IMS Young Mathematical Scientists Forum — Statistics and Data Science

24 Nov 2025–28 Nov 2025

November 10, 2025

Abstracts

Extreme Value Theory for Singular Subspace Estimation in the Matrix Denoising Model

Joshua Cape

University of Wisconsin-Madison, USA

This talk considers fine-grained singular subspace estimation in the matrix denoising model where a deterministic low-rank signal matrix is additively perturbed by a stochastic matrix of Gaussian noise. We establish that the maximum Euclidean row norm (i.e., the two-to-infinity norm) of the aligned difference between the leading sample and population singular vectors approaches the Gumbel distribution in the large-matrix limit, under suitable signal-to-noise conditions and after appropriate centering and scaling. We apply our novel asymptotic distributional theory to test hypotheses of low-rank signal structure encoded in the leading singular vectors and their corresponding principal subspace. We provide de-biased estimators for the corresponding nuisance signal singular values and show that our proposed plug-in test statistic has desirable properties. Notably, compared to using the Frobenius norm subspace distance, our test statistic based on the two-to-infinity norm has higher power to detect structured alternatives that differ from the null in only a few matrix entries or rows. Our main results are obtained by a novel synthesis of and technical analysis involving entrywise matrix perturbation analysis, extreme value theory, saddle point approximation methods, and random matrix theory. Our contributions complement the existing literature for matrix denoising focused on minimaxity, mean squared error analysis, unitarily invariant distances between subspaces, component-wise asymptotic distributional theory, and row-wise uniform error bounds. Numerical simulations illustrate our main results and demonstrate the robustness properties of our testing procedure to non-Gaussian noise distributions.

Combining Evidence Across Filtrations

Yo Joong Choe
INSEAD Singapore, Singapore

In sequential anytime-valid inference, any admissible procedure must be based on e-processes: generalizations of test martingales that quantify the accumulated evidence against a composite null hypothesis at any stopping time. We propose a method for combining e-processes constructed in different filtrations but for the same null. Although e-processes in the same filtration can be combined effortlessly (by averaging), e-processes in different filtrations cannot because their validity in a coarser filtration does not translate to a finer filtration. This issue arises in sequential tests of randomness and independence, as well as in the evaluation of sequential forecasters. We establish that a class of functions called adjusters can lift arbitrary e-processes across filtrations. The result yields a generally applicable "adjust-then-combine" procedure, which we demonstrate on the problem of testing randomness in real-world financial data. Furthermore, we prove a characterization theorem for adjusters that formalizes a sense in which using adjusters is necessary. There are two major implications. First, if we have a powerful e-process in a coarsened filtration, then we readily have a powerful e-process in the original filtration. Second, when we coarsen the filtration to construct an e-process, there is a logarithmic cost to recovering validity in the original filtration. This is joint work with Aaditya Ramdas (Carnegie Mellon University).

The Computational Advantage of Depth: Learning High-Dimensional Hierarchical Functions with Gradient Descent

Yatin Dandi
École Polytechnique Fédérale de Lausanne, Switzerland

Understanding the advantages of deep neural networks trained by gradient descent (GD) compared to shallow models remains an open theoretical challenge. In this paper, we introduce a class of target functions (single and multi-index Gaussian hierarchical targets) that incorporate a hierarchy of latent subspace dimensionalities. This framework enables us to analytically study the learning dynamics and generalization performance of deep networks compared to shallow ones in the high-dimensional limit. Specifically, our main theorem shows that feature learning with GD successively reduces the effective dimensionality, transforming a high-dimensional problem into a sequence of lower-dimensional ones. This enables learning the target function with drastically less samples than with shallow networks. While the results are proven in a controlled training setting, we also discuss more common training procedures and argue that they learn through the same mechanisms.

Generalized Data Thinning Using Sufficient Statistics

Ameer Dharamshi
University of Washington, USA

Our goal is to develop a general strategy to decompose a random variable X into multiple independent random variables, without sacrificing any information about unknown parameters. A recent paper showed that for some well-known natural exponential families, X can be thinned into independent random variables X1,...,XK, such that X=X1+...+XK. These independent random variables can then be used for various model validation and inference tasks, including in contexts where traditional sample splitting fails. In this article, we generalize their procedure by relaxing this summation requirement and simply asking that some known function of the independent random variables exactly reconstruct X. This generalization of the procedure serves two purposes. First, it greatly expands the families of distributions for which thinning can be performed. Second, it unifies sample splitting and data thinning, which on the surface seem to be very different, as applications of the same principle. This shared principle is sufficiency. We use this insight to perform generalized thinning operations for a diverse set of families.

