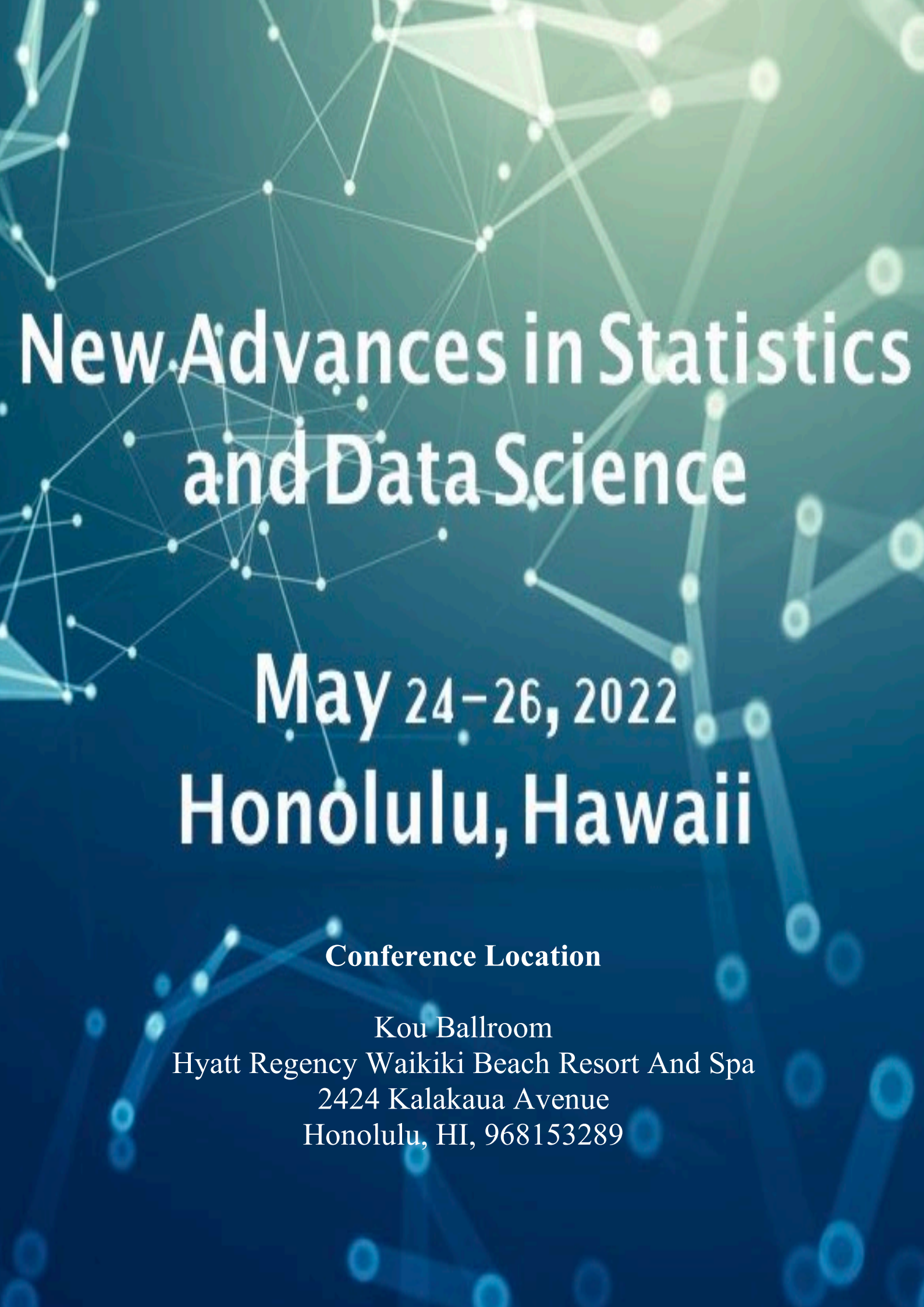


Conference Location

Kou Ballroom
Hyatt Regency Waikiki Beach Resort And Spa
2424 Kalakaua Avenue
Honolulu, HI, 968153289

Conference Program

Day 1 (May 24): New Advances in Statistical Inference

Time	People	Titles /Activities
8:30-8:40	Opening remarks	
Session 1	Chair: Weijie Su	
8:40-9:05	Yang Feng	Random Subspace Ensemble
9:05-9:30	Arian Maleki	Asymptotic analysis of SLOPE
9:30-9:55	Philip Ernst	New frontiers in statistical inference for stochastic processes
9:55-10:20	Jiashun Jin	Counting Cycles in Networks
10:20-10:40	Coffee Break	
Session 2	Chair: Jiashun Jin	
10:40-11:05	Cun-Hui Zhang	Tensor PCA in High Dimensional CP Models
11:05-11:30	Heping Zhang	Tensor Quantile Regression for Neuroimage Study of Human Intelligence
11:30-11:55	Yuting Wei	Minimum L1-norm interpolators: Precise asymptotics and multiple descent
11:55-13:40	Lunch Break	
Session 3	Chair: Peter Song	
13:40-14:05	Jinchi Lv	High-Dimensional Knockoffs Inference for Time Series Data
14:05-14:30	Debashis Paul	Estimation of spectra of high-dimensional separable covariance matrices
14:30-14:55	Weichen Wang	Volatility prediction comparison via robust volatility proxies: An empirical deviation perspective
14:55-15:20	Chunming Zhang	Maximum independent component analysis with application to non-linear temporal signals
15:20-15:40	Coffee Break	
Session 4	Chair: Weichen Wang	
15:40-16:05	Wanjie Wang	Covariate-Associated Community Detection on Social Networks

16:05-16:30	Peter Song	Real-time regression analysis of streaming clustered data with possible abnormal data batches
16:30-16:55	Tracy Ke	Estimating trajectories of statisticians in co-citation networks
16:55-17:20	Weijie Su	When Will You Become the Best Reviewer of Your Own Papers? An Owner-Assisted Approach to Mechanism Design
17:30	**Reception**	

Day 2 (May 25): New Advances in Biostatistics, Health and Causal Inference

Time	People	Titles /Activities
Session 1	**Chair: Lucas Janson**	
8:40-9:05	Mladen Kolar	Confidence sets for Causal Discovery
9:05-9:30	Jelena Bradic	Dynamic Causal Learning: excursions in double robustness
9:30-9:55	Rajarshi Mukherjee	A new central limit theorem for the augmented IPW estimator: variance inflation, cross-fit covariance and beyond
9:55-10:20	David Choi	Causal inference in experiments with interference
10:20-10:40	**Coffee Break**	
Session 2	**Chair: Mladen Kolar**	
10:40-11:05	Hongyu Zhao	Estimating cell-type-specific gene co-expression networks from bulk gene expression data with an application to Alzheimer's disease
11:05-11:30	Genevera Allen	Large-Scale Graph Learning with Latent Variables
11:30-11:55	Lucas Janson	Controlled Discovery and Localization of Signals via Bayesian Linear Programming (BLiP)
11:55-13:40	**Lunch Break**	
Session 3	**Chair: Boxiang Wang**	
13:40-14:05	Zongming Ma	Matching of datasets and its applications in single-cell biology
14:05-14:30	Eric Laber	Safe Reinforcement Learning in mHealth
14:30-14:55	Hongzhe Li	Estimation and Inference with Proxy Data and its Genetic Applications
14:55-15:20	Xiufan Yu	Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing

