Abstracts

Contents

Updated: August 20, 2024

Giles Hooker
University of Pennsylvania, USA

Functional Data Analysis as Nonparametric ODEs

This talk will examine the recent developments in model selection for description of system dynamics through the lens of functional data analysis. In particular, the observation of replicate process motivates the development of a class of smoothly-forced stochastic systems, which in turn have implications for the statistical properties of our estimators and attendant uncertainty quantification methods. The same models have implication for what we can expect from our observations that also limit how well we can resolve quantities such as global stability properties. We will explore an analysis of biomechanics as a natural application of these methods.

Back to Contents Page →

Tailen Hsing
University of Michigan, USA

Variable Selection in High-dimensional Functional Linear Models

---

High-dimensional functional data has become increasingly prevalent in modern applications such as high-frequency financial data and neuroimaging data analysis. We investigate a class of high-dimensional linear regression models, where each predictor is a random element in an infinite-dimensional function space, and the number of functional predictors p can potentially be much greater than the sample size n. Assuming that each of the unknown coefficient functions belongs to some reproducing kernel Hilbert space (RKHS), we regularized the fitting of the model by imposing a group elastic-net type of penalty on the RKHS norms of the coefficient functions. We show that our loss function is Gateaux sub-differentiable, and our functional elastic-net estimator exists uniquely in the product RKHS. Under suitable sparsity assumptions and a functional version of the irrepresentible condition, we establish the variable selection consistency property of our approach. As a byproduct, we present a two-stage estimation procedure that achieves the minimax optimal rate for a class of models.

Back to Contents Page →

Jian Huang
The Hong Kong Polytechnic University, Hong Kong

Conditional Stochastic Interpolation for Generative Learning

We propose a conditional stochastic interpolation (CSI) method for learning conditional distributions. CSI is based on estimating probability flow equations or stochastic differential equations that transport a reference distribution to the target conditional distribution. This is achieved by first learning the conditional drift and score functions based on CSI, which are then used to construct a deterministic process governed by an ordinary differential equation or a diffusion process for conditional sampling. In our proposed approach, we incorporate an adaptive diffusion term to address the instability issues arising in the diffusion process. We derive explicit expressions of the conditional drift and score functions in terms of conditional expectations, which naturally lead to an nonparametric regression approach to estimating these functions. Furthermore, we establish nonasymptotic error bounds for learning the target conditional distribution. We illustrate the application of CSI on image generation using a benchmark image dataset.

Back to Contents Page →

Jialiang Li
National University of Singapore, Singapore

Robust Model Averaging Prediction of Longitudinal Response with Ultrahigh-dimensional Covariate

---

Model averaging is an attractive ensemble technique to construct fast and accurate prediction. Despite of having been widely practiced in cross-sectional data analysis, its application to longitudinal data is rather limited so far. We consider model averaging for longitudinal response when the number of covariates is ultrahigh. To this end, we propose a novel two-stage procedure in which variable screening is first conducted and then followed by model averaging. In both stages, a robust rank-based estimation function is introduced to cope with potential outliers and heavy-tailed error distributions, while the longitudinal correlation is modeled by a modified Cholesky decomposition method and properly incorporated to achieve efficiency. Asymptotic properties of our proposed methods are rigorously established, including screening consistency and convergence of the model averaging estimates. Extensive simulation studies demonstrate that our method outperforms existing competitors, resulting in significant improvements in screening and prediction performance. Finally, we apply our proposed framework to analyze a human microbiome dataset, showing the capability of our procedure in resolving robust prediction using massive metabolites.

Back to Contents Page →

Eftychia Solea
Queen Mary University of London, UK

Robust Inverse Regression for Multivariate Elliptical Functional Data

Functional data have received significant attention as they frequently appear in modern applications, such as functional magnetic resonance imaging (fMRI) and natural language processing. The infinite-dimensional nature of functional data makes it necessary to use dimension reduction techniques. Most existing techniques, however, rely on the covariance operator, which can be affected by heavy-tailed data and unusual observations. Therefore, in this paper, we consider a robust sliced inverse regression for multivariate elliptical functional data. For that reason, we introduce a new statistical linear operator, called the conditional spatial sign Kendall's tau covariance operator, which can be seen as an extension of the multivariate Kendall's tau to both the conditional and functional settings. The new operator is robust to heavy-tailed data and outliers, and hence can provide a robust estimate of the sufficient predictors. We also derive the convergence rates of the proposed estimators for both completely and partially observed data. Finally, we demonstrate the finite sample performance of our estimator using simulation examples and a real dataset based on fMRI.

Back to Contents Page →

Zhigang Yao
National University of Singapore, Singapore

Manifold Learning: An Invitation to Data Science

The manifold fitting problem can go back to H. Whitney's work in the early 1930s (Whitney (1992)), and finally has been answered in recent years by C. Fefferman's works (Fefferman, 2006, 2005). The solution to the Whitney extension problem leads to new insights for data interpolation and inspires the formulation of the Geometric Whitney Problems (Fefferman et al. (2020, 2021a)): Assume that we are given a set $Y \subset \mathbb{R}^D$. When can we construct a smooth d-dimensional submanifold $\widehat{M} \subset \mathbb{R}^D$ to approximate Y, and how well can $\widehat{M}$ estimate Y in terms of distance and smoothness? To address these problems, various mathematical approaches have been proposed (see Fefferman et al. (2016, 2018, 2021b)). However, many of these methods rely on restrictive assumptions, making extending them to efficient and workable algorithms challenging. As the manifold hypothesis (non-Euclidean structure exploration) continues to be a foundational element in data science, the manifold fitting problem, merits further exploration and discussion within the modern data science community.

This talk will be partially based on recent works of Yao and Xia (2019), Yao, Su, Li and Yau (2022), Yao, Su, and Yau (2023) and Yao, Li, Lu, and Yau (2023).

Back to Contents Page →