

CSFAI Talk Abstract Booklet

Statistical Foundations of AI

November 13–14, 2026 • University of California, Santa Barbara

This detailed companion follows the current conference program. It contains the full submitted abstracts, titles, presenter names, and affiliations for all 131 scheduled parallel-session talks, along with plenary and banquet entries and both panel rosters.

Sessions follow program order. Each room’s two-hour talk block is divided equally among its speakers. Times are Pacific Standard Time. Click a session code in the guide or a name in the presenter index to navigate.

Program guide

Day and time	Corwin	Flying-A	Harbor	State Street
Fri Nov 13 8:30–10:30 am	<u>1C</u> Talks Statistical computation and structured modeling	<u>1F</u> Talks AI agents and adaptive evaluation	<u>1H</u> Talks Statistical learning for biomedical data	<u>1S</u> Talks Generative models and learning dynamics
Fri Nov 13 1:00–3:00 pm	<u>2C</u> Talks Trustworthy AI and structured inference	<u>2F</u> Talks Regression and predictive inference	<u>2H</u> Talks Statistical learning and assisted inference	<u>2S</u> Talks Network models, representations, and experiments
Fri Nov 13 3:30–5:30 pm	<u>3C</u> Panel Industry perspectives on trustworthy AI	<u>3F</u> Talks Online learning and decision making	<u>3H</u> Talks Statistical inference after learning	<u>3S</u> Talks Causal discovery and treatment effects
Sat Nov 14 8:30–10:30 am	<u>4C</u> Talks Biomedical modeling and AI-assisted inference	<u>4F</u> Talks Evaluation and inference with AI predictions	<u>4H</u> Talks Structured representations and clustering	<u>4S</u> Talks Reinforcement learning and adaptive decisions
Sat Nov 14 1:00–3:00 pm	<u>5C</u> Panel AI influence and the next generation of research	<u>5F</u> Talks Statistical learning and computational inference	<u>5H</u> Talks Multi-view learning and data integration	<u>5S</u> Talks Statistical learning with heterogeneous data
Sat Nov 14 3:30–5:30 pm	<u>6C</u> Talks Reliable learning and structured inference	<u>6F</u> Talks Foundations of AI and distributed learning	<u>6H</u> Talks Learning from distributions and complex measurements	<u>6S</u> Talks Learning from sequential and process data

Plenary and banquet talks

Opening remarks: Shelly Gable, Dean of Mathematical, Life, and Physical Sciences at UCSB • Friday November 13, 11:00–11:05 am • Corwin

Susan Murphy • Friday 11:05 am–12:05 pm • Corwin • Chair: Annie Qu

Xihong Lin • Saturday 11:00 am–12:00 pm • Corwin • Chair: Xiaowu Dai

Sudipto Banerjee • Friday banquet • 7:00–8:30 pm • Corwin

Poster session: Friday November 13 • 5:30–6:30 pm • Corwin. The poster directory is in the program book.

Abstract status. All parallel-talk abstracts are reproduced from supplied submissions, with light typographical and mathematical formatting. Susan Murphy and Sudipto Banerjee have clearly labeled website summaries and links to their published abstracts. Xihong Lin’s full submitted plenary abstract is included. Panelists appear in their panel roles; this booklet covers oral presentations.

Plenary and banquet talks

Reinforcement Learning for Digital Health Interventions in the Dyadic Setting

Susan Murphy • Harvard University

Friday November 13 • 11:05 am–12:05 pm • Corwin

Chair: Annie Qu

Abstract summary. This talk describes online reinforcement learning for digital care involving a patient and carepartner. Three cooperating agents support health behaviors, caregiving, and the relationship between the two people. Causal domain knowledge helps learning when observations are limited. The motivating ADAPTS-HCT trial supports medication adherence among young adults following blood and bone marrow transplantation.

[Full abstract on the conference speaker page](#)

From Generative AI to Statistical Inference: Harnessing Synthetic Data for Validity and Power

Xihong Lin

Department of Biostatistics and Department of Statistics
Harvard University

Saturday November 14 • 11:00 am–12:00 pm • Corwin

Chair: Xiaowu Dai

The integration of statistics and generative AI plays a pivotal role in accelerating trustworthy, cross-domain scientific discovery. Recent advances in generative models have dramatically expanded the availability and use of synthetic data across scientific domains. While these developments create exciting opportunities to enhance data analysis, they also raise fundamental statistical challenges regarding how synthetic data can be used in a valid, reliable, and principled manner. In this talk, we first discuss the generation of high-fidelity, high-dimensional synthetic data for real-world applications using generative AI models, including transformers, diffusion models, and latent flow-matching models. We then present a principled framework for integrating observed and synthetic data in downstream analyses to ensure valid statistical inference, even when generative AI models are misspecified. Specifically, we introduce synthetic-data-assisted methods for continuous and discrete outcomes based on an augmented synthetic data likelihood that integrates observed and synthetic data. We show that these methods are robust to misspecification of black-box generative models and can improve statistical power when the generative models provide informative synthetic data. We demonstrate the utility of these methods through applications to UK Biobank data, including genome-wide association studies (GWAS) of proteomic data, whole-genome sequencing (WGS) analyses of brain imaging phenotypes, and analyses of binary mental health survey outcomes. These applications involve substantial missingness, with approximately 70–90% of the relevant data missing, highlighting the potential of synthetic-data-assisted methods to enable more powerful and reliable scientific discovery in challenging real-world settings.

Carbs, Calories, and Computation: Serving Up Statistical Inference in the AI Ecosystem

Sudipto Banerjee • University of California, Los Angeles

Friday November 13 • Banquet • 7:00–8:30 pm • Corwin

Abstract summary. This talk considers statistical inference within the expanding AI ecosystem. It connects amortized inference, transfer learning, and predictive stacking as tools for efficient probabilistic analysis. A spatial-energetics example uses streaming wearable-device data from UCLA's PASTA-LA project to relate movement to metabolic activity and illustrate public-health applications.

[Full abstract on the conference speaker page](#)

Panel discussions

Panel rosters are listed below. David Talby also speaks in 1F, Hao Helen Zhang in 1C, George Michailidis in 6C, Bertrand Clarke in 1C, and Michael Daniels and Sam Behseta in 1H. David and Sam join the Friday industry panel.

3C Industry perspectives on trustworthy AI

Friday November 13 • 3:30–5:30 pm • Corwin

Moderator: Hao Helen Zhang — University of Arizona

Panelists

David Talby — John Snow Labs

Sam Behseta — Kaiser Permanente Bernard J. Tyson School of Medicine

Guannan Zhai — Amazon

Yujia Deng — Zoom Communications Inc.

Haoda Fu — Amgen

Louis Ehwerhemuepha — CHOC

5C AI influence and the next generation of research

Saturday November 14 • 1:00–3:00 pm • Corwin

Moderator: Regina Liu — Rutgers University

Panelists

George Michailidis — UCLA

Bertrand Clarke — U. Nebraska-Lincoln

Heping Zhang — Yale University

Peter X. K. Song — University of Michigan

Radu Craiu — University of Toronto

Michael Daniels — Department of Statistics, University of Florida

Xiao Wang — Purdue University

Ying Wei — Columbia University

Michele Guindani — University of California, Los Angeles

1C Statistical computation and structured modeling

Friday November 13 • 8:30–10:30 am • Corwin

Chair: Grace Yi • 6 speakers

Theoretical and Empirical Evaluation of a Grouped Gibbs Sampler for Parallel Computation in the Latent Dirichlet Allocation Model

Hani Doss

University of Florida

Latent Dirichlet Allocation is very heavily used for topic modelling. In this Bayesian model, there is a latent variable for every word in every document. Two general approaches for estimating the posterior distribution in the model are Markov chain Monte Carlo (MCMC) and variational inference. Traditional MCMC methods give estimates which converge to the truth as the number of iterations goes to infinity, but because they update, sequentially, the latent variables in the model, they are slow and cannot handle very large corpora. On the other hand, variational methods are fast, so can handle very large corpora; but the estimates they produce do not converge to the truth as the number of variational steps goes to infinity. We describe and analyze a grouped Gibbs sampling algorithm for estimating the posterior distribution which is amenable to parallelization. The algorithm gives estimates that are exact in the limit, and we prove that the chain it produces is uniformly ergodic. We compare it to other parallel MCMC algorithms and to stochastic variational inference. Our criteria for comparisons include speed of execution, and accuracy of estimation as measured by mean squared error, which balances bias and variance, the former being the most important source of error for stochastic variational inference. Our empirical studies show that, for a given computational budget, the grouped Gibbs sampler is comparable to other MCMC algorithms in terms of speed of execution and ergodic properties, but that it greatly outperforms stochastic variational inference and all other MCMC algorithms in terms of accuracy of estimation.

Dynamic Supervised Principal Component Analysis for High-Dimensional Classification

Hao Helen Zhang

University of Arizona

High-dimensional data classification is challenged by distributions that shift dynamically over time, making static subspace definitions and decision boundaries inadequate. We propose a novel framework for dynamic classification in high-dimensional spaces, designed to accommodate evolving class distributions across time or other index variables. The framework employs a supervised dimension reduction technique based on kernel smoothing to identify an optimal subspace and construct adaptive classification boundaries that respond to distributional changes. We develop theory and computational algorithms for both linear and quadratic discriminant analysis, and illustrate effectiveness of the proposed approach through simulation studies and real data applications.

Learning Relationship between Quantum Walks and Underdamped Langevin Dynamics

Yazhen Wang

University of Wisconsin-Madison

Fast computational algorithms are in constant demand, and their development has been driven by advances such as quantum speedup and classical acceleration. This paper intends to study search algorithms based on quantum walks in quantum computation and sampling algorithms based on Langevin dynamics in classical computation. On the quantum side, quantum walk-based search algorithms can achieve quadratic speedups over their classical counterparts. On the classical side, considerable research has focused on gradient acceleration, with gradient-adjusted algorithms derived from underdamped Langevin dynamics providing quadratic acceleration over conventional Langevin algorithms. Because both search and sampling algorithms are designed to address learning tasks, we investigate learning relationship between coined quantum walks and underdamped Langevin dynamics. Specifically, we show that, in terms of the Le Cam deficiency distance, a quantum walk with randomization is asymptotically equivalent to underdamped Langevin dynamics, whereas the quantum walk without randomization is not asymptotically equivalent due to its high-frequency oscillatory behavior. We further discuss the implications of these equivalence and nonequivalence results for the computational and inferential properties of the associated algorithms in machine learning tasks. Our findings offer new insight into the relationship between quantum walks and underdamped Langevin dynamics, as well as the intrinsic mechanisms potentially underlying quantum speedup and classical gradient acceleration.

Quantum Field Theory as Generative Models, Emergent Gravity as Latent Variable Model

Ying Nian Wu

University of California, Los Angeles

Quantum field theory on a finite lattice is a multivariate distribution over field configurations. This talk presents it using three equivalent statistical models. Model 1: a recurrent network with a softmax output layer (the Born rule). Model 2: an energy-based model (exponential family) obtained by Wick rotation, sampleable by MCMC. Model 3: a latent variable model obtained by Hubbard-Stratonovich transformation, which decouples the quartic interaction into two quadratics so that both conditionals are Gaussian — amenable to Gibbs sampling and EM.

We introduce a tensor-square interaction that seeds a latent tensor field. Integrating out the microscopic field (a Gaussian integral) yields the effective action of the latent field, with a Tr-log term from the determinant. Minimizing the effective action (EM / saddle point) gives the Einstein field equation: the emergent spacetime metric is the mode of the latent variable. Sampling from the effective action (Gibbs sampler) includes quantum gravitational fluctuations beyond the saddle point.

The emergent metric is not a mysterious new ontology. It is an organization of the readings of clocks and rulers — macroscopic composites of the microscopic field — no more mysterious than temperature emerging from molecular motion.

The lattice is kept finite throughout. The tools are familiar: exponential families, Gaussian Markov random fields, latent variables, EM, Gibbs sampling, and Tr-log determinants. The physics is general relativity.

Generation-Powered Inference for Distribution-Valued Outcomes

Hongzhe Li

University of Pennsylvania

Modern generative models increasingly produce distribution-valued outputs, such as predicted cellular responses to genetic perturbations in single-cell genomics. While these models provide valuable auxiliary information, they are inherently imperfect, creating a need for statistical methods that leverage their predictions without relying on their correctness. We propose generation-powered inference (GPI), a general framework for improving inference on distribution-valued parameters using auxiliary generative models. Focusing on Wasserstein barycenters and related distributional functionals, we introduce a function-valued bridge representation that transforms inference in the nonlinear Wasserstein space into estimation of a mean function in a Hilbert space, enabling an augmented estimation framework analogous to prediction-powered inference. We develop a family of GPI estimators with optimal information borrowing, establish consistency, asymptotic normality, and simultaneous confidence bands, and derive valid inference for linear functionals and Wasserstein distances. Simulation studies demonstrate efficiency gains over labeled-data-only methods and robust performance under generative model misspecification. We illustrate the proposed framework using a Perturb-seq study of K562 cells, where synthetic perturbation responses generated by the State foundation model are used to improve inference for pathway-level consensus gene expression distributions associated with perturbations of the 40S ribosome module.

Uncertainty Quantification via the Posterior Predictive Variance

Bertrand Clarke

U. Nebraska-Lincoln

We use the law of total variance to generate multiple expansions for the posterior predictive variance. These expansions are sums of terms involving conditional expectations and conditional variances and provide a quantification of the sources of predictive uncertainty. Since the posterior predictive variance is fixed given the model, it represents a constant quantity that is conserved over these expansions. The terms in the expansions can be assessed to understand the main contributors to the length of prediction intervals. We examine the term-wise uncertainty across expansions varying in the number of terms and the order of conditioning. In particular, given that a specific term in one expansion is small or zero, we identify the other terms in other expansions that must also be small or zero. We illustrate a computational technique for assessing whether a term should be retained or set to zero.

1F AI agents and adaptive evaluation

Friday November 13 • 8:30–10:30 am • Flying-A

Chair: Jin Zhou • 6 speakers

Agentifying AI Evaluation: What to Evaluate, When to Stop, and How to Estimate

Jingshen Wang

UC Berkeley

Evaluating modern AI agents is increasingly costly and time-consuming, as benchmarking tasks can require long sequences of model calls and tool interactions. Can we reduce evaluation time and token cost without sacrificing model evaluation accuracy? We introduce an agentic evaluation framework that jointly and adaptively learns what to evaluate and when to stop, with a tailored estimator that preserves evaluation accuracy. The system jointly learns what tasks to evaluate and when to stop execution, while probabilistically randomizing and adapting these decisions to enable statistical correction. By “learning what tasks to evaluate,” it uses historical evaluation traces to identify which benchmark tasks are most informative for estimating agent performance. By “deciding when to stop,” it actively monitors each trajectory and probabilistically continues only when further execution is expected to provide sufficient additional information about the eventual outcome. We pair this agentic system with a tailored estimator that corrects for adaptive task selection and stopping. Our theoretical analysis shows that the resulting framework preserves model evaluation accuracy while improving precision for a given evaluation budget. Experiments on agent benchmarks show that our approach reduces both evaluation time and token cost while preserving accurate model evaluation results.

Statistical Evaluation of LLMs

Linjun Zhang

Rutgers University

Evaluating mathematical reasoning in LLMs is constrained by limited benchmark sizes and inherent model stochasticity, yielding high-variance accuracy estimates and unstable rankings across platforms. On difficult problems, an LLM may fail to produce a correct final answer, yet still provide reliable pairwise comparison signals indicating which of two candidate solutions is better. We leverage this observation to design a statistically efficient evaluation framework that combines standard labeled outcomes with pairwise comparison signals obtained by having models judge auxiliary reasoning chains. Treating these comparison signals as control variates, we develop a semiparametric estimator based on the efficient influence function (EIF) for the setting where auxiliary reasoning chains are observed. This yields a one-step estimator that achieves the semiparametric efficiency bound, guarantees strict variance reduction over naive sample averaging, and admits asymptotic normality for principled uncertainty quantification. Across simulations, our one-step estimator substantially improves ranking accuracy, with gains increasing as model output noise grows. Experiments on GPQA Diamond, AIME 2025, and GSM8K further demonstrate more precise performance estimation and more reliable model rankings, especially in small-sample regimes where conventional evaluation is pretty unstable.

Statistical design for medical AI safety benchmarks: precision targets, joint balance, and perturbation validity

David Talby

John Snow Labs

Evaluation of clinical language models rests on benchmarks that are rarely designed as measurement instruments. Current benchmarks report point estimates without intervals, derive their size from what a generation pipeline produced, and leave their design dimensions confounded. We describe PacificMed, a framework of over 70 benchmarks in eight suites for evaluating bias, fairness, safety and reliability in clinical language models, and the statistical design underlying the 38 built benchmarks that hold over 25,000 cases. First, sample sizes are derived by inverting the confidence interval half-width at a stated precision target. Second, per-subgroup precision is treated as a separate design guarantee, with an interval on every reported figure. Third, cases are allocated on the joint distribution of deployment context and harm severity, so that a per-context comparison is not partly a comparison of harm mix: measured association falls from a Cramér's V of 0.123 to 0.134 in the inherited corpus to 0.044 or below across all ten built safety benchmarks. Fourth, distinctness is verified at two near-duplicate thresholds. Fifth, adversarial rewrites of a base prompt are treated as clustered rather than independent observations, where ignoring the design effect narrows a reported interval from approximately $\pm 3.8\%$ to $\pm 2.0\%$. Sixth, perturbation tests carry a validity criterion, which excluded six candidate benchmarks: varying an attribute measures fairness only where that attribute does not change the correct answer. The test suites are contributed to the open-source MedHELM framework and available to academic researchers under a data use agreement that keeps test cases unpublished, which limits data contamination. We report the most recent model comparison results from the automated suites, using paired tests on common items instead of overlapping marginal intervals. All built suites remain unsaturated: no evaluated frontier model approaches ceiling on any built benchmark, so each retains discriminating power among current frontier LLMs.

Statistical Attribute Alignment for Black-Box Generative AI

Jason Klusowski

Princeton University

Generative AI offers a powerful way to create synthetic data for simulation, training, and evaluation. But what if a model generates outputs with an attribute distribution that differs from what a downstream task requires? For example, medical research may require image collections with specified distributions of demographic groups or disease categories. This talk develops a statistical framework for attribute alignment through output post-processing, using only query access to the generative model. After observing the attributes of generated outputs, we select a batch whose attributes jointly follow a target distribution, exactly or approximately. We develop sampling algorithms and establish their asymptotic optimality in expected model query cost as the desired batch size grows.

The Future of AI Scientists: Emerging Directions and Fundamental Challenges

Jie Ding

University of Minnesota

Modern AI systems are rapidly evolving from passive tools into agentic systems that can reason, learn, collaborate, and interact with human scientists in the research process. This talk discusses emerging directions in AI scientist systems, drawing inspiration from human-centered principles such as lifelong learning, teaching, collaboration, objective alignment, and auditing. The goal is to identify fundamental research challenges and stimulate new perspectives on building AI systems that can operate reliably, adaptively, and transparently in complex real-world research settings.

