Minghui Yu Memorial Conference

March 27, 2021

Doctoral Student Body
Department of Statistics
Columbia University

About

The 2021 Minghui Yu Memorial Conference, organized by doctoral students in the Statistics Department of Columbia University, will take place on Saturday, March 27th over Zoom. Minghui Yu was a doctoral student at the statistics department, who passed away in a tragic accident in the spring of 2008. Since then, doctoral students in the Statistics Department have been organizing a conference each year to honor his memory. The conference will feature talks by doctoral students at the Statistics Department, ranging from those just beginning a research program to those who are about to defend dissertations. In addition to being an occasion to remember our friend and colleague, this event will be an opportunity to learn about exciting new research areas emerging from our department. We would like to thank the Department of Statistics for their continued support.

Contact

If you have any questions, please do not hesitate to contact Casey Bradshaw at cb3431@columbia.edu.

Schedule

9:30am-9:40am	Welcome
Session 1	
9:40am-9:55am	Long Zhao- *Entropic Martingale Transport*
9:55am-10:10am	Miguel Ángel Garrido- *Large Deviations Principle for the Linear Schrödinger Equation on the Torus*
10:10am-10:25am	Alejandra Quintos Lima- *Optimal Group Size in Microlending*
10:25am-10:40am	Milad Bakhshizadeh- *Sharp Concentration Results for Heavy-Tailed Distributions*
10:40am-11:00am	Break
Session 2	
11:00am-11:15am	Leon Fernandes- *Independent Component Analysis with Heavy Tails using Distance Covariance*
11:15am-11:30am	Arnab Auddy- *Perturbation Bounds for Orthogonally Decomposable Tensors*
11:30am-11:45am	Jialin Ouyang- *PCA for Tensor-Valued Data*
11:45am-12:00pm	Yilin Guo- *Learning Tensor Representations for Meta-Learning*
12:00pm-1:00pm	Lunch Break
1:00pm-2:00pm	Keynote Address- Professor Amy Herring, Duke University
2:00-2:15pm	Break
Session 3	
2:15pm-2:30pm	Ye Tian- *RaSE: Random Subspace Ensemble Classification*
2:30pm-2:45pm	Reed Palmer- *Count-Valued Time Series Models for COVID-19 Daily Death Dynamics*
2:45pm-3:00pm	Charles Margossian- *Bayesian Inference for Latent Gaussian Models: MCMC, Approximate Methods, and Hybrids*
3:00pm-3:15pm	Jitong Qi- *Differential Item Functioning Detection and Removal by Process Features*
3:15pm-3:30pm	Chengliang Tang- *Wasserstein Distributional Learning for Modeling Conditional Densities*
3:30pm-3:45pm	Break
Session 4	
3:45pm-4:00pm	Ding Zhou- *Disentangled Sticky Hierarchical Dirichlet Process Hidden Markov Model*
4:00pm-4:15pm	Joe Suk- *Self-Tuning Bandits Over Unknown Covariate-Shifts*
4:15pm-4:30pm	Nick Galbraith- *Relative Dimensions of Measures, with Applications to Transfer Learning*
4:30pm-4:45pm	Elliott Gordon Rodriguez- *Learning Sparse Log-Ratios for High-Throughput Sequencing Data*
5:00pm	Virtual Happy Hour

Keynote - Professor Amy Herring

Amy Herring is the Sara and Charles Ayres Distinguished Professor of Statistical Science at Duke University, and co-director of Duke Forge, the university's center for actionable health data science. Professor Herring received her doctorate in biostatistics from Harvard University, where she developed methods for dealing with missing data in clinical trials and observational studies of survival outcomes. She came to Duke from UNC-Chapel Hill, where she was a Distinguished Professor and Associate Chair in the Department of Biostatistics.

Her research interests include development of statistical methodology for longitudinal or clustered data, Bayesian methods, latent class and latent variable models, missing data, complex environmental mixtures, and applications of statistics in population health and medicine. Her research program has been supported by funding from, among others, the National Institutes of Health, Environmental Protection Agency, and Maternal and Child Health Bureau.

Professor Herring has received numerous awards for her work, including the Mortimer Spiegelman Award from the American Public Health Association as the best applied public health statistician under age 40. She holds leadership positions at the national and international level, including as Chair-Elect of the American Statistical Association Section on Bayesian Statistical Science and Director of the Eastern North American Region of the International Biometric Society, and is an elected fellow of the American Statistical Association and the International Statistical Institute.

