

Workshop Program

	Friday, May 26	Saturday, May 27
7:30-8:30	Registration, Breakfast Provost Patrick Wolfe - Open Remark	Breakfast
Chair	Jen Tang	Alex Wang
8:30-9:30	Zhiliang Ying, <i>in Memory of Professor Tze Leung Lai</i> <i>Encounters with Martingales in Statistics and Stochastic Optimization</i>	Stephen Wright <i>Optimization in Data Science</i>
9:30-10:00	Wei Biao Wu <i>Concentration bounds for statistical learning for time-dependent data</i>	Michael Mahoney <i>Multiplicative noise and heavy tails in stochastic optimization and machine learning</i>
10:00-10:30	Jinchi Lv <i>Optimal Nonparametric Inference with Two-Scale Distributional Nearest Neighbors</i>	Mert Gürbüzbalaban <i>Heavy-tail Phenomenon in SGD</i>
10:30-11:00	Coffee Break	Coffee Break
Chair	Weibin Mo	Thanh Nguyen
11:00-11:30	Huixia Judy Wang <i>Estimation for Quantile Spatially Varying Coefficient Models over Complicated Domains</i>	Peter Frazier <i>Grey-box Bayesian Optimization</i>
11:30-12:00	Wen-xin Zhou <i>Some Recent Developments in Expected Shortfall Regression</i>	Rahul Mazumder <i>Computational Lenses for learning under structural constraints: sparsity, decision trees, and conditional computing</i>
12:00-12:30	Hamsa Bastani <i>Decision-Aware Learning for Global Health Supply Chains</i>	Kaizheng Wang <i>Pseudo-Labeling for Kernel Ridge Regression under Covariate Shift</i>
12:30-2:00	Lunch on your own (not provided)	Lunch Panel - Mohit Tawarmalani Peter Frazier, Steve Hanneke, Faming Liang, George Shanthikumar, Paul Valiant, Stephen Wright
Chair	Wei Sun	Yanjun Li

	Friday, May 26	Saturday, May 27
2:00-2:30	Lan Wang <i>Fairness-oriented Learning for Optimal Individualized Treatment Rules</i>	Alexandre Belloni <i>Neighborhood Adaptive Estimators for Causal Inference under Network Interference</i>
2:30-3:00	George Lan <i>First Order Policy Optimization for Robust Markov Decision Process</i>	Roberto Oliveira <i>Trimmed sample means for robust uniform mean estimation and regression</i>
3:00-3:30	Yuxin Chen <i>Towards Minimax-Optimal Multi-Agent RL in Markov Games</i>	Gesualdo Scutari <i>Some Reflections on Decentralized Optimization for High-Dimensional Statistical Inference</i>
3:30-4:00	Coffee Break	Hari Sundaram <i>Learning tails from heads: Item prototype representations for few-shot recommendations</i>
Chair	Tongseok Lim	Poster Awards - Yanjun Li Closing Remark - Mohit Tawarmalani
4:00-4:30	Moulinath Banerjee <i>Tackling Posterior Drift Via Linear Adjustments and Exponential Tilts</i>	
4:30-5:00	Jelena Diakonikolas <i>Advances in Cyclic Block Coordinate Methods: Gradient Extrapolation, Acceleration, and Variance Reduction</i>	
5:00-5:30	Ilias Diakonikolas <i>Non-Gaussian Component Analysis and its Applications</i>	
5:30-6:00	Break - Walk to Purdue Memorial Union	
6:00-7:00	Poster Session Faculty Lounges, Purdue Memorial Union	
7:00-9:00	Banquet - Dean David Hummels <i>Separate registration required</i>	

Regrettably, we received the heartbreaking news that Professor Tze Leung Lai, an esteemed plenary speaker of our workshop from Stanford Statistics, passed away unexpectedly on Sunday, May 21, 2023. Our deepest condolences go out to Lai's family during this difficult time. In light of this unfortunate incident, we are grateful to have Professor Zhiliang Ying from Columbia University to deliver a plenary talk during our workshop on May 26. Professor Ying's talk will serve as a heartfelt tribute to honor the memory of Professor Lai, allowing us to remember his remarkable contributions.

