

# Titles and Abstracts

1. Hao Chen, University of California, Davis, USA

**Title:** Change-point Detection for Locally Dependent Data

**Abstract:** Local dependence is common in high-dimension and non-Euclidean data sequences. We consider the testing and estimation of change-points in such sequences. A new way of permutation, circular block permutation with a random starting point, is proposed and studied for a scan statistic utilizing graphs representing the similarity between observations. The proposed permutation approach could correctly address for local dependence and make it possible the theoretical treatments for the non-parametric graph-based scan statistic for locally dependent data. We derive accurate analytic approximations to the significance of graph-based scan statistics under the circular block permutation framework, facilitating its application to locally dependent multivariate or object data sequences.

2. Yuguo Chen, Department of Statistics, University of Illinois at Urbana-Champaign, USA

**Title:** Community Detection in Multi-relational Data through Restricted Multi-layer Stochastic Blockmodel

**Abstract:** In recent years there has been an increased interest in statistical analysis of data with multiple types of relations among a set of entities. Such multi-relational data can be represented as multi-layer graphs where the set of vertices represents the entities and multiple types of edges represent the different relations among them. For community detection in multi-layer graphs, we consider two random graph models, the multi-layer stochastic blockmodel and a model with a restricted parameter space. We derive consistency results for community assignments of the maximum likelihood estimators in both models. We also derive minimax rates of error and sharp thresholds for achieving consistency of community detection in both models, which are then used to compare the multi-layer models with a baseline model, the aggregate stochastic block model. The simulation studies and real data applications confirm the superior performance of the multi-layer approaches in comparison to the baseline procedures.

3. Jianfeng Feng, Fudan University, China& University of Warwick, UK

**Title:** Integrating Genetic, Imaging and Phenotype Data for Predictions: Theory and Applications

**Abstract:** With the availability of big data at the molecular, cellular, imaging and phenotype scales, we are able to explore many issues in cognitive neuroscience such as the genetic and neuronal basis of happiness, IQ, EQ, creativity etc., and in psychiatric diseases such as locate the roots of various mental disorders etc. In my talk, a summary of our recent efforts on these issues are provided, focusing on integrating

multi-scale data for predictions. Working closely with our colleagues as clinical doctors and psychologists, our approach aims for directly addressing clinical problems and developing various novel statistical and machine learning approaches

4. Wensheng Guo, Department of Biostatistics and Epidemiology, University of Pennsylvania, USA

**Title:** Dynamic Functional Classification and Clustering

**Abstract:** We propose a dynamic functional clustering algorithm where each group of curves are modeled by a functional mixed effects model, and the posterior probability is used to iteratively classify each subject into different subgroups. The functional mixed effects model allows flexible designs and nested structures. The classification takes into account both group-average trajectories and between-subject variability. We propose an equivalent dynamic state space model to calculate the likelihood in fitting the model, and to efficiently compute the posterior probability in classifying a new subject. The resultant sequential algorithm is  $O(n)$  and can be implemented online. We also propose a leave-one-subject-out cross-validation Kullback-Leibler information criterion to choose the number of clusters. The performance is assessed through a simulation study, and we apply the proposed methods to longitudinal urological chronic pelvic pain syndrome symptom data collected in the Epidemiology and Phenotyping Study conducted by the NIDDK-funded Multidisciplinary Approach to the Study of Chronic Pelvic Pain (MAPP) Research Network and identify three subgroups.

5. Jaroslaw Harezlak, Indiana University School of Public Health-Bloomington, USA

**Title:** Laplacian-based Regularized Statistical Approach: Association of Gray Matter Imaging Markers with Alcoholism Incorporating Structural Connectivity Information

**Abstract:** Majority of multimodal neuroimaging studies are analyzed separately for each modality. Importantly, statistical methods that simultaneously assess multimodal data provide a more integrative and comprehensive understanding of the brain. We propose an extension to the statistical regularization methods in the linear model setting with a Laplacian-based penalty operator. Model parameters are estimated by a unified approach directly incorporating structural connectivity information into the estimation by exploiting the joint eigenproperties of the predictors and the penalty operator. We present the closed-form solution for the estimators, test its properties via a simulation study and apply it to find the best predictive imaging markers of alcohol drinking phenotypes. Introducing a priori information minimized spurious findings by assigning penalty weights in such a way that highly connected regions associated with the outcome were less penalized than other regions that had no association with the outcome. Future work will incorporate functional connectivity and finer cortical brain parcellation.