Student Talks

Long Zhao

Entropic Martingale Transport

Through the lens of math finance, we study the convex optimization problem $\inf H(Q|P)$ over $Q \in \mathcal{M}(\mu, \nu)$, the collection of calibrated equivalent martingale measure for a given stochastic process S on the probability space $(\Omega, \mathcal{F}, P)$. We show that the minimizer Q_* is attained in $\mathcal{M}(\mu, \nu)$ and can be characterized by its log-density via semi-static trading strategies. Moreover, we establish a duality to the utility maximization problem with $u(x) = -e^{-x}$.

Miguel Ángel Garrido

Large Deviations Principle for the Linear Schrödinger Equation on the Torus

Recently, the cubic NLS equation and the Dysthe equation, a third and fourth order approximation for the incompressible Navier-Stokes equation in the context of deep water waves, have been used to simulate rare phenomena on large water bodies such as rogue waves. These numerical experiments are based on the conjectured existence of a Large Deviations Principle, when considering initial conditions with independent, but not identically distributed, complex Gaussian Fourier coefficients. In our work we present an LDP for the linear Schrödinger equation where the random Fourier coefficients have exponential decay, as a first step to understand the nonlinear problem.

Alejandra Quintos Lima

Optimal Group Size in Microlending

Microlending, where a bank lends to a small group of people without credit histories, began with the Grameen Bank in Bangladesh, and is widely seen as the creation of Muhammad Yunus, who received the Nobel Peace Prize in recognition of his largely successful efforts. Since that time the modeling of microlending has received a fair amount of academic attention. One of the issues not yet addressed in full detail, however, is the issue of the size of the group. Some attention has nevertheless been paid using an experimental and game theory approach. We, instead, take a mathematical approach to the issue of an optimal group size, where the goal is to minimize the probability of default of the group. To do this, one has to create a model with interacting forces, and to make precise the hypotheses of the model. We show that the original choice of Muhammad Yunus, of a group size of five people, is, under the right, and, we believe, reasonable hypotheses, either close to optimal, or even at times exactly optimal, i.e., the optimal group size is indeed five people.

Milad Bakhshizadeh

Sharp Concentration Results for Heavy-Tailed Distributions

We obtain concentration and large deviation for the sums of independent and identically distributed random variables with heavy-tailed distributions. Our main theorem can not only recover some of the existing results, such as the concentration of the sum of subWeibull random variables, but it can also produce new results for the sum of random variables with heavier tails. We show that the concentration inequalities we obtain are sharp enough to offer large deviation results for the sums of independent random variables as well. Our analyses which are based on standard truncation arguments simplify, unify and generalize the existing results on the concentration and large deviation of heavy-tailed random variables.

Leon Fernandes

Independent Component Analysis with Heavy Tails using Distance Covariance

Independent Component Analysis (ICA) is a popular tool used for blind source separation and has found application in fields like financial time series, signal processing, feature extraction, brain imaging, and so on. Inspired by modeling a macro economic time series that have components with heavy tails, we consider the ICA problem with an infinite variance source. We develop methodology and prove consistency for estimation in this context using distance covariance. Distance covariance is a measure of independence developed by Székely, et al., 2007 and Székely & Rizzo, 2009. We prove a uniform convergence result for distance covariance over a compact space and use it to obtain consistency for our method. This is based on joint work with Richard A. Davis.

Arnab Auddy

Perturbation Bounds for Orthogonally Decomposable Tensors

We develop deterministic perturbation bounds for singular values and vectors of orthogonally decomposable tensors, in a spirit similar to the classical Davis-Kahan theorem for matrices. Our bounds exhibit intriguing differences between matrices and higher-order tensors. Most notably, they indicate that for higher-order tensors perturbation affects each singular value/vector in isolation. In particular, its effect on a singular vector does not depend on the multiplicity of its corresponding singular value or its distance from other singular values. Our results can be readily applied and provide a unified treatment to many problems involving tensor methods, namely latent

variable models, tensor PCA, and independent component analysis.

Jialin Ouyang

PCA for Tensor-Valued Data

Principal Component Analysis (PCA) is one of the most commonly used methods to extract the lower-dimensional subspace that contains the signals from the higher-dimensional data. One of the most investigated models for PCA is the so-called spiked covariance model. We extend the spiked covariance model to the case of tensor-valued data. We show that even when the rank is allowed to diverge and the magnitudes of the signal strengths are allowed to differ, the optimal rates of estimating the principle components will still be the same as if we observe data from a rank-1 vector-valued spiked covariance model with the same principle component. Then a computational tractable algorithm is proposed such that under initialization conditions, the estimators reach the optimal rates.

