein distance. In particular, the approach leads to Principal Observable Analysis that may be viewed as a metric counterpart to standard PCA. Among other things, beyond dimension reduction, principal observables yield basis functions for analysis of signals on metric-measure spaces and a vectorization technique to enable integration of metric data analysis with machine learning. Several illustrations for networked data will be presented.

Back to Contents Page →

Sayan Mukherjee
Max Planck Institute for Mathematics in the Sciences, Germany

Modeling shapes and surfaces - Geometry meets machine learning

We will consider modeling shapes and fields via topological and lifted-topological transforms. Specifically, we show how the Euler Characteristic Transform and the Lifted Euler Characteristic Transform can be used in practice for statistical analysis of shape and field data. We also state a moduli space of shapes for which we can provide a complexity metric for the shapes. We also provide a sheaf theoretic construction of shape space that does not require diffeomorphisms or correspondence. A direct result of this sheaf theoretic construction is that in three dimensions for meshes, 0-dimensional homology is enough to characterize the shape. We will also discuss Gaussian processes on fiber bundles and applications to evolutionary questions about shapes. Applications in biomedical imaging and evolutionary anthropology will be stated throughout the talk. Bio: Born in India, Professor Dr Sayan Mukherjee received his academic training in the United States. He completed his doctorate at MIT, Cambridge, United States, in 2001 and initially worked there and at the nearby Broad Institute supported by a Sloan Postdoctoral Fellowship. Since 2004, he has been at Duke University, Durham, United States, becoming a full professor in 2015. As a professor of mathematics, applied statistics and computer science he has always been associated with several departments. In 2011, he spent a year as a visiting researcher in Chicago. He received the International Indian Statistical Association's Young Researcher Award in 2008 and became a Fellow of the Institute of Mathematical Statistics in 2018. He is also a member of various international scientific associations.

Back to Contents Page →

Susovan Pal
Vrije Universiteit Brussel, Belgium

Optimal lifts and their roles in the asymptotics of inference on shapes.

We study Lie group actions on manifolds and random optimal lifts from the quotient to the manifold as occur in practice while conducting inference on shape data.

Back to Contents Page →

Xavier Pennec
Université Côte d'Azur and Inria, Sophia-Antipolis, France

Advances in Geometric statistics with Flag spaces

Flags are sequences of properly embedded linear subspaces. They appear in multiscale dimension reduction methods such as principal or independent component analysis. Non-linear flags also appear to be the right geometric objects to work with for generalizations of PCA to manifolds such as principal nested spheres or barycentric subspace analysis. One can actually show that extracting order principal components actually optimizes a criterion on the flag space and not on Grassmannians as usually thought. Flag spaces are Riemannian homogeneous spaces that generalize Grassmmann and Steifel manifolds. However, they are usually not symmetric. In this talk, I will present an extension of PCA based on flag spaces called Principal subspace analysis that may turn out to be much more stable that the classical PCA decomposition into unidimensional modes. I will also expose a method to obtain confidence regions on the resulting subspaces based on a geometric formulation of the central limit theorem directly in the space of flags.

This is joint work with Tom Szwagier for the first part and Dimbihery Rabenoro for the second part.

Back to Contents Page →

Stephen M. Pizer
University of North Carolina at Chapel Hill, USA

Object Correspondence for Statistics via Interior Geometry

Locational correspondence between positions in objects in a population is critical for performance of statistical objectives on the objects. First, one must obtain correspondence not only on the boundary of the objects but also on their interior. Second, the method of obtaining correspondence must avoid the pollution of pre-alignment. Third, producing the object-based coordinates describing alignment in each object in the population must involve a consistent process recognizing rich geometric properties in the closure of the object interior. I describe a skeletal representation targeted for anatomic objects which is designed to provide all of these capabilities. The method generates fitted frames on the boundary and in the interior of objects and produces alignment-free curvature and inter-point vector features from them at the corresponding object coordinates it provides. It accomplishes this from each object's boundary in the population by generating a skeletal model fit using a diffeomorphic deformation of an ellipsoid in a way recognizing rich interior and boundary geometric features throughout the deformation. We call this model the evolutionary s-rep. Via classification performance on hippocampi shape between individuals with a disorder vs. others, I compare our method to state-of-the-art methods for producing object representations that are intended to capture geometric correspondence across a population of objects and to yield geometric features useful for statistics, and I show notably improved classification performance by this new representation.

Back to Contents Page →

Yvo Pokern
University College London, UK

Inference for a Riemannian Ornstein-Uhlenbeck Process on Covariance Matrices

A generalization of the Ornstein–Uhlenbeck process to the cone of covariance matrices endowed with the log-Euclidean and the affine-invariant metrics is presented. The development exploits the Riemannian structure of symmetric positive definite matrices viewed as a smooth manifold. Bayesian inference for discretely observed diffusion processes of covariance matrices is then carried out based on an MCMC algorithm which requires sampling diffusion bridges. Following a review of simulation algorithms for diffusions and their associated bridge processes on manifolds, a novel diffusion bridge sampler is proposed. Our proposed algorithm is illustrated with a real data financial application.