15:20-15:40	Coffee Break	
Session 4	Chair: Xiufan Yu	
15:40-16:05	Gang Li	Efficient Algorithms and Implementation of a Semiparametric Joint Model for Longitudinal and Competing Risks Data: With Applications to Massive Biobank Data
16:05-16:30	Haoyu Zhang	Novel Methods for Multi-ancestry Polygenic Prediction and their Evaluations in 3.7 Million Individuals of Diverse Ancestry
16:30-17:20	Poster Session	

Day 3 (May 26): New Advances in Machine Learning

Time	People	Titles /Activities
Session 1	Chair: Tracy Ke	
8:40-9:05	Yingying Fan	Asymptotic properties of high-dimensional random forests
9:05-9:30	Edgar Dobriban	T-Cal: An optimal test for the calibration of predictive models
9:30-9:55	Jun S. Liu	Statistics Meet Neural Networks: Bootstrap, Cross-Validations, and Beyond
9:55-10:20	Jianqing Fan	How do noise tails impact on Deep ReLU Networks?
10:20-10:40	Coffee Break	
Session 2	Chair: Jason Lee	
10:40-11:05	Tony Cai	Transfer Learning: Optimality and Adaptive Algorithms
11:05-11:30	Patrick Rubin-Delanchy	Manifold structure in graph embeddings
11:30-11:55	Tengyu Ma	Understanding self-supervised learning
11:55-13:40	Lunch Break	
Session 3	Chair: Qi Lei	
13:40-14:05	Jason Lee	Offline Reinforcement Learning with only Realizability
14:05-14:30	Yuejie Chi	Offline Reinforcement Learning: Towards Optimal Sample Complexities
14:30-14:55	Adel Javanmard	The curse of overparametrization in adversarial training
14:55-15:20	Simon Du	When is Offline Two-Player Zero-Sum Markov Game Solvable?

15:20-15:40	**Coffee Break**	
Session 4	**Chair: Yujie Chi**	
15:40-16:05	Quan Zhou	Informed MCMC sampling for high-dimensional model selection problems
16:05-16:30	Krishna Balasubramanian	Towards a Theory of Non-Log-Concave Sampling
16:30-16:55	Boxiang Wang	Sparse Convoluted Rank Regression in High Dimensions
16:55-17:20	Yuqi Gu	Blessing of Latent Dependence and Identifiable Deep Modeling of Discrete Latent Variables

Abstract

Day 1 (May 24): New Advances in Statistical Inference

1. **Random Subspace Ensemble**
   Yang Feng, New York University
   Abstract: We propose a flexible ensemble framework, Random Subspace Ensemble (RaSE). In the RaSE algorithm, we aggregate many weak learners, where each weak learner is trained in a subspace optimally selected from a collection of random subspaces using a base method. In addition, we show that in a high-dimensional framework, the number of random subspaces needs to be very large to guarantee that a subspace covering signals is selected. Therefore, we propose an iterative version of the RaSE algorithm and prove that under some specific conditions, a smaller number of generated random subspaces are needed to find a desirable subspace through iteration. We study the RaSE framework for classification where a general upper bound for the misclassification rate was derived, and for screening where the sure screening property was established. An extension called Super RaSE was proposed to allow the algorithm to select the optimal pair of base method and subspace during the ensemble process. The RaSE framework is implemented in the R package RaSEn on CRAN.

2. **Asymptotic analysis of SLOPE**
   Arian Maleki, Columbia University
   Abstract: SLOPE estimator has been shown to adaptively achieve the minimax estimation rate under high-dimensional sparse linear regression models. Such minimax optimality holds in the regime where the sparsity level k, sample size n, and dimension p, satisfy k/p->0, klogp/n->0. In this paper, we characterize the estimation error of SLOPE under the complementary regime where both k and n scale linearly with p , and provide new insights into the performance of SLOPE estimators. We first derive a concentration inequality for the finite sample mean square error (MSE) of SLOPE. The quantity that MSE concentrates around takes a complicated and implicit form. With delicate analysis of the quantity, we prove that among all SLOPE estimators, LASSO is optimal for estimating k-sparse parameter vectors that do not have tied nonzero components in the low noise scenario. On the other hand, in the large noise scenario, the family of SLOPE estimators are sub-optimal compared with bridge regression such as the Ridge estimator.

3. **New frontiers in statistical inference for stochastic processes**
   Philip Ernst, Rice University
   Abstract: I will first briefly review “Yule’s `nonsense correlation’ solved!” (The Annals of Statistics, 2017). Consider the standard empirical correlation ρ_n, which is defined for two related series of data of length n using the standard Pearson correlation statistic. This empirical correlation is known as Yule's "nonsense correlation" in honor of the British statistician G. Udny Yule, who in 1926 described the phenomenon by which empirical correlation fails to gauge independence of data series for random walks and for other time series. For the case of two independent and identically distributed random walks independent from each other, Yule empirically observed that the distribution of the empirical correlation is not concentrated around 0; rather, it is “volatile” in the sense that its distribution is heavily dispersed and is frequently large in absolute value. This well-documented effect was ignored by many scientists over the decades, up to the present day, even sparking recent controversies in climate-change attribution. Since the 1960s, some probabilists have wanted to eliminate any possible ambiguity about the issue by computing the variance of the continuous-time version ρ of Yule's nonsense correlation, based on the paths

of two independent Wiener processes. The problem would remain open for over ninety years until we finally closed it in our 2017 paper.

I will then turn to speaking about our subsequent success in explicitly calculating all moments of ρ for two independent Brownian motions. Our solution leads to the first approximation to the density of Yule's nonsense correlation. We are also able to explicitly compute higher moments of Yule's nonsense correlation when the two independent Wiener processes are replaced by two correlated Wiener processes, two independent Ornstein-Uhlenbeck processes, and two independent Brownian bridges. We then consider extending the definition of ρ to the time interval $[0,T]$ for any $T>0$ and prove a Central Limit Theorem for the case of two independent Ornstein-Uhlenbeck processes. All of these aforementioned results appear in our preprint entitled "The distribution of Yule's nonsense correlation" (https://arxiv.org/pdf/1909.02546.pdf).

Finally, I will then discuss present work in building asymptotically exact and powerful tests of independence for pairs of independent Ornstein-Uhlenbeck processes and for other stationary Gaussian processes. Many of the methods of proof are drawn from Wiener chaos analysis, a simplified way of implementing the so-called Malliavin calculus for random variables depending in a polynomial way on finite or infinite dimensional Gaussian vectors such as Wiener processes. Time permitting, I also hope to speak about some initial leads in building tests of independence for pairs of nonstationary processes, in particular processes with long memory such as the so-called fractional Brownian motion. I will conclude with some concrete applications of our work to the study of weather and climate extremes.

4. **Counting cycles in networks**

   Jiashun Jin, Carnegie Mellon University

   Abstract: For many years, researchers have been looking for a testing statistic that is powerful and also has a tractable limiting null. We show that such a statistic can be constructed using the network cycle counts. We show that the statistic can be useful for network goodness-of-fit and for estimating the number of network communities. We also discuss how to properly address the minimax lower bound in network testing, using Sinkhorn's matrix scaling theorem. For applications, we show the statistic can be useful in measuring research diversity and in building co-authorship tree.