6. Feifang Hu, George Washington University, USA

**Title:** Balance Many Important Covariates in Causal Inference

**Abstract:** This article introduces a new randomization procedure to improve the covariate balance across treatment groups. Covariate balance is one of the most important concerns for successful comparative studies, such as causal inference and clinical trials, because it reduces bias and improves the accuracy of inference. However, chance imbalance may still exist in traditional randomized experiments, in which units are randomly allocated without using their covariate information. To address this issue, the proposed method allocates the units sequentially and adaptively, using information on the current level of imbalance and the incoming unit's covariate. With a large number of covariates or a large number of units, the proposed method shows substantial advantages over the traditional methods in terms of the covariate balance and computational time, making it an ideal technique in the era of big data. Furthermore, the proposed method attains the optimal covariate balance, in the sense that the estimated average treatment effect under the proposed method attains its minimum variance asymptotically. Numerical studies and real data analysis provide further evidence of the advantages of the proposed method.

7. Satish Iyengar, University of Pittsburgh, USA

**Title:** Big Data Challenges in Psychiatry

**Abstract:** Current psychiatric diagnoses are based primarily on self-reported experiences. Unfortunately, treatments for the diagnoses are not effective for all patients. One hypothesized reason is that “artificial grouping of heterogeneous syndromes with different pathophysiological mechanisms into one disorder.” To address this problem, the US National Institute of Mental Health instituted the Research Domain Criteria framework in 2009. This research framework calls for integrating data from many levels of information: genes, cells, molecules, circuits, physiology, behavior, and self-report. Clustering comes to the forefront as a key tool in this big-data effort. In this talk, we present a case study of the use of mixture models to cluster older adults based on measures of sleep from three domains: diary, actigraphy, and polysomnography. Challenges in this effort include the use of mixtures of asymmetric (skewed) distributions, a large number of potential clustering variables, and seeking clinically meaningful solutions. We present novel variable selection algorithms, study them by simulation, and demonstrate our methods on the sleep data.

8. Bei Jiang, University of Alberta, Edmonton, Alberta, Canada

**Title:** A Joint Modeling Approach for Treatment Response and Baseline Imaging Data

**Abstract:** There have been a lot of interests in using neuroimaging approaches to help guide clinicians in selecting treatment for patients of major depressive disorders and understand placebo response. In this talk, we discuss a unified Bayesian framework for modeling treatment outcome while exploiting image-based features as predictors. Traditional methods usually take two separate steps. First, a dimension reduction procedure is conducted to reduce high-dimensional images to low dimensional features. Second, a regression analysis is carried out to investigate the relationship between a treatment response of interest and the extracted low dimensional imaging features such that the effect of treatment depends on these imaging

features. In contrast, our method performs these tasks simultaneously to ultimately take into account uncertainty in both steps. We also illustrate the application of the method on a large placebo-controlled depression clinical trial using baseline EEG measurements. This is a joint work with Eva Petkova, Thaddeus Tarpey and R. Todd Ogden.

9. Hui Jiang, Department of Biostatistics, University of Michigan, USA

**Title:** Computational Methods for Accelerating Resampling-based Hypothesis Testing Procedures

**Abstract:** Resampling-based hypothesis testing procedures such as bootstrapping and permutations tests are widely used when the distribution of the test statistic is analytically intractable. However, these test procedures are often computationally intensive, especially when the dataset is large, the desired significant level is very small or there are many tests to perform, all of which are very common in modern genetics and other studies in life sciences. In this talk, I will discuss several computational methods for accelerating such testing procedures while having theoretical justification and empirical evidence of achieving substantial speedup and high accuracy. The methods will be demonstrated with simulated and real data experiments. This talk is mainly based on the joint works with Julia Salzman, Brian Segal, Yang Shi and other collaborators.