This is joint work with Prof. Petros Dellaportas (UCL) and Dr Bui Ngoc Mai (British University Vietnam)

Back to Contents Page →

Armin Schwartzmann
University of California, San Diego, USA

Toward a scaling-rotation geometry of symmetric positive definite matrices

Symmetric positive definite (SPD) matrices are familiar in statistics as covariance matrices. They also appear as data objects, particularly in brain imaging, such as in Diffusion Tensor Imaging and Tensor Based Morphometry. A natural representation of such matrices is by their eigenvalue-eigenvector decomposition, because changes in that coordinate system can be directly interpreted as scaling and rotation in space. In this talk, I give an overview of the geometry of SPD matrices given by their eigenvalue-eigenvector decomposition, specifically as a Whitney-stratified manifold equipped with a semi-metric. When restricted to the top stratum, this geometry can be used to define and compute Fréchet means of PSD matrices that can be interpreted in terms of scaling and rotation.

Back to Contents Page →

Stefan Sommer
Københavns Universitet, Germany

Kunita flows, shape stochastics, and phylogenetic inference

I will discuss constructions of stochastic processes on shape spaces in different settings and which properties of such processes are relevant in applications in morphological analysis in evolutionary biology. Particularly, Kunita flows of diffeomorphisms and their actions on shapes can be used as a shape equivalents of the standard Brownian motion model often used in evolutionary studies. I will detail this application, including how statistical inference of properties of the stochastic flow can be performed from observed shapes, in both finite and infinite dimensional settings.

Back to Contents Page →

Zhigang Yao
National University of Singapore, Singapore

TBA

TBA

Back to Contents Page →

Wilderich Tuschmann
Karlsruher Institut für Technologie, Germany

Moduli Spaces of Metrics and locally symmetric spaces

I will provide a gentle introduction to the study of moduli spaces of metrics on Riemannian manifolds and more singular spaces and, in the flat and Ricci flat case, relate upon their connection to certain locally symmetric spaces and orbifolds of potential interest in applications.

Back to Contents Page →

Guowei Wei
Michigan State University, USA

Topological deep learning: The past, present, and future

In the past few years, topological deep learning (TDL), a term coined by us in 2017, has become an emerging paradigm in artificial intelligence (AI) and data science. TDL is built on persistent homology (PH), an algebraic topology technique that bridges the gap between complex geometry and abstract topology through multiscale analysis. While TDL has made huge strides in a wide variety of scientific and engineering disciplines, its most compelling success was observed in biosciences with intrinsically high dimensional and intricately complex data. I will discuss the achievements of TDL in drug discovery and accurate forecasting of emerging viral variants. I will further discuss the limitations and challenges of TDL, and how new approaches based on algebraic topology, geometric topology and differential geometry may address these challenges. I will also discuss how topology is enabling AI and how AI is assisting topological reasoning.

Back to Contents Page →

Jie Wu
BIMSA, China

GLMY Theory and Topological Statistics

The interactions between topology and statistics have achieved various successful applications. In this talk, we will give an introduction to a new approach in data analytics given as a combination of the path homology (GLMY theory) on digraphs introduced by Shing-Tung Yau and the statistical modeling of idopNetwork introduced by Rongling Wu that has admited important applications in sciences. The talk will consist of three sections. In the first section, we will give a brief review on algebraic topology and classical topological data analysis. In the second section, we will discuss topological approaches to more abstract data given by (di-)graphs and (di-)hypergraphs. In the last section, we will discuss idopNetwork modeling and GLMY theory as well as some applications.

Back to Contents Page →

Yingying Wu
University of Houston, USA

Comparison Theorems of Phylogenetic Spaces and Algebraic Fans

The moduli space of an object contains rich information about that object and, consequently, provides insights into discovering or constructing the object being parameterized by the moduli space. In this talk, I will introduce the moduli space of phylogenetic trees and the moduli space of phylogenetic networks, which are homeomorphic to algebraic fans spanned by root subsystems of type D that arise in the moduli space of smooth marked del Pezzo surfaces. This correspondence offers promising insights into the study of mathematical biology, particularly for understanding and discovering evolutionary models. I will also discuss how these spaces are tied to the moduli space of algebraic curves. I will then continue in the context of supersymmetry, focusing on algebraic curves with additional supersymmetric or supergeometric structures called SUSY curves. I will conclude my talk with an exposition on the construction of dual graphs of SUSY curves with Neveu–Schwarz and Ramond punctures, describing how the moduli space of genus 0 SUSY graphs coincides with the aforementioned moduli space of phylogenetic trees.

Back to Contents Page →

Jun Zhang
University of Michigan Ann Arbor and SIMIS, USA

Information Geometry: An Invitation

Information Geometry is the differential geometric study of the manifold of probability models, and promises to be a unifying geometric framework for investigating statistical inference, information theory, machine learning, etc. Central to such manifolds are "divergence functions" (in place of distance) for measuring proximity of two points, for instance Kullback-Leibler divergence, Bregman divergence, etc. Such divergence functions are known to induce a beautiful geometric structure of the set of parametric probability models. The fundamental duality e/m duality is explained in terms of two most popular parametric statistical families: the exponential and the mixture families. This talk will use two examples to introduce some basic ingredients of this geometric framework: the univariate normal distributions (a case with continuous support) and the probability simplex (a case with discrete support). The newly discovered Statistical Mirror Symmetry will also be discussed for the former case.

Back to Contents Page →