Abstracts

Contents

Updated: August 21, 2024

Karthik Bharath
University of Nottingham, UK

Rolling-without-slipping models for manifold-valued functional data

Given a curve in R^d, it is possible to uniquely induce, or develop, a curve on a manifold M by rolling the manifold along the Euclidean curve without slipping. The technique lays the foundation for stochastic analysis on M. The focus of the talk will be on describing how the rolling-without-slipping strategy can be profitably used to define stochastic process models for M-valued functional data, with particular emphasis on models induced by a Gaussian process (GP) in R^d. The modelling technique is general enough to accommodate intrinsic and extrinsic approaches, as dictated by practical considerations.

Theoretical and computational challenges in estimation of and inference for parameters of the GP, and their relationship to Frechet means on M, using discretely observed M-valued functional data will be discussed, aided by numerical examples. TBA

Back to Contents Page →

Anne Van Delft
Columbia University, USA

A statistical framework for analyzing shape in a time series of random geometric objects

---

We introduce a new framework to analyze shape descriptors that capture the geometric features of an ensemble of point clouds. At the core of our approach is the point of view that the data arises as sampled recordings from a metric space-valued stochastic process, possibly of nonstationary nature, thereby integrating geometric data analysis into the realm of functional time series analysis. We focus on the descriptors coming from topological data analysis.

Our framework allows for natural incorporation of spatial-temporal dynamics, heterogeneous sampling, and the study of convergence rates. Further, we derive complete invariants for classes of metric spacevalued stochastic processes in the spirit of Gromov, and relate these invariants to so-called ball volume processes. Under mild dependence conditions, a weak invariance principle in $D([0,1] \times [0,R])$ is established for sequential empirical versions of the latter, assuming the probabilistic structure possibly changes over time. Finally, we use this result to introduce novel test statistics for topological change, which are distribution free in the limit under the hypothesis of stationarity.

Back to Contents Page →

Hannah Lai
National University of Singapore, Singapore

Neural Tangent Kernel in Implied Volatility Forecasting: A Nonlinear Functional Autoregression Approach

Implied volatility (IV) forecasting is inherently challenging due to its high dimensionality across various moneyness and maturity, and nonlinearity in both spatial and temporal aspects. We utilize implied volatility surfaces (IVS) to represent comprehensive spatial dependence and model the nonlinear temporal dependencies within a series of IVS. Leveraging advanced kernel-based machine learning techniques, we introduce the functional Neural Tangent Kernel (fNTK) estimator within the Nonlinear Functional Autoregression framework, specifically tailored to capture intricate relationships within implied volatilities. We establish the connection between fNTK and kernel regression, emphasizing its role in contemporary nonparametric statistical modeling. Empirically, we analyze S&P 500 Index options from January 2009 to December 2021, encompassing more than 6 million European calls and puts, thereby showcasing the superior forecast accuracy of fNTK.We demonstrate the significant economic value of having an accurate implied volatility forecaster within trading strategies. Notably, short delta-neutral straddle trading, supported by fNTK, achieves a Sharpe ratio ranging from 1.45 to 2.02, resulting in a relative enhancement in trading outcomes ranging from 77% to 583%.

Back to Contents Page →

James Stephen Marron
University of California, Davis, USA

Data Integration Via Analysis of Subspaces (DIVAS)

A major challenge in the age of Big Data is the integration of disparate data types into a single data analysis. That is tackled here in the context of data blocks measured on a common set of experimental cases. Joint variation is defined in terms of modes of variation having identical scores across data blocks. That allows mathematically rigorous formulation of individual variation within each data block in terms of individual modes. These are mathematically defined through modes of variation with common scores. DIVAS improves earlier methods using a novel random direction approach to statistical inference, and by treating partially shared blocks. Usefulness is illustrated using mortality, cancer and neuroimaging data sets.

Back to Contents Page →

Hans-Georg Muller
University of California, Davis, USA

Modeling Distribution-Valued Random Trajectories With Optimal Transports

Statistical models for samples of (one-dimensional) distribution-valued stochastic processes are of interest for longitudinally observed distributions such as age-at-death distributions in demography and are part of the emerging area of Distributional Data Analysis. While functional data analysis provides a toolbox for the analysis of samples of processes that take values in Euclidean spaces, there is at present no coherent statistical methodology available for samples of distribution-valued processes.

To address this challenge, we introduce a transport model for samples of distribution-valued stochastic processes that implements an intrinsic approach, whereby the observed distributions are converted to optimal transports from the respective barycenter.