Yilin Guo

Learning Tensor Representations for Meta-Learning

We introduce a tensor-based model of shared representation for meta-learning a diverse set of tasks. Prior works on learning linear representations assume that there is a common shared representation across different tasks, and do not consider the additional task specific observable side information. Multi-task learning with observable side features is prevalent in the recommender system, where the items (tasks) that the users rate, often come with observable features. For example, when we try to predict the user rating on a different restaurant, we should consider not only the customers' (users) features, such as the age, alcohol, w/wo children, but also the restaurant (task) features, such as the location, servants, and accepted payments. We estimated such a representation with theoretically guarantees by introducing a two-stage method, which is motivated by the method of moments

---

Ye Tian

RaSE: Random Subspace Ensemble Classification

We propose a flexible ensemble classification framework, Random Subspace Ensemble (RaSE), for sparse classification. In the RaSE algorithm, we aggregate many weak learners, where each weak learner is a base classifier trained in a subspace optimally selected from a collection of random subspaces. To conduct subspace selection, we propose a new criterion, ratio information criterion

(RIC), based on weighted Kullback-Leibler divergence. The theoretical analysis includes the risk and Monte-Carlo variance of the RaSE classifier, establishing the screening consistency and weak consistency of RIC, and providing an upper bound for the misclassification rate of the RaSE classifier. 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. An array of simulations under various models and real-data applications demonstrate the effectiveness and robustness of the RaSE classifier and its iterative version in terms of low misclassification rate and accurate feature ranking. The RaSE algorithm is implemented in the R package RaSEn on CRAN.

Reed Palmer

Count-Valued Time Series Models for COVID-19 Daily Death Dynamics

---

Charles Margossian

Bayesian Inference for Latent Gaussian Models: MCMC, Approximate Methods, and Hybrids

Latent Gaussian models are a class of hierarchical models, which can be used to pool information between groups of data. Examples include Gaussian processes and general linear models with a sparsity inducing prior. The posterior distribution of these models often induces a geometry that can frustrate Markov chains Monte Carlo sampler, even gradient-based methods such as Hamiltonian Monte Carlo (HMC). Approximate methods, while fast, may be inaccurate and their failures can be difficult to diagnose. We study a hybrid method, where we use a well-motivated Laplace approximation to marginalize out certain parameters and run HMC on the remaining, well-behaved, parameters. We'll discuss how to efficiently compute the approximation and its gradient, by leveraging insights on the application of automatic differentiation to implicit functions. Past iterations of the method required users to provide derivatives for the covariance matrix and up to third-order derivatives for the likelihood. Our method removes this requirement, offering modelers a more flexible tool; not only that, we even find the new differentiation method to be faster than computation using analytical derivatives.

Jitong Qi

Differential Item Functioning Detection and Removal by Process Features

While early research on differential item functioning (DIF) primarily focused on DIF detection for item quality control purposes, there has been a surging interest in identifying the causes that underly the systematic differences in subgroup performances. The collection of process data, which can document the problem-solving processes that examinees go through to arrive at a final response, brings new possibilities for understanding the differences in problem-solving processes between subgroups, pinpointing the construct-irrelevant factors that contribute to DIF, and, further, removing DIF by explicitly taking these construct-irrelevant factors into account in item response modelling. In this talk, we propose a novel framework for latent trait estimation, where items with final response DIF could be "debiased" using the process data. Process features are first extracted using multidimensional scaling or recurrent neural network from action sequences. Initial latent trait estimates are then obtained by a two-stage conditional expectation method. Using transfer learning, we characterize the relationship that latent trait and process features have with corresponding score that is independent of the examinee group affiliation. Finally, we re-estimate the latent trait based on this functional form given process features. This framework is applied to the process data from the PIAAC PSTRE assessment for evaluation of its effectiveness.

Chengliang Tang

Wasserstein Distributional Learning for Modeling Conditional Densities

Density functions are the target outcome of interest in many data-driven applications. Conventional learning algorithms work mostly with summary statistics or numerical features of the density outcomes, rather than directly learning the density functional objects. Having density functions as the response variable, in a supervised learning setting, enables a more comprehensive and effective investigation into the factors that impact the entire distribution function of interest. There have been recent developments on functional regression models for density curves. One bottleneck of fully deploying such models in practice lies in the inherent constraint of non-negativity and unit integral for the space of density functions. Conventional regression algorithms would fail as this functional space is not closed for linear operators. To overcome this challenge, we propose the Wasserstein distributional learning, a flexible density-on-scalar regression framework for modeling density function outcomes. The proposed framework is constructed using the Wasserstein distance W_2 as a proper metric for the space of density functional objects. We focus on the Conditional Gaussian Mixture Models (CGMM) as the model class $\mathfrak{F}$ to account for a wide range of heterogeneity in the density responses. The resulting metric space $(\mathfrak{F}, W_2)$ of Wasserstein distributional learning not only, by definition, naturally satisfies the required constraints, but also

offers a dense and closed functional subspace for approximating the functional outputs. For fitting the CGMM under the Wasserstein loss, we develop an efficient algorithm based on Majorization-Minimization optimization with boosted trees. We demonstrate the effectiveness of the proposed modeling framework through simulations and real-world applications.