10. Tianzi Jiang, Institute of Automation, The Chinese Academy of Sciences, China

**Title:** Mathematics for the Brainnetome Atlas

**Abstract:** Brain atlas is considered to be the cornerstone of basic neuroscience and clinical researches. Brainnetome atlas is constructed with brain connectivity profiles. It is in vivo, with finer-grained brain subregions, and with anatomical and functional connection profiles. Using the brainnetome atlas, researchers could simulate and model brain networks using informatics and simulation technologies to elucidate the basic organizing principles of the brain. Others could use this same atlas to design novel neuromorphic systems that are inspired by the architecture of the brain. Therefore, this cutting-edge brainnetome atlas paves the way for constructing an even more fine-grained atlas of the human brain and offers the potential for applications in brain-inspired computing. In this lecture, we will summarize the advance of the human brainnetome atlas and its applications, especially mathematics which is needed for the brainnetome atlas studies. We first give a brief introduction on the history of the brain atlas development. Then we present the basic ideas of the human brainnetome atlas and the computational methodologies to construct this atlas. After that, some parcellation results of representative brain areas will be presented. We also give a brief presentation on how to use the human brainnetome atlas to address issues in neuroscience and clinical research. Finally, we will give a brief perspective on monkey brainnetome atlas and the related neurotechniques.

11. Timothy Duane Johnson, Department of Biostatistics, University of Michigan, USA

**Title:** Bayesian Semi-Parametric Modeling of Near Infra-Red Spectroscopy Data

**Abstract:** Functional near-infrared spectroscopy (fNIRS) is a relatively new neuroimaging technique. It is a low cost, portable, and non-invasive method to monitor brain activity. Similar to fMRI, it measures changes in the level of blood oxygen in the brain. Its time resolution is much finer than fMRI, however its spatial resolution is much coarser—similar to EEG or MEG. fNIRS is finding widespread use on young children whom cannot remain still in the MRI magnet and it can be used in situations where fMRI is contraindicated—such as with patients whom have cochlear implants. Since it measure the hemodynamic response to stimulus, similar to fMRI, statistical methods that are in use simply use modifications to existing fMRI packages. In this talk, I show that these methods are inadequate and propose a fully Bayesian semi-parametric hierarchical model to analyze fNIRS data. The hemodynamic response function is modeled using a cubic B-spline basis while nuisance signals (e.g. vasomotor signal and heart beat) are modeled using a Gaussian process. We assume the residual time-series is a high-order AR process and adopt a spike-and-slab prior to shrink unnecessary AR parameters to zero. Our current work is a model for a single time-series of fNIRS data. However, our model is easily adapted to handle the bivariate fNIRS time-series data at a single detector (oxygenated-and deoxygenated-hemoglobin). It can also easily be adapted to handle the spatial aspects of an array of detectors as well as a population-level analysis.

12. Jian Kang, Department of Biostatistics, University of Michigan, China

**Title:** Bayesian Feature Screening for Big Neuroimaging Data via Massively Parallel Computing

**Abstract:** Motivated by the needs of selecting important features from big neuroimaging data, we develop a Bayesian variable screening algorithm for ultra-high dimensional data consisting of two steps: Step 1: compute a multivariate variable screening statistic based on marginal posterior moments; Step 2: perform the mixture model-based cluster analysis on screening statistics to identify the unimportant variables. Step 1 only requires a computational complexity on the order of  $O(n^2 p_n)$  and it is straightforward to be parallelized. It has a close connection with sure independent screening (SIS) statistics and high-dimensional ordinary least-squares projection (HOLP) methods. Step 2 is an extension of the local false discovery rate (FDR) analysis. We implement our method using massively parallel computing techniques based on the general-purpose computing on graphics processing units (GPGPU), leading