Algorithmic Pursuit of Causality

Yihong Gu
Princeton University, USA

The past few decades have witnessed remarkable advances in modern machine learning, particularly in deep learning and large language models, which now lead state-of-the-art prediction systems. However, from a statistical view, these methods often inherit a fundamental limitation: the learning target is to find the most predictive solution in population, which inevitably reflects intrinsic biases present in the collected data. This limitation results in fitted models that misrepresent causal relationships and so hinders the further intelligence of the system.

In this talk, I would present a solution to this challenge, particularly when domain knowledge is unavailable. The refined objective is designed to find a set of variables that yield invariant predictions across diverse environments while minimizing prediction error. I will discuss the proposed scalable and sample-efficient estimation framework, the causal interpretation of the refined target, the fundamental computational limits in theory, and practical approaches for efficient computation.

Efficient Data Integration Under Prior Probability Shift

Ming-Yueh Huang
Academia Sinica, Taipei

Conventional supervised learning usually operates under the premise that data are collected from a homogeneous underlying population. However, challenges may arise when integrating new data from different populations, resulting in a phenomenon known as dataset shift.

In this talk, we will focus on prior probability shift, a specific form of dataset shift, where the distribution of the outcome varies across different datasets but the conditional distribution of features given the outcome remains the same.

To tackle the challenges posed by this shift, we propose a maximum likelihood estimation method that efficiently amalgamates information from multiple sources under prior probability shift. Unlike existing methods that are restricted to discrete outcomes, the proposed approach accommodates both discrete and continuous outcomes.

It also handles high-dimensional covariate vectors through variable selection using an adaptive LASSO penalty, producing efficient estimates that possess the oracle property. Moreover, a novel semiparametric likelihood ratio test is proposed to check the validity of prior probability shift assumptions by embedding the null conditional density function into Neyman's smooth alternatives and testing study-specific parameters.

The proposed methods serve as a useful addition to the repertoire of tools for addressing challenges that arise from dataset shifts in machine learning.

A Computable Measure of Suboptimality for Entropy-Regularised Variational Objectives

Heishiro Kanagawa
Newcastle University, UK

Several emerging post-Bayesian methods target a probability distribution for which an entropy-regularised variational objective is minimised. This increased flexibility introduces a computational challenge, as one loses access to an explicit unnormalised density for the target. To mitigate this difficulty, we introduce a novel measure of suboptimality called gradient discrepancy, and in particular a kernel gradient discrepancy (KGD) that can be explicitly computed. In the standard Bayesian context, KGD coincides with the kernel Stein discrepancy (KSD), and we obtain a novel characterisation of KSD as measuring the size of a variational gradient. Outside this familiar setting, KGD enables novel sampling algorithms to be developed and compared, even when unnormalised densities cannot be obtained.

When Does Adaptive Experimentation Permit Valid Inference?

Koulik Khamaru

Rutgers University, USA

Modern decision-making increasingly relies on adaptive experimentation, particularly in settings such as A/B testing, multi-armed bandits, and reinforcement learning. While these methods enable more efficient learning and allocation of resources, they fundamentally challenge traditional statistical inference. Classical i.i.d.-based tools often break down under such adaptive data collection, resulting in biased estimators and misleading confidence intervals.

This talk offers an overview of statistical inference in these adaptive environments. We highlight the pitfalls of naive inference through concrete examples and introduce the concept of stability—originally formulated by Lai and Wei (1982)—as a unifying principle for valid inference under adaptivity. We show how certain algorithms, such as the Upper Confidence Bound (UCB), achieve stability and thus permit the use of classical inferential methods, even in the absence of independence. In contrast, we highlight how other popular algorithms, such as Thompson Sampling, do not satisfy the stability condition, which can lead to invalid inference; we also propose fixes for special cases. Finally, we demonstrate the utility of such stability results through central limit theorems for both the stochastic multi-armed bandit and contextual bandit problems.