Substituting transports for distributions addresses the challenge of centering distribution-valued processes and leads to a useful and interpretable representation of each realized process by an overall time-independent transport and a real-valued trajectory. This representation is obtained by utilizing a recently introduced scalar multiplication for transports {Zhu, M. JRSSB 2023) and facilitates a connection to Gaussian processes. This connection makes it possible to include longitudinal scenarios where the distributionvalued processes are only observed on a sparse grid of randomly situated time points. The proposed approach is supported by convergence results and its practical utility in applications.

This talk is based on joint work with Hang Zhou.

Back to Contents Page →

Byeong Park
Seoul National University, Korea

High-dimensional Hilbert-Schmidt linear regression for Hilbert manifold variables

---

In this paper we propose a novel high-dimensional linear regression technique for variables taking values in Hilbert manifolds. Our approach offers a flexible framework where the response and covariates originate from distinct spaces and are interconnected by Hilbert-Schmidt operators. Our methodology is designed for situations where the number of the covariates grows as the sample size increases and some of the covariates together with the response take values in infinite dimensional spaces. It is formulated under a general penalization scheme that includes various non-convex penalty functions.

Leveraging modern statistical theory for data residing on Hilbert manifolds, we establish oracle property and derive error bounds for the proposed estimators in various convergence modes. We also provide an efficient computational algorithm to solve the associated constrained optimization problem. The practical performance of the proposed method is demonstrated via numerical simulation and real data applications.

Back to Contents Page →

Xinghao Qiao
London School of Economics and Political Science, UK

Large-scale multiple testing of cross-covariance functions with applications to functional network models

In this talk, we reframe the functional covariance model estimation as a tuning-free problem of simultaneously testing $p(p\text{-}1)/2$ hypotheses for cross-covariance functions. Our procedure begins by constructing a Hilbert--Schmidt-norm-based test statistic for each pair, and employs normal quantile transformations for all test statistics, upon which a multiple testing step is proposed. We then explore the multiple testing under a general error-contamination framework and establish that our procedure can control false discoveries asymptotically. Additionally, we show the seamless integration of our proposed methods for two concrete examples: the functional covariance model with partial observations and the functional graphical model, into the general error-contamination framework, and, with verifiable conditions, achieve theoretical guarantees on false discovery control. Finally, we demonstrate the superiority of our proposal through extensive simulations and functional connectivity analysis of two neuroimaging datasets.

Back to Contents Page →

Jane-Ling Wang
University of California, Davis. USA

FDA in the age of AI

Estimation and prediction in functional linear models with scalar response and fully or densely observed functional covariates have been widely studied in the literature. However, the extension to sparsely observed functional covariates remains an open problem. The key obstacle in prediction is the approximation of the index in the functional linear model, which involves the inner product of the functional covariates with the corresponding coefficient function. However, such an approximation cannot be consistent if the functional covariates are observed sparsely at only a few time points. We provide a solution by imputing the sample functional path and substituting the imputed path into the original functional linear model. We show why such a substitution method works and propose two estimation approaches: one through the normal equation, which targets the population quantities, and the other through estimation of the slope function using a reproducing kernel Hilbert space approach. Asymptotic properties of these estimators and their corresponding prediction methods are established. Numerical performance of the proposed methods is evaluated through simulations and real data from the Framingham Heart Study.

Back to Contents Page →

Fang Yao
Peking University, China

FPCA with Diverging Dimensions: From Sparse to Dense Designs

TBA Functional data analysis is an important research field in statistics which treats data as random functions drawn from some infinite-dimensional functional space, and functional principal component analysis (FPCA) based on eigen-decomposition plays a central role for data reduction and representation. After nearly three decades of research, there remains a key problem unsolved, namely, the perturbation analysis of covariance operator for diverging number of eigencomponents obtained from noisy and discretely observed data. This is fundamental for studying models and methods based on FPCA, while there has not been substantial progress since the result obtained by Hall et al. (2006) for a fixed number of eigenfunction estimates. In this work, we establish a unified theory for this problem, obtaining upper bounds for eigenfunctions with diverging indices in both the L2 and supremum norms, and deriving the asymptotic distributions of eigenvalues for a wide range of sampling schemes. Our results provide insight into the phenomenon when the L2 bound of eigenfunction estimates with diverging indices is minimax optimal as if the curves are fully observed, and reveal the transition of convergence rates from nonparametric to parametric regimes in connection to sparse or dense sampling. The technical arguments in this work are useful for handling the perturbation series with noisy and discretely observed data and can be applied in models or those involving inverse problems based on FPCA as regularization, such as functional linear regression.

Back to Contents Page →