5. **Tensor PCA in High Dimensional CP Models**

   Cun-Hui Zhang, Rutgers University

   Abstract: The CP decomposition for high dimensional non-orthogonal spike tensors is an important problem with broad applications across many disciplines. However, previous works with theoretical guarantee typically assume restrictive incoherence conditions on the basis vectors for the CP components. We propose new computationally efficient composite PCA and concurrent orthogonalization algorithms for tensor CP decomposition with theoretical guarantees under mild incoherence conditions. The composite PCA applies the principal component or singular value decompositions twice, first to a matrix unfolding of the tensor data to obtain singular vectors and then to the matrix folding of the singular vectors obtained in the first step. It can be used as an initialization for any iterative optimization schemes for the tensor CP decomposition. The concurrent orthogonalization algorithm iteratively estimates the basis vector in each mode of the tensor by simultaneously applying projections to the orthogonal complements of the spaces generated by others CP components in other modes. It is designed to improve the alternating least squares estimator and other forms of the high order orthogonal iteration for tensors with low or moderately high CP ranks. Our theoretical investigation provides estimation accuracy and convergence rates for the two proposed algorithms. Our implementations on synthetic data demonstrate significant practical superiority of our approach over existing methods.

6. **Tensor Quantile Regression for Neuroimage Study of Human Intelligence**
   Heping Zhang, Yale University
   Abstract: Human intelligence is usually measured by well-established psychometric tests through a series of problem solving. The recorded cognitive scores are continuous but usually heavy-tailed with potential outliers and violating the normality assumption. Meanwhile, magnetic resonance imaging provides an unparalleled opportunity to study brain structures and cognitive ability. Motivated by association studies between MRI images and human intelligence, we propose a tensor quantile regression model, which is a general and robust alternative to the commonly used scalar-on-image linear regression. Moreover, we take into account rich spatial information of brain structures, incorporating low-rankness and piece-wise smoothness of imaging coefficients into a regularized regression framework. We formulate the optimization problem as a sequence of penalized quantile regressions with a generalized Lasso penalty based on tensor decomposition, and develop a computationally efficient alternating direction method of multipliers algorithm estimate the model components. Extensive numerical studies are conducted to examine the empirical performance of the proposed method and its competitors. Finally, we apply the proposed method to a large-scale important dataset: The Human Connectome Project. We find that the tensor quantile regression can serve as a prognostic tool to assess future risk of cognitive impairment progression. More importantly, with the proposed method, we are able to identify the most activated brain subregions associated with quantiles of human intelligence. The prefrontal and anterior cingulate cortex are found to be mostly associated with lower and upper quantile of fluid intelligence. The insular cortex associated with median of fluid intelligence is a rarely reported region.

7. **Minimum L1-norm interpolators: Precise asymptotics and multiple descent**
   Yuting Wei, University of Pennsylvania
   Abstract: An evolving line of machine learning works observe empirical evidence that suggests interpolating estimators --- the ones that achieve zero training error --- may not necessarily be harmful. In this talk, we pursue theoretical understanding for an important type of interpolators: the minimum L1-norm interpolator, which is motivated by the observation that several learning algorithms favor low L1-norm solutions in the over-parameterized regime. Concretely, we consider the noisy sparse regression model under Gaussian design, focusing on linear sparsity and high-dimensional asymptotics (so that both the number of features and the sparsity level scale proportionally with the sample size).
   We observe, and provide rigorous theoretical justification for, a curious multi-descent phenomenon; that is, the generalization risk of the minimum L1-norm interpolator undergoes multiple (and possibly more than two) phases of descent and ascent as one increases the model capacity. This phenomenon stems from the special structure of the minimum L1-norm interpolator as well as the delicate interplay between the over-parameterized ratio and the sparsity, thus unveiling a fundamental distinction in geometry from the minimum L2-norm interpolator. Our finding is built upon an exact characterization of the risk behavior, which is governed by a system of two non-linear equations with two unknowns.

8. **High-Dimensional Knockoffs Inference for Time Series Data**
   Jinchi Lv, University of Southern California
   Abstract: The recently introduced framework of model-X knockoffs provides a flexible tool for exact finite-sample false discovery rate (FDR) control in variable selection in arbitrary dimensions without assuming any dependence structure of the response on covariates. It also completely bypasses the use of conventional p-values, making it especially appealing in high-dimensional nonlinear models. Existing works have focused on the setting of independent and identically distributed observations. Yet time series data is prevalent in practical applications in various fields such as economics and

social sciences. This motivates the study of model-X knockoffs inference for time series data. In this paper, we make some initial attempt to establish the theoretical and methodological foundation for the model-X knockoffs inference for time series data. We suggest the method of time series knockoffs inference (TSKI) by exploiting the idea of subsampling to alleviate the difficulty caused by the serial dependence. We establish sufficient conditions under which the original model-X knockoffs inference combined with subsampling still achieves the asymptotic FDR control. Our technical analysis reveals the exact effect of serial dependence on the FDR control. To alleviate the practical concern on the power loss because of reduced sample size cause by subsampling, we exploit the idea of knockoffs with copies and multiple knockoffs. Under fairly general time series model settings, we show that the FDR remains to be controlled asymptotically. To theoretically justify the power of TSKI, we further suggest the new knockoff statistic, the backward elimination ranking (BE) statistic, and show that it enjoys both the sure screening property and controlled FDR in the linear time series model setting. The theoretical results and appealing finite-sample performance of the suggested TSKI method coupled with the BE are illustrated with several simulation examples and an economic inflation forecasting application. This is a joint work with Chien-Ming Chi, Yingying Fan and Ching-Kang Ing.

9. **Estimation of spectra of high-dimensional separable covariance matrices**
   Debashis Paul, University of California, Davis
   Abstract: We consider the problem of estimating the joint spectra of high-dimensional time series for which the observed data matrix is assumed to have a separable covariance structure. The primary interest is in estimating the distribution of the eigenvalues of the marginal covariance of the observation vectors, under partial information -- such as stationarity or sparsity -- on the temporal covariance structure. We develop a method that utilizes random matrix theory to estimate the unknown population spectra by repressing the spectrum of the dimensional covariance matrix on a simplex. We prove the consistency of the proposed estimator under the dimension proportional to the sample size setting. Furthermore, we develop a resampling-based method for statistical inference on low-dimensional functionals of the joint spectrum of the population covariance matrix. This is a joint work with Lili Wang (Zhejiang Gongshang University).

10. **Volatility prediction comparison via robust volatility proxies: An empirical deviation perspective**
    Weichen Wang, University of Hong Kong
    Abstract: Volatility forecasting is crucial to risk management and portfolio construction. One particular challenge of assessing volatility forecasts is how to construct a robust proxy for the unknown true volatility. In this work, we show that the empirical loss comparison between two volatility predictors hinges on the deviation of the volatility proxy from the true volatility. We then establish non-asymptotic deviation bounds for three robust volatility proxies, two of which are based on clipped data, and the third of which is based on exponentially weighted Huber loss minimization. In particular, in order for the Huber approach to adapt to non-stationary financial returns, we propose to solve a tuning-free weighted Huber loss minimization problem to jointly estimate the volatility and the optimal robustification parameter at each time point. We then inflate this robustification parameter and use it to update the volatility proxy to achieve optimal balance between the bias and variance of the global empirical loss. We also extend this Huber method to construct volatility predictors. Finally, we exploit the proposed robust volatility proxy to compare different volatility predictors on the Bitcoin market data. It turns out that when the sample size is limited, applying the robust volatility proxy gives more consistent and stable evaluation of volatility forecasts.