Building AI Agents for Better Human-Agent Interaction

Ruoqing Zhu

University of Illinois at Urbana-Champaign

Agentic AI tools are increasingly being engineered to automate workflows and complete tasks on behalf of the user. Yet in research and many other complex settings, the decisions that matter most often involve human judgment, preferences, and choices among scientifically reasonable alternatives. In this talk, I will discuss this perspective and illustrate, through several examples, how I have designed agentic tools and research harnesses that deliberately incorporate user interaction and decision-making into the workflow. I will first demonstrate an interactive causal consulting skill that turns causal analysis into a structured dialogue rather than a one-shot recommendation or report. I will then present a research harness that organizes literature review, methodological development, theory, computation, and manuscript writing and revision, while keeping the researcher involved in key decisions throughout the process. Together with several other examples, I will argue that agentic AI is not only something we use, but increasingly something we can design. And thoughtfully constructed human-agent interactions may be essential for making these systems scientifically useful, productive, and trustworthy.

1H Statistical learning for biomedical data

Friday November 13 • 8:30–10:30 am • Harbor

Chair: Yang Feng • 6 speakers

Structured Self-Supervised Learning for Heterogeneous Wearable Data

Jingjing Zou

UC San Diego

Wearable sensors provide rich multivariate time series, yet limited labels and variation across individuals, sensor placements, and acquisition protocols challenge reliable prediction. In this talk, I will present three complementary approaches to representation learning and transfer for wearable data. MoCA uses cross-modality masking and reconstruction to learn temporal and cross-sensor dependencies from accelerometer and gyroscope signals. W-JEPA organizes signals into wavelet tokens indexed by scale, sensor axis, and time, and predicts masked latent representations to capture multiscale motion structure. WearableBridge addresses deployment shifts through reference-anchored transport, using reference recordings available during training to adapt wrist representations to a fixed activity classifier while retaining wrist-only inference. Results across activity-recognition benchmarks and free-living recordings demonstrate gains in predictive performance, with evaluations emphasizing transfer across datasets and generalization to unseen individuals. Together, these methods illustrate how structured self-supervision and task-guided representation transport can support wearable learning under heterogeneous measurement conditions.

Decoding Aging Neurons

Andrew Holbrook

UCLA Biostatistics

I present a novel graph convolutional generative model for learning multiple orders of dependencies from longitudinally observed multivariate functional data. I use this model to analyze and visualize marked arrays of local field potentials (LFP) recorded during weekly behavioral experiments involving both wild type mice and THY-Tau22 mice subject to tau aggregation. I compare the longitudinal change in LFP connectivity to that of high-dimensional neuronal spikes data, paying close attention to the relationship between neuronal silencing and connectivity dynamics within THY-Tau22 mice.

A case study of causal mediation using Bayesian nonparametrics and semiparametric corrections

Michael Daniels

Department of Statistics, University of Florida

We propose a Bayesian nonparametric approach using a truncated Enriched Dirichlet Process mixture (EDPM) model to estimate natural direct (NDE) and indirect (NIE) effects in causal mediation analyses in the presence of post-treatment confounders. We implement a one-step posterior correction based on the efficient influence function for our setting. This post-processing step solves a critical problem in Bayesian nonparametrics: how to obtain reliable estimates and posteriors for a specific causal estimand of interest (the NDE and NIE) with excellent frequentist properties, such as correct coverage, from a model designed for complex joint distributions. We conduct simulation studies to assess our method's performance and apply it to evaluate causal mediation effects in a weight management clinical trial.

The Specification Gap: What Artificial Intelligence Leaves Unresolved in Statistical Inference

Sam Behseta

Kaiser Permanente Bernard J. Tyson School of Medicine

Every statistical analysis rests on assumptions, but not all are treated alike. Some are compared using information criteria, cross-validation, or simulation; others simply enter. Artificial intelligence is making comparison and computation inexpensive, but does not establish that the specified target, the data-generating mechanism, and the key assumptions match the scientific problem. We call this mismatch the specification gap, extending Fisher's sense of specification, and we call the mechanism that sustains it the assessment gradient: effort flows to the choices a criterion can rank, so cheaper comparison improves those choices and leaves the rest as first specified. We examine four such gaps from our research on privacy-preserving synthetic clinical data. Validation showed strong source-to-synthetic agreement, including a very high coefficient correlation, while an important clinical relationship was lost. A privacy bound was applied to repeated measurements rather than individuals. Under an illustrative dependence model, assuming independence among repeated measurements costs roughly 400 times the difference between the methods compared. Formal verification strengthens reasoning from stated assumptions, but scientific adequacy also requires evidence that those assumptions represent the data and the question. The risk of full automation is not nonsense but fluent, well-validated answers to an unexamined question. We draw implications for research and teaching.

Cluster Analysis of Longitudinal Associations Between Microbial Abundance and Environmental Factors

Lan Xue

Oregon State University

Microbial communities are dynamic systems that shift in composition in response to host and environmental factors. Understanding these temporal patterns is essential for identifying taxa with shared biological roles. However, standard clustering approaches for microbiome data often neglect its longitudinal structure and fail to account for external variables that may influence microbial trajectories. We introduce a novel clustering framework named *cleverly*, which incorporates external variables into the clustering of multivariate responses represented as longitudinal and compositional counts. Our method enables the generation of clusters based either on the responses' baseline trajectories or on their reactions to external perturbations, as well as accommodates the compositional, discrete, and time-correlated nature of the data. Using simulated data, we demonstrate that *cleverly* recovers true groupings under a range of biologically realistic conditions, including varying correlation structures and compositional constraints. We further apply *cleverly* to longitudinal fecal microbiome data from wild African buffalo (*Syncerus caffer*), revealing distinct clusters of taxa that either maintain consistent trajectories or respond sharply to changes in resource availability. These findings underscore the importance of accounting for environmental covariates in microbiome clustering and illustrate how *cleverly* enables a deeper understanding of microbial function and host-microbiome interactions in dynamic ecosystems.

Keywords: Compositional count data, External variables, Generalized estimating equations, Pairwise fusion penalty, Temporal trajectories.

Scalable Randomized Kernel Methods for Multiview Data Integration and Prediction

Sandra Safo

University of Minnesota

Many diseases are complex, heterogeneous conditions influenced by molecular, environmental, and clinical factors, motivating integrative approaches for understanding disease pathobiology. Existing multiview learning methods often rely on linear models that fail to capture nonlinear relationships, while available nonlinear approaches are limited in interpretability, scalability, or applicability to more than two data sources. We propose scalable randomized kernel methods for jointly associating data from multiple views and simultaneously predicting clinical outcomes or classifying individuals into multiple classes. Our approach models nonlinear relationships across views, identifies variables or groups of variables that drive shared structure, and learns view-independent low-dimensional representations using random Fourier feature approximations of shift-invariant kernels. Through simulations, we show that the proposed methods outperform several linear and nonlinear competitors. Applications to gene expression, metabolomics, proteomics, and lipidomics data from COVID-19 patients reveal molecular signatures associated with disease status and severity. Results from simulations and real data suggest the methods are effective even in small-sample settings. We also note that this framework naturally connects to emerging attention-based architectures, which we are exploring as a direction for future work.

1S Generative models and learning dynamics

Friday November 13 • 8:30–10:30 am • State Street

Chair: Jacob Bien • 6 speakers

Beyond Prediction: World Models for Understanding, Planning, and Action

Ning Ning

Texas A&M University

AI systems can now generate language, recognize images, and solve increasingly complex problems. Yet they still struggle with abilities that humans and animals acquire naturally: understanding how the physical world works, anticipating the consequences of actions, and planning efficiently. A promising direction is to build AI systems that learn world models—internal representations that predict how the world will evolve. Joint-Embedding Predictive Architectures (JEPAs) offer a powerful framework for learning such representations. But accurate prediction alone is not enough: the learned representation must also make it easy to reason about possible actions. In this talk, I will discuss recent progress by my group toward making learned world models more suitable for planning and action. I will show how structuring representations around the effects of actions can make planning substantially more efficient, and discuss what this suggests about the representations needed for AI systems that can understand, predict, and act in the physical world.

Optimal Convergence Analysis of DDPM for General Distributions

Yuchen Zhou

University of Illinois Urbana-Champaign

Score-based diffusion models have achieved remarkable empirical success in generating high-quality samples from target data distributions. Among them, the Denoising Diffusion Probabilistic Model (DDPM) is one of the most widely used samplers, generating samples via estimated score functions. Despite its empirical success, a tight theoretical understanding of DDPM — especially its convergence properties — remains limited. In this paper, we provide a refined convergence analysis for the DDPM sampler and establish near-optimal convergence rates under general distributional assumptions. Specifically, we consider a relaxed smoothness condition parameterized by L , which is small for many practical distributions (e.g., Gaussian mixture models). We prove that, to approximate a target distribution on $\mathbb{R}^d$ to accuracy ϵ in total variation distance and ϵ^2 in KL-divergence, the DDPM sampler with accurate score estimates requires at most

$$T = \tilde{O}\left(\frac{\sqrt{d} \min\{\sqrt{d}, L\}}{\epsilon}\right)$$

iterations, where $\tilde{O}$ hides polylogarithmic factors in d and $\frac{1}{\epsilon}$. This result substantially improves upon the best-

known $\tilde{O}\left(\frac{d}{\epsilon}\right)$ iteration complexity when $L < \sqrt{d}$. By establishing a matching lower bound, we show that our convergence analysis is tight for a wide array of target distributions. Moreover, it reveals that DDPM and Denoising Diffusion Implicit Models (DDIM) share the same dependence on d , raising an interesting question of why DDIM often appears empirically faster.

Statistical Interpretation of AI Architecture and AI Agents for Statistical Research

Shihao Yang

Georgia Institute of Technology

This talk explores a two-way exchange between statistics and AI: using statistical models to interpret AI architectures and using AI agents to support statistical research. Taking attention mechanisms as the main example, I will describe how linear attention can be interpreted as a mixture of vector autoregressive models. This connection makes classical time-series ideas available for understanding and designing Transformer architectures. I will discuss how autoregressive and moving-average structure motivates new attention mechanisms, and how modeling time warp leads to new position encoding architectures.

I will then describe how AI agents support our own research on these models. Drawing on our experience developing automated research workflows, I will discuss how a research plan can guide agents through method implementation, computational experiments, and iterative evaluation. The statistical questions extend to the agents themselves: how to assess variability across runs, measure sensitivity to prompts, and separate exploratory development from held-out evaluation.

Gradient-Flow Optimization as Dynamic Random-Effects Inference with Applications to Deep Learning

Zhonghua Liu

Columbia University

Gradient-flow optimization is usually viewed as an algorithmic procedure for minimizing empirical loss, with training duration selected by validation or heuristic early stopping rules. We develop a statistical inference framework for gradient-flow training. We show that whenever fitted values evolve through a time-invariant positive semidefinite training operator, the output at each time is equivalent to the best linear unbiased predictor under a corresponding random-effects model. Training time then becomes a variance-component parameter governing variance reallocation from residual noise to structured signal. This turns two training decisions into inferential problems: whether training is needed becomes a variance-component test for signal beyond initialization, and how long to train becomes restricted maximum likelihood (REML) estimation of the training-time variance component. We show that the REML-guided early stopping rule selects the time at which optimized spectral losses become decorrelated from the training-operator eigenvalues. The asymptotic prediction optimality of the REML-guided early stopping time is established for fixed-design in-sample risk and random-design out-of-sample risk. Deep learning models in fixed-kernel gradient regimes provide canonical instantiations for our results. Numerical experiments and a UK Biobank proteomics application show competitive accuracy of the REML-guided early stopping time with reduced reliance on validation splits and repeated checkpoint evaluation.

ComposeCFM: Flow-Matching Synthetic Generation and Harmonization

Mingzhao Hu

Mayo Clinic

Background: Quantitative MRI biomarkers differ across sites, scanners, and acquisition protocols, complicating dementia studies. This is particularly important for understanding cerebrovascular contributions to dementia: absolute white matter hyperintensity volume (WMHV) and percentage of WMHV over total intracranial volume (WMHTIV) measure visible lesion burden, along with fractional anisotropy (FA), free water (FW), peak width of skeletonized mean diffusivity (PSMD) and ARTerioSclerosis score (ARTS) which capture diffuse white matter injury. They differ in scale, skewness, and distributional shape. Most multisite studies also lack paired scans of the same participants on multiple scanners. **Methods:** We develop a flexible harmonization method based on conditional flow matching (ComposeCFM), that can be learned from unpaired observations while incorporating paired crossover scans when available. Site, scanner, and protocols are represented as nonlinear transformations based on age and sex. In contrast to the prevalent method, ComBat, which uses parametric mean and variance, ComposeCFM aligns the biomarker distributions. The transformations were refined jointly across all observed scanner and software groups. Model selection used held-out participants. **Results:** Without crossover supervision, ComposeCFM produced lower Wasserstein-1 distance (W1) and squared maximum mean discrepancy (MMD2) than ComBat for WMHV, WMHTIV, FA, and PSMD with reductions 2-17% for W1 and from 21% to 54% for MMD2. In an independent subject-level half-split, ComposeCFM improved both measures for these four biomarkers along with FW, with reductions of 4-21% and 17-47%, respectively. When true crossover scans were incorporated, both distances were lower than ComBat for all six biomarkers in the half-split evaluation. In the reference-scanner analysis, crossover-supervised ComposeCFM improved MMD2 for every biomarker except ARTS and improved W1 for FW, FA, and PSMD. **Conclusions:** Flexible harmonization can be learned without requiring paired rescans using our flow matching framework. ComposeCFM improved distributional comparability for markers of both visible lesion burden and white matter injury. This may support more reliable pooling of emerging cerebrovascular MRI biomarkers across separate studies at different institutions.

SportD: How do VLMs physically strategize?

Weining Shen

UC Irvine

Vision–language models (VLMs) can describe a scene, but can they act well within one? We study whether VLMs can make sound strategic decisions, using soccer as an objective testbed with actions that carry measurable outcomes. We introduce SportD, a dataset of decision scenarios from professional matches, where a VLM observes the seconds before a decision and chooses the next action. Models select the optimal action only around 30% of the time, less frequently than humans. They also favor safer, lower-variance, lower-value choices that make less physical progress toward goal. Frontier VLMs are better at estimating whether an action will succeed, placing the highest-success-probability action among their top choices in most cases. Yet they systematically conflate likelihood with value, assigning higher value to more likely actions, despite no such relationship in ground truth. SportD opens a new direction for rigorously evaluating physical strategic decision-making in VLMs, showing that careful decomposition of their choices can reveal mechanisms underlying systematic biases such as risk aversion.

2C Trustworthy AI and structured inference

Friday November 13 • 1:00–3:00 pm • Corwin

Chair: Jin Zhou • 5 speakers

Alignment in Large Language Models: Statistical and Game-Theoretic Perspectives

Weijie Su

University of Pennsylvania

Large language models (LLMs) are predominantly aligned with human preferences through reinforcement learning from human feedback (RLHF). In this talk, we explore the theoretical foundations of LLM alignment through the intertwined lenses of statistics and game theory. First, we show how the current formulation of RLHF induces a systematic bias we call preference collapse, and how this can be mitigated by introducing a tailored regularization term into the reward function. Next, we expose a fundamental bottleneck of reward-based alignment, demonstrating that cyclic human preferences cannot be faithfully represented by scalar reward models such as the Bradley-Terry model. More precisely, we establish that such cyclic inconsistencies give rise to a lower bound on the approximation error of any scalar reward fitting. Shifting to a game-theoretic perspective, we focus on Nash learning from human feedback and establish several social choice desiderata for this approach to alignment, including the preservation of preference diversity through the emergence of mixed strategies. Finally, we show that the zero-sum game approach generally cannot perfectly match a target preference distribution as a unique Nash equilibrium.

Controlling the Probability of Unknown Unknowns in Pre-Deployment Testing of AI Systems

Aaditya Ramdas

Stanford University

In pre-deployment testing of AI systems such as AI agents, a developer can often deal with *seen* failure modes — leaking API keys, deleting important databases, or installing a malicious package — by implementing fixes or making contingency plans for when the same issue might occur again. What is more challenging are the *unknown unknowns*: the *unseen* failure modes that haven't surfaced during testing. The decision to stop testing and deploy therefore faces an inherent tradeoff: continue testing to observe more behaviors, at the cost of time and resources, or stop testing and deploy, at the risk of failure modes remaining undiscovered. How then does the developer decide when they have tested enough? To answer this, we develop an algorithm that certifies when the *probability of an unseen failure mode* is sufficiently small. The algorithm constructs an anytime-valid upper confidence sequence on this quantity, so the developer can repeatedly test a system, inspect the bound, and decide to stop when it falls below a chosen threshold without invalidating the statistical guarantee. Our problem can be interpreted as a sequential variant of the missing-mass estimation problem which gave rise to the classic Good-Turing estimator. On SWE-Chat, a dataset of real-world coding agent traces, we implement our procedure and show that it even dominates a fixed-horizon (non-multiple-testing-corrected) Good-Turing upper confidence bound, which is not anytime-valid.

Prediction, Confidence, and Calibration in the Age of Deep Learning

Padhraic Smyth

University of California, Irvine

A natural question in the deployment of machine learning models (such as language models) is whether they can self-assess confidence in their predictions in a reliable way: do they know what they don't know? Using this question as a starting point, this talk will discuss uncertainty quantification and calibration of probabilistic predictions for modern deep learning models. The talk will begin with a brief review of how statistical ideas for calibration of prediction models have been adapted to machine learning contexts. We will then discuss recent work on calibration for deep learning models, including classification and regression models, time-series foundation models, and large language models, with some concluding comments on the role of confidence and calibration in the context of human-AI interaction more broadly.

Federated Feature Selection with False Discovery Rate Control

Runze Li

Penn State University

Selecting a set of universally relevant features associated with a given response variable across multiple distributed data sites is an important problem in numerous scientific fields. However, performing this federated feature selection task becomes challenging when individual-level data cannot be shared due to privacy concerns. The problem is further complicated by potential heterogeneity in both feature distributions and model parameters across sites. In this paper, we propose Fed-FDR, a federated feature selection framework that simultaneously identifies important features while controlling the false discovery rate (FDR). To ensure privacy preservation and reduce communication costs, the Fed-FDR shares only lower-dimensional coefficient estimates instead of transmitting summary statistics for all features, with the dimensionality shown to be of the same order as the number of relevant features. The coordinating center then leverages these lower-dimensional coefficient estimates to construct a generalized mirror statistic to identify the important features. The Fed-FDR is robust to the heterogeneity of feature distribution and model parameters, easy to implement, and computationally efficient. We further demonstrate that Fed-FDR effectively controls the FDR while achieving strong statistical power in our simulation studies. The results of empirical study also demonstrate that the method is both valid and implementation-ready.