Welcome

We are pleased to welcome you to the *Statistics and Optimization in Data Science Workshop* hosted by Purdue University's Daniels School of Business. This is a joint effort of the school's Quantitative Methods Area and the Krenicki Center for Business Analytics and Machine Learning. We believe this workshop will interest a broad community of people interested in data science and its applications. It will consist of speakers from business and decision sciences, computer science, operations research, optimization, and statistics.

It is exciting to announce the final program which features 22 outstanding speakers, including 2 keynote lectures. This will be only possible with your participation. The workshop starts on May 26 with an opening remark from Purdue's Provost, Prof. Patrick J. Wolfe. The first day ends with a poster session followed by a foreword by Prof. David Hummels and a dinner. There will be also a lunch panel on May 27 featuring discussions and questions.

Regrettably, we received the heartbreaking news that Professor Tze Leung Lai, an esteemed plenary speaker of our workshop from Stanford Statistics, passed away unexpectedly on Sunday, May 21, 2023. Our deepest condolences go out to Lai's family during this difficult time. In light of this unfortunate incident, we are grateful to have Professor Zhiliang Ying from Columbia University to deliver a plenary talk during our workshop on May 26. Professor Ying's talk will serve as a heartfelt tribute to honor the memory of Professor Lai, allowing us to remember his remarkable contributions.

We are truly grateful for your willingness to participate in this conference and enrich all in attendance with your thoughts and wisdom.

Enjoy the conference!

Location

Workshop is at Rawls Hall (100 S Grant St, West Lafayette, IN 47907) Room 1086.

Poster Session is at Purdue Memorial Union 2nd Floor Faculty Lounges.

From Hotel to Rawls Hall - Use the main entrance/exit of the hotel, turn right, and walk 1 minute south to the State St and Grant St intersection. The Rawls Hall is on the southeast corner.



Registration covers breakfast on both days and lunch on May 27 only – Please show your badges.

Breakfast starts 7:30 am. The regular sessions start at 8:15 am. Lunch on May 26 is on your own. There are a great variety of [dining options](#) in the area for lunch.

Talk title and abstract

May 26, 2023

Zhiliang Ying, in memory of **Tze Leung Lai**

Encounters with Martingales in Statistics and Stochastic Optimization

Ying's talk will serve as a heartfelt tribute to honor the memory of Professor Lai, allowing us to remember his remarkable contributions. Below is the original talk abstract by Lai.

In this intellectual memoir, Professor Lai describes how martingales came into his world of mathematical statistics, first in sequential tests and confidence intervals, then in time series, stochastic approximation, sequential design of experiments, and stochastic optimization. He sketches the trajectories of many other statisticians that he met along the way. He emphasizes the roles of Harold Hotelling, Abraham Wald, and Herbert Robbins in their creation of the environment for the study of martingales at Columbia University and then his own subsequent work at Stanford University. At Stanford, he came to see stochastic optimization as a unifying theme for the use of martingales in statistical modeling. In conclusion, he describes the multifaceted applications of martingales to statistics and stochastic optimization in the BigData and MultiCloud era.

Wei Biao Wu

Concentration bounds for statistical learning for time dependent data

Classical statistical learning theory primarily concerns independent data. In comparison, it has been much less investigated for time dependent data, which are commonly encountered in economics, engineering, finance, geography, physics and other fields. In this talk, we focus on concentration inequalities for suprema of empirical processes which plays a fundamental role in the statistical learning theory. We derive a Gaussian approximation and an upper bound for the tail probability of the suprema under conditions on the size of the function class, the sample size, temporal dependence and the moment conditions of the underlying time series. Due to the dependence and heavy-tailedness, our tail probability bound is substantially different from those classical exponential bounds obtained under the independence assumption in that it involves an extra polynomial decaying term. We allow both short- and long-range dependent processes, where the long-range dependence case has never been previously explored. We showed our tail probability inequality is sharp up to a multiplicative constant. These bounds work as theoretical guarantees for statistical learning applications under dependence.