11. **Maximum independent component analysis with application to non-linear temporal signals**
    Chunming Zhang, University of Wisconsin-Madison

Abstract: In many scientific disciplines, finding hidden influential factors behind observational data is essential but challenging. The majority of existing approaches rely on linear transformation, i.e., true signals are linear combinations of hidden components. Motivated from analyzing non-linear temporal signals in neuroscience, genetics, and finance, this paper proposes the “maximum independent component analysis" (MaxICA), based on max-linear combinations of components. In contrast to existing methods, MaxICA benefits from focusing on significant major components while filtering out ignorable components. A major tool for parameter learning of MaxICA is an augmented genetic algorithm. Extensive empirical evaluations demonstrate the effectiveness of MaxICA in either extracting max-linearly combined essential sources in many applications or supplying a better approximation for non-linearly combined source signals, such as EEG recordings analyzed in this paper.

12. **Covariate-Associated Community Detection on Social Networks**
    Wanjie Wang, National University of Singapore
    Abstract: In social networks, besides the connection information, the nodes also have some node-specific covariates or attributes. These covariates also provide some information to detect the connection community of nodes, but community detection based on the covariates only may not coincident with the network communities. For example, individuals in the same community may have separate profiles. Further, the node covariates may suffer sparse entries in the high dimensional setting. In this talk, I will present a method to exploit the covariate information on the network community detection.

13. **Real-time regression analysis of streaming clustered data with possible abnormal data batches**
    Peter Song, University of Michigan
    Abstract: In this talk I will introduce an incremental learning algorithm to analyze streaming datasets with correlated outcomes such as longitudinal data and clustered data. We develop a renewable QIF (RenewQIF) method within a paradigm of renewable estimation and incremental inference, in which statistical results are recursively renewed with a current data batch and summary statistics of historical data batches, but with no use of any historical subject-level raw data. We compare our renewable estimation method with both offline QIF method and offline generalized estimating equations (GEE) approach that process the entire cumulative subject-level data all together. We show theoretically and numerically that the RenewQIF enjoys statistical and computational efficiency. In addition, we propose an approach to diagnose the homogeneity assumption of regression coefficients via a sequential goodness-of-fit test as a screening procedure on occurrences of potential abnormal data batches. We implement the proposed methodology by expanding the existing Spark’s Lambda architecture for the operation of statistical inference and data quality monitoring. We illustrate the proposed methodology by extensive simulation studies and an analysis of streaming car crash datasets from the National Automotive Sampling System-Crashworthiness Data System (NASS CDS).

14. **Estimating trajectories of statisticians from the co-citation networks**
    Tracy Ke, Harvard University
    Abstract: We are interested in characterizing the evolvement of research interests of individual authors (i.e., the research trajectory). We approach this with a data set we collected and cleaned with 2+ years of efforts. The data set consists of the citation and bibtex (author, title, abstract, reference) information of over 83K papers published in 36 statistical journals from 1975 to 2015. Using the data set, we constructed 21 co-citation networks, each for a time window between 1990 and 2015. We propose a dynamic Degree-Corrected Mixed- Membership (dynamic-DCMM) model, where we model the research interests of an author by a low-dimensional weight vector (called the

network memberships) that evolves slowly over time. We propose dynamic-SCORE as a new spectral approach to estimating the memberships.

We discover a triangle in the spectral domain which we call the Statistical Triangle, and use it to visualize the research trajectories of individual authors. We interpret the three vertices of the triangle as the three primary research areas in statistics: "Bayes", "Biostatistics" and "Nonparametrics". The Statistical Triangle further splits into 15 sub- regions, which we interpret as the 15 representative sub-areas in statistics. These results provide useful insights over the research trend and behavior of statisticians.

15. **When Will You Become the Best Reviewer of Your Own Papers? An Owner-Assisted Approach to Mechanism Design**
    Weijie Su, University of Pennsylvania
    Abstract: Alice submits a number of papers to a machine learning conference and has knowledge of the quality of her papers. Given noisy grades rated by independent reviewers, can Bob obtain accurate estimates of the ground-truth quality of the papers by asking Alice a question about the ground truth? In this talk, we address this when the payoff of Alice is additive convex utility over all her papers. First, if Alice would truthfully answer the question because by doing so her payoff is maximized, we show that the questions must be formulated as pairwise comparisons between her papers. Moreover, if Alice is required to provide a ranking of her papers, which is the most fine-grained question via pairwise comparisons, we prove that she would be truth-telling. By incorporating the ground-truth ranking, we show that Bob can obtain an estimator with the optimal squared error in certain regimes based on any possible ways of truthful information elicitation. Moreover, the estimated grades are substantially more accurate than the raw grades when the number of papers is large and the raw grades are very noisy. Finally, we conclude the talk with several extensions and some refinements for practical considerations. This is based on a working paper (tinyurl.com/4f7pnfk6) and arXiv:2110.14802.

Day 2 (May 25): New Advances in Biostatistics, Health and Causal Inference

16. **Confidence sets for Causal Discovery**
    Mladen Kolar, University of Chicago
    Abstract: Causal discovery procedures are popular methods for discovering causal structure across the physical, biological, and social sciences. However, most procedures for causal discovery only output a single estimated causal model or single equivalence class of models. We propose a procedure for quantifying uncertainty in causal discovery. Specifically, we consider linear structural equation models with non-Gaussian errors and propose a procedure which returns a confidence set of causal orderings which are not ruled out by the data. We show that asymptotically, the true causal ordering will be contained in the returned set with some user specified probability. Joint work with Sam Wang and Mathias Drton.

17. **Dynamic Causal Learning: excursions in double robustness**
    Jelena Bradic, UC San Diego: Math and Halicioglu Data Science Institute
    Abstract: Recent progress in machine learning provides many potentially effective tools to learn estimates or make predictions from datasets of ever-increasing sizes. Can we trust such tools in clinical settings? If a learning algorithm predicts an effect of a new policy to be positive, what guarantees do we have concerning the accuracy of such prediction? The talk introduces new statistical ideas to ensure that the learned estimates in dynamic treatment settings satisfy some

fundamental properties. The talk will discuss potential connections and departures between causality and robustness.

18. **A new central limit theorem for the augmented IPW estimator: variance inflation, cross-fit covariance and beyond**
    Rajarshi Mukherjee, Harvard University
    Abstract: In recent times, inference for the ATE in the presence of high-dimensional covariates has been extensively studied. Among the diverse approaches that have been proposed, augmented inverse propensity weighting (AIPW) with cross-fitting has emerged as a popular choice in practice. In this work, we study this cross-fit AIPW estimator under well-specified outcome regression and propensity score models in a high-dimensional regime where the number of features and samples are both large and comparable. Under assumptions on the covariate distribution, we establish a new CLT for the suitably scaled cross-fit AIPW that applies without any sparsity assumptions on the underlying high-dimensional parameters. Our CLT uncovers two crucial phenomena among others: (i) the AIPW exhibits a substantial variance inflation that can be precisely quantified in terms of the signal-to-noise ratio and other problem parameters, (ii) the asymptotic covariance between the pre-cross-fit estimates is non-negligible even on the root-n scale. In fact, these cross-covariances turn out to be negative in our setting. These findings are strikingly different from their classical counterparts. On the technical front, our work utilizes a novel interplay between three distinct tools—approximate message passing theory, the theory of deterministic equivalents, and the leave-one-out approach. We believe our proof techniques should be useful for analyzing other two-stage estimators in this high-dimensional regime. Finally, we complement our theoretical results with simulations that demonstrate both the finite sample efficacy of our CLT and its robustness to our assumptions.