Bias-Corrected Inference for Edge-Covariate Effects in Latent Space Network Models

Ji Zhu

University of Michigan

Latent space models (LSMs) provide a powerful framework for analyzing network data by embedding nodes in a latent space. Incorporating covariate information via edge covariates offers an important generalization that strengthens both the interpretability and practical utility of the model. However, we show that coefficient estimates for edge covariate effects obtained through maximum likelihood estimation exhibit asymptotic bias due to high-order geometric effects and errors in latent variable estimation. To address this issue, we propose a plug-in bias-correction estimator that enables asymptotically valid and unbiased statistical inference for the effects of edge covariates. We establish theoretical guarantees, including consistency and asymptotic normality, under various network structures. Extensive simulations and real-world data examples demonstrate that our method effectively reduces estimation bias and improves the accuracy of inference. Our findings contribute to the statistical methodology of LSMs by providing a principled framework for unbiased parameter estimation in network models with edge covariates.

2F Regression and predictive inference

Friday November 13 • 1:00–3:00 pm • Flying-A

Chair: Yazhen Wang • 6 speakers

Risk Bounds For Distributional Regression

Oscar Hernan Madrid Padilla

University of California, Los Angeles

This work examines risk bounds for nonparametric distributional regression estimators. For convex-constrained distributional regression, general upper bounds are established for the continuous ranked probability score (CRPS) and the worst-case mean squared error (MSE) across the domain. These theoretical results are applied to isotonic and trend filtering distributional regression, yielding convergence rates consistent with those for mean estimation. Furthermore, a general upper bound is derived for distributional regression under non-convex constraints, with a specific application to neural network-based estimators. Comprehensive experiments on both simulated and real data validate the theoretical contributions, demonstrating their practical effectiveness.

Residual Correction and Representation Learning for Robust Regression with Continuous Noisy Labels

Grace Yi

University of Western Ontario

Many regression problems involve outcomes that are measured or recorded with error. When they are used as training labels, the errors can vary considerably across observations. Some labels differ only slightly from the underlying values, while others differ substantially. Such errors can affect individual predictions and make it harder for a model to learn which observations have similar outcomes. In this talk, I will present a method for regression with noisy continuous labels. The proposed method addresses the problem at two levels. First, it uses a robust loss to estimate a limited correction for each observation, reducing the influence of moderately noisy labels on the fitted regression function. Second, it groups outcomes into ordered intervals and encourages observations with nearby outcomes to have similar learned representations. This broader structure helps when an individual correction is insufficient. The two components are linked through updated estimates of the outcomes during training. Experiments with both simulated and real-world label noise show improvements in predictive performance across several regression tasks.

ABC: Certified Selective Evaluation with a Panel of LLM Judges

Boxiang Wang

University of Iowa

LLM-as-a-judge has reshaped evaluation by reducing the human effort required to assess response quality, yet LLM judgments can disagree with human ratings. More capable models may narrow this gap, but they can also make the evaluation substantially more expensive. In practice, users may therefore rely on a panel of affordable judges and, for each response, decide whether to accept the panel's judgment or escalate it to a stronger model or a human. This raises two questions: how should the judges' scores be combined, and which resulting judgments can be accepted automatically at a controlled error rate. In this paper, we propose the ABC panel: Admit, Blend, and Certify. ABC constructs a statistical test based on cross-fitted residual association, admitting judges whose scores remain associated with human ratings beyond observed surface features, such as response length and style. It then blends the admitted judges using penalized logistic regression fitted to human ratings. Finally, it selects and certifies an acceptance rule such that, with a high probability, the population error rate among admitted judgments is below a threshold. Experiments on extensive human-rated benchmarks show that the ABC panel automates a larger share of evaluation than single-judge and ensemble baselines while controlling the error rate.

Noise-Adaptive Conformal Classification with Marginal Coverage

Matteo Sesia

University of Southern California

Conformal inference provides a rigorous statistical framework for uncertainty quantification in machine learning, enabling well-calibrated prediction sets with precise coverage guarantees for any classification model. However, its reliance on the idealized assumption of perfect data exchangeability limits its effectiveness in the presence of real-world complications, such as low-quality labels—a widespread issue in modern large-scale data sets. This work tackles this open problem by introducing an adaptive conformal inference method capable of efficiently handling deviations from exchangeability caused by random label noise, leading to informative prediction sets with tight marginal coverage guarantees even in those challenging scenarios. We validate our method through extensive numerical experiments demonstrating its effectiveness on synthetic and real data sets, including BigEarthNet and CIFAR-10H.

ATCG: Learning Structural Traits for Auditable Deep Classification

Hao Ho/Ker-Chau Li

University of California, Los Angeles

Deep classifiers map learned embeddings to class logits through a flat prediction head and achieve high accuracy. However, when classes contain heterogeneous latent subpopulations, this mapping is difficult to audit. A natural question is whether the structure of a prediction head can explicitly represent decision-related within-class heterogeneity. This talk explores this question through ATCG, a framework that retrofits pretrained backbones with class-specific hierarchies of learnable structural traits. Within each hierarchy, a geometric regularizer promotes complementary specialization between sibling traits during training, discouraging trait collapse and redundancy. Thresholded trait activations form combinatorial barcodes that encode decision-related within-class heterogeneity. These barcodes support case-level inspection of decision evidence and cohort-level analysis of distribution shifts through global and local diagnostics. We illustrate these uses on a real-world histopathology dataset and in controlled synthetic experiments.

Estimation of the sub-Gaussian Parameter

Min Xu

Rutgers University

The sub-Gaussian parameter (also called the variance proxy) of a mean-zero random variable X is defined as

$\xi_*^2 = \sup_{\lambda \in \mathbb{R}} L(\lambda)$ where $L(\lambda) = \frac{2}{\lambda^2} \log \mathbb{E} e^{\lambda X}$ is a weighted cumulant generating function. We study the estimation

of ξ_*^2 and prove that the minimax risk is governed by a non-increasing function

$\delta_P(C) = \sup_{|\lambda| \geq C} L(\lambda) - \sup_{|\lambda| \leq C} L(\lambda)$ which captures the influence of the tail behavior of the distribution P . Over

the class of distributions with $\delta_P \leq r$ for a non-increasing function r , the minimax risk is, up to a multiplicative constant, lower bounded by $r(\sqrt{\log n}) + n^{-1/2}$ and upper bounded by $r((\log n)^{1/2-\epsilon}) + n^{-1/2+\epsilon}$ for any $\epsilon > 0$. Our estimator for the upper bound is based on constrained maximization of the empirical analogue of L .

In addition to being almost minimax optimal and adaptive, we further prove that the estimator is asymptotic normal under suitable conditions and that, if the underlying distribution is not sub-Gaussian, the estimator diverges with a rate determined by the heaviness of the distributional tail.

2H Statistical learning and assisted inference

Friday November 13 • 1:00–3:00 pm • Harbor

Chair: Ying Nian Wu • 6 speakers

Tree-Guided Feature Aggregation as Structured Representation Learning

Kun Chen

University of Connecticut

Modern learning problems often involve fine-resolution features organized by a known hierarchy. Estimating all leaf-level effects separately can be unstable, while aggregation at a fixed resolution may obscure meaningful heterogeneity. We study tree-guided feature aggregation as a form of structured representation learning and propose CARAT, which combines two complementary regularization geometries on the original feature effects: an atomic penalty that rewards economical representations through subtree aggregates and a nested contrast penalty that suppresses unnecessary within-branch refinements. This formulation adaptively learns the resolution at which features contribute while preserving the original feature effects as the scientific target. We develop non-asymptotic prediction guarantees and recovery results for the underlying homogeneous tree partition. Numerical studies illustrate when hierarchical synthesis and local refinement are complementary and how their roles differ from explicit sparsity. More broadly, this perspective suggests hierarchical aggregation as a structured and interpretable representation layer for modern learning systems.

Beyond Point Predictions: Distribution-Informed Prediction-Powered Inference

Emma Jingfei Zhang

Emory University

Prediction-powered inference (PPI) typically relies on point predictions on unlabeled data. When predictive distributions are available as in a wide range of applications, including predictions from LLMs, we introduce distribution-informed prediction-powered inference (DiPPI), a general framework for further improving statistical efficiency by using predictive distributions as auxiliary information in the spirit of PPI. We characterize the optimal use of this information through score calibration, and provide theoretical guarantees for positive learning with the cross-fitted DiPPI estimator. Through simulations and three real-data applications, we show that DiPPI achieves better efficiency than PPI methods based on point predictions. These gains arise when the predictive distribution contains score-relevant information that is partially lost in point predictions.

Federated Transfer Learning with Differential Privacy

Yang Feng

New York University

Federated learning has emerged as a powerful framework for analysing distributed data, yet two challenges remain pivotal: heterogeneity across sites and privacy of local data. In this paper, we address both challenges within a federated transfer learning framework, aiming to enhance learning on a target data set by leveraging information from multiple heterogeneous source data sets while adhering to privacy constraints. We rigorously formulate the notion of federated differential privacy, which offers privacy guarantees for each data set without assuming a trusted central server. Under this privacy model, we study four statistical problems: univariate mean estimation, low-dimensional linear regression, high-dimensional linear regression, and M-estimation. By investigating the minimax rates and quantifying the cost of privacy, we show that federated differential privacy is an intermediate privacy model between the well-established local and central models of differential privacy. Our analyses account for data heterogeneity and privacy, highlighting the fundamental costs associated with each factor and the benefits of knowledge transfer in federated learning.

BELIEF in Dependence

Kai Zhang

University of North Carolina at Chapel Hill

Two linearly uncorrelated binary variables must be also independent because nonlinear dependence cannot manifest with only two possible states. This inherent linearity is the atom of dependency constituting any complex form of relationship. Inspired by this observation, we develop a framework called binary expansion linear effect (BELIEF) for understanding arbitrary relationships with a binary outcome. Models from the BELIEF framework are easily interpretable because they describe the association of binary variables in the language of linear models, yielding convenient theoretical insight and striking Gaussian parallels. With BELIEF, one may study generalized linear models (GLM) through transparent linear models, providing insight into how the choice of link affects modeling. For example, setting a GLM interaction coefficient to zero does not necessarily lead to the kind of no-interaction model assumption as understood under their linear model counterparts. Furthermore, for a binary response, maximum likelihood estimation for GLMs paradoxically fails under complete separation, when the data are most discriminative, whereas BELIEF estimation automatically reveals the perfect predictor in the data that is responsible for complete separation. We explore these phenomena and provide related theoretical results. We also provide preliminary empirical demonstration of some theoretical results.

Algorithmic inference via nonconvex gradient descent

Qiyang Han

Rutgers University

Conventional statistical inference methods are typically developed for models simple enough to admit tractable estimators through carefully designed iterative algorithms. In contrast, modern deep learning models are enormously complex, yet are trained by simple gradient-descent-type algorithms, often without any provable guarantee of algorithmic convergence to global/local optima.

Can we reconcile classical inference principles with these highly complicated, modern learning paradigms? In this talk, we will present a new inference framework addressing this question, by showing that valid statistical inference can be performed along the entire gradient descent trajectory, iteration by iteration, without requiring convexity of the loss landscape or convergence of the algorithm.

To illustrate this concept, we begin with a single-index (one-layer neural network) regression model and demonstrate how gradient descent iterates can be "debiased", at each iteration, to yield valid confidence intervals for the underlying signal and consistent estimates of generalization errors. We then extend this paradigm to the much more challenging setting of learning with general multi-layer neural networks in their full complexity, where the loss landscape can be arbitrarily complex. Crucially, the proposed method remains valid without requiring either algorithmic convergence or oracle knowledge of the unknowns, and may therefore inform practical decisions such as early stopping and hyperparameter tuning.

The key technical ingredient underlying this new inference paradigm is a recent entrywise dynamical mean-field theory for a broad class of first-order algorithms developed by the speaker.

Mean-Median gaps in real line and Inference in Infinite-dimensions

Arun Kuchibhotla

Carnegie Mellon University

In this talk, I will present some interesting results for the gap between mean and median of sums of independent real-valued random variables. Based on this result in the univariate case, I will present several interesting methods of statistical inference for functionals in high or infinite-dimensional spaces without any dimension/complexity assumptions. Time permitting, some applications to problems in generative AI will be discussed.

2S Network models, representations, and experiments

Friday November 13 • 1:00–3:00 pm • State Street

Chair: Bertrand Clarke • 6 speakers

HODGEDEPTH: SPATIAL REPRESENTATION LEARNING THROUGH GLOBAL GEOMETRY

Xiwei Tang

University of Texas at Dallas

A central challenge in spatial representation learning is to connect local feature patterns with global geometric patterns. For example, understanding layered tissues requires connecting local molecular variation to global spatial organization. We introduce *HodgeDepth*, a global-geometry-aware framework that learns an interpretable depth coordinate describing progression across layers and corresponding positions along curved tissue contours. The method aligns noisy local directions of molecular change and integrates them into a coherent coordinate, making global geometry an explicit representation-learning target. The learned depth provides a spatial axis for inspecting tissue contours and gene-expression profiles, with a complementary construction for recurrent layers. Our theory characterizes geometric bias and contour recovery, establishes conditions for correct global ordering from local estimates, and bounds recovery of a canonical depth scale based on cumulative spatial mass. We evaluate HodgeDepth on 23 human cortical Visium sections across two cohorts. On the 12-section DLPFC benchmark, mean absolute depth–layer Spearman correlation increases from 0.690 for GASTON to 0.885. Downstream clustering achieves mean ARI 0.618 versus 0.533 for SEDR, the strongest of nine tested baselines, a 16% relative improvement. Controlled synthetic examples further illustrate recovery of layered geometry. Together, these results support global spatial depth as an interpretable representation of tissue organization with measurable benefits for downstream analysis.

Testing for correlation between network structure and high-dimensional node covariates

Keith Levin

University of Wisconsin—Madison

In many application domains, networks are observed with node-level features. In such settings, a common problem is to assess whether or not nodal covariates are correlated with the network structure itself. We present four novel methods for addressing this problem. Two of these are based on a linear model relating node-level covariates to latent node-level variables that drive network structure. The other two are based on applying canonical correlation analysis to the node features and network structure, avoiding the linear modeling assumptions. We provide theoretical guarantees for all four methods when the observed network is generated according to a low-rank latent space model endowed with node-level covariates, which we allow to be high-dimensional. Our methods are computationally cheaper and require fewer modeling assumptions than previous approaches to network dependency testing. We demonstrate and compare the performance of our novel methods on both simulated and real-world data.

Hierarchical Clustering With Confidence

Jacob Bien

University of Southern California

Hierarchical clustering is one of the most widely used approaches for exploring data. However, its greedy nature makes it highly sensitive to small perturbations, blurring the lines between genuine structure and spurious patterns. In this work, we show how randomizing hierarchical clustering can be useful not just for assessing clustering stability but also for designing valid hypothesis testing procedures based on clustering results. In particular, we propose a method for constructing a valid p-value at each node of the hierarchical clustering dendrogram that quantifies evidence against performing the greedy merge. Furthermore, we show how our p-values can be used to estimate the number of clusters, with a probabilistic guarantee on overestimation of the number of clusters.

A Dyadic Latent Space Model for Directed Networks

Xiaoyue Maggie Niu

Pennsylvania State University

Reciprocity, the tendency of two nodes to form mutual ties, is a defining feature of directed networks, yet most models capture it only through a single global parameter, community membership, or observed covariates. In many applications the propensity to reciprocate varies across node pairs and is driven by latent node characteristics. We propose a latent space model for directed networks that models each pair of reciprocal edges jointly. The model allows the strength of reciprocity to vary flexibly across node pairs while retaining the interpretability and low-dimensional structure of latent space models, and it nests the classical p_1 model as a special case. We establish identifiability, develop a scalable projected gradient descent algorithm for maximum likelihood estimation, derive finite-sample error bounds, and extend the algorithm to link prediction. Applied to the international trade network for precision and medical instruments, the model improves link prediction and uncovers strongly heterogeneous reciprocity across country pairs.

Identifiability and Spectral Estimation in Multilayer Degree-Corrected Mixed Membership Models

Jesús Arroyo

Texas A&M University

Multilayer network data, arising from multiple sources, repeated measurements, or multi-view interactions, are increasingly common, motivating methods that jointly leverage information across layers while accommodating layer heterogeneity. A multilayer degree-corrected mixed membership model (MLDCMM) is studied, in which nodes share a common mixed membership structure while degree effects and block connectivity patterns vary by layer. The model encompasses several existing multilayer models while preserving identifiability and interpretation. A spectral estimation procedure integrates layer-wise spectral embeddings with corner estimation and connects the resulting representation to a PARATUCK2 tensor decomposition. Theoretical guarantees show that the probability of accurate corner recovery increases with the number of layers and that the resulting initial membership estimator is consistent. A minimax lower bound is established for membership estimation, and under homogeneous degree correction, the convergence rate of the spectral estimator matches this lower bound up to a factor of $\log(n)$. Empirical results in simulations and real data demonstrate the benefits of the proposed method.

How to Experiment When You Don't Yet Know What Matters

Alexander Volfovsky

Duke University

Every experiment hides a few expensive decisions: which data are worth collecting, which people or locations affect one another and where limited resources should go. We usually commit to these decisions before the experiment begins, effectively treating them as facts.

Using two vignettes this talk asks what happens when the experiment is allowed to learn the answers to these questions as it runs. In the first, a sequential randomized trial learns which costly baseline measurements are worth collecting for future study participants. In the second, an experiment learns an uncertain network of spillovers while deciding where to allocate treatment. In both cases, Thompson sampling turns uncertainty into a decision rule: act on one plausible picture of the world, observe what happens, update and repeat.

The broader lesson is simple: we do not need to know in advance what matters most. With the right randomization and design safeguards, an experiment can learn what matters as it unfolds — answering today's question while improving the design of the next round.

3F Online learning and decision making

Friday November 13 • 3:30–5:30 pm • Flying-A

Chair: Lan Xue • 6 speakers

Optimal Auditing of Sequential Decision Systems under Dynamical Constraints

Xuan Bi

University of Minnesota

How can we audit an AI system before it is safe or practical to deploy? This question is difficult because historical operational logs, generated by a behavior policy, may fail to cover the states visited by the target policy. We study this problem by proposing a hybrid auditing framework that combines fixed operational logs, generated by a behavior policy, with a small number of actively collected audit trials. We formulate probe design as a constrained convex optimization problem that minimizes the mismatch between the combined sampling distribution and the distribution induced by the target policy. The resulting probe allocates trials toward audit blind spots, namely regions frequently visited by the target policy but poorly covered by operational logs. Then, we develop an iterative algorithm that estimates the environment dynamics and distribution induced by the target policy from data while refining the probe design. Theoretically, we prove that minimizing this mismatch bounds the estimator's mean squared error and asymptotic variance. Experiments on GridWorld and LunarLander environments show that the proposed method substantially outperforms baseline methods.