Jinchi Lv

Optimal Nonparametric Inference with Two-Scale Distributional Nearest Neighbors

The weighted nearest neighbors (WNN) estimator has been popularly used as a flexible and easy-to-implement nonparametric tool for mean regression estimation. The bagging technique is an elegant way to form WNN estimators with weights automatically generated to the nearest neighbors; we name the resulting estimator as the distributional nearest neighbors (DNN) for easy reference. Yet, there is a lack of distributional results for such estimator, limiting its application to statistical inference. Moreover, when the mean regression function has higher-order smoothness, DNN does not achieve the optimal nonparametric convergence rate, mainly because of the bias issue. In this work, we provide an in-depth technical analysis of the DNN, based on which we suggest a bias reduction approach for the DNN estimator by linearly combining two DNN estimators with different subsampling scales, resulting in the novel two-scale DNN (TDNN) estimator. The two-scale DNN estimator has an equivalent representation of WNN with weights admitting explicit forms and some being negative. We prove that, thanks to the use of negative weights, the two-scale DNN estimator enjoys the optimal nonparametric rate of convergence in estimating the regression function under the fourth-order smoothness condition. We further go beyond estimation and establish that the DNN and two-scale DNN are both asymptotically normal as the subsampling scales and sample size diverge to infinity. For the practical implementation, we also provide variance estimators and a distribution estimator using the jackknife and bootstrap techniques for the two-scale DNN. These estimators can be exploited for constructing valid confidence intervals for nonparametric inference of the regression function. The theoretical results and appealing finite-sample performance of the suggested two-scale DNN method are illustrated with several simulation examples and a real data application. This is a joint work with Emre Demirkaya, Yingying Fan, Lan Gao, Patrick Vossler and Jingbo Wang.

Huixia Judy Wang

Estimation for Quantile Spatially Varying Coefficient Models over Complicated Domains

Regression analysis is a widely used tool for analyzing spatial data. In this talk, I will introduce a flexible quantile spatially varying coefficient model to assess how conditional quantiles of the response depend on covariates, allowing the coefficient function to vary with spatial location. This model can be used to explore spatial non-stationarity of regression relationships for heterogeneous spatial data distributed over a domain of complex or irregular shape. To estimate the model, we propose a quantile regression method that adopts the bivariate penalized spline technique to approximate the unknown functional coefficients. I will also discuss the L_2 convergence of the proposed estimator with an optimal convergence rate under some regularity conditions. Additionally, I will present an efficient algorithm based on the alternating direction method of multipliers to solve the optimization problem. Finally, I will illustrate the practical applicability of our method through simulation studies and an analysis of mortality data in the United States. This is joint work with Myungjin Kim from Kyungpook National University and Lily Wang from George Mason University.

Wen-xin Zhou

Some Recent Developments in Expected Shortfall Regression

Expected Shortfall (ES), also known as superquantile or Conditional Value-at-Risk, has been recognized as an important measure in risk analysis and stochastic optimization. In finance, it refers to the conditional expected return of an asset given that the return is below some quantile of its distribution. In this work, we consider a joint regression framework that simultaneously models the conditional quantile and ES of a response variable given a set of covariates, for which the state-of-the-art approach is based on minimizing a joint loss function that is non-differentiable and non-convex. Motivated by the idea of using Neyman-orthogonal scores to reduce sensitivity with respect to nuisance parameters, we propose a statistically robust and computationally efficient two-step procedure for fitting joint quantile and ES regression models. With increasing covariate dimensions, we establish explicit non-asymptotic bounds on estimation and Gaussian approximation errors, which lay the foundation for statistical inference. Under high-dimensional sparse models, we consider a lasso-penalized two-step approach and its robust counterpart, and propose debiased estimators for inference. We further discuss a more general weighted-average quantile regression framework, including ES regression as a special case. This talk is based on joint works with Xuming He, Kean Ming Tan and Shushu Zhang.

Hamsa Bastani

Decision-Aware Learning for Global Health Supply Chains

The combination of machine learning (for prediction) and optimization (for decision-making) is increasingly used in practice. However, a key challenge is the need to align the loss function used to train the machine learning model with the decision loss associated with the downstream optimization problem. Traditional solutions have limited flexibility in the model architecture and/or scale poorly to large datasets. We propose a principled decision-aware learning algorithm that uses a novel Taylor expansion of the optimal decision loss to derive the machine learning loss. Importantly, our approach only requires a simple re-weighting of the training data, allowing it to flexibly and scalably be incorporated into complex modern data science pipelines, yet producing sizable efficiency gains. We apply our framework to optimize the distribution of essential medicines in collaboration with policymakers at the Sierra Leone National Medical Supplies Agency; highly uncertain demand and limited budgets currently result in excessive unmet demand. We leverage random forests with meta-learning to learn complex cross-correlations across facilities, and apply our decision-aware learning approach to align the prediction loss with the objective of minimizing unmet demand. Out-of-sample results demonstrate that our end-to-end approach significantly reduces unmet demand across 1000+ health facilities throughout Sierra Leone.