19. **Causal inference in experiments with interference**
    David Choi, Carnegie Mellon University
    Abstract: In experiments that study social phenomena, such as peer influence or herd immunity, the treatment of one unit may influence the outcomes of others. Such “interference between units” violates traditional approaches for causal inference, so that additional assumptions are often imposed to model or limit the underlying social mechanism; for example, one might assume that the units can be partitioned into non-interfering groups, or that an underlying dependency graph is unknown but sparse so that most units do not interfere with each other. For binary outcomes, we propose an approach that does not require such assumptions, allowing for interference that is both unmodeled and arbitrarily strong, with confidence intervals derived using only the randomization of treatment. However, the estimates will have wider confidence intervals and weaker causal implications than those attainable under stronger assumptions, essentially showing only that effects exist and are associated with specified measures of treatment exposure, such as the number of treated friends or neighborhood treatment rate. The approach allows for the usage of regression, matching, or weighting, as may best fit the application at hand. Inference is done by bounding the distribution of the estimation error over all possible values of the unknown counterfactual, using an integer program. Examples are shown using a vaccination trial and two experiments investigating the effects of social influence.

20. **Estimating cell-type-specific gene co-expression networks from bulk gene expression data with an application to Alzheimer's disease**
    Hongyu Zhao, Yale University
    Abstract: Inferring and characterizing gene co-expression networks has led to important insights on the molecular mechanisms of complex diseases. Most co-expression analyses to date have been performed on gene expression data collected from bulk tissues with different cell type

compositions across samples. As a result, the co-expression estimates only offer an aggregate view of the underlying gene regulations and can be confounded by heterogeneity in cell type compositions, failing to reveal gene coordination that may be distinct across different cell types. In this talk, we describe a flexible framework for estimating cell-type specific gene co-expression networks from bulk sample data, without making specific assumptions on the distributions of gene expression profiles in different cell types. We develop a novel sparse least squares estimator, referred to as CSNet, that is efficient to implement and has good theoretical properties. Using CSNet, we analyzed the bulk gene expression data from a cohort study on Alzheimer's disease and identified previously unknown cell-type-specific co-expressions among Alzheimer's disease risk genes, suggesting cell-type-specific disease pathology for Alzheimer's disease. This is joint work with Chang Su and Emma Jingfei Zhang.

21. **Large-Scale Graph Learning with Latent Variables**
    Genevera Allen, Rice University
    Abstract: Graph learning, often called graphical model selection, is a well-studied problem. But motivated by challenges arising in estimating functional neuronal activity, we seek to learn the graph structure in the presence of latent variables and for extremely large-scale data with tens- to hundreds-of-thousands of nodes. To solve this, we propose a simple solution: hard thresholding existing graph selection estimators. We show that this approach is graph selection consistent in the presence of latent variables and at better statistical rates than previous approaches. Further, we leverage this result to yield a computationally fast and memory efficient method for learning large-scale graphs. We learn thresholded graphs on minipatches, or tiny subsets of both observations and nodes, and ensemble selection events to yield a provably consistent yet fast learner for large-scale graphs. We demonstrate our approaches via simulations on real examples to estimate functional connectivity from large-scale calcium imaging data. This is joint work with Minjie Wang and Tianyi Yao.

22. **Controlled Discovery and Localization of Signals via Bayesian Linear Programming (BLiP)**
    Lucas Janson, Harvard University
    Abstract: In many statistical problems, it is necessary to simultaneously discover signals and localize them as precisely as possible. For instance, genetic fine-mapping studies aim to discover causal genetic variants, but the strong local dependence structure of the genome makes it hard to identify the exact locations of those variants. So the statistical task is to output as many regions as possible and have those regions be as small as possible, while controlling how many outputted regions contain no signal. The same problem arises in any application where signals cannot be perfectly localized, such as locating stars in astronomical sky surveys and change point detection in time series data. However, there are two competing objectives: maximizing the number of discoveries and minimizing the size of those discoveries (all while controlling false discoveries), so our first contribution is to propose a single unified measure we call the resolution-adjusted power that formally trades off these two objectives and hence, in principle, can be maximized subject to a constraint on false discoveries. We take a Bayesian approach, but the resulting posterior optimization problem is intractable due to its non-convexity and high-dimensionality. Thus our second contribution is Bayesian Linear Programming (BLiP), a method which overcomes this intractability to jointly detect and localize signals in a way that verifiably nearly maximizes the expected resolution-adjusted power while provably controlling false discoveries. BLiP is very computationally efficient and can wrap around any Bayesian model and algorithm for approximating the posterior distribution over signal locations. Applying BLiP on top of existing state-of-the-art Bayesian analyses of UK Biobank data (for genetic fine-mapping) and the Sloan Digital Sky Survey (for astronomical point source detection) increased the resolution-adjusted

power by 30-120% with just a few minutes of computation. BLiP is implemented in the new packages pyblip (Python) and blipr (R). This is joint work with Asher Spector.

23. **Matching of datasets and its applications in single-cell biology**
    Zongming Ma, University of Pennsylvania
    Abstract: In this talk, we discuss theory and methods for matching of datasets with low rank signals. They are motivated by examples arising from single-cell biology, especially spatial single-cell data analysis. We demonstrate the prowess of the proposed methods on several real data examples.

24. **Safe Reinforcement Learning in mHealth**
    Eric Laber, Duke University
    Abstract: An optimal mHealth strategy for type I diabetes (T1D) maximizes longterm patient health by tailoring recommendations for diet, exercise, and insulin to the unique biology and evolving health status of each patient. We develop a response-adaptive randomization method that learns an optimal intervention strategy while controlling the risk of adverse events. The method, which uses a variant of Thompson Sampling (TS) to facilitate learning, maximizes efficiency while providing strict controls on the probability of an adverse event. We illustrate the application of NP-TS using data from a pilot mHealth study on T1D.

25. **Estimation and Inference with Proxy Data and its Genetic Applications**
    Hongzhe Li, University of Pennsylvania
    Abstract: Existing high-dimensional statistical methods are largely established for analyzing individual-level data. In this work, we study estimation and inference for high-dimensional linear models where we only observe “proxy data”, which include the marginal statistics and sample covariance matrix that are computed based on different sets of individuals. We develop a rate optimal method for estimation and inference for the regression coefficient vector and its linear functionals based on the proxy data. Moreover, we show the intrinsic limitations in the proxy-data based inference: the minimax optimal rate for estimation is slower than that in the conventional case where individual data are observed; the power for testing and multiple testing does not go to one as the signal strength goes to infinity. These interesting findings are illustrated through simulation studies and an analysis of a dataset concerning the genetic associations of hindlimb muscle weight in a mouse population.

26. **Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing**
    Xiufan Yu, University of Notre Dame
    Abstract: Power-enhanced tests with high-dimensional data have received growing attention in theoretical and applied statistics in recent years. Existing tests possess their respective high-power regions, and we may lack prior knowledge about the alternatives when testing for a problem of interest in practice. There is a critical need of developing powerful testing procedures against more general alternatives. This paper studies the joint test of two-sample mean vectors and covariance matrices for high-dimensional data. We first expand the high-power regions of high-dimensional mean tests or covariance tests to a wider alternative space and then combine their strengths together in the simultaneous test. We develop a new power-enhanced simultaneous test that is powerful to detect differences in either mean vectors or covariance matrices under either sparse or dense alternatives. We prove that the proposed testing procedures align with the power enhancement principles introduced by Fan et al. (2015) and achieve the accurate asymptotic size and consistent asymptotic power. We demonstrate the finite-sample performance using simulation studies and a real application to find differentially expressed gene-sets in cancer studies.