Demystifying Group Relative Policy Optimization: Its Policy Gradient is a U-Statistic

Chengchun Shi

London School of Economics and Political Science

Group relative policy optimization (GRPO), a core methodological component of DeepSeekMath and DeepSeek-R1, has emerged as a cornerstone for scaling reasoning capabilities of large language models. Despite its widespread adoption and the proliferation of follow-up works, the theoretical properties of GRPO remain less studied. This paper provides a unified framework to understand GRPO through the lens of classical U-statistics. We demonstrate that the GRPO policy gradient is inherently a U-statistic, allowing us to characterize its mean squared error (MSE), derive the finite-sample error bound and asymptotic distribution of the suboptimality gap for its learned policy. Our findings reveal that GRPO is asymptotically equivalent to an oracle policy gradient algorithm – one with access to a value function that quantifies the goodness of its learning policy at each training iteration – and achieves asymptotically optimal performance within a broad class of policy gradient algorithms. Furthermore, we establish a universal scaling law that offers principled guidance for selecting the optimal group size. Empirical experiments further validate our theoretical findings, demonstrating that the optimal group size is universal, and verify the oracle property of GRPO.

Reinforcement Learning in the Physical World

Yuhua Zhu

University of California, Los Angeles

This talk addresses continuous-time reinforcement learning (RL) in settings where the system dynamics are governed by a stochastic differential equation but remains unknown, with only discrete-time observations available. We introduce Optimal-PhiBE, an equation that integrates discrete-time information into a PDE, combining the strengths of both RL and PDE formulations. In linear-quadratic control, Optimal-PhiBE can even achieve accurate continuous-time optimal policy with only discrete-time information. In general dynamics, Optimal-PhiBE is less sensitive to reward oscillations, leading to smaller discretization errors. Furthermore, we extend Optimal-PhiBE to higher orders, providing increasingly accurate approximations.

Estimating Dynamic Marginal Policy Effects under Sequential Unconfoundedness

Stefan Wager

Stanford University

We develop methods for estimating how infinitesimal policy changes affect long-term outcomes in dynamic systems. We show that dynamic marginal policy effects (MPEs) can be identified via tractable reduced-form expressions, and can be estimated under a general sequential unconfoundedness assumption. We also propose a doubly robust estimator for dynamic MPEs. Our approach does not require observing full dynamic state information (as is typically assumed for off-policy evaluation in Markov decision processes), and does not incur an exponential curse of horizon (as is typical in non-Markovian off-policy evaluation). We demonstrate practicality and robustness of our approach in a number of simulations, including one motivated by a dynamic pricing application where people use past prices to form a reference level for current prices.

Structured Learning of Individualized Decision Rules with Complex Options

Yufeng Liu

University of Michigan

Advances in medical research and clinical practice have greatly expanded the range of treatment options available for many diseases. While this increased flexibility creates opportunities for personalized care, it also poses significant challenges for learning optimal individualized decision rules (IDRs), particularly when the number of treatment options is large. In this talk, I will present a unified framework for learning IDRs by exploiting latent structure among treatment options. By identifying and leveraging similarities and shared patterns across treatments, the proposed approach improves statistical efficiency and leads to more accurate, stable, and interpretable decision rules. I will discuss the theoretical properties of the framework and illustrate its performance through simulation studies and real-data applications. The results demonstrate the potential of structured learning approaches to enhance personalized decision-making in complex treatment settings.

Two Sample Sizes in Test-Time LLM Ensembling: Buffered ERM and Minimax Theory

Arash Amini

University of California, Los Angeles

Test-time scaling can improve the performance of large language models by generating multiple answers to a question and aggregating them. Weighted ensembles extend this idea by drawing answers from several models, with weights determining how often each is used. Learning these weights involves two sample sizes: Q labeled questions and N generations per model and question. We study their interplay through the clean risk, the population error of the ensemble's infinite-generation voting rule. Under a margin condition with exponent β , we show that empirical risk minimization (ERM) with a calibrated margin buffer achieves excess risk of order $Q^{-1/2} + N^{-\beta/2}$, up to logarithmic factors, for fixed numbers of models and candidate answers. To investigate optimality, we analyze a submodel with two models and two possible answers per question. We characterize its minimax excess risk up to constants by a Hellinger modulus for a binomial-mixture experiment. Its lower bounds extend to general ensembles. Polynomial approximation and spectral analysis reveal three regimes. When questions are relatively scarce, the $Q^{-1/2}$ term governs learning and buffered ERM is optimal. In an intermediate regime, pooling generation counts across questions improves on estimating each answer distribution separately, but additional questions give only logarithmic gains. With sufficiently many questions, the minimax excess risk reaches an $N^{-\beta}$ identification floor that persists even with infinitely many questions.

3H Statistical inference after learning

Friday November 13 • 3:30–5:30 pm • Harbor

Chair: Hongzhe Li • 6 speakers

LOCO-AdaMP: Built-in LOCO Inference for Adaptive Minipatch Ensembles with Enhanced Prediction

Lili Zheng

University of Illinois Urbana - Champaign

As black-box machine learning models become increasingly common, extracting interpretations with uncertainty quantification has become a critical challenge. One popular type of interpretation is the leave-one-covariate-out (LOCO) feature importance, while prior LOCO inference methods often require data-splitting or model-refitting. A recent ensemble framework, LOCO-MP, addresses these challenges using minipatches that subsample both observations and features, but massive feature subsampling can hurt prediction in high-dimensional sparse settings. Motivated by this limitation, we consider minipatch ensembles with adaptive feature sampling guided by LOCO importance, and propose LOCO-AdaMP, which enables free LOCO importance inference for the resulting adaptive minipatch ensemble. We show that LOCO-AdaMP yields substantially improved predictive models while retaining asymptotically valid feature importance inference without data splitting, despite the complex dependence between the adaptive sampling distribution and the LOCO importance statistics. Our analysis relies on a careful leave-two-out perturbation bound for the iteratively updated sampling probabilities together with the stability of LOCO scores induced by observation subsampling. Empirical results on synthetic and real datasets demonstrate the advantages of LOCO-AdaMP over existing methods in predictive performance, inferential power, and stability. Overall, LOCO-AdaMP provides a flexible ensemble framework (agnostic to base models) that delivers both strong predictive performance and asymptotically valid, powerful feature importance inference for regression.

Consensus Tree Estimation with False Discovery Control via Partially Ordered Sets

Armeen Taeb

University of Washington

Trees are data objects that hierarchically organize categories. Collections of trees arise in a diverse variety of fields, including evolutionary biology, machine learning, social sciences and anatomy. Summarizing a collection of trees by a single representative is challenging, in part due to the dimensions of both the sample space and the parameter space. We frame consensus tree estimation as a structured feature-selection problem, where leaves and edges are the features. We introduce a partial order on trees, use it to define false discoveries for a candidate summary tree, and develop novel estimation algorithms that control the false discovery rate at a nominal level for a broad class of generative models. We also use the partial order to assess the stability of features in a selected tree. Importantly, our method accommodates unequal leaf sets and non-binary trees, which commonly arise in modern datasets. Our feature-selection perspective yields finite-sample and model-free guarantees and provides a foundation for integrating multiple testing tools into tree estimation. We apply the method to study the origins of complex life. Our estimated tree recapitulates well-known divisions but highlights that there is insufficient data to determine the most recent archaean ancestor of eukaryotic life.

Inference After Exploration: A Grammar for Adaptive Data Science

Dan Kessler

University of North Carolina at Chapel Hill

There is a fundamental tension between scientific discovery, which is often sequential and adaptive, and classical statistical methodology, whose inferential guarantees assume adherence to a pre-specified analytic plan. This friction is widely appreciated, and various areas of statistics, including multiple testing and selective inference, have developed methods to narrow this gap. However, this interface is difficult to navigate for scientists and statisticians alike. These challenges stymie the application of cutting-edge statistical methods to scientific analyses, and they limit the development of tools that can support the complex and multifaceted analytic pipelines increasingly employed in practice.

To address these challenges, we introduce a Grammar of Selective Inference (GoSI). GoSI is a framework for describing and reasoning about adaptive data science workflows. The central object of GoSI is the **query**, which formalizes the ways in which scientists extract information from data. Queries can be composed into sequences, and inference is a special query whose statistical validity depends on the structure of its ancestor queries. GoSI provides an accessible, expressive language for articulating adaptive workflows, applying existing methods, and identifying gaps where methodological development is needed. We illustrate GoSI with a reference implementation in Python and show how it can improve communication between scientists and statisticians. Finally, we discuss the role of GoSI in agentic workflows where artificial intelligence guides analyses.

Global Inference for High-Dimensional Quantile Regression with Heterogeneity

Tate Jacobson

Oregon State University

In recent decades high-dimensional data have become commonplace in domains from economics to genomics. Among methods extended and adapted for the high-dimensional setting, quantile regression has received significant attention due to its well-known robustness to outliers and ability to accommodate heteroscedastic errors. Despite this considerable interest, the inferential tools that have been developed for high-dimensional quantile regression have significant limitations—being limited to testing simple hypotheses, only making inferences for a single quantile at a time, or assuming a homoscedastic error model. To broaden the scope of statistical questions that can be answered in high-dimensional quantile regression, we introduce partial penalized tests for the quantile regression process which can be used to test any linear hypothesis, provide inferences over a set of quantiles, and allow for heteroscedastic errors. We propose a novel globally adaptive local linear approximation (LLA) algorithm to compute the partial penalized estimators used in our tests and prove that the LLA solutions possess the strong oracle property with respect to “testing oracles”. Using this strong oracle result, we establish asymptotic guarantees for the partial penalized tests evaluated at the LLA solutions.

Towards Efficient Inference under Nonmonotone Missingness with General Imputation

Qi Xu

University of Minnesota

Missing data are ubiquitous in classical survey and longitudinal studies as well as modern multi-modality data analysis. A longstanding challenge arises under nonmonotone missingness, where different units may observe arbitrary subsets of all variables. We study parameter estimation and inference problem under this setting. Semiparametric efficiency theory characterizes the efficient estimator through inversion of an operator constructed from pattern-specific conditional expectations. However, this estimator is generally not tractable due to compositions of conditional expectations across patterns.

We introduce the Restricted ANOVA hierarchy (RAY), a functional decomposition that reveals an almost-eigen structure of the operator under missing completely at random. This structure yields a closed-form, computable approximation to the efficient estimator. RAY estimator is applicable to general Z-estimation problems, and it remains unbiased for arbitrary independent imputation functions. In theory, we establish verifiable sufficient conditions where RAY attains the efficiency lower bound, and offer a general bound for the efficiency gap otherwise. We further develop adaptive RAY estimator, which attains the minimal asymptotic variance within a broader class containing RAY and other existing estimators. Finally, we investigate the extension of RAY under missing at random mechanism. Simulations and a single-cell multi-omics application demonstrate the efficiency gains of the proposed estimators.

Network Signflip Parallel Analysis for Selecting the Embedding Dimension

Joshua Cape

University of Wisconsin-Madison

This talk considers the problem of selecting the embedding dimension for large heterogeneous networks that have weakly distinguishable community structure. For a broad family of embeddings based on normalized adjacency matrices, we introduce a novel spectral method that compares the eigenvalues of the normalized adjacency matrix to those obtained after randomly signflipping its entries. The proposed method, called network signflip parallel analysis (NetFlipPA), is interpretable, simple to implement, data driven, and does not require users to carefully tune parameters. For large random graphs arising from degree-corrected stochastic blockmodels with weakly distinguishable community structure (and consequently, non-diverging eigenvalues), NetFlipPA provably recovers the spectral noise floor (i.e., the operator norm of the noise component of the normalized adjacency matrix). NetFlipPA thus provides a statistically rigorous randomization-based method for selecting the embedding dimension by keeping the eigenvalues whose magnitudes rise above the recovered spectral noise floor. Compared to traditional cutoff-based methods, the data-driven threshold used in NetFlipPA is provably effective under milder assumptions on the node degree heterogeneity and the number of node communities. Our main results rely on careful non-asymptotic perturbation analysis and leverage recent progress on local laws for nonhomogeneous Wigner-type random matrices. These findings serve as a potential baseline and foundation for understanding and developing principled, interpretable AI systems.

3S Causal discovery and treatment effects

Friday November 13 • 3:30–5:30 pm • State Street

Chair: Hani Doss • 6 speakers

Discrete Causal Representation Learning

Yuqi Gu

Columbia University

Causal representation learning seeks to uncover causal relationships among high-level latent variables from low-level, entangled, and noisy observations. Existing approaches often either rely on deep neural networks, which lack interpretability and formal guarantees, or impose restrictive assumptions like linearity, continuous-only observations, and strong structural priors. These limitations particularly challenge applications with a large number of discrete latent variables and mixed-type observations. To address these challenges, we propose discrete causal representation learning (DCRL), a generative framework that models a directed acyclic graph among discrete latent variables, along with a sparse bipartite graph linking latent and observed layers. This design accommodates continuous, count, and binary responses through flexible measurement models while maintaining interpretability. Under mild conditions, we prove that both the bipartite measurement graph and the latent causal graph are identifiable from the observed data distribution alone. We further propose a three-stage estimate-resample-discovery pipeline: penalized estimation of the generative model parameters, resampling of latent configurations from the fitted model, and score-based causal discovery on the resampled latents. We establish the consistency of this procedure, ensuring reliable recovery of the latent causal structure. Empirical studies on educational assessment and synthetic image data demonstrate that DCRL recovers sparse and interpretable latent causal structures. In the end, I will also discuss an extension of DCRL to the interpretability and steering of large language models.

Challenges in Causality: Sensitivity Analysis and Automation

Carlos Cinelli

University of Washington

Causal inference has emerged as a central research area at the intersection of statistics and computer science, with numerous applications in empirical fields such as biomedicine and the social sciences. However, despite significant progress over the past three decades, many important problems remain open. This talk will highlight two critical research directions: (i) drawing valid conclusions when key assumptions, such as the absence of unobserved confounding, are violated, and (ii) automating the causal inference pipeline. We will outline key open challenges in the areas of sensitivity analysis and partial identification, as well as the automation of such tasks. Addressing these challenges is essential for making robust causal inferences under realistic settings.

On regression with estimated covariates and conditional effects given the propensity score

Matteo Bonvini

Rutgers University

Machine-learning methods can successfully learn low-dimensional summaries or representations of high-dimensional data that are subsequently used as covariates in downstream statistical analyses. In this work, we study the problem of estimating a regression function when some of its covariates are themselves estimated in a previous step. Our motivating application is the study of heterogeneous returns to education in Brand and Xie (2010), which considers how the effect of completing college varies with the (unknown) probability of completing college. Our analysis, however, applies more generally to regression functions indexed by generated covariates obtained using potentially black-box first-stage learners.

Plug-in estimators that treat the estimated covariates as known generally inherit non-negligible first-stage estimation error. To mitigate this problem, we analyze two debiasing approaches within a framework that is agnostic to the choice of first-stage estimation method and accommodates both local-smoothing and sieve-based methods for the second-stage regression. We consider (i) influence-function-based estimators of pathwise-differentiable approximations to the target regression function and (ii) a variant of plug-in estimators that directly aims to correct their bias. For each method, we derive upper bounds on the estimation error and characterize conditions under which oracle rates can be approached, highlighting potential gains in convergence rates relative to plug-in estimators.

Simulation studies illustrate the finite-sample behavior of the methods. We apply our methodology to data from the National Longitudinal Survey of Youth 1997 and find that completing college yields the largest reductions in unemployment among individuals least likely to complete college, consistent with earlier evidence of negative selection in the returns to college.

Causal inference under interference with incomplete exposure mappings

Yubai Yuan

The Pennsylvania State University

Causal analyses under network interference commonly summarize neighbors' treatment assignments using a prespecified exposure mapping. When this mapping omits outcome-relevant features, conventional exposure contrasts may conflate the effect of the target exposure with changes in the broader treatment environment, even under randomized assignment. We separate these roles: the target exposure defines the causal estimand, while a distributional environment summary adjusts for omitted interference. We introduce permutation-invariant path and heat-diffusion summaries that encode the structural positions of treated nodes, together with kernel embeddings and random-feature approximations for estimation. Under omitted-exposure sufficiency and conditional positivity, we establish identification of direct and spillover effects for subpopulations with common support. We develop partial-linear and flexible response-surface estimators, establishing consistency and asymptotic normality for the former under network dependence. Simulations and applications to insurance and school-network experiments demonstrate the value of environment adjustment for estimating interpretable spillover effects under incomplete exposure mappings.

When Can We Trust Synthetic Data for Statistical Inference?

Didong Li

University of North Carolina at Chapel Hill

Generative artificial intelligence (AI) provides a promising approach to sharing sensitive data: a data owner can learn the underlying data distribution, generate a synthetic dataset, and release the synthetic data for downstream analysis without directly sharing the original observations. At the same time, modern generative models increasingly come with theoretical guarantees that the learned distribution is close to the true data distribution. However, distributional fidelity alone does not guarantee that statistical analyses performed on synthetic data will reproduce those based on the original data. In this talk, I will study when and how distributional guarantees for synthetic data translate into guarantees for downstream statistical inference. We show that the answer depends critically on the metric used to quantify distributional fidelity, the underlying data-generating model, the downstream inferential task, and the statistical estimator. Focusing on semiparametric regression and Cox regression, we investigate when distributional convergence guarantees accurate parameter estimation, hypothesis testing, and uncertainty quantification, and when these guarantees can fail despite strong distributional approximation. Finally, we discuss three complementary strategies for bridging this gap and enabling valid statistical inference using synthetic data.

Learning from the Unseen: Offline Reinforcement Learning with Hidden Actions

Ying Zhou

University of Connecticut

Standard offline reinforcement learning (RL) algorithms typically assume that the actions in the dataset are observed without error. However, in many real-world applications, the true actions are unobserved and only noisy proxies are available, causing existing RL methods to yield biased and potentially misleading conclusions. We study off-policy evaluation in infinite-horizon discounted Markov decision processes with hidden actions. By leveraging the next-state variable as a natural proxy for the unobserved action, we establish identification of the policy value and propose an influence-function-based estimator: LURE (Learning from the Unseen: Robust Estimator). LURE is multiply robust—remaining consistent under several combinations of correctly specified nuisance components—and is asymptotically normal, enabling valid statistical inference. To our knowledge, this is the first work to address offline RL with hidden actions. We demonstrate LURE’s effectiveness via simulations and a sepsis management application using the MIMIC-III database.