Specialization after Generalization: Towards Understanding Test-Time Training in Foundation Models

Gil Kur

ETH Zürich, Switzerland

Recent empirical studies have explored the idea of continuing to train a model at test-time for a given task, known as test-time training (TTT), and have found it to yield significant performance improvements. However, there is limited understanding of why and when TTT is effective.

Earlier explanations mostly focused on the observation that TTT may help when applied to out-of-distribution adaptation or used with privileged data. However, the growing scale of foundation models with most test data being in-distribution questions these explanations.

We instead posit that foundation models remain globally underparameterized, with TTT providing a mechanism for specialization after generalization — focusing capacity on concepts relevant to the test task. Specifically, under the linear representation hypothesis, we propose a model in which TTT achieves a substantially smaller in-distribution test error than global training.

We empirically validate our model's key assumptions by training a sparse autoencoder on ImageNet, showing that semantically related data points are explained by only a few shared concepts. Finally, we conduct scaling studies across image and language tasks, confirming the practical implications of our model and identifying the regimes where specialization is most effective.

Based on a joint work with Jonas Hübotter, Patrick Wolf, Alexander Shevchenko, Dennis Jüni, and Andreas Krause.

HeteroJIVE: Weighted Spectral Estimation for Shared Subspace Recovery under Multi-View Heteroskedasticity

Jingyang Li

University of Michigan, USA

Many modern datasets consist of multiple related matrices measured on a common set of units, where the goal is to recover the shared low-dimensional subspace across views. The widely used AJIVE algorithm performs this task via a two-stage spectral procedure and serves as the foundation for numerous integrative analyses. However, existing theoretical results for AJIVE remain unsatisfactory: even under homogeneous noise, current bounds are not rate-sharp, and the bias term identified in recent studies does not always persist.

In this talk, I will present a refined analysis that closes this gap and establishes sharp non-asymptotic error bounds for AJIVE under both homogeneous and heterogeneous noise. The key insight is that the so-called non-diminishing term vanishes in regimes satisfying mild orthogonality or random-orientation conditions, revealing a faster rate than previously thought. Building on this understanding, we propose HeteroJIVE, a weighted spectral estimator that adapts to view-specific noise levels while retaining AJIVE's computational simplicity. I will discuss the oracle-optimal weighting rule, its plug-in implementation, and how the interplay between weighting and subspace geometry determines statistical efficiency in large-view integration.

Root Cause Discovery via Permutations and Cholesky Decomposition

Jinzhou Li

National University of Singapore, Singapore

Although the statistical literature on causality has largely focused on forward causal problems concerning the effects of causes, reverse causal questions about identifying the causes of effects are equally important. In this talk, we discuss one such reverse causal question, known as root cause discovery, which aims to identify the root cause of an observed effect. This work is motivated by the problem of identifying the disease-causing gene (i.e., the root cause) in a patient affected by a monogenic disorder, using the gene expression data of healthy individuals as reference. We consider a linear structural equation model where the causal ordering is unknown. We first show that simply comparing marginal squared z-scores cannot identify the root cause in general. We then prove, without additional assumptions, that the root cause is identifiable even when the causal ordering is not. Two key ingredients of this identifiability result are the use of permutations and Cholesky decomposition, which allow us to exploit an invariant property across different permutations to discover the root cause. Furthermore, we characterize permutations that yield the correct root cause and, based on this, propose a valid method for root cause discovery. We also adapt this approach to high-dimensional settings. Finally, we evaluate the performance of our methods through simulations and apply the high-dimensional method to identify disease-causing genes in the gene expression dataset that motivates this work.

Unifying Regression-based and Design-based Causal Inference in Time-series Experiments

Zhexiao Lin

University of California, Berkeley, USA

Time-series experiments, also called switchback experiments or N-of-1 trials, play increasingly important roles in modern applications in medical and industrial areas. Under the potential outcomes framework, recent research has studied time-series experiments from the design-based perspective, allowing the randomness in the design to drive the statistical inference. Focusing on simpler statistical methods, we examine the design-based properties of regression-based methods for estimating treatment effects in time-series experiments. We demonstrate that the treatment effects of interest can be consistently estimated using ordinary least squares with an appropriately specified working model with transformed regressors. Our analysis allows for estimating a diverging number of treatment effects simultaneously, and establishes the asymptotic normality of the resulting estimators. Additionally, we show that asymptotically, the heteroskedasticity and autocorrelation consistent variance

estimators provide conservative estimates of the true, design-based variances. Importantly, although our approach relies on regression, our design-based framework allows for misspecification of the regression model.