Lan Wang

Fairness-oriented Learning for Optimal Individualized Treatment Rules

A notable drawback of the standard approach to optimal individualized treatment rule (ITR) estimation is that the estimated optimal ITR may be suboptimal or even detrimental to certain disadvantaged subpopulations. Motivated by the importance of incorporating an appropriate fairness constraint in optimal decision-making (e.g., assign treatment with protection to those with shorter survival time), we propose a

new framework that aims to estimate an optimal ITR to maximize the average value with the guarantee that its tail performance exceeds a prespecified threshold. The optimal fairness-oriented ITR corresponds to a solution to a nonconvex optimization problem. Furthermore, we extend the proposed method to dynamic optimal ITRs. (Joint work by Ethan Fang and Zhaoran Wang)

George Lan

First-order Policy Optimization for Robust Markov Decision Process

We consider the problem of solving robust Markov decision process (MDP), which involves a set of discounted, finite state, finite action space MDPs with uncertain transition kernels. The goal is to find a robust policy that optimizes the worst-case values against the transition uncertainties, and thus encompasses the standard MDP planning as a special case. For (s,a)-rectangular uncertainty sets, we develop a policy-based first-order method, namely the robust policy mirror descent (RPMD), and establish its iteration complexity for finding a nearly optimal policy. We further develop a stochastic variant of the robust policy mirror descent method, named SRPMD, when the first-order information is only available through online interactions with the nominal environment. We establish the best known so far sample complexity for solving RMDP problems.

Yuxin Chen

Towards Minimax-Optimal Multi-Agent RL in Markov Games

Abstract: This talk is concerned with multi-agent reinforcement learning in Markov games, with the goal of learning Nash equilibria or coarse correlated equilibria (CCE) sample-optimally. All prior results suffer from at least one of the two obstacles: the curse of multiple agents and the barrier of long horizon, regardless of the sampling protocol in use. We take a step towards settling this problem, assuming access to a flexible sampling mechanism: the simulator. Focusing on non-stationary finite-horizon Markov games, we develop a fast learning algorithm and an adaptive sampling scheme that leverage the optimism principle in online adversarial learning (particularly the Follow-the-Regularized-Leader (FTRL) method). Our algorithm achieves minimax-optimal sample complexity when the number of players is fixed. Our results emphasize the prolific interplay between high-dimensional statistics, online learning, and game theory. This work is based on joint work with Gen Li, Yuejie Chi and Yuting Wei; more details can be found in <https://arxiv.org/abs/2208.10458>.

Mouline Banerjee

Tackling Posterior Drift Via Linear Adjustments and Exponential Tilts

I will speak on some of our recent work on transfer learning from a source to a target population in the presence of ‘posterior drift’: i.e. the regression function/Bayes classifier in the target population is different

from that in the source. In the situation where labeled samples from the target domain are available, by modeling the posterior drift through a linear adjustment (on an appropriately transformed scale), we are able to learn the nature of the posterior drift using relatively few samples from the target population as compared to the source population, which provides an abundance of samples. The other (semi-supervised) case, where labels from the target are unavailable, is addressed by connecting the probability distribution in the target domain to that in the source domain via an exponential family formulation, and learning the corresponding parameters. Both approaches are motivated by ideas originating in classical statistics. I will present theoretical guarantees for these procedures as well as applications to real data from the UK Biobank study (mortality prediction) and the Waterbirds dataset (image classification).