4C Biomedical modeling and AI-assisted inference

Saturday November 14 • 8:30–10:30 am • Corwin

Chair: Hua Zhou • 6 speakers

Synthetic Surrogate Functional Regression for Neuroimaging Analysis

Lily Wang

George Mason University

Modern generative models can synthesize missing neuroimaging modalities from auxiliary data, creating new opportunities for large-scale brain imaging studies. However, using synthetic images directly in downstream regression can lead to misleading inference if discrepancies between synthetic and observed images are not properly addressed. In this talk, I will present a two-stage Synthetic Surrogate Functional Regression (SSFR) framework for image-on-scalar regression with partially observed imaging outcomes. SSFR provides a statistical safeguard for incorporating synthetic neuroimages when estimating spatially varying associations between demographic, clinical, or genetic covariates and complex imaging responses. The framework integrates observed and surrogate images while remaining robust to discrepancies between surrogate and target regression functions. I will also introduce a functional canonical correlation coefficient for assessing surrogate informativeness and a weighted bootstrap procedure for constructing simultaneous confidence corridors for spatially varying coefficient functions. Simulations demonstrate strong performance across varying surrogate quality and model misspecification. An application to PET imaging data from the Alzheimer's Disease Neuroimaging Initiative shows that SSFR improves estimation efficiency and strengthens localization of age-related metabolic changes, illustrating the synergy between modern data synthesis and rigorous statistical inference.

Integrating Statistics and AI for Neuroscience Research

Lexin Li

University of California, Berkeley

Understanding the inner workings of the human brain, and their connections to neurological disorders, cognition, and normal development, is one of the most intriguing scientific questions. Recent advances in artificial intelligence (AI), machine learning, and statistical methodology have created unprecedented opportunities, while many challenges remain. In this talk, I will provide an overview of several research directions pursued by our group integrating statistics and AI for neuroscience research. I will discuss how modern AI methods can facilitate statistical analysis of brain imaging data, as well as how statistical principles can improve the development of AI models. I will illustrate with a number of case studies, highlighting the synergistic relationship between statistics and AI in advancing our understanding of the human brain.

Task-Aware Statistical AI for Learning Directed Brain Connectomes from Neuroimaging Data

Ying Guo

Emory University

In recent years, connectome-based research has become a major focus in neuroscience, providing important insights into brain organization and advancing the prediction of cognitive, behavioral, and mental health outcomes. However, most existing approaches focus on undirected brain connectivity, overlooking the directionality of interactions and causal influences between brain regions. To address this limitation, we propose a task-aware statistical AI framework for learning directed brain connectomes from neuroimaging data. Our approach integrates principled statistical modeling with graph neural networks to jointly estimate sparse, interpretable directed connectivity graphs and learn low-dimensional graph representations. A key innovation is the task-aware learning strategy, which incorporates downstream prediction objectives into connectivity estimation, enabling the framework to identify directed brain interactions that are both biologically interpretable and informative for predicting demographic characteristics and clinical phenotypes. Applications to neuroimaging data demonstrate the ability of the proposed method to uncover whole-brain directed connectivity patterns and identify previously unrecognized subpopulation-specific differences in brain network organization. By integrating statistical modeling, task-aware directed connectivity estimation, and predictive graph representation learning, the proposed framework offers a unified approach to advancing our understanding of brain network organization and improving predictive modeling in neuroscience.

Statistical and AI Approaches for Investigating Viruses and Virus-host Interactions

Fengzhu Sun

University of Southern California

Viruses play important roles in controlling bacterial population size and altering host metabolism, and have broader impacts on the functions of microbial communities, such as human gut, soil, and ocean microbiomes. However, the investigations of viruses and their functions were vastly underdeveloped. Metagenomic studies provide enormous resources for the identifications of novel viruses and their hosts. We developed several statistical and AI tools for the identification of new viruses and virus-host interactions such as VirFinder, DeepVirFinder, DeepMicroClass, VirHostMatcher, VirHostMatcherNet, and ContigNet. Applications of these tools to metagenomic data will help explore mobile genetic elements and their interactions with their hosts.

SAPA: Safe and Adaptive Prediction Augmentation for Survival Analysis

Gang Li

Department of Biostatistics, University of California, Los Angeles

We develop a Safe and Adaptive Prediction Augmentation (SAPA) framework for survival analysis that leverages black-box predictions from labeled and unlabeled data to improve statistical efficiency. The Cox proportional hazards model fitted to the labeled data is treated as a working model and need not correctly specify the underlying survival process; its regression parameters nevertheless remain well-defined targets for estimation and inference, and the fitted model can be used for prediction. SAPA adaptively incorporates auxiliary predictions from unlabeled data while safeguarding against unreliable auxiliary information. We establish theoretical guarantees and extend the framework to accommodate covariate shift between labeled and unlabeled populations. Simulations demonstrate efficiency gains, improved prediction, and valid inference, while real-data applications illustrate the practical utility of SAPA.

Scalable Joint Modeling of Multiple Longitudinal Biomarkers and Competing Risks Time-to-Event Data: Applications to Mega-Scale Health Research

Xinping Cui

University of California, Riverside

Joint modeling has become increasingly popular for the analysis of longitudinal and time-to-event data. However, semiparametric multivariate joint modeling for large-scale datasets still faces substantial statistical and computational challenges, primarily due to the high dimensionality of random effects and the complexity of estimating nonparametric baseline hazards. These issues often result in prolonged computation time and excessive memory usage, limiting the practical utility of joint modeling for biobank-scale analyses. In this article, we develop an efficient implementation for a semiparametric multivariate joint modeling framework that substantially reduces computation time and requires only moderate memory usage, enabling analyses involving large numbers of subjects and multiple longitudinal biomarkers. The scalability and estimation accuracy of our approach are demonstrated through simulation studies and an application to the UK Biobank primary care (UKB-PC) data, jointly modeling six longitudinal biomarkers and competing risks in over ten thousand subjects. We also illustrate that leveraging biobank-scale UKB data can improve dynamic prediction in both discrimination and calibration, while models trained on smaller samples may experience reduced accuracy, particularly for low-event or rare outcomes. A user-friendly R package, FastJM, has been developed and is publicly available on the Comprehensive R Archive Network: <https://CRAN.R-project.org/package=FastJM>

4F Evaluation and inference with AI predictions

Saturday November 14 • 8:30–10:30 am • Flying-A

Chair: Veronica Berrocal • 6 speakers

Performance guarantees for LLMs

Carey Priebe

Johns Hopkins University

Performance guarantees for LLMs are essential to myriad critical applications, enabling for any application for which quantifiable confidence in LLM performance is nonnegotiable. We develop theory and methods for providing performance guarantees for LLMs in the context of a map between vector embeddings. The Data Kernel Perspective Space (DKPS) provides a framework to generate empirically validated theoretical predictions allowing guaranteed performance quantification for these transformative models. I will present probability, statistics and linear algebra for DKPS, and illustrative examples.

Biography. Carey E. Priebe is Professor in the Department of Applied Mathematics & Statistics and Director of the Mathematical Institute for Data Science (MINDS) at Johns Hopkins University. He is an Elected Member of the International Statistical Institute, Senior Member of the IEEE, Fellow of the Institute of Mathematical Statistics, and Fellow of the American Statistical Association. His current work on foundations of AI is funded by DARPA, ONR, AFOSR, Microsoft, Amazon.

Language Model Preference Inference via Fisher Random Walk

Junwei Lu

Harvard University

Motivated by the need for rigorous and scalable evaluation of large language models, we study contextual preference inference for pairwise comparison functionals of context-dependent preference score functions across domains. Focusing on the contextual Bradley-Terry-Luce model, we develop a semiparametric efficient estimator that automates the debiased estimation through aggregating weighted residual balancing terms across the comparison graph. We show that the efficiency is achieved when the weights are derived from a novel strategy called Fisher random walk.

Low Rank for Rank: Uncertainty-Aware Task-Specific LLM Ranking from Sparse Pairwise Comparisons

Will Wei Sun

Purdue University

Modern LLM leaderboards increasingly rely on pairwise human preferences, yet reliable task-specific ranking remains challenging when comparisons are sparse and unevenly distributed across tasks and models. In this talk, I will present a low-rank framework for task-specific LLM ranking that shares information across related tasks while preserving task-specific heterogeneity. We develop a ranking procedure that combines low-rank initialization with local pairwise-logistic refinement to achieve uniform entrywise accuracy, leading to task-wise top- K recovery guarantees. Beyond point estimation, we construct efficient one-step estimators for task-specific score gaps and use multiplier-bootstrap calibration to account for the dependence among many ranking comparisons. This yields simultaneous rank confidence intervals and three-way top- K certificates. The framework provides a principled way to move from sparse pairwise feedback to uncertainty-aware, task-specific LLM leaderboards.

SADA: Safe and Adaptive Aggregation of Multiple Black-Box Predictions in Semi-Supervised Learning

Jiwei Zhao

University of Wisconsin-Madison

Semi-supervised learning (SSL) arises in practice when labeled data are scarce or expensive to obtain, while large quantities of unlabeled data are readily available. With the growing adoption of machine learning techniques, it has become increasingly feasible to generate multiple predicted labels using a variety of models and algorithms, including deep learning, large language models, and generative AI. In this paper, we propose a novel approach that safely and adaptively aggregates multiple black-box predictions of uncertain quality for both inference and prediction tasks. Our method provides two key guarantees: (i) it never performs worse than using the labeled data alone, regardless of the quality of the predictions; and (ii) if any one of the predictions (without knowing which one) perfectly fits the ground truth, the algorithm adaptively exploits this to achieve either a faster convergence rate or the semiparametric efficiency bound. We demonstrate the effectiveness of the proposed algorithm through small-scale simulations and two real-data analyses with distinct scientific goals. A user-friendly R package, *sada*, is provided to facilitate practical implementation.

Building Statistical AI Tools for Comparative Analysis

Peng Wang

Lindner College of Business, University of Cincinnati

Qualitative Comparative Analysis (QCA) and fuzzy set Qualitative Comparative Analysis (fsQCA) are established tools for causal inference in sociology, political science, and management research. In these fields, causality is viewed as configurational, with an outcome explained by combinations of conditions, whereas statistics usually frames causal inference through graphical models or counterfactuals. When only a few conditions are under consideration, researchers can rely on theory to decide which ones enter the truth table. When many attributes are available, this scoping step can no longer rest on theory alone, since the number of possible configurations grows exponentially, and automating the analysis requires a data driven scoping procedure backed by valid inference. We propose such a procedure based on the repro samples framework. It forms interactions among the raw conditions and their negations, screens them, and applies repro samples inference for model selection to determine both the scope and the candidate configurations, each of which receives a valid finite sample p value. The output is a confidence set of plausible systems of conditions rather than a single answer, so researchers can choose among them using theory. The set can also open new perspectives, since it may contain a configuration that points to a model or explanation the researcher had not considered. In addition, researchers can direct the scoping itself by fixing or excluding conditions on theoretical grounds before the procedure is run. Simulations and a real data application illustrate the approach. This is joint work with Yanran Liu, Jiawei Huang, and Zewei Lin.

Prompt engineering via order-of-addition experiments and screening

Hongquan Xu

UCLA

Large language models (LLMs) are increasingly deployed across scientific and technical fields to convert natural language prompts into executable code, complex workflows, and structured problem formulations. However, many LLMs suffer from order dependency, where the ordering and specific formulation of prompt elements strongly affect model performance. To overcome this issue, we introduce a systematic prompt engineering framework based on order-of-addition (OofA) experiments to quantify ordering effects, screen for critical components, and identify optimal prompt configurations. By incorporating screening techniques and statistical modeling, our approach efficiently evaluates permutations and phrasing variations to isolate key prompt elements that drive performance while eliminating unnecessary components. We demonstrate our methodology through the construction of two-level fractional factorial designs using state-of-the-art LLMs. The results show that screening and quantifying order-of-addition effects effectively illuminate order dependency in LLMs and enable the identification of high-quality prompt configurations that dramatically increase task success rates.

4H Structured representations and clustering

Saturday November 14 • 8:30–10:30 am • Harbor

Chair: Ali Shojaie • 6 speakers

Learning What to Share: Adaptive Spectral Integration and Clustering of Multimodal Data

Shujie Ma

University of California, Riverside

Multimodal data provide complementary views of the same subjects and can reveal population structures that are difficult to detect from any single source. Effective integration, however, requires more than pooling features because modalities may differ in signal strength, noise level, and relevance while containing both shared and modality-specific information. This talk presents MUSIC (Multimodal Spectral Integration and Clustering), an adaptive framework that extracts a low-dimensional singular subspace from each modality and decomposes it into common and modality-specific components. MUSIC iteratively estimates the common subspace and reweights modalities according to their agreement with the evolving common structure, thereby downweighting heterogeneous or uninformative sources while preserving modality-specific variation. The resulting spectral embeddings recover a coarse clustering shared across modalities together with finer, nested modality-specific refinements. Under a general low-rank model, we establish finite-sample guarantees for common and modality-specific subspace estimation and derive corresponding misclustering bounds. An application to multimodal molecular data from The Cancer Genome Atlas shows that MUSIC identifies clinically meaningful glioma subtypes and interpretable modality-specific refinements. The framework offers a principled approach to learning what information should be shared and how strongly each modality should be trusted.

Borrowing Strength Across Studies: Structure-Aware Transfer Learning for High-Dimensional Models

Mladen Kolar

University of Southern California & MBZUAI

Modern scientific applications often provide only a small sample from the population, tissue, or task of primary interest, while related studies offer substantially more information. Transfer learning can improve estimation by borrowing strength from these studies, but its success depends on choosing a notion of similarity that reflects the structure of the problem. In this talk, I will present SMART, a spectral transfer method for low-rank multi-task linear regression. Rather than requiring the source and target coefficient matrices to be globally close, SMART leverages shared latent structure: the target singular subspaces are contained in the source subspaces and sparsely aligned with the source singular bases. SMART incorporates this information through structured regularization and requires only a fitted source model, not the original source data. We establish non-asymptotic error bounds and, in the noiseless-source regime, a minimax lower bound; under additional regularity conditions, SMART achieves near-minimax Frobenius error rates up to logarithmic factors. Simulations demonstrate improved accuracy in small-target-sample regimes and empirical robustness to negative transfer. An application to multimodal single-cell data shows improved prediction of protein abundance from gene expression across related cell types. These results highlight a broader principle: effective transfer can arise from stable latent structure even when source and target effect magnitudes differ substantially.

This is joint work with Boxin Zhao and Jinchi Lv.

Beyond Vintage Rotation: Bias-Free Sparse Representation Learning with Oracle Inference

Gongjun Xu

University of Michigan

Learning low-dimensional latent representations is a central topic in statistics and machine learning, and rotation methods have long been used to obtain sparse and interpretable representations. Despite nearly a century of widespread use across many fields, rigorous guarantees for valid inference for the learned representation remain lacking. In this paper, we identify a surprisingly prevalent phenomenon that suggests a reason for this gap: for a broad class of vintage rotations, the resulting estimators exhibit a non-estimable bias. Because this bias is independent of the data, it fundamentally precludes the development of valid inferential procedures, including the construction of confidence intervals and hypothesis testing. To address this challenge, we propose a novel bias-free rotation method within a general representation learning framework based on latent variables. We establish an oracle inference property for the learned sparse representations: the estimators achieve the same asymptotic variance as in the ideal setting where the latent variables are observed. To bridge the gap between theory and computation, we develop an efficient computational framework and prove that its output estimators retain the same oracle property. Our results provide a rigorous inference procedure for the rotated estimators, yielding statistically valid and interpretable representation learning.

Majorization-Minimization Algorithms in Clustering

Jason Xu

University of California Los Angeles

Majorization-minimization (MM) is a general principle for converting difficult optimization tasks, often involving non smooth or non convex objectives, into a sequence of tractable surrogate problems. While the EM algorithm for Gaussian mixture models is a classical special case of MM applied to cluster analysis, we show how leveraging a similar geometry can overcome long-standing challenges in classical clustering tasks. This includes annealing to avoid local minima in k-means clustering, as well as robust clustering under the L2E criterion. We show how the MM principle can be used to elicit efficient algorithms that outperform peer methods on real and simulated examples.

GEM-KMeans: Memory-Efficient and Accurate Clustering on Massive Scale with GPU Optimization

Xiaohui Chen

University of Southern California

Memory-efficient scaling on clustering problems without sacrificing statistical accuracy is of central interest for large-scale data analysis and machine learning problems. Nonnegative low-rank (NLR) matrix factorization for K -means is a scalable clustering method, which connects to semidefinite relaxations with optimal average-case exact recovery guarantees. However, a direct GPU implementation of NLR requires multiple large factor-sized buffers and substantial data movements that are essentially memory-bound. In this paper, we introduce GEM-KMeans, a spectrally normalized yet mathematically equivalent NLR formulation that fuses the gradient update, nonnegative projection, and sufficient statistics for normalization and iterate movement into a matrix-multiplication epilogue. Instead of retaining three massive factor-sized arrays, our IO-aware GPU implementation materializes only one single factor with small tile-reduction arrays as additional storage in the High Bandwidth Memory (HBM). We derive explicit memory costs and spectrally normalized smoothness bounds for optimizing the clustering objective function. Accurate clustering is demonstrated at massive scales on synthetic and real datasets, where performance gains of GEM-KMeans over existing GPU-accelerated Lloyd's algorithms involve data-dependent runtime tradeoffs.

A study on kernels for discrete random variables

Marianthi Markatou

University at Buffalo

Kernels are essential elements in the construction of learning systems and have received considerable attention in machine learning. In statistics, kernels are used as tools for achieving specific data analytic goals, such as density estimation or goodness of fit testing. In this talk, we first offer a scoping review of the literature on the topic of kernel construction for discrete random variables. We then consider the problem of constructing kernels for discrete random variables and propose an algorithm for obtaining kernels appropriate for use with data that are realizations of multinomial random variables. Finally, we discuss their use in clustering.

4S Reinforcement learning and adaptive decisions

Saturday November 14 • 8:30–10:30 am • State Street

Chair: Minge Xie • 6 speakers

AI-Powered Adaptive Experiments: From Sequential Analysis to Intelligent Clinical Trials.

Feifang Hu

George Washington University

Artificial intelligence is transforming the design and conduct of experiments by enabling real-time learning and decision-making from accumulating data. In clinical trials and other scientific studies, AI-driven methods can improve treatment allocation, personalize interventions, and accelerate discovery through adaptive experimentation. Recent advances in machine learning and reinforcement learning have expanded the scope of adaptive clinical trials, creating new opportunities to make studies more efficient, informative, and patient-centered.

This talk will discuss how modern AI methods extend the principles of sequential experimentation and adaptive clinical trial design, highlighting both the opportunities and challenges of integrating AI into evidence generation. The convergence of statistical theory, adaptive experimentation, and artificial intelligence is poised to play a central role in the future of clinical research and data-driven decision-making.

Quantile Temporal Difference Algorithm for Distributional Reinforcement Learning

Lan Wang

University of Miami

Distributional reinforcement learning (RL) estimates the complete distribution of returns, enabling effective risk-sensitive decision-making. Quantile temporal difference (QTD) learning is a cornerstone algorithm for distributional RL with strong empirical performance, yet its theoretical guarantees remain limited. Analyzing QTD is challenging due to quantile semi-gradient bias, discretization errors from approximating return distributions, and non-smooth update dynamics. We present the first non-asymptotic analysis for QTD in both synchronous (generative model) and asynchronous (single Markovian trajectory) settings. We formulate QTD as a nonlinear stochastic approximation scheme for a quantile Bellman equation and establish an alignment property. We then derive finite-time bounds on the expected squared ℓ_∞ error of the last iterate, uniformly over states and quantile levels, for both the synchronous setting and the asynchronous setting. We obtain finite-time bounds on the expected supremum 1-Wasserstein distance between the learned and true return distributions and analyze the sample complexity bounds. These results advance the theoretical foundation of distributional RL.