27. **Efficient Algorithms and Implementation of a Semiparametric Joint Model for Longitudinal and Competing Risks Data: With Applications to Massive Biobank Data**
    Gang Li, University of California at Los Angeles
    Abstract: Semiparametric joint models of longitudinal and competing risks data are computationally costly and their current implementations do not scale well to massive biobank data. This paper identifies and addresses some key computational barriers in a semiparametric joint model for longitudinal and competing risks survival data. By developing and implementing customized linear scan algorithms, we reduce the computational complexities from O(n^2) or O(n^3) to O(n) in various steps including numerical integration, risk set calculation, and standard error estimation, where n is the number of subjects. Using both simulated and real world biobank data, we demonstrate that these linear scan algorithms can speed up existing methods by a factor of up to hundreds of thousands when n>10^4, often reducing the runtime from days to minutes. We have developed an R-package, FastJM, based on the proposed algorithms for joint modeling of longitudinal and competing risks time-to-event data and made it publicly available on the Comprehensive R Archive Network (CRAN).

28. **Novel Methods for Multi-ancestry Polygenic Prediction and their Evaluations in 3.7 Million Individuals of Diverse Ancestry**
    Haoyu Zhang, Harvard University
    Abstract: Polygenic risk scores are becoming increasingly predictive of complex traits, but subpar performance in non-European populations raises concerns about their potential clinical applications. We develop a powerful and scalable method to calculate PRS using GWAS summary statistics from multi-ancestry training samples by integrating multiple techniques, including clumping and thresholding, empirical Bayes and super learning. We evaluate the performance of the proposed method and a variety of alternatives using large-scale simulated GWAS on ~19 million common variants and large 23andMe Inc. datasets, including up to 800K individuals from four non-European populations, across seven complex traits. Results show that the proposed method can substantially improve the performance of PRS in non-European populations relative to simple alternatives and has comparable or superior performance relative to a recent method that requires a higher order of computational time. Further, our simulation studies provide novel insights to sample size requirements and the effect of SNP density on multi-ancestry risk prediction.

Day 3 (May 26): New Advances in Machine Learning

29. **Asymptotic properties of high-dimensional random forests**
    Yingying Fan, University of Southern California
    Abstract: As a flexible nonparametric learning tool, random forests algorithm has been widely applied to various real applications with appealing empirical performance, even in the presence of high-dimensional feature space. Unveiling the underlying mechanisms has led to some important recent theoretical results on the consistency of the random forests algorithm and its variants. However, to our knowledge, all existing works concerning random forests consistency in high dimensional setting were established for various modified random forests models where the splitting rules are independent of the response. In light of this, in this paper we derive the consistency rates for the random forests algorithm associated with the sample CART splitting criterion, which is the one used in the original version of the algorithm (Breiman2001), in a general high-dimensional nonparametric regression setting through a bias-variance decomposition analysis. Our new theoretical results show that random forests can indeed adapt to high dimensionality and allow for discontinuous regression function. Our bias analysis characterizes

explicitly how the random forests bias depends on the sample size, tree height, and column subsampling parameter. Some limitations on our current results are also discussed.

30. T-Cal: An optimal test for the calibration of predictive models

Edgar Dobriban, University of Pennsylvania

Abstract: The prediction accuracy of machine learning methods is steadily increasing, but the calibration of their uncertainty predictions poses a significant challenge. Numerous works focus on obtaining well-calibrated predictive models, but less is known about reliably assessing model calibration. This limits our ability to know when algorithms for improving calibration have a real effect, and when their improvements are merely artifacts due to random noise in finite datasets. In this work, we consider detecting mis-calibration of predictive models using a finite validation dataset as a hypothesis testing problem. The null hypothesis is that the predictive model is calibrated, while the alternative hypothesis is that the deviation from calibration is sufficiently large. We find that detecting mis-calibration is only possible when the conditional probabilities of the classes are sufficiently smooth functions of the predictions. When the conditional class probabilities are Hölder continuous, we propose T-Cal, a minimax optimal test for calibration based on a debiased plug-in estimator of the ℓ2-Expected Calibration Error (ECE). We further propose Adaptive T-Cal, a version that is adaptive to unknown smoothness. We verify our theoretical findings with a broad range of experiments, including with several popular deep neural net architectures and several standard post-hoc calibration methods. T-Cal is a practical general-purpose tool, which -- combined with classical tests for discrete-valued predictors -- can be used to test the calibration of virtually any probabilistic classification method.

31. Statistics Meet Neural Networks: Bootstrap, Cross-Validations, and Beyond

Jun S. Liu, Harvard University

Abstract: Inspired by the great successes of neural networks (NN) for various AI tasks such as image recognition, machine translation, etc., we examine how recent ideas in NN research may be effectively employed in classical statistical problems. Many statistical estimation problems can be formulated as solving for an M-estimator, and their uncertainties can be quantified by multiple copies of weighted M-estimators, such as in bootstrap methods. Incidentally, the problem of tuning parameter selection via cross-validation can also be formulated as putting weights onto samples and obtaining different solutions under different sets of weights and different specifications of the tuning parameters. In this talk, we discuss ways of setting up flexible neural networks (a) to receive inputs as different weights and to give out outputs that we desire so as to achieve either uncertainty quantification or tuning parameter selection, and (b) to form nonparametric prior to enable nonparametric Bayes analysis. This is based on the joint work with Minsuk Shin, Shijie Wang, Zhirui Hu, and Tracy Ke.

32. How do noise tails impact on Deep ReLU Networks?

Jianqing Fan, Princeton University

Abstract: This paper investigates the stability of deep ReLU neural networks for nonparametric regression under the assumption that the noise has only a finite p-th moment. We unveil how the optimal rate of convergence depends on p, the degree of smoothness, and the intrinsic dimension in a class of nonparametric regression functions with hierarchical composition structure when both the adaptive Huber loss and deep ReLU neural networks are used. This optimal rate of convergence cannot be obtained by the ordinary least squares but can be achieved by the Huber loss with a properly chosen parameter that adapts to the sample size, smoothness and moment parameters. A concentration inequality for the adaptive Huber ReLU neural network estimators with allowable optimization errors is also derived. To establish a matching lower bound within the class of neural network estimators using the Huber loss, we employ a different strategy from the traditional route:

constructing a deep ReLU network estimator that has a better empirical loss than the true function and the difference between these two functions furnishes a low bound. This step is related to the Huberization bias, yet more critically to the approximability of deep ReLU networks. As a result, we also contribute some new results on the approximation theory of deep ReLU neural networks. (Joint with Yihong Gu and Wen-Xin Zhou)

33. **Transfer Learning: Optimality and Adaptive Algorithms**
    Tony Cai, University of Pennsylvania
    Abstract: Human learners have the natural ability to use knowledge gained in one setting for learning in a different but related setting. This ability to transfer knowledge from one task to another is essential for effective learning. However, in statistical learning, most procedures are designed to solve one single task, or to learn one single distribution, based on observations from the same setting. In this talk, we discuss statistical transfer learning in various settings with a focus on nonparametric classification based on observations from different distributions under the posterior drift model, which is a general framework and arises in many practical problems. The results show that significant benefit of incorporating data from the source distributions for learning under the target distribution.