Jelena Diakonikolas

Advances in Cyclic Block Coordinate Methods: Gradient Extrapolation, Acceleration, and Variance Reduction

Abstract: In this talk, I will present new results advancing the theory of cyclic block coordinate methods, enabled by a novel Lipschitz condition w.r.t. a Mahalanobis norm. The newly introduced Lipschitz condition is motivated by addressing block coordinate updates in variational inequality (equilibrium) problems, where traditional block Lipschitz conditions appear insufficient. I will first present results that address generalized variational inequality problems using gradient extrapolation and cyclic block coordinate updates. I will then discuss further developments of this theory, including variance reduction, gradient norm guarantees in smooth nonconvex minimization settings, and acceleration in composite (smooth + nonsmooth) convex settings.

Ilias Diakonikolas

Non-Gaussian Component Analysis and its Applications

Abstract: Non-Gaussian component analysis (NGCA) is the following statistical task: Given sample access to a high-dimensional distribution that is non-Gaussian in a hidden direction and a standard multivariate Gaussian in the orthogonal complement, approximate the hidden direction. In a 2017 paper with Kane and Stewart, we established tight lower bounds for this problem in the Statistical Query (SQ) model and derived similar implications for several, seemingly unrelated, statistical tasks. Since then, a number of works have leveraged this framework to obtain tight computational lower bounds for a range of learning problems. In this work, we will survey these developments and discuss opportunities for future work.

May 27, 2023

Stephen Wright

Optimization in Data Science

Optimization has always played a key role in solving problems in data science, and the engagement between these areas continues to grow. In this talk, we discuss how formulation and algorithmic tools from optimization have been used to address problems in computational statistics, machine learning, and AI, starting over 200 years ago with least squares and continuing today with neural networks. We discuss the nature of current research at the interface of optimization and data science, which has both theoretical aspects (concerning the convergence properties of algorithms, the presence of "benign nonconvexity," the effectiveness of the optimization formulation in solving the underlying statistical problem, and other issues) and practical aspects (for example, the choice of algorithms and algorithmic parameters for matrix optimization problems and neural network training). Finally, we briefly survey some areas of ongoing research and open issues.

Michael Mahoney

Multiplicative noise and heavy tails in stochastic optimization and machine learning

Stochastic optimization is widely-used throughout machine learning; and the theory that guides its use and development comes mostly from convex optimization. For example, stochastic gradient descent and its variants are most often viewed within the theory as trying to approximate deterministic gradient descent, and one tries to establish certain types of convergence that parallel what is obtained in convex situations. In many parts of modern machine learning, perhaps most notably in the training of neural networks, this theory often does not provide even a qualitative guide to practice. Here, we view stochastic optimization as a Markov chain via the lens of iterated random functions. From this perspective, heavy tails arise naturally. We describe the widespread phenomena of heavy tails in stochastic optimization and machine learning, we describe one mechanism by which they arise, and we outline their implications more generally.

Mert Gurbuzbalaban

Heavy-tail Phenomenon in SGD

In the first part of the talk, we focus on the gradient noise (GN) in the stochastic gradient descent (SGD) algorithm for training deep neural networks (DNNs). In this setting, GN is often considered to be Gaussian in the large data regime by assuming that the classical central limit theorem (CLT) kicks in. This assumption is often made for mathematical convenience, since it enables SGD to be analyzed as a stochastic differential equation (SDE) driven by a Brownian motion. We argue that the Gaussianity assumption might fail to hold in deep learning settings and hence render the Brownian motion-based

analyses inappropriate. We conduct extensive experiments on common deep learning architectures and show that in all settings, the GN is highly non-Gaussian and admits heavy-tails. We further investigate the tail behavior in varying network architectures and sizes, loss functions, and datasets. By exploiting connections to Levy-driven differential equations and their metastability properties, our results open up a different perspective and shed more light on the belief that SGD prefers wide minima. Joint work with Umut Simsekli and Levent Sagun.