Principled Fast and Meta Knowledge Learners for Continual Reinforcement Learning

Linglong Kong

University of Alberta

Inspired by the human learning and memory system, particularly the interplay between the hippocampus and cerebral cortex, this study proposes a dual-learner framework comprising a fast learner and a meta learner to address continual Reinforcement Learning (RL) problems. These two learners are coupled to perform distinct yet complementary roles: the fast learner focuses on knowledge transfer, while the meta learner ensures knowledge integration. In contrast to traditional multi-task RL approaches that share knowledge through average return maximization, our meta learner incrementally integrates new experiences by explicitly minimizing catastrophic forgetting, thereby supporting efficient cumulative knowledge transfer for the fast learner. To facilitate rapid adaptation in new environments, we introduce an adaptive meta warm-up mechanism that selectively harnesses past knowledge. We conduct experiments in various pixel-based and continuous control benchmarks, revealing the superior performance of continual learning for our proposed dual-learner approach relative to baseline methods.

Uncertainty-Guided Control: Statistical Foundations for Brain-Inspired Decision-Making

Babak Shahbaba

UCI

Intelligent agents must decide not only what to do, but how to decide: whether to act on a learned habit, retrieve a past experience, or plan ahead. This talk presents a brain-inspired reinforcement learning framework in which habitual, episodic, and goal-directed controllers operate in parallel, with principled uncertainty quantification guiding their relative influence. A key idea is to distinguish epistemic, aleatoric, and latent-state uncertainty, since each has different implications for learning and control. We explore extensions of conformal inference to sequential, partially observed, and multi-agent settings, focusing on the assumptions needed for valid coverage guarantees when an agent's own choices determine the data it collects. Behavioral and neuroimaging studies provide a way to test whether humans coordinate decision strategies through similar mechanisms and whether imbalances in this coordination contribute to maladaptive decision-making.

Successive classification learning for estimating quantile optimal treatment regimes

Dehan Kong

University of Toronto

Quantile optimal treatment regimes (OTRs) aim to assign treatments that maximize a specified quantile of patients' outcomes. Compared to treatment regimes that target the mean outcomes, quantile OTRs offer fairer regimes when a lower quantile is selected, as it improves outcomes for vulnerable patients. In this paper, we propose a novel method for estimating quantile OTRs by reformulating the problem as a successive classification task, solvable via training a sequence of classifiers, each successive classifier built on the output of its predecessors. This reformulation enables us to leverage the powerful machine learning technique to enhance computational efficiency and handle complex decision boundaries. We also investigate the estimation of quantile OTRs when outcomes are discrete, a setting that has received limited attention in the literature. A key challenge is that direct extensions of existing methods to discrete outcomes often lead to inconsistency and ineffectiveness issues. To overcome this, we introduce a smoothing technique that maps discrete outcomes to continuous surrogates, enabling consistent and effective estimation. We provide theoretical guarantees to support our methodology, and demonstrate its superior performance through comprehensive simulation studies and real-data analysis.

Future-Path Incremental Excursion Effects in Sequentially Randomized Trials

Walter Dempsey

University of Michigan

In sequentially randomized trials, researchers are often interested in assessing time-varying effects of the interventions on both proximal and distal outcomes. When considering impact on distal outcomes, it is common to either estimate the value function under strong assumptions such as time-homogeneous Markov dynamics or to assess a localized contrast of an intervention versus no intervention at a single decision point on the distal outcome which requires minimal assumptions but averages over the remaining treatment path. In this talk, we introduce a future-path incremental excursion effect defined through the local derivative of expected downstream utility under an infinitesimal log-odds tilt of future treatment probabilities from a decision time onward which bypasses the curse of the horizon problem without strong assumptions on the underlying dynamics. The parameter is anchored to the randomized trial policy while allowing analysis of both cumulative long-term proximal outcomes and distal. We will introduce likelihood-ratio, score, outcome-regression, and recursive representations, yielding direct and cross-fitted augmented estimators. The same local gradient supports KL-calibrated pessimistic policy updates that recommend changes only when a lower confidence bound is positive. Simulations show that the estimand captures future-path heterogeneity and that recursive augmentation can substantially improve precision. An application to LowSalt4Life 2, a micro-randomized trial, illustrates long-run engagement and a secondary clinical endpoint. Time permitting, we will discuss extensions of this work to AI evaluation.

5F Statistical learning and computational inference

Saturday November 14 • 1:00–3:00 pm • Flying-A

Chair: Marianthi Markatou • 6 speakers

Model-Free Inference for High-Dimensional Binary Classification Using Repro Samples

Minge Xie

Rutgers University

In high-dimensional settings, when the underlying data-generating regression structure is completely unknown and unspecified, conducting statistical inference is a particularly challenging task. In this talk, we develop a model-free inference approach for high-dimensional binary classification, leveraging a class of (likely misspecified) sparsity generalized models. Our method builds on the so-called repro samples framework, which utilizes artificial samples that are generated by mimicking the true data or noises. The proposed approach facilitates inference by targeting both the model support (discrete model parameter) and the regression coefficients (continuous model parameter) of the oracle model, defined as the working model closest to the unknown true model. The method offers three key advantages: (1) it is fully model-free, requiring neither correct model specification nor sparsity of the true model; (2) it constructs a candidate set of influential covariates with guaranteed coverage under weak signal conditions; and (3) it provides confidence sets for any linear transformation of the oracle coefficients. When the oracle model coincides with the true underlying model, the inference results directly apply to the true model, allowing simultaneous quantification of uncertainty in (discrete) model selection and (continuous) parameter estimation. Simulation studies illustrate the effective performance of the proposed method. An application to single-cell RNA-seq immune response data demonstrates that it identifies key genes, offering new insights into cellular immune response mechanisms.

An Introduction to Stochastic Deep Learning

Faming Liang

Purdue University

Deep neural networks (DNNs) have achieved remarkable success in prediction, but their deterministic formulation makes many statistical inference tasks difficult. StoNet, short for stochastic neural network, addresses this limitation by reformulating a DNN as a probabilistic latent-variable model, in which the outputs of selected hidden layers or units are treated as latent variables. The article describes the formulation of StoNet and its asymptotic equivalence to a conventional DNN, showing how StoNet can serve both as a deep learning model and as an analytical device for studying DNNs. It reviews training algorithms for StoNet and discusses its applications to statistical inference problems, including nonlinear sufficient dimension reduction, causal inference with observed and unobserved confounders, Granger-causality learning for nonlinear time series, nonlinear variable selection, and prediction uncertainty quantification for large-scale DNNs. The article also discusses two related architectures: kernel-expanded StoNet and sublinearly structured DNNs, which can be viewed as structural or theoretical extensions of the StoNet framework. Under suitable architectural conditions, the former is immune to local traps in training and tends to outperform conventional DNNs in prediction; whereas the latter achieves feature-learning consistency in over-parameterized regimes when the training sample size is sufficiently large. Together, these developments show that StoNet broadens deep learning from prediction-oriented function approximation to a richer framework for statistical inference.

Online Kernel-Based Mode Learning

Weixin Yao

University of California, Riverside

Modern big data streams routinely present dual challenges: massive volume that overwhelms single-machine memory, and severe data contamination—such as vertical outliers and heavy-tailed distributions—that undermines traditional mean-based methods like Ordinary Least Squares (OLS). To address these issues simultaneously, we propose a novel online kernel-based mode learning framework designed for robust and memory-efficient parameter estimation over continuous data streams. Processing data sequentially in blocks, our approach leverages a specialized Modal Expectation-Maximization (MEM) algorithm to localize conditional density peaks within each chunk. By extracting and aggregating local empirical Hessians as directional precision weights, the global parameter vector is updated recursively through a single-step algebraic combination, enabling the immediate deletion of historical raw records. Crucially, we establish the asymptotic normality of the resulting online estimator and prove its statistical equivalence to the global batch estimator, confirming that computational acceleration introduces no efficiency penalty. Extensive simulation studies and an empirical application involving over 7.2 million American Airline flight records demonstrate that our method achieves superior resistance to asymmetric contamination, matches the predictive accuracy of batch workflows, and delivers substantial speedups over standard benchmarks.

Black Box Variational Inference with a Deterministic Objective: Faster, More Accurate, and Even More Black Box

Ryan Giordano

University of California, Berkeley

Automatic differentiation variational inference (ADVI) offers fast and easy-to-use posterior approximation in multiple modern probabilistic programming languages. However, its stochastic optimizer lacks clear convergence criteria and requires tuning parameters. Moreover, ADVI inherits the poor posterior uncertainty estimates of mean-field variational Bayes (MFVB). We introduce "deterministic ADVI" (DADVI) to address these issues. DADVI replaces the intractable MFVB objective with a fixed Monte Carlo approximation, a technique known in the stochastic optimization literature as the "sample average approximation" (SAA). By optimizing an approximate but deterministic objective, DADVI can use off-the-shelf second-order optimization, and, unlike standard mean-field ADVI, is amenable to more accurate posterior covariances via linear response (LR). In contrast to existing worst-case theory, we show that, on certain classes of common statistical problems, DADVI and the SAA can perform well with relatively few samples even in very high dimensions, though we also show that such favorable results cannot extend to variational approximations that are too expressive relative to mean-field ADVI. We show on a variety of real-world problems that DADVI reliably finds good solutions with default settings (unlike ADVI) and, together with LR covariances, is typically faster and more accurate than standard ADVI.

Probabilistic Symbolic Regression for Equation Discovery via Operator-induced and Regularized Symbolic Forests

Debdeep Pati

University of Wisconsin-Madison/Boston University

Symbolic regression has emerged as a powerful tool for artificial intelligence-driven scientific discovery by learning interpretable analytical expressions that reveal governing relationships directly from data. Existing methods, however, often rely on heuristic search, struggle to balance predictive accuracy with expression complexity in noisy settings, and offer limited characterization of symbolic uncertainty. Probabilistic approaches that address these challenges in a unified manner remain underexplored. We introduce a probabilistic symbolic regression framework that represents mathematical expressions as ensembles of symbolic trees. A regularizing prior over tree topology controls expression complexity, while an Occam's window-based posterior summary captures uncertainty across multiple plausible symbolic models. Given the limited existing theoretical treatment of symbolic regression, we develop posterior concentration guarantees when symbolic expressions approximate the underlying relationship arbitrarily well, with a near-parametric rate when an exact finite formula exists. Additionally, we establish a sharp oracle concentration result under symbolic misspecification. Comparisons of our proposed framework with state-of-the-art competitors demonstrate superior predictive accuracy, optimal symbolic complexity, and stable structural recovery when learning benchmark scientific equations, together with the identification of scientifically interpretable descriptor formulas in a challenging materials discovery application.

GLM Inference with AI-Generated Synthetic Data

Ali Shojaie

University of Washington

Data privacy concerns have led to the growing interest in synthetic data and digital twins, which strive to preserve the statistical properties of the original dataset while ensuring privacy by excluding real records. Recent advances in deep neural networks and generative artificial intelligence have facilitated the generation of synthetic data. However, valid statistical inference with synthetic data remains under studied. In particular, in many settings, including generalized linear models (GLMs), the estimator obtained using synthetic data converges much more slowly than in standard settings. To address these limitations, we propose a method that leverages summary statistics from the original data and use a misspecified linear regression estimator to develop inference that greatly improves the convergence rate and restores the standard root-behavior for GLMs.

5H Multi-view learning and data integration

Saturday November 14 • 1:00–3:00 pm • Harbor

Chair: Hua Zhou • 6 speakers

Finding Shared Signals Across Many Datasets: Statistical Insights from RMT

Rong Ma

Harvard University

Modern data analysis increasingly requires identifying shared latent structure across multiple high-dimensional datasets. A commonly used model assumes that the data matrices are noisy observations of low-rank matrices with a shared singular subspace. In this case, two primary methods have emerged for estimating this shared structure, which vary in how they integrate information across datasets. The first approach, termed Stack-SVD, concatenates all the datasets, and then performs a singular value decomposition (SVD). The second approach, termed SVD-Stack, first performs an SVD separately for each dataset, then aggregates the top singular vectors across these datasets, and finally computes a consensus amongst them. While these methods are widely used, they have not been rigorously studied in the proportional asymptotic regime, which is of great practical relevance in today's world of increasing data size and dimensionality. Consequently, it remains unclear when one method should be preferred over another. In this work, we derive exact expressions for the asymptotic performance and phase transitions of these two methods and develop optimal weighting schemes to further improve both methods. Our analysis reveals that while neither method uniformly dominates the other in the unweighted case, optimally weighted Stack-SVD dominates optimally weighted SVD-Stack when the low rank signal is fully shared across the datasets. We then analyze multiple, partially shared components per dataset and demonstrate that SVD-Stack can yield improved performance without requiring estimation of subspace alignment. Finally, we provide practical algorithms for estimating optimal weights from data, offering theoretical guidance for method selection in practical data integration problems. Extensive numerical simulations and semi-synthetic experiments on genomic data corroborate our theoretical findings.

Bias-Corrected Subspace Intersection: Minimax-Optimal Shared Subspace Estimation in Multi-View Data

Cong Ma

University of Chicago

Estimating a low-dimensional subspace shared across noisy data matrices is a fundamental problem in multi-view matrix estimation. We study this problem under the two-view JIVE model, where each data matrix contains shared and view-specific low-rank components. We demonstrate that standard plug-in subspace intersection, including AJIVE, suffers from a second-order bias caused by direction-dependent leakage of the empirical singular vectors. We propose *bias-corrected subspace intersection* (BCSI), which removes this bias before estimating the shared subspace. We establish finite-sample risk bounds for BCSI that accommodate unequal view dimensions, signal strengths, and view-specific ranks and require no condition-number assumptions on the signal matrices. When the shared and view-specific ranks are comparable, these bounds match our minimax lower bounds up to universal constants. The resulting minimax rate contains a new second-order term, arising from quadratic leakage perturbations relative to the shrinking spectral gap when the view-specific subspaces are nearly aligned. This term is absent from previous JIVE minimax lower bounds. Numerical experiments demonstrate the advantage of BCSI over AJIVE when the leakage bias is pronounced. Along the way, we establish a nonasymptotic concentration result for the bias-corrected leakage Gram matrix of a rectangular spiked matrix, which may be of independent interest. This is joint work with Xianwen Song and Yuepeng Yang.

Consensus dimension reduction via multi-view learning

Tiffany Tang

University of Notre Dame

A plethora of dimension reduction methods have been developed to visualize high-dimensional data in low dimensions. However, different dimension reduction methods often output different and possibly conflicting visualizations of the same data. This problem is further exacerbated by the choice of hyperparameters, which may substantially impact the resulting visualization. To obtain a more robust and trustworthy dimension reduction output, we advocate for a consensus approach, which summarizes multiple visualizations into a single consensus dimension reduction visualization. Here, we leverage ideas from multi-view learning in order to identify the patterns that are most stable or shared across the many different dimension reduction visualizations, or views, and subsequently visualize this shared structure in a single low-dimensional plot. We demonstrate that this consensus visualization effectively identifies and preserves the shared low-dimensional data structure through both simulated and real-world case studies. We further highlight our method's robustness to the choice of dimension reduction method and hyperparameters—a highly-desirable property when working towards trustworthy and reproducible data science.

Meta Fusion: A Unified Framework For Multimodal Fusion with Adaptive Mutual Learning

Ziyi Liang

University of California, Santa Barbara

Developing effective multimodal data fusion strategies has become increasingly essential across a wide range of applications, from autonomous driving to medical diagnosis. Traditional fusion methods, including early, intermediate, and late fusion, integrate data at different stages, each offering distinct advantages and limitations. In this paper, we introduce Meta Fusion, a flexible and principled framework that unifies these strategies as special cases. Motivated by deep mutual learning and ensemble learning, Meta Fusion constructs a cohort of models based on various latent representation combinations, and further boosts predictive performance through soft information sharing within the cohort. Our approach is model-agnostic in learning the latent representations, allowing it to flexibly adapt to the unique characteristics of each modality. Theoretically, our soft information sharing mechanism reduces the generalization error. Furthermore, we develop an uncertainty quantification framework that provides rigorous coverage guarantees with minimal assumptions. Empirically, Meta Fusion consistently outperforms conventional strategies across extensive numerical studies, including Alzheimer's disease detection and neural decoding.

Stein-Encoder: A White-Box Supervised Encoder via Stein Identities in Multi-Modal Studies

Xinzhou Guo

HKUST

In multi-modal biomedical research, integrating high-dimensional genomic data with clinical baselines is essential for precision medicine. However, standard deep neural network approaches often entangle these modalities, obscuring the specific predictive impact of genetic features and leading to possibly suboptimal predictive performance. Motivated by the landmark METABRIC cohort primary breast tumors study, we propose the Stein-Encoder, a white-box supervised framework designed to isolate the genetic signal driving clinical outcomes conditional on nuisance covariates. By leveraging Stein's method and residualization techniques, our approach constructs an interpretable single index that summarizes relevant biological heterogeneity while flexibly incorporating clinical factors and can be used to improve downstream prediction. We establish theoretical guarantees for identification, consistency and efficiency improvement. Applied to the METABRIC cohort, the Stein-Encoder outperforms unsupervised benchmarks in predictive accuracy. Crucially, it achieves structural disentanglement by revealing response-specific biological mechanisms: we find that tumor size is driven primarily by mitotic networks, whereas prognostic indices rely on a distinct proliferation-versus-immune axis. This work contributes a unified, computationally efficient framework that bridges statistical rigor with the representational power of neural networks, enabling interpretable, task-specific and efficient compression of multi-modal health data for a wide range of precision medicine applications, beyond biomarker discovery.

Whom to trust and what to keep: adaptive Bayesian transfer learning from pretrained models

Tian Gu

Columbia University

Pretrained models offer information for a new study without requiring access to their training data. The challenge is that source studies may differ from the target in their populations, effect scales, and observed features. This talk presents two Bayesian approaches for high-dimensional regression that use fitted source coefficients to improve estimation and inference in a small target study. The approaches address two borrowing decisions: which sources to trust and which signals to keep. One approach centers a horseshoe prior on a learned mixture of rescaled source estimates, with similarity-informed weights and an explicit no-transfer component. Its contraction bounds improve with an informative source center and retain the target-only rate when borrowing is unhelpful. A related approach combines source weighting with a sparse projection of posterior draws, using coordinate-level agreement to determine which coefficients to retain. Source calibration permits borrowing across partially overlapping feature sets and missing modalities. Applications to GTEx fine-mapping and ADNI multimodal prediction illustrate the approaches' feasibility and advantages over comparisons in signal selection, prediction, and uncertainty quantification.