34. **Manifold structure in graph embeddings**
    Patrick Rubin-Delanchy, University of Bristol
    Abstract: Statistical analysis of a graph often starts with embedding, the process of representing its nodes as points in space. How to choose the embedding dimension is a nuanced decision in practice, but in theory a notion of true dimension is often available. In spectral embedding, this dimension may be very high. However, this paper shows that existing random graph models, including graphon and other latent position models, predict the data should live near a much lower dimensional manifold. One may therefore circumvent the curse of dimensionality by employing methods which exploit hidden manifold structure.

35. **Understanding self-supervised learning**
    Tengyu Ma, Stanford University
    Abstract: Self-supervised learning has made empirical breakthroughs in producing representations that can be applied to a wide range of downstream tasks. In this talk, I will primarily present a recent work that analyzes contrastive learning algorithms under realistic assumptions on the data distributions for vision applications. We prove that contrastive learning can be viewed as a parametric version of spectral clustering on a so-called population augmentation graph, and analyze the linear separability of the learned representations and provide sample complexity bounds. I will also briefly discuss two follow-up works studying self-supervised representations' performance under imbalanced training datasets and for shifting test distributions. The talk is based on recent works: https://arxiv.org/abs/2106.04156**,** https://arxiv.org/abs/2110.05025, https://arxiv.org/abs/2204.00570, https://arxiv.org/abs/2204.02683.

36. **Offline Reinforcement Learning with only Realizability**
    Jason Lee, Princeton University
    Abstract: Sample-efficiency guarantees for offline reinforcement learning (RL) often rely on strong assumptions on both the function classes (e.g., Bellman-completeness) and the data coverage (e.g., all-policy concentrability). Despite the recent efforts on relaxing these assumptions, existing works are only able to relax one of the two factors, leaving the strong assumption on the other factor intact. As an important open problem, can we achieve sample-efficient offline RL with weak assumptions on both factors?

In this paper we answer the question in the positive. We analyze a simple algorithm based on the primal-dual formulation of MDPs, where the dual variables (discounted occupancy) are modeled using a density-ratio function against offline data. With proper regularization, we show that the algorithm enjoys polynomial sample complexity, under only realizability and single-policy concentrability. We also provide alternative analyses based on different assumptions to shed light on the nature of primal-dual algorithms for offline RL.

37. Offline Reinforcement Learning: Towards Optimal Sample Complexities

Yuejie Chi, Carnegie Mellon University

Abstract: Offline or batch reinforcement learning seeks to learn a near-optimal policy using history data without active exploration of the environment. To counter the insufficient coverage and sample scarcity of many offline datasets, the principle of pessimism has been recently introduced to mitigate high bias of the estimated values. However, prior algorithms or analyses either suffer from suboptimal sample complexities or incur high burn-in cost to reach sample optimality, thus posing an impediment to efficient offline RL in sample-starved applications. In this talk, we demonstrate that the model-based (or "plug-in") approach achieves minimax-optimal sample complexity without burn-in cost for tabular Markov decision processes (MDPs). Our algorithms are "pessimistic" variants of value iteration with Bernstein-style penalties, and do not require sophisticated variance reduction.

38. The curse of overparametrization in adversarial training

Adel Javanmard, University of Southern California

Abstract: Successful deep learning models often involve training neural network architectures that contain more parameters than the number of training samples. Such overparametrized models have been extensively studied in recent years, and the virtues of overparametrization have been established from both the statistical perspective, via the double-descent phenomenon, and the computational perspective via the structural properties of the optimization landscape. Despite the remarkable success of deep learning architectures in the overparametrized regime, it is also well known that these models are highly vulnerable to small adversarial perturbations in their inputs. Even when adversarially trained, their performance on perturbed inputs (robust generalization) is considerably worse than their best attainable performance on benign inputs (standard generalization). In this talk we will discuss a precise characterization of the role of overparametrization on robustness by focusing on random features regression models in a regime where the sample size, the input dimension and the number of parameters grow in proportion to each other, and derive an asymptotically exact formula for the robust generalization error when the model is adversarially trained. Our developed theory reveals the nontrivial effect of overparametrization on robustness and indicates that for adversarially trained random features models, high overparametrization can hurt robust generalization.

39. When is Offline Two-Player Zero-Sum Markov Game Solvable?

Simon Du, University of Washington

Abstract: We study what dataset assumption permits solving offline two-player zero-sum Markov game. In stark contrast to the offline single-agent Markov decision process, we show that the single strategy concentration assumption is insufficient for learning the Nash equilibrium (NE) strategy in offline two-player zero-sum Markov games. On the other hand, we propose a new assumption named unilateral concentration and design a pessimism-type algorithm that is provably efficient under this assumption. In addition, we show that the unilateral concentration assumption is necessary for learning an NE strategy. Furthermore, our algorithm can achieve minimax sample complexity without any modification for two widely studied settings: dataset with uniform

concentration assumption and turn-based Markov game. Our work serves as an important initial step towards understanding offline multi-agent reinforcement learning.

40. Informed MCMC sampling for high-dimensional model selection problems

Quan Zhou, Texas A&M University

Abstract: Informed Markov chain Monte Carlo (MCMC) methods have been proposed as scalable solutions to Bayesian posterior computation on high-dimensional discrete state spaces, but theoretical results about their convergence behavior in general settings are lacking. In this talk, we introduce a generally applicable mixing time bound for Markov chains on discrete spaces, which can be used to prove the rapid mixing of both random walk and informed Metropolis-Hastings algorithms for high-dimensional model selection problems. On the algorithmic side, we propose two novel informed MCMC algorithms, one based on Metropolis-Hastings sampling and the other based on importance weighting. Theoretical results specific to these two algorithms will also be discussed, which justify the practical advantages of informed MCMC methods and provide guidance on how to choose an informed proposal weighting scheme.

41. Towards a Theory of Non-Log-Concave Sampling

Krishna Balasubramanian, University of California, Davis

Abstract: The task of sampling from a given density is a fundamental computational task with numerous applications in statistics, machine learning and applied mathematics. In the last decade, the iteration complexity of sampling from a smooth and (strongly) log-concave density has been well-studied. However, a general theory of sampling when the above mentioned assumptions are not satisfied is lacking. In this talk, taking motivation from the theory of non-convex optimization, I will discuss a recently proposed framework for establishing the iteration complexity of sampling when the target density satisfies only the relatively milder Holder-smoothness assumption. I will also discuss several extensions and applications of our result; in particular, it yields a new state-of-the-art guarantee for sampling from distributions which satisfy a Poincar\'e inequality.