In the second part of the talk, we question the origins of heavy tails in SGD iterations further and their link to various notions of capacity and complexity that have been proposed for characterizing the generalization properties of SGD in deep learning. Some of the popular notions that correlate well with the performance on unseen data are (i) the ‘flatness’ of the local minimum found by SGD, which is related to the eigenvalues of the Hessian, (ii) the ratio of the stepsize η to the batch size b , which essentially controls the magnitude of the stochastic gradient noise, and (iii) the ‘tail-index’, which measures the heaviness of the tails of the eigenspectra of the network weights. We argue that these three seemingly unrelated perspectives for generalization are deeply linked to each other. We claim that depending on the structure of the Hessian of the loss at the minimum, and the choices of the algorithm parameters η and b , the SGD iterates will converge to a heavy-tailed stationary distribution. We rigorously prove this claim in the setting of linear regression: we show that even in a simple quadratic optimization problem with independent and identically distributed Gaussian data, the iterates can be heavy-tailed with infinite variance. We further characterize the behavior of the tails with respect to algorithm parameters, the dimension, and the curvature. We then translate our results into insights about the behavior of SGD in deep learning. We finally support our theory with experiments conducted on both synthetic data and neural networks. Joint work with Umut Simsekli and Lingjiong Zhu.

In the third part of the talk, if time permits, we will talk about a new class of initialization schemes for fully-connected neural networks that can improve the performance of SGD training by inducing a particular form of heavy-tailed behavior in the stochastic gradients. Joint work with Yuanhan Hu.

Peter Frazier

Grey-box Bayesian Optimization

Bayesian optimization (BayesOpt) is a powerful tool for optimizing time-consuming-to-evaluate non-convex noisy derivative-free objective functions. It combines a machine-learning-based surrogate for the objective function with a decision-theoretic acquisition function that quantifies the value of objective function evaluations. It is widely in industry for optimizing social media platforms, tuning hyperparameters in deep neural networks, designing new drugs and materials, and beyond. While BayesOpt has historically been deployed as a black-box optimizer, recent advances show orders-of-magnitude improvement is possible by "peeking inside the box". For example, when tuning hyperparameters in deep neural networks to minimize validation error, state-of-the-art BayesOpt tuning methods leverage the ability to stop training early, restart previously paused training, perform training and testing on a strict subset of the available data, and warm-start from previously tuned network architectures. We describe new "grey box" Bayesian optimization methods that selectively exploit problem structure to deliver state-of-the-art performance. We then briefly applications of these methods to tuning deep neural

networks, inverse reinforcement learning, calibrating physics-based simulators to observational data, and molecular design.

Rahul Mazumder

Computational Lenses for learning under structural constraints: sparsity, decision trees, and conditional computing

Optimization problems involving discrete or combinatorial structure frequently arise in statistics and machine learning: best subset selection, classification, robust statistics, decision tree ensembles, among others. Combinatorial modules also occur within large neural networks (e.g., sparse mixture of experts and compression of large models). I will discuss how such combinatorial problems have inspired new recent developments in mathematical optimization and how these algorithms can inform our computational and statistical understanding of these problems. For example, I will discuss how tools in convex and discrete optimization can be used to solve, to global optimality, instances of the best subset selection problem with one million features and thousands of samples. Time permitting, I will also discuss approaches for learning decision trees and shuffled regression.

Kaizheng Wang

Pseudo-Labeling for Kernel Ridge Regression under Covariate Shift

We develop and analyze a principled approach to kernel ridge regression under covariate shift. The goal is to learn a regression function with small mean squared error over a target distribution, based on unlabeled data from there and labeled data that may have a different feature distribution. We propose to split the labeled data into two subsets and conduct kernel ridge regression on them separately to obtain a collection of candidate models and an imputation model. We use the latter to fill the missing labels and then select the best candidate model accordingly. Our non-asymptotic excess risk bounds show that in quite general scenarios, our estimator adapts to the structure of the target distribution and the covariate shift. It achieves the minimax optimal error rate up to a logarithmic factor. The use of pseudo-labels in model selection does not have major negative impacts.