5S Statistical learning with heterogeneous data

Saturday November 14 • 1:00–3:00 pm • State Street

Chair: Hongtu Zhu • 6 speakers

Subsampling for the analysis of large data

Zhezhen Jin

Department of Biostatistics, Columbia University

Analysis of large-scale data is challenging due to data storage and computational complexity. When analyzing large-scale data, subsampling methods and divide-and-conquer procedures are appealing, because they ease the computational burden, while preserving the validity of inferences. In this talk, several challenges and issues will be reviewed and a perturbation subsampling approach will be presented based on independent and identically distributed stochastic weights for analyzing large-scale data. The method can be justified based on optimizing convex objective functions by establishing the asymptotic consistency and normality of the resulting estimators. The method simultaneously provides consistent point and variance estimators. We demonstrate the finite-sample performance of the proposed method using simulation studies and a real-data analysis.

Kernelized Stein Discrepancy for Learning from Heterogeneous Data

Gourab Mukherjee

University of Southern California

Modern data-driven systems increasingly require learning across large collections of heterogeneous populations, where related tasks share information but differ in unknown nuisance characteristics. A fundamental statistical challenge is to determine how to pool information adaptively across such tasks while accounting for heterogeneity and avoiding systematic selection biases. We address this problem through Nonparametric Empirical Bayes Structural Tweedie (NEST), a framework for simultaneous estimation that combines empirical Bayes learning with kernelized Stein discrepancy. NEST learns the score quantities underlying a generalized Tweedie formula by minimizing a kernelized Stein discrepancy, yielding a flexible, data-adaptive approach that avoids explicit estimation of the unknown mixing distribution. For the heterogeneous normal means problem, NEST addresses two intertwined sources of selection bias arising from heterogeneity: bias in the estimated mean and bias in the estimated variance. Correcting the former generally requires accounting for the latter. We establish that NEST asymptotically attains the performance of the optimal Bayes rule under a weighted squared-error loss, providing a theoretical foundation for adaptive information sharing across heterogeneous samples. The results demonstrate how Stein-based learning and empirical Bayes principles can provide statistically grounded mechanisms for adaptive learning across heterogeneous tasks.

Model positive and unlabeled data with a generalized additive density ratio model

Peijun Sang

University of Waterloo

This paper focuses on learning from positive and unlabeled (PU) data, where only some positives are labeled and the rest are mixed with negatives. Classical exponential tilting models guarantee identifiability by imposing a linear structure, but they can be severely misspecified when the true relationships are nonlinear. We propose a generalized additive density-ratio framework that retains identifiability while allowing nonlinear and feature-specific effects. The approach comes with a practical fitting algorithm and supporting theory that enable estimation and inference for the mixture proportion and other quantities of interest. In simulations and analyses of real datasets, the proposed method matches the standard exponential tilting method when the linear model is correct and delivers clear gains when it is not. Overall, the framework strikes a useful balance between flexibility and interpretability for PU data and provides principled tools for estimation, prediction, and uncertainty assessment.

fDA: A data-augmented predictive framework for generating fair synthetic labels with controlled fairness-faithfulness trade-offs

Bei Jiang

University of Alberta

Algorithmic decision systems are increasingly used in socially sensitive domains, raising concerns about bias inherited from historical data. Fair synthetic data generation provides a pre-processing strategy for bias mitigation, but existing methods often rely on black-box generative models whose tuning parameters have limited theoretical interpretation. We propose fDA, a Data-Augmented predictive framework for generating Fair synthetic labels. The framework combines a fairness-enforcing model, which specifies a fair reference conditional law satisfying the desired fairness constraint, with a faithfulness-preserving model, which generates auxiliary variables from observed labels to retain controlled original-label information. Synthetic labels are sampled from the predictive distribution induced by jointly modeling these components, coupled with a tuning mechanism. For continuous labels, the Gaussian working specification yields explicit calibration of the effective noise level to a target population upper bound on unfairness. For both continuous and ordinal labels, the predictive distribution is fully specified for each tuning value, enabling empirical calibration of the achieved fairness--faithfulness trade-off. Theoretically, when the auxiliary variable becomes fully informative, synthetic labels converge to the original labels in probability and distribution; when it becomes non-informative, the mechanism reduces to the fair reference law. Experiments on simulated and real datasets show interpretable trade-offs and improved faithfulness over GAN-based baselines.

Fairness within Clinical Context: Stratified Tilted Risk Minimization for Medical Prediction

Lu Tang

University of Pittsburgh

Medical AI systems are often evaluated for fairness at the population level, even when patient groups differ in clinically meaningful ways. In such settings, an aggregate group-parity objective can confound demographic disparity with case-mix variation and miss local failures that matter in practice. We propose Stratified Tilted Empirical Risk Minimization (StraTERM), a training objective that applies group-level tilted optimization within clinically coherent strata rather than across the full population at once. The method is designed to align fairness optimization with within-context comparisons while retaining the simplicity of tilted-risk training. We evaluate StraTERM on UCI Diabetes readmission prediction and MIMIC-CXR Lung Opacity classification using five patient-level cross-validation folds. StraTERM preserves AUROC and obtains the lowest mean conditional equalized-odds gap on UCI Diabetes and within the TERM-family comparison on MIMIC-CXR; external baselines remain competitive on MIMIC-CXR, where GroupDRO has the lowest fixed-threshold mean and worst-stratum conditional EO. The gains are not uniform across strata: MIMIC-CXR ED and UCI Musculoskeletal worsen relative to ERM. These results support StraTERM as a mechanism for improving average within-context fairness, while showing that deployment-oriented evaluation should report per-stratum and worst-stratum behavior rather than relying only on a mean gap.

A Bayesian scalar-on-function kernel regression model for crop yield

Veronica Berrocal

UC Irvine

Agriculture in the United States is critical for food security across the world. Additionally, the agricultural industry contributes substantially to the United States' GDP. Any deterioration in the U.S. agricultural yield could have considerable consequences for worldwide economy and food security. To better understand climate's impact on crop yield and devise successful climate adaptation strategies, it is fundamental to know at what stages of a plant's growth crops are most vulnerable to the harmful effects of certain environmental factors. In this talk, we present a modeling strategy to determine how two important climatic conditions, heat stress and water stress, individually and jointly affect crop's health during the course of a growing season. Our model is specified within a statistical functional framework, estimated under the Bayesian paradigm, and postulates a non-linear interaction between heat and water stress. Additionally, the model envisions the effect of heat and water stress to be received by the crops differently depending on the plant's developmental stage. We illustrate the capability of the model by applying it to maize yield in the US Midwest over a 4-year period.

6C Reliable learning and structured inference

Saturday November 14 • 3:30–5:30 pm • Corwin

Chair: Shujie Ma • 6 speakers

MLFM: Manifold-Based Latent Factor Discovery in Functional Brain Networks

Hongtu Zhu

University of North Carolina at Chapel Hill

Large-scale functional MRI studies generate subject-specific brain connectivity networks for tens of thousands of participants, creating high-dimensional matrix datasets that strain standard vector-based statistical methods. To address this challenge, we propose a manifold-based latent factor model (MLFM) that decomposes each connectivity matrix into a shared loading matrix on the Stiefel manifold, subject-specific symmetric cores capturing individual variation, and an explicit isotropic noise term. This noise separation is absent from conventional principal component analysis competitors, a feature that sharpens eigenvalue gaps and reduces subspace estimation bias. We develop two complementary estimation frameworks, profiled likelihood and a generalized method of moments, solved via Stiefel-manifold optimization, alongside a manifold-adapted Bayesian information criterion that consistently selects the true latent dimension. For the loading matrix, we establish root-n consistency and asymptotic normality on the Grassmann manifold, characterizing their joint distribution via the Riemannian logarithmic map to enable principled, coordinate-free inference for latent network structures. Applied to resting-state fMRI data from 37,794 UK Biobank subjects across 360 brain regions, MLFM identifies eight heritable functional networks, with two exceeding a 0.35 heritability threshold. A subsequent genome-wide association study yields 108 significant genetic loci across 27 phenotypes, providing a dimension-reduced pathway for large-scale neuroimaging genetics.

BROWNIE: Bayesian Random Weight Neural Networks as a Simple Alternative to Deep Learning

Jian Kang

University of Michigan

We propose a simple and scalable alternative to deep neural networks based on shallow neural networks with random hidden-layer weights and Bayesian inference. We refer to the resulting framework as the Bayesian Random Weight Neural Network Inference Engine (BROWNIE). BROWNIE combines the expressive power of rich nonlinear feature representations induced by random weights with principled Bayesian inference, while avoiding the computational burden of training all network parameters. We develop an efficient algorithm for posterior computation and establish theoretical properties of the proposed framework. Through extensive numerical experiments, we demonstrate that BROWNIE can achieve competitive or improved performance compared to deep neural networks, while exhibiting greater robustness across a range of settings. These results suggest that simple neural architectures with random weights, when coupled with Bayesian inference, can provide an effective and computationally attractive alternative to fully trained deep networks.

Beyond Flat Minima: Sharpness-Aware Bilevel and Minimax Optimization.

George Michailidis

UCLA

Many modern learning procedures are defined through nested optimization problems, including hyperparameter tuning, meta-learning, adversarial training, and reinforcement learning. At the same time, the geometry of the loss landscape—and in particular the preference for flat over sharp solutions—has emerged as an important perspective on generalization in overparameterized models. While sharpness-aware methods have been successful for single-level optimization, considerably less is understood about how to control sharpness when the learning problem itself is bilevel or minimax.

This talk introduces a sharpness-aware framework for nested optimization that explicitly accounts for the geometry of both levels of the problem. For bilevel optimization, this leads to a formulation that seeks solutions that are simultaneously robust to perturbations of the outer and inner objectives. A penalty-based reformulation yields a practical first-order algorithm that avoids implicit differentiation and second-order computations. The same perspective leads to a simpler sharpness-aware method for minimax problems. We discuss convergence guarantees under Polyak–Łojasiewicz conditions, connections between sharpness and generalization through PAC-Bayesian bounds, and empirical results on data hypercleaning and generative adversarial learning. The broader goal is to understand how loss-landscape geometry can be incorporated into the statistical and computational foundations of nested learning.

LIDS: LLM Summary Inference Under the Layered Lens

Jinchi Lv

University of Southern California

Large language models (LLMs) have attracted substantial attention from researchers and practitioners in natural language processing (NLP) since the release of ChatGPT in 2022. A prominent capability of ChatGPT and related LLMs is prompt-based summarization, yet the linguistic complexity of generated text makes summary quality difficult to evaluate in a statistically principled manner. To this end, we propose LLM summary inference with a BERT--SVD-based direction metric and SOFARI (LIDS), a framework that jointly assesses summary accuracy and identifies reproducible, interpretable key words associated with layered latent themes. LIDS exploits BERT token embeddings and a latent SVD-based direction metric to measure the similarity between a summary and its source text, while repeated prompts quantify prompt-level statistical uncertainty. The induced direction vector also provides a natural text-level embedding of each summary for large text reduction. In the second step, LIDS combines the SOFARI p-values with dependence-aware false discovery rate (FDR) procedures—ordinary Benjamini--Hochberg (BH) under PRDS-type positive dependence and Benjamini--Yekutieli (BY) harmonic reshaping under arbitrary dependence—to recover important key words within each latent theme. Our theory establishes perturbation rates for the empirical LIDS directions, consistency of the LIDS similarity measure and the selected number of latent SVD layers, stability of the induced summary embeddings, prompt-level uncertainty quantification from repeated summaries, high-probability benchmark separation, and FDR control and power guarantees for layered key word recovery under positive or arbitrary p-value dependence. Comprehensive empirical studies demonstrate the practical utility and robustness of LIDS through human validation, comparisons to established similarity metrics, paired SOFARI--BH and SOFARI--BY visualizations, cross-LLM comparisons, and robustness analyses across text domains. This is a joint work with Yingying Fan and Dylan Park.

The Gold Rush in AI4Math: Structural Signatures of AI Involvement

Tracy Ke

Harvard University

Can we identify AI involvement in a paper's mathematics, rather than only its writing? I will introduce ProofTrace, which turns mathematical papers into result-dependency graphs and extracts interpretable features of how their arguments are organized. We compare five cohorts, from historical human papers to AI-generated manuscripts, including paired designs that probe the effects of research topic and prose rewriting. AI-reworded papers remain close to human papers in this feature space; disclosed-AI papers occupy an intermediate position; and several structural differences recur across cohorts. An exploratory detection study also finds that ProofTrace's performance remains relatively stable after rewording. I will discuss both the insights and the limits of these signatures.

Causal Discovery on Dependent Data

Qing Zhou

Department of Statistics & Data Science, UCLA

Learning causal structures from observational data is a central challenge in statistics and machine learning. A common assumption in causal discovery is that data samples are i.i.d., yet this assumption is often violated in real-world settings such as social networks, brain imaging, and cell differentiation, where dependencies exist across observational units. In this talk, I present a unified framework for causal structure learning from dependent data through decorrelation. For continuous data, we extend the Gaussian DAG model by incorporating sample dependence over an undirected graph. For discrete data, we introduce a latent utility model with correlated errors across units for local conditional distributions in a DAG. In both cases, we develop methods to estimate the between-unit covariance structure and use it to decorrelate the data, enabling the application of standard DAG learning algorithms. Numerical experiments on both simulated and single-cell RNA-seq data show that the decorrelation approach significantly improves the accuracy of causal graph recovery compared to directly applying standard methods to dependent data.

6F Foundations of AI and distributed learning

Saturday November 14 • 3:30–5:30 pm • Flying-A

Chair: Faming Liang • 6 speakers

An Interpretable and Scalable Framework for Evaluating Large Language Models

Xiaoqian Liu

University of California, Riverside

Evaluation of large language models (LLMs) is increasingly critical, yet standard benchmarking methods rely on average accuracy, overlooking both the inherent stochasticity of LLM outputs and the heterogeneity of benchmark items. Item Response Theory (IRT) offers a principled framework for modeling latent model abilities and item characteristics, but conventional methods are computationally expensive and numerically unstable, limiting large-scale implementations. To address these challenges, we propose an interpretable and scalable framework for LLM evaluation based on the majorization–minimization principle. Our approach reformulates the problem as a sequence of constrained matrix factorization subproblems, enabling stable and efficient parameter estimation with theoretical guarantees for identifiability and convergence. Experiments on synthetic and real-world datasets, including MATH-500 and six Open LLM Leaderboard benchmarks, demonstrate that our method achieves superior scalability and interpretability. It delivers orders-of-magnitude speedups over competing methods while maintaining comparable or even higher estimation accuracy. Our results align with established scaling laws and provide interpretable insights into item difficulty and discrimination, informing more principled benchmark design than conventional average-accuracy evaluation.

Sample complexity bounds for categorical Markov random fields via Discrete Diffusions

Nabarun Deb

University of Chicago Booth School of Business

Many applications in statistics, economics, and physics require sampling from high-dimensional categorical distributions with local dependence structures. Examples include finite memory language models, Ising and Potts systems, and spatial models with local interactions. In modern machine learning, discrete diffusions have emerged as a flexible approach for sampling such data, with strong empirical performance across structured discrete domains. Motivated by this, we develop learning methods with end-to-end sample complexity bounds for discrete diffusion under local dependence, which we model through low order Markov random fields (MRFs). Our main technical insight is a new *pinning decomposition* of the discrete score, which captures how local interactions propagate through the noising process. Building on this decomposition, we propose a *weight-sharing neural score learner* and combine it with τ -leaping to obtain an end-to-end sampling procedure. Rather than treating score-learning error as a black-box input, as is common in existing sampling analyses, we study the score learning error from finite data and derive optimal sampling guarantees with explicit dependence on the dictionary size, the interaction order of the MRF, and the sample size. Moreover, our strategy trains a single score network across uniform noise levels while leaving the sampling discretization to be chosen at inference-time. This allows the same trained model to trade accuracy for computational cost as inference-time budgets vary. This is joint work with Shivam Kumar.

Beyond Marginal Coverage: Efficient Localized Conformal Prediction via Residual Rank Calibration

Wenqing He

University of Western Ontario

Conformal prediction provides finite-sample marginal coverage guarantees under exchangeability, but marginal validity alone does not ensure that prediction intervals adapt to local structure in the conditional distribution. Existing adaptive methods such as conformalized quantile regression (CQR) and conformal histogram regression (CHR) improve conditional coverage by incorporating estimated quantiles or conditional densities into the conformity score, but doing so requires direct estimation of the conditional distribution of response Y given covariates X , which can be statistically and computationally demanding. We propose RLR-CHR, a residual-based local rank conformal histogram regression method that sidesteps full conditional distribution estimation by decoupling location and scale effects via a learned noise proxy and constructing histogram-based conformity scores on the standardized residuals. Local calibration is then performed through rank comparisons rather than weighted conformal quantiles, yielding a procedure that is both localized and computationally efficient. We establish finite-sample marginal coverage for RLR-CHR under exchangeability and asymptotic conditional coverage under local stability of the residual score distribution. We further prove that RLR-CHR is asymptotically equivalent to its weighted-localization counterpart RBC-CHR under suitable regularity conditions, providing a precise theoretical justification for the reduction in per-query computational cost without sacrificing asymptotic accuracy. Experiments on synthetic and real datasets demonstrate that RLR-CHR consistently improves worst-slab coverage and produces more compact prediction intervals than existing conformal approaches.

TOPPO: Rethinking PPO for Multi-Task Reinforcement Learning with Critic Balancing

Rui Miao

University of Texas at Dallas

Soft Actor-Critic (SAC) and its variants dominate Multi-Task Reinforcement Learning (MTRL) due to their off-policy sample efficiency, while on-policy methods such as Proximal Policy Optimization (PPO) remain underexplored. We diagnose that PPO in MTRL suffers from a previously overlooked issue: critic-side gradient ill-conditioning, which may cause tail tasks to stall while easy tasks dominate the value function's updates. To address this, we propose TOPPO (Tail-Optimized PPO), a reformulation of PPO via Critic Balancing -- a set of modules that improve gradient conditioning and balance learning dynamics across tasks. Unlike prior approaches that rely on modular architectures or large models, TOPPO targets the optimization bottleneck within PPO itself. Empirically, TOPPO achieves stronger mean and tail-task performance than published SAC-family and ARS-family baselines while using substantially fewer parameters and environment steps on Meta-World+ benchmark. Notably, TOPPO matches or surpasses strong SAC baselines early in training and maintains superior performance at full budget. Ablations confirm the effectiveness of each module in TOPPO and provide insights into their interactions. Our results demonstrate that, with proper optimization, on-policy methods can rival or exceed off-policy approaches in MTRL, challenging the prevailing reliance on SAC and highlighting critic-side gradient conditioning as the central bottleneck.

Enhancing Federated Learning Methods using Third Order Optimization

Jeffrey Zhang

UC Riverside

We investigate the use of third-order methods for federated statistical learning. Higher-order approximations provide more accurate representations of an objective function, and incorporating them into federated procedures may lead to more accurate statistical methods. We study two existing federated algorithms that rely on second-order approximations and develop third-order analogues by replacing their second-order components with higher-order ones. We show theoretically that the resulting procedures provide improved approximation to the corresponding pooled-data estimators, and demonstrate empirically that these improvements translate into better finite-sample performance.