42. Sparse Convoluted Rank Regression in High Dimensions

Boxiang Wang, University of Iowa

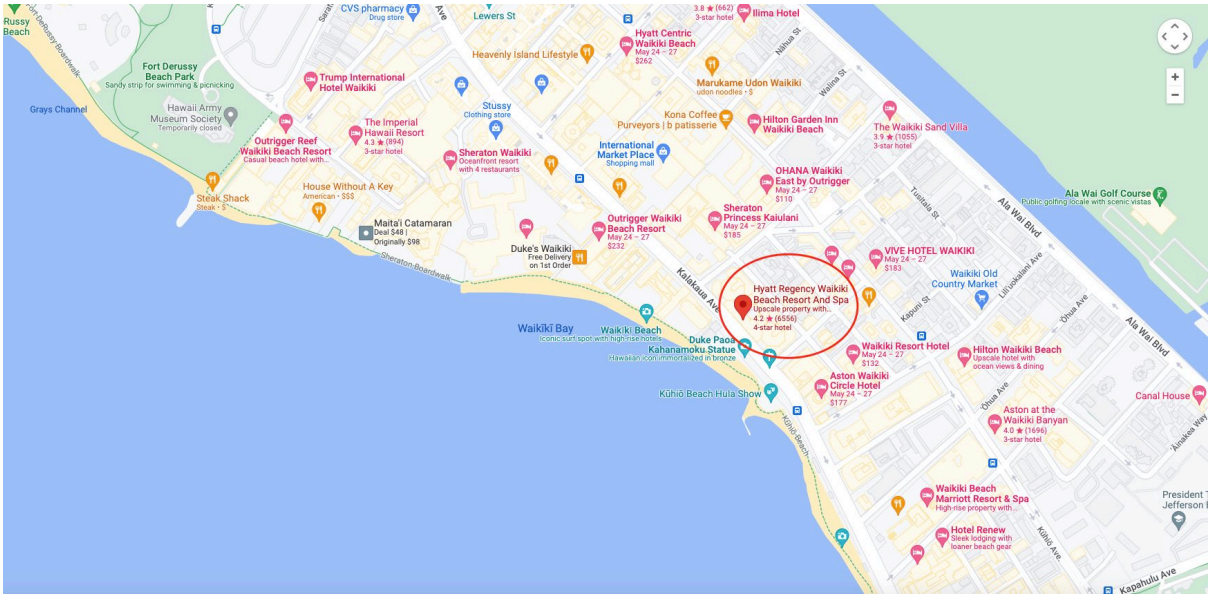
Abstract: Wang et al. (2020, JASA) studied the high-dimensional sparse penalized rank regression and established its nice theoretical properties. Compared with the least squares, rank regression can have a substantial gain in estimation efficiency while maintaining a minimal relative efficiency of 86.4%. However, the computation of penalized rank regression can be very challenging for high-dimensional data, due to the highly non-smooth rank regression loss. In this work we view the rank regression loss as a non-smooth empirical counterpart of a population level quantity, and a smooth empirical counterpart is derived by substituting a kernel density estimator for the true distribution in the expectation calculation. This view leads to the convoluted rank regression loss and consequently the sparse penalized convoluted rank regression (CRR) for high-dimensional data. Under the same key assumptions for sparse rank regression, we establish the rate of convergence of the l1-penalized CRR for a tuning free penalization parameter and prove the strong oracle property of the folded concave penalized CRR. We further propose a high-dimensional Bayesian information criterion for selecting the penalization parameter in folded concave penalized CRR and prove its selection consistency. We derive an efficient algorithm for solving sparse convoluted rank regression that scales well with high dimensions. Numerical examples demonstrate the promising performance of the sparse convoluted rank regression over the sparse rank regression. Our theoretical and numerical results suggest that sparse convoluted rank regression enjoys the best of both sparse least squares regression and sparse rank regression.

43. Blessing of Latent Dependence and Identifiable Deep Modeling of Discrete Latent Variables

Yuqi Gu, Columbia University

Abstract: In the first part, we present a general algebraic technique to investigate the identifiability of complicated discrete models with latent and graphical components. Specifically, motivated by diagnostic tests collecting multivariate categorical data, we focus on discrete models with multiple binary latent variables. In the considered model, the latent variables can have arbitrary dependencies among themselves while the latent-to-observed measurement graph takes a "star-forest" shape. We establish necessary and sufficient graphical criteria for identifiability, and reveal an interesting and perhaps surprising phenomenon of blessing-of-dependence: under the minimal conditions for generic identifiability, the parameters are identifiable if and only if the latent variables are not statistically independent. In the second part, partly motivated by the blessing-of-dependence geometry, we propose a class of identifiable deep discrete latent structure models. We establish the identifiability of these models by developing transparent conditions on the sparsity structure of the pyramid-shaped directed graph. The proposed identifiability conditions can ensure Bayesian posterior consistency under suitable priors. As an illustration, we consider the two-latent-layer model and propose a Bayesian shrinkage estimation approach. Simulation results for this model corroborate identifiability and estimability of the model parameters. Applications of the methodology to DNA nucleotide sequence data uncover useful discrete latent features that are highly predictive of sequence types.

Venue Map


Russy Beach
Fort Derussy Beach Park
Sandy strip for swimming & picnicking
Grays Channel
Hawaii Army Museum Society
Temporarily closed
CVS pharmacy
Drug store
Lewers St
Heavenly Island Lifestyle
Trump International Hotel Waikiki
Stüssy
Clothing store
The Imperial Hawaii Resort
4.3 ★ (894)
3-star hotel
Outrigger Reef Waikiki Beach Resort
Casual beach hotel with...
Sheraton Waikiki
Oceanfront resort with 4 restaurants
Steak Shack
Steak · $
House Without A Key
American · $$$
Maita'i Catamaran
Deal $48 | Originally $98
Duke's Waikiki
Free Delivery on 1st Order
Sheraton Boardwalk
Waikīkī Bay
Iconic surf spot with high-rise hotels
Waikiki Beach
Duke Paoa Kahanamoku Statue
Hawaiian icon immortalized in bronze
Kūhiō Beach Hula Show
Kūhiō Beach
Kalakaua Ave
Hyatt Centric Waikiki Beach
May 24 – 27
$262
Ilima Hotel
3.8 ★ (662)
3-star hotel
Nahua St
Marukame Udon Waikiki
udon noodles · $
Walina St
Kona Coffee Purveyors | b patisserie
Hilton Garden Inn Waikiki Beach
The Waikiki Sand Villa
3.9 ★ (1055)
3-star hotel
International Market Place
Shopping mall
OHANA Waikiki East by Outrigger
May 24 – 27
$110
Sheraton Princess Kaiulani
May 24 – 27
$185
Outrigger Waikiki Beach Resort
May 24 – 27
$232
Tusitala St
Ala Wai Golf Course
Public golfing locale with scenic vistas
Ala Wai Blvd
VIVE HOTEL WAIKIKI
May 24 – 27
$183
Waikiki Old Country Market
Lili'uokalani Ave
Hyatt Regency Waikiki Beach Resort And Spa
Upscale property with...
4.2 ★ (6556)
4-star hotel
Kapuni St
Ohua Ave
Waikiki Resort Hotel
May 24 – 27
$132
Hilton Waikiki Beach
Upscale hotel with ocean views & dining
Aston Waikiki Circle Hotel
May 24 – 27
$177
Aston at the Waikiki Banyan
4.0 ★ (1696)
3-star hotel
Canal House
Waikiki Beach Marriott Resort & Spa
High-rise property with...
Hotel Renew
Sleek lodging with loaner beach gear
Kūhiō Ave
Kapahulu Ave
President Jefferson

Other Information

Reception location:

Na Lea Terrace, Hyatt Regency Waikiki Beach Resort And Spa
(In case of extreme weather) Regency Club Terrance

Wifi:

Connect to Hyatt_WiFi using email and last name when you log in for the first time.

Organizing Committee:

Tracy Ke (Harvard University)
Weichen Wang (University of Hong Kong)

Sponsors:

Statistics

ims

NISS | National Institute of Statistical Sciences