Alexandre Belloni

Neighborhood Adaptive Estimators for Causal Inference under Network Interference

Authors: Alexandre Belloni, Fei Fang, Alexander Volfovsky

Abstract: Estimating causal effects has become an integral part of most applied fields. Solving these modern causal questions requires tackling violations of many classical causal assumptions. In this work we consider the violation of the classical no-interference assumption, meaning that the treatment of one

individuals might affect the outcomes of another. To make interference tractable, we consider a known network that describes how interference may travel. However, unlike previous work in this area, the radius (and intensity) of the interference experienced by a unit is unknown and can depend on different sub-networks of those treated and untreated that are connected to this unit. We study estimators for the average direct treatment effect on the treated in such a setting. The proposed estimator builds upon a Lepski-like procedure that searches over the possible relevant radii and treatment assignment patterns. In contrast to previous work, the proposed procedure aims to approximate the relevant network interference patterns. We establish oracle inequalities and corresponding adaptive rates for the estimation of the interference function. We leverage such estimates to propose and analyze two estimators for the average direct treatment effect on the treated. We address several challenges stemming from the data-driven creation of the patterns (i.e. feature engineering) and the network dependence. In addition to rates of convergence, under mild regularity conditions, we show that one of the proposed estimators is asymptotically normal and unbiased.

Roberto I. Oliveira

Trimmed sample means for robust uniform mean estimation and regression

Modern results on High-Dimensional Robust Statistics require a combination of algorithmic and mathematical insights. The most fundamental mathematical question in this area is this: how well can one estimate the expectations of a family of functions from a random sample with a heavy-tailed distribution and possibly containing outliers? This talk presents the best known results on this problem. In a nutshell, we will see that simple sample trimming leads to near-optimal results. We also discuss some statistical methods for linear regression that are based on this idea. This is joint work with Lucas Resende, a PhD student at IMPA; see arXiv:2302.06710 for details.

Gesualdo Scutari

Some Reflections on Decentralized Optimization for High-Dimensional Statistical Inference

Abstract: There is growing interest in solving large-scale inference problems over decentralized networks, where data are distributed across nodes and no centralized coordination exists (we termed these systems as “mesh” networks). While statistical-computation guarantees have been largely explored in the centralized setting, our understanding over mesh networks is limited. In particular, decentralized schemes that work well in the classical low-dimensional regime may break down in high-dimension, and existing convergence studies may fail to predict their behaviors. This is because the majority of distributed algorithms have been designed and studied only from an optimization perspective, without considering the statistical dimension. This talk will discuss some vignettes from high-dimensional statistical inference suggesting new analyses aimed at incorporating statistical thinking into decentralized optimization.

Hari Sundaram

Developing Recommender systems for heavy-tailed data distributions

In this talk, we'll examine recent work at Illinois on building sophisticated recommender systems for heavy-tailed data distributions. Heavy-tailed distributions—a small fraction of the population accounts for most of the behavioral data—are an essential feature of online networks. We observe a paradox when training neural models on this data: while the overall item recommendation or action identification accuracy is high, accuracy levels are poor for a significant chunk of the target audience. Since the data are heavy-tailed, the neural models are biased towards the behavioral patterns of a small fraction of individuals responsible for most of the activity. The neural models perform well at predicting the probability of item purchase (since a small fraction of individuals produces most of the activity), but not at predicting how most individuals would behave. In this talk, I'll cover some of our technical contributions in addressing these challenges.

Poster Session

Riddhiman Bhattacharya, *Purdue University*, General Framework for Threshold Estimation Using Bandits

Biao Cai, *Yale University*, Jointly Modeling and Clustering Tensors in High Dimensions

Qiyu Han, *Purdue University*, Online Statistical Inference for Matrix Contextual Bandit

Yue Jiang, *Purdue University*, An Adaptive Sampling Strategy for Real-time Anomaly Detection with Unmanned Sensing Vehicles

Yuantong Li, *University of California, Los Angeles*, Two-sided Matching Markets with Implicit, Dynamic, and Complementary Preferences

Pangpang Liu, *Purdue University*, Contextual Dynamic Pricing with Strategic Buyers

Jorge Loría, *Purdue University*, SURE-tuned Bridge Regression

Subha Maity, *University of Michigan*, Understanding New Tasks Through the Lens of Training Data via Exponential Tilting

Xinran Miao, *University of California, Los Angeles*, Efficient Estimation for the Transportability Index Using Neural Networks

Haoyu Song, *Purdue University*, Learning Non-parametric Choice Models with Discrete Fourier Analysis

Danyang Wang, *Purdue University*, Offline Causal Reinforcement Learning for Confounded Markov Decision Process

Zhanyu Wang, *Purdue University*, Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies

Shirong Xu, *University of California, Los Angeles*, An Analytic Framework for Synthetic Data in Supervised Learning