Ding Zhou

Disentangled Sticky Hierarchical Dirichlet Process Hidden Markov Model

The Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) has been used widely as a natural Bayesian nonparametric extension of the classical Hidden Markov Model for learning from sequential and time-series data. A sticky extension of the HDP-HMM has been proposed to strengthen the self-persistence probability in the HDP-HMM. However, the sticky HDP-HMM entangles the strength of the self-persistence prior and transition prior together, limiting its expressiveness. Here, we propose a more general model: the disentangled sticky HDP-HMM (DS-HDP-HMM). We develop novel Gibbs sampling algorithms for efficient inference in this model. We show that the disentangled sticky HDP-HMM outperforms the sticky HDP-HMM and HDP-HMM on both synthetic and real data, and apply the new approach to analyze neural data and segment behavioral video data.

Joe Suk

Self-Tuning Bandits over Unknown Covariate-Shifts

Bandits with covariates, a.k.a.contextual bandits, address situations where optimal actions (or arms) at a given time t, depend on a context x_t, e.g., a new patient's medical history, a consumer's past purchases. While it is understood that the distribution of contexts might change over time, e.g., due to seasonalities, or deployment to new environments, the bulk of studies concern the most adversarial such changes, resulting in regret bounds that are often worst-case in nature. Covariate-shift on the other hand has been considered in classification as a middle-ground formalism that can capture mild to relatively severe changes in distributions. We consider non-parametric bandits under such middle-ground scenarios, and derive new regret bounds that tightly capture a continuum of changes in context distribution. Furthermore, we show that these rates can be adaptively attained without knowledge of the time of shift (change point) nor the amount of shift.

Nick Galbraith

Relative Dimensions of Measures, with Applications to Transfer Learning

In many practical settings, we collect nominally high-dimensional data with intrinsically low-dimensional structure. Many statistical learning methods adaptively leverage this, in the sense that they suffer the curse of dimensionality according to the intrinsic dimension rather than the ambient dimension. We a) propose an extension of a notion of intrinsic dimension for a probability measure to a relative dimension between measures and b) show that in the setting of nonparametric transfer learning under covariate shift, the minimax learning rate is determined by the relative dimension between the 'source' and 'target' feature distributions.

Elliott Gordon Rodriguez

Learning Sparse Log-Ratios for High-Throughput Sequencing Data

The automatic discovery of interpretable features that are associated with an outcome of interest is a central goal of bioinformatics. In the context of high-throughput genetic sequencing data, and Compositional Data more generally, an important class of features are the log-ratios between subsets of the input variables. However, the space of these log-ratios grows combinatorially with the dimension of the input, and as a result, existing learning algorithms do not scale to increasingly common high-dimensional datasets. Building on recent literature on continuous relaxations of discrete latent variables, we design a novel learning algorithm that identifies sparse log-ratios several orders of magnitude faster than competing methods. As well as dramatically reducing runtime, our method outperforms its competitors in terms of sparsity and predictive accuracy, as measured across a wide range of benchmark datasets.

About Minghui Yu

Minghui was born in Shandong, China in 1983. In 2002, he entered the Special Class for the Gifted Young at the University of Science and Technology of China (USTC), one of the most prestigious universities in China. Minghui possessed the rare quality of being not only smart, but also diligent, versatile, modest and easy-going. He was the type of friend who would stand by you no matter the situation. Minghui breezed through the challenging undergraduate program at USTC, ranking at the top of his class. Minghui was well liked by his fellows students having served as the class president from his sophomore year. Although under enormous academic pressure, he still found time to organize a series of student activities, such as hiking, art performances, and athletic contests for his fellow students.

After graduating summa cum laude in 2006 from USTC, Minghui entered the PhD program at the Physics Department of Columbia University. After one year, he transferred to the doctorate program in statistics. During his time at Columbia, Minghui served as the public relations head of the Columbia University's Chinese Students and Scholars Association (2007-2008), and was a member of the Columbia Chinese Basketball Association and the Columbia Graduate Student Consulting Club. His biography on the CUCSSA website mentioned his love of "movies, photography and delicacies". Minghui described himself in his blog as a boy who wants to combine art and science together.

On April 4, 2008, after attending a student- organized conference, Minghui escorted his girlfriend home on the west side of campus. On his return, he was accosted as he was crossing 122nd and Broadway and in his attempt to flee, he was struck by an automobile on Broadway. Minghui was taken to St. Luke's Hospital where he passed away a short time later.