Federated Learning for Feature Mismatch

Fei Xue

Purdue University

We consider a framework that involves multiple sites with privacy constraints and heterogeneity, where different sites collect data from independent subjects, and site-specific, mismatched feature sets (e.g., hospitals with differing equipment and workflows). Most existing methods assume either aligned covariates or overlapping cohorts, and thus are inapplicable when both samples are disjoint and the covariates are mismatched. We propose Fed-PATE, a three-stage framework: local estimation at each site using its own observed features; a cross-site projection aggregation that recovers a consensus full-feature coefficients among similar sites; and a penalized transfer estimator that feeds the aggregated information back to each site. To accommodate heterogeneity and prevent negative transfer, we introduce a weighting scheme for model selection that adapts to inter-site similarity. Our proposed approach shows a lower prediction risk in theoretical analyses. Simulations show significant reductions in prediction MSE and robustness to inter-site heterogeneity. In a real-world application to Serum Creatinine (SCr) prediction using EHR variables, Fed-PATE improves performance over existing approaches for the majority of sites.

6H Learning from distributions and complex measurements

Saturday November 14 • 3:30–5:30 pm • Harbor

Chair: Lily Wang • 6 speakers

Estimating Wasserstein Barycenters from Sparse Deep Sequencing

Linbo Wang

University of Toronto

We study distributional data under sparse sampling where each unit is represented by a probability distribution on the real line observed only through a small i.i.d. sample. A natural notion of central tendency for one-dimensional distributional data is the Wasserstein barycenter, whose quantile function is the pointwise average of the unit-level quantile functions. We focus on pointwise estimation of the Wasserstein barycenter quantile function: at a given quantile level, the target is the population mean of the corresponding unit-level quantiles. A naive plug-in estimator is the empirical Wasserstein barycenter, which treats observed unit-level empirical distributions as the true latent unit-level distributions. Under sparse sampling, however, this estimator can be severely biased. We propose an approach that avoids directly estimating either the unit-level distributions or the full population law of distributions. We start with the more ambitious goal of characterizing the distribution of latent unit-level quantiles at a given quantile level. We show that this distribution can be written in terms of the marginal distributions of the unit-level CDF values, which can be estimated using binomial mixture methods. This motivates our estimator, the marginal-constructed barycenter (MCB) estimator, obtained by taking the mean of the estimated distribution of latent unit-level quantiles. We establish conditions under which the MCB estimator is pointwise consistent and asymptotically normal, and show through simulations that it can substantially outperform the empirical Wasserstein barycenter under sparse sampling. We illustrate the method in an analysis of HIV-1 sequence data from the HVTN 502/503 vaccine efficacy trials, using the barycenter to summarize and compare within-participant distributions of viral sequence features when only a small number of sequences are available per participant.

Learning from multivariate distributions with nonparanormal transport

Irina Gaynanova

University of Michigan

Many modern data sources record thousands of measurements per subject, as with wearable devices such as continuous glucose monitors and accelerometers, yet analyses typically compress these recordings into a handful of scalar summaries, while emerging foundation models replace them with high-dimensional learned embeddings that are difficult to interpret. An appealing alternative is to treat each subject's measurements as a sample from a subject-specific distribution and to take that distribution itself as the unit of analysis. Equipped with an optimal transport metric, the same geometry that underlies many modern generative models, the space of distributions is a metric space rather than a vector space, so standard tools for regression and dimension reduction no longer apply directly. In this talk I will discuss statistical and computational challenges that arise in this setting, focusing on extensions beyond the univariate case, where optimal transport becomes computationally prohibitive. I will introduce a nonparanormal transport metric that decomposes the distance between multivariate distributions into univariate Wasserstein distances between marginals and a Bures–Wasserstein distance between latent correlation matrices, enabling efficient and interpretable regression that mitigates the curse of dimensionality. The methods will be illustrated on data from continuous glucose monitoring studies.

Scalable four-level nested functional principal component analysis for large dense functional data

Hyunkeun Ryan Cho

UCSD

Modern neuroscience studies increasingly produce dense functional data indexed simultaneously by multiple nested and crossed factors rather than a single replication index. Standard functional principal component analysis, and even existing multilevel extensions, are not designed for this four-way structure. At this scale, direct score prediction is not merely slow but computationally infeasible, since it requires inverting the covariance matrix of the full observation vector. We introduce a four-level multilevel functional principal component framework that decomposes such data into interpretable participant-, stimulus-, and electrode-level sources of variation through a nested covariance estimator that exactly identifies each level's covariance function. Combined with scalable covariance smoothing and efficient score prediction, the method remains tractable for very large functional datasets where conventional approaches cannot be applied at all, while preserving statistical interpretability. We establish convergence rates and consistency for the resulting eigenvalues and eigenfunctions, and illustrate the framework on a large auditory neuroscience application.

Modeling with speckle noise in high-dimensional image sensing

Soham Jana

University of Notre Dame

Speckle noise is an integral part in several modern imaging paradigms encompassing ultrasound, synthetic aperture radar, and optical coherence tomography. Deep learning-based methods have been proven fruitful for analyzing sensor data and reconstructing images when the signal resides in a low-dimensional space. However, the existing guarantees for such methods are far from being optimal. In addition, the problem becomes challenging when the number of sensors is very small compared to the image dimension, and it is often necessary to take multiple images of the scene to enable consistent reconstruction. Our recent work establishes estimation thresholds for such problems when the signal can arise from either low- or high-dimensional regime. Real and simulated data analyses support the efficiency of our estimation techniques.

Unsupervised Aggregation of Multiple Learning Algorithms

Rui Duan

Harvard T.H. Chan School of Public Health

Across various domains, the growing advocacy for open science and open-source machine learning has made an increasing number of models publicly available. These models allow practitioners to integrate them into their own contexts, reducing the need for extensive data labeling, training, and calibration. However, selecting the best model for a specific target population remains challenging due to issues like limited transferability, data heterogeneity, and the difficulty of obtaining true labels or outcomes in real-world settings. In this paper, we propose an unsupervised model aggregation method, U-aggregation, designed to integrate multiple pre-trained models for enhanced and robust performance in new populations. Unlike existing supervised model aggregation or super learner approaches, U-aggregation assumes no observed labels or outcomes in the target population. Our method addresses limitations in existing unsupervised model aggregation techniques by accommodating more realistic settings, including heteroskedasticity at both the model and individual levels, and the presence of adversarial models. Drawing on insights from random matrix theory, U-aggregation incorporates a variance stabilization step and an iterative sparse signal recovery process. These steps improve the estimation of individuals' true underlying risks in the target population and evaluate the relative performance of candidate models. We provide a theoretical investigation and systematic numerical experiments to elucidate the properties of U-aggregation. We demonstrate its potential real-world application by using U-aggregation to enhance genetic risk prediction of complex traits, leveraging publicly available models from the PGS Catalog.

A Latent Semiparametric Mixture Model for Robust Heterogeneity Learning in High Dimensions

Wen Zhou

New York University

Learning heterogeneity from high-dimensional data has become a fundamental task across many data-driven fields. Cluster analysis is a widely used unsupervised tool for this purpose, and many existing methods have been adapted to high-dimensional settings through sparsity constraints. In contrast to algorithmic approaches, model-based clustering offers clean theoretical guarantees and computational simplicity, but can be restrictive in practice due to its reliance on specific parametric distributions. We propose a novel latent semiparametric mixture model for clustering, in which the observed data follow a Gaussian mixture after latent coordinate-wise monotone transformations. Furthermore, assuming a sparse optimal clustering rule in the latent space, we develop a new method, CESME, tailored for high-dimensional clustering. By allowing unspecified transformations, our framework becomes substantially more flexible and robust than traditional model-based approaches. The clustering error of CESME is shown to approach the optimal clustering error at a rate matching that obtained when the true transformations are known. Extensive simulations and real data analysis confirm that CESME outperforms a broad range of existing high-dimensional clustering methods, including those designed for single-cell RNA sequencing data. The online supplementary file contains technical proofs, additional discussions, and more numerical results.

6S Learning from sequential and process data

Saturday November 14 • 3:30–5:30 pm • State Street

Chair: Radu Craiu • 6 speakers

Transformers for Time Series: Statistical Optimality, Structural Exploration and Gradient Descent Consistency

Xiucui Ding

University of California, Davis

Transformers have achieved remarkable empirical success in time-series forecasting, yet the statistical mechanisms through which attention learns temporal structure remain poorly understood. In this talk, we introduce a stationary time-series Transformer (StsT) that combines time-delayed embeddings with a single-head attention mechanism. We show that the overlap structure of the embedded tokens connects the Transformer directly to classical best linear prediction. Under finite-moment, short-range-dependence, and covariance nondegeneracy conditions, the attention mechanism concentrates on the most recent token, while the learned value matrix recovers the companion matrix associated with the best linear predictor. We establish uniform approximations of both population and empirical risks and characterize their minimizers. We then analyze joint gradient descent on the attention-score and value parameters, showing that the value matrix recovers the population-optimal predictor within $O(\log d)$ iterations, where d is the embedding dimension. For empirical training on a single dependent time series, we provide sample-size conditions under which gradient descent remains statistically consistent. These results offer a rigorous explanation of how attention performs state selection while the value matrix learns the underlying forecasting rule, thereby connecting Transformer training dynamics with classical time-series prediction theory.

Multiple Changepoint Detection in Non-Gaussian Time Series

Robert Lund

University of California, Santa Cruz

This paper shows how to estimate multiple changepoint configurations in non-Gaussian autocorrelated time series. A transformation is used to convert a Gaussian series into a non-Gaussian series, enabling penalized likelihood methods to handle non-Gaussian scenarios. When the marginal distribution of the data is continuous, the methods essentially reduce to the change of variables formula for probability densities. When the marginal distribution is count-oriented, Hermite expansions and particle filtering techniques are used to quantify the scenario. Simulations demonstrating the efficacy of the methods are given and two data sets are analyzed: 1) the proportion of home runs hit by Major League Baseball batters from 1920-2023, and 2) a six-dimensional series of tropical cyclone counts from the Earth's basins of generation from 1980-2023. In the first series, Beta marginal distributions are used to describe the proportions; in the second, Poisson marginal distributions seem appropriate.

Dependence in Networks: Extremal Vulnerability and Granger Causality Beyond Linear Models

Hernando Ombao

Statistics, Bioengineering, and Applied Mathematics, KAUST

Characterizing dependence between the nodes of a network raises statistical challenges that standard correlation-based measures do not address. This talk presents two contributions. The first is motivated by electroencephalograms (EEGs), where dependence between nodes (channels) is often strongest in the extremes. We introduce the notion of *extremal vulnerability*, defined through tail conditional probabilities. Inspired by the PageRank algorithm, the approach identifies the most vulnerable nodes in a network: those with the highest probability of exhibiting an extreme event given that other nodes in the system are also extreme. The second contribution, which is ongoing work, concerns Granger causality that goes beyond linear conditional moments. We propose an inferential procedure that quantifies the causal influence of one node on another by the increase in predictive uncertainty when the past of the first node is omitted. Both methods are illustrated on financial stock data and on EEG recordings from a study of Alzheimer's disease.

Finite-Step Training of Deep ReLU Networks

Chunming Zhang

University of Wisconsin-Madison

Understanding why training succeeds in deep neural networks remains a central theoretical challenge, particularly when both the network activations and the loss function are nonsmooth. This talk considers sufficiently wide deep ReLU networks trained directly with hinge loss for classification. Under suitable conditions on the data and network width, we show that full-batch training reaches exactly zero hinge loss on the training data after finitely many updates, with high probability over Gaussian initialization. The key idea is geometric: a progress argument for a linear approximation of the network can be transferred to the actual nonlinear training process. I will explain this connection, discuss extensions to other classification and regression losses, and present numerical experiments illustrating finite-step termination and the role of network width.

Process data and educational measurement

Zhiliang Ying

Columbia University

Measurement theory has played a foundational role in educational assessment. Traditional methods are primarily based on IRT latent variable models, which may not be applicable to modern complex data such as process data obtained from PSTRE (Problem solving in technology-rich environments) items. This talk will discuss a number of statistical and machine learning methods for handling process data from PSTRE items.

(Mis)Understanding Benign Overfitting in Equity Return Prediction

Yan Yu

University of Cincinnati

Highly overparameterized models often predict well despite interpolating training data in complex domains, challenging the classical bias--variance tradeoff. We investigate whether this "benign overfitting" phenomenon extends to equity return prediction. Consistent with recent statistical theory, we document two key phenomena: first, a double descent pattern in the ridgeless model's prediction risk; and second, that while the optimal ridge model consistently outperforms its ridgeless counterpart, this performance gap becomes negligible at large parameter-to-observation ratios. Ultimately, however, both models fail to outperform a simple historical average. This empirical evidence aligns with our asymptotic results under the null hypothesis of zero slope coefficients, suggesting that standard equity predictors lack true forecasting power—even within highly flexible, nonlinear machine learning architectures. These findings reconcile modern and classical machine learning in asset pricing: in the absence of a true signal, they asymptotically collapse to the historical average benchmark.

Presenter index

Oral presenters are listed alphabetically by surname. Click a presenter's name to go to their title and abstract. Session codes match the compact program.

Presenter	Session	Page	Presenter	Session	Page
Arash Amini	3F	24	George Michailidis	6C	49
Jesús Arroyo	2S	22	Gourab Mukherjee	5S	46
Sudipto Banerjee	Friday banquet	2	Susan Murphy	Friday plenary	2
Sam Behseta	1H	10	Ning Ning	1S	11
Veronica Berrocal	5S	48	Xiaoyue Maggie Niu	2S	22
Xuan Bi	3F	23	Hernando Ombao	6S	57
Jacob Bien	2S	21	Oscar Hernan Madrid Padilla	2F	16
Matteo Bonvini	3S	29	Debdeep Pati	5F	42
Joshua Cape	3H	27	Carey Priebe	4F	33
Kun Chen	2H	18	Aaditya Ramdas	2C	15
Xiaohui Chen	4H	37	Sandra Safo	1H	11
Hyunkeun Ryan Cho	6H	54	Peijun Sang	5S	46
Carlos Cinelli	3S	28	Matteo Sesia	2F	17
Bertrand Clarke	1C	6	Babak Shahbaba	4S	38
Xinping Cui	4C	33	Weining Shen	1S	14
Michael Daniels	1H	10	Chengchun Shi	3F	23
Nabarun Deb	6F	51	Ali Shojaie	5F	42
Walter Dempsey	4S	39	Padhraic Smyth	2C	15
Jie Ding	1F	8	Weijie Su	2C	14
Xiucan Ding	6S	56	Fengzhu Sun	4C	32
Hani Doss	1C	4	Will Wei Sun	4F	34
Rui Duan	6H	55	Armeen Taeb	3H	25
Yang Feng	2H	19	David Talby	1F	8
Irina Gaynanova	6H	54	Lu Tang	5S	47
Ryan Giordano	5F	41	Tiffany Tang	5H	44
Tian Gu	5H	45	Xiwei Tang	2S	21
Yuqi Gu	3S	28	Alexander Volfovsky	2S	22
Xin Zhou Guo	5H	45	Stefan Wager	3F	24
Ying Guo	4C	32	Boxiang Wang	2F	17
Qiyang Han	2H	20	Jingshen Wang	1F	7
Wenqing He	6F	52	Lan Wang	4S	38
Hao Ho/Ker-Chau Li	2F	17	Lily Wang	4C	31
Andrew Holbrook	1H	9	Linbo Wang	6H	53

Presenter	Session	Page	Presenter	Session	Page
<u>Feifang Hu</u>	4S	37	<u>Peng Wang</u>	4F	34
<u>Mingzhao Hu</u>	1S	13	<u>Yazhen Wang</u>	1C	5
<u>Tate Jacobson</u>	3H	26	<u>Ying Nian Wu</u>	1C	5
<u>Soham Jana</u>	6H	54	<u>Minge Xie</u>	5F	40
<u>Bei Jiang</u>	5S	47	<u>Gongjun Xu</u>	4H	36
<u>Zhezhen Jin</u>	5S	46	<u>Hongquan Xu</u>	4F	35
<u>Jian Kang</u>	6C	49	<u>Jason Xu</u>	4H	36
<u>Tracy Ke</u>	6C	50	<u>Min Xu</u>	2F	18
<u>Dan Kessler</u>	3H	26	<u>Qi Xu</u>	3H	27
<u>Jason Klusowski</u>	1F	8	<u>Fei Xue</u>	6F	53
<u>Mladen Kolar</u>	4H	36	<u>Lan Xue</u>	1H	10
<u>Dehan Kong</u>	4S	39	<u>Shihao Yang</u>	1S	12
<u>Linglong Kong</u>	4S	38	<u>Weixin Yao</u>	5F	41
<u>Arun Kuchibhotla</u>	2H	20	<u>Grace Yi</u>	2F	16
<u>Keith Levin</u>	2S	21	<u>Zhiliang Ying</u>	6S	57
<u>Didong Li</u>	3S	30	<u>Yan Yu</u>	6S	57
<u>Gang Li</u>	4C	32	<u>Yubai Yuan</u>	3S	29
<u>Hongzhe Li</u>	1C	6	<u>Chunming Zhang</u>	6S	57
<u>Lexin Li</u>	4C	31	<u>Emma Jingfei Zhang</u>	2H	18
<u>Runze Li</u>	2C	15	<u>Hao Helen Zhang</u>	1C	4
<u>Faming Liang</u>	5F	40	<u>Jeffrey Zhang</u>	6F	52
<u>Ziyi Liang</u>	5H	44	<u>Kai Zhang</u>	2H	19
<u>Xihong Lin</u>	Saturday plenary	2	<u>Linjun Zhang</u>	1F	7
<u>Xiaoqian Liu</u>	6F	51	<u>Jiwei Zhao</u>	4F	34
<u>Yufeng Liu</u>	3F	24	<u>Lili Zheng</u>	3H	25
<u>Zhonghua Liu</u>	1S	13	<u>Qing Zhou</u>	6C	50
<u>Junwei Lu</u>	4F	33	<u>Wen Zhou</u>	6H	55
<u>Robert Lund</u>	6S	56	<u>Ying Zhou</u>	3S	30
<u>Jinchi Lv</u>	6C	50	<u>Yuchen Zhou</u>	1S	12
<u>Cong Ma</u>	5H	43	<u>Hongtu Zhu</u>	6C	48
<u>Rong Ma</u>	5H	43	<u>Ji Zhu</u>	2C	16
<u>Shujie Ma</u>	4H	35	<u>Ruoqing Zhu</u>	1F	9
<u>Marianthi Markatou</u>	4H	37	<u>Yuhua Zhu</u>	3F	23
<u>Rui Miao</u>	6F	52	<u>Jingjing Zou</u>	1H	9