Recent Advances of the Fingerprinting Method in Private and Adaptive Data Analysis

Xin Lyu

University of California, Berkeley, USA

Fingerprinting codes were originally introduced as a cryptographic primitive (Boneh and Shaw, 1995; Tardos, 2003), motivated by the goal of defending against digital content piracy. The seminal work of Bun, Ullman, and Vadhan (2013) revealed a foundational connection between fingerprinting codes and the sample complexity of publishing counting statistics under data privacy constraints. Since then, fingerprinting codes—and the methods underlying their construction—have become a cornerstone technique of data privacy, with numerous applications across diverse domains including distribution learning and estimation, streaming, and sublinear-time algorithms. Notably, their influence extends even beyond privacy, finding applications in adaptively robust statistics and data analysis.

In this talk, I will describe our recent research developing a geometry-aware fingerprinting method, partly inspired by advances in the differential privacy literature (Hardt and Talwar, 2010; Nikolov, Talwar, and Zhang, 2013). Our approach has led to substantial progress on several long-standing open problems. Among other results, we resolved a central question in the emerging area of adaptive data analysis (Dwork et al., 2015) and characterized the optimal privacy–accuracy trade-off for answering ensembles of randomly chosen counting queries—refuting a folklore conjecture and advancing a well-studied problem dating back to the inception of differential privacy.

This work is joint with Kunal Talwar (Apple). A preliminary version appeared at STOC 2025.

Adaptive Transfer Clustering: A Unified Framework

Zhongyuan Lyu

The University of Sydney, Australia

We propose a general transfer-learning framework for clustering when a main dataset and an auxiliary dataset describe the same subjects but may exhibit related—yet distinct—latent group structures. Our Adaptive Transfer Clustering (ATC) method automatically leverages shared structure while accommodating unknown discrepancies by optimizing an estimated bias–variance trade-off. ATC applies broadly, including Gaussian mixture models, stochastic block models, and latent class models. We establish optimality guarantees for ATC under Gaussian mixtures and explicitly quantify the gains from transfer. Extensive simulations and real-data examples demonstrate strong and robust performance across a range of scenarios.

Sampling as Bandits: Evaluation-Efficient Design for Black-Box Densities

Takuo Matsubara

The University of Edinburgh, UK

We introduce bandit importance sampling (BIS), a new class of importance sampling methods designed for settings where the target density is expensive to evaluate. In contrast to adaptive importance sampling, which optimises a proposal distribution, BIS directly designs the samples through a sequential strategy that combines space-filling designs with multi-armed bandits. Our method leverages Gaussian process surrogates to guide sample selection, enabling efficient exploration of the parameter space with minimal target evaluations. We establish theoretical guarantees on convergence and demonstrate the effectiveness of the method across a broad range of sampling tasks. BIS delivers accurate approximations with fewer target evaluations, outperforming competing approaches across multimodal, heavy-tailed distributions, and real-world applications to Bayesian inference of computationally expensive models.

Regularizing Extrapolation in Causal Inference

Harsh Parikh
Yale University, USA

Many common estimators in machine learning and causal inference are linear smoothers, where the prediction is a weighted average of the training outcomes. Some estimators, such as ordinary least squares and kernel ridge regression, allow for arbitrarily negative weights, which improve feature imbalance but often at the cost of increased dependence on parametric modeling assumptions and higher variance. By contrast, estimators like importance weighting and random forests (sometimes implicitly) restrict weights to be non-negative, reducing dependence on parametric modeling and variance at the cost of worse imbalance. In this paper, we propose a unified framework that directly penalizes the level of extrapolation, replacing the current practice of a hard non-negativity constraint with a soft constraint and corresponding hyperparameter. We derive a worst-case extrapolation error bound and introduce a novel "bias-bias-variance" tradeoff, encompassing biases due to feature imbalance, model misspecification, and estimator variance; this tradeoff is especially pronounced in high dimensions, when positivity is poor. We then develop an optimization procedure that regularizes this bound while minimizing imbalance and outline how to use this approach as a sensitivity analysis for dependence on parametric modeling assumptions. We demonstrate the effectiveness of our approach through synthetic experiments and a real-world application, involving the generalization of randomized controlled trial estimates to a target population of interest.

Bagging Regularized M-estimators: Precise Asymptotics and Cross-validation

Pratik Patil
University of California, Berkeley, USA

Ensemble methods can improve model stability by averaging predictions from multiple base learners. A canonical ensemble method is bagging [bootstrap aggregating] (which trains models on resampled datasets and averages their outputs), and its sibling, subagging [subsampled bootstrap aggregating] (which uses subsampled datasets). Beyond computational benefits, subagging can also improve generalization, especially in overparameterized regimes near interpolation thresholds (e.g., by smoothing out the double descent peaks). This talk will present theoretical results on subagging of regularized M-estimators under proportional asymptotics, where the sample size, feature size, and subsample sizes all grow with fixed limiting ratios.

Risk asymptotics: Precise risk formulas for heterogeneous ensembles (each component may have a different subsample size, loss, and regularizer), derived via a

provably contractible nonlinear system that captures limiting correlations between estimator errors and residuals on overlapping subsamples. {Theorem 2 of [1]}

Optimal (oracle) tuning properties: In the homogeneous case (common loss, penalty, and subsample size), optimally tuning the subsample size k results in monotonic risk behavior in n, and as the number of ensembles $M \to \infty$, the optimal $k^\star$ lies in the overparameterized regime ($k^\star \leq \min(n, p)$) when bagging estimators with vanishing explicit regularization. {Proposition 7 of [1]}

Data-dependent tuning: A corrected GCV (CGCV) procedure, adding a simple degrees-of-freedom adjustment, yields consistent risk estimates for any finite ensemble without sample splitting or refitting. {Theorem 2 of [2]}

Of independent interest, in the non-ensemble setting (M = 1), our analysis also establishes convergence of trace-based degrees-of-freedom functionals, extending previous results for square loss and ridge, lasso regularizers. {Table 2 of [1]}

This is joint work with the following collaborators (in alphabetical order): Pierre Bellec, Jin-Hong Du, Takuya Koriyama, and Kai Tan, and will feature the following papers (in presentation order): [1] subagging asymptotics (https://pratikpatil.io/papers/subagging-asymptotics.pdf); [2] corrected generalized cross-validation (https://pratikpatil.io/papers/cgcv.pdf).

Optimal Assortment Inference within an Online Learning Framework

Shuting Shen

National University of Singapore, Singapore

The modern retailing system is witnessing fast updating in product features and customer behaviors, entailing adaptive policies that can effectively capture the dynamics of customer preferences. To optimize potential revenues and manage the risks associated with changing customer preferences, it is important to develop an online framework that quantifies the uncertainty of the optimal assortment adaptively. We study the combinatorial inference of the optimal assortment within the framework of the contextual multinomial logit model. In this setting, customer choice outcomes are actively collected over a series of T time points, where the contextual information for n products, including embedding vectors that capture the customer-product dynamics and evolving market trends, as well as revenue parameters, varies over time. Using a dynamic policy, the offer set is adaptively selected at each time point based on historical data. We propose an inferential procedure that constructs a discrete confidence set for the true optimal assortment at the end of the time series, facilitating inference on key properties of the optimal assortment, such as the number of product categories to include in the offer set.

The temporal dependency and combinatorial structure of the Hessian matrix of the log-likelihood function create challenges for convergence analysis. To address these, we develop new probabilistic results on anti-concentration bounds for the difference between the maxima of two Gaussian random vectors. Furthermore, we address the

high dimensionality of the combinatorial inference problem by employing discretization via epsilon-covering and subspace projection techniques. We provide theoretical guarantees on both the validity and power of our inferential procedure, and establish information-theoretic lower bounds for the required signal strength, which match the upper bounds of our procedure up to logarithmic factors.

Bridging Root-n and Non-standard Asymptotics: Adaptive Inference in M-Estimation

Kenta Takatsu

Carnegie Mellon University, USA

We study a general approach to construct confidence sets for the solution of population-level optimization, commonly referred to as M-estimation. Statistical inference for M-estimation poses significant challenges due to the non-standard limiting behaviors of the corresponding estimator, which arise in settings with increasing dimension of parameters, non-smooth objectives, or constraints. We propose a simple and unified method that guarantees validity in both regular and irregular cases. Moreover, we provide a comprehensive width analysis of the proposed confidence set, showing that the convergence rate of the diameter is adaptive to the unknown degree of instance-specific regularity. We apply the proposed method to several high-dimensional and irregular statistical problems.

Estimating Generalization Error for Iterative Algorithms in High-Dimensional Regression

Kai Tan

Stanford University, USA

In this talk, I will present recent progress on understanding the generalization error of iterates from iterative algorithms in high-dimensional regression. I will introduce estimators for Gradient Descent-type algorithms, which consistently track the error and allow data-driven selection of the optimal iteration number. I will then discuss extensions to Stochastic Gradient Descent and its proximal variants in robust regression, where the noise may have infinite variance. Simulations on synthetic data demonstrate the practical utility of these methods.

Divide-and-shrink: An Efficient and Heterogeneity-agnostic Approach for Transfer Estimation

Ruoyu Wang

Harvard T.H. Chan School of Public Health, USA

Knowledge transfer across data sources holds great promise for improving the estimation of target population parameters by leveraging the growing availability of data from different sources. However, the effectiveness of knowledge transfer is often challenged by the complex and pervasive heterogeneity between data sources and the lack of access to individual-level data. This paper proposes the divide-and-shrink (dShrink) method, a transfer estimation method that estimates target population parameters in a closed form using summary statistics from a target population and an external source population while accounting for population heterogeneity. dShrink is guaranteed to outperform the estimator using the target population under arbitrary population heterogeneity. Moreover, dShrink significantly reduces the estimation error when the target and source populations are similar or the underlying true parameter values are near zero. Notably, it is tuning-free, model-free, robust to various types of heterogeneity between data sources, and applies to a broad range of parameter estimation problems. dShrink also offers flexibility in incorporating side information and remains effective even when the covariance matrix is not accessible for the external summary statistics. Simulations and real data analyses demonstrate the superior performance of the dShrink estimator and its potential as a robust tool for transfer estimation.

Universal Log-optimality for General Classes of E-processes and Sequential Hypothesis Tests

Ian Waudby-Smith

University of California, Berkeley, USA

We consider the problem of sequential hypothesis testing by betting. For a general class of composite testing problems – which include bounded mean testing, equal mean testing for bounded random tuples, and some key ingredients of two-sample and independence testing as special cases – we show that any e-process satisfying a certain sublinear regret bound is adaptively, asymptotically, and almost surely log-optimal for a composite alternative. This is a strong notion of optimality that has not previously been established for the aforementioned problems and we provide explicit test supermartingales and e-processes satisfying this notion in the more general case. Furthermore, we derive matching lower and upper bounds on the expected rejection time for the resulting sequential tests in all of these cases. The proofs of these results make weak, algorithm-agnostic moment assumptions and rely on a general-purpose proof technique involving the aforementioned regret and a family of numeraire

portfolios. Finally, we discuss how all of these theorems hold in a distribution-uniform sense, a notion of log-optimality that is stronger still and seems to be new to the literature.

Doubly Robust Calibration of Prediction Sets under Covariate Shift

Yachong Yang
University of Pennsylvania, USA

Conformal prediction has received tremendous attention in recent years and has offered new solutions to problems in missing data and causal inference; yet these advances have not leveraged modern semiparametric efficiency theory for more efficient uncertainty quantification. We consider the problem of obtaining well-calibrated prediction regions that can data adaptively account for a shift in the distribution of covariates between training and test data. Under a covariate shift assumption analogous to the standard missing at random assumption, we propose a general framework based on efficient influence functions to construct well-calibrated prediction regions for the unobserved outcome in the test sample without compromising coverage.

Shifted Composition IV: Toward Ballistic Acceleration for Log-Concave Sampling

Matthew S. Zhang
University of Toronto, Canada

Acceleration is a celebrated cornerstone of convex optimization, enabling gradient-based algorithms to converge sublinearly in the condition number. A major open question is whether an analogous acceleration phenomenon is possible for log-concave sampling. Underdamped Langevin dynamics (ULD) has long been conjectured to be the natural candidate for acceleration, but a central challenge is that its degeneracy necessitates the development of new analysis approaches, e.g., the theory of hypocoercivity. Although recent breakthroughs established ballistic acceleration for the (continuous-time) ULD diffusion via space-time Poincare inequalities, (discrete-time) algorithmic results remain entirely open: the discretization error of existing analysis techniques dominates any continuous-time acceleration.

In this paper, we give a new coupling-based local error framework for analyzing ULD and its numerical discretizations in KL divergence. This extends the framework in Shifted Composition III from uniformly elliptic diffusions to degenerate diffusions, and shares its virtues: the framework is user-friendly, applies to sophisticated discretization schemes, and does not require contractivity. Applying this framework to

the randomized midpoint discretization of ULD establishes the first ballistic acceleration result for log-concave sampling (i.e., sublinear dependence on the condition number). Along the way, we also obtain the first $d^{1/3}$ iteration complexity guarantee for sampling to constant total variation error in dimension d.

Expected Shortfall Random Forest for Heterogeneous Treatment Effect

Shushu Zhang

University of Michigan, USA

Understanding heterogeneous treatment effects in the tail of a response distribution is crucial in many applications. As a comprehensive summary of the tail distribution, the expected shortfall (ES) is defined as the average over the tail above (or below) a certain quantile of a response distribution. Under the joint quantile and ES framework, we propose a novel expected shortfall random forest (ESRF) to model the nonlinear relationship between covariates and the ES of the response. The proposed ESRF approach integrates subsampling and data-splitting schemes to construct a nonparametric ensemble that jointly estimates conditional quantiles and expected shortfalls. Building upon this framework, we further develop the expected shortfall causal random forest (ESCRF) to estimate the conditional expected shortfall treatment effect, defined as the difference between the conditional ES of potential outcomes. We establish the pointwise consistency and the asymptotic normality for both the ESRF and the ESCRF estimators. We illustrate the finite-sample performance of the proposed methods through simulation studies and an empirical application examining health disparities among low-birthweight infants.

Representation Learning for Healthcare Data Analysis

Doudou Zhou

National University of Singapore, Singapore

The growing availability of electronic health records (EHRs) presents unprecedented opportunities to advance biomedical research and improve patient care. However, effectively leveraging such data across diverse healthcare systems remains challenging due to differences in coding schemes, data modalities, patient demographics, and clinical practices. This talk will discuss recent advances in representation learning that enable the extraction of informative, transferable, and interpretable features from heterogeneous healthcare data. These representations support a range of downstream applications, including clinical code mapping, feature selection, and knowledge graph construction, providing a unified foundation for robust and scalable healthcare analytics.

Geodesic Causal Inference

Yidong Zhou
University of California, Davis, USA

Adjusting for confounding and imbalance when establishing statistical relationships is an increasingly important task, and causal inference methods have emerged as the most popular tool to achieve this. Causal inference has been developed mainly for regression relationships with scalar responses and also for distributional responses. We introduce here a general framework for causal inference when responses reside in general geodesic metric spaces, where we draw on a novel geodesic calculus that facilitates scalar multiplication for geodesics and the quantification of treatment effects through the concept of geodesic average treatment effect. Using ideas from Fréchet regression, we obtain a doubly robust estimation of the geodesic average treatment effect and results on consistency and rates of convergence for the proposed estimators. We also study uncertainty quantification and inference for the treatment effect. Examples and practical implementations include simulations and data illustrations for responses corresponding to compositional responses as encountered for U.S. statewise energy source data, where we study the effect of coal mining, network data corresponding to New York taxi trips, where the effect of the COVID-19 pandemic is of interest, and the studying the effect of Alzheimer's disease on connectivity networks.

Doubly Robust Inference on Causal Derivative Effects for Continuous Treatments

Yikun Zhang
University of Washington, USA

Statistical methods for causal inference with continuous treatments mainly focus on estimating the mean potential outcome function, commonly known as the dose-response curve. However, it is often not the dose-response curve but its derivative function that signals the treatment effect. In this talk, we investigate nonparametric inference on the derivative of the dose-response curve with and without the positivity condition. Under the positivity and other regularity conditions, we propose a doubly robust (DR) inference method for estimating the derivative of the dose-response curve using kernel smoothing. When the positivity condition is violated, we demonstrate the inconsistency of conventional inverse probability weighting (IPW) and DR estimators, and introduce novel bias-corrected IPW and DR estimators. In all settings, our DR estimator achieves asymptotic normality at the standard nonparametric rate of convergence. Additionally, our approach reveals an interesting connection to nonparametric support and level set estimation problems. Finally, we demonstrate the applicability of our proposed estimators through simulations and a case study of evaluating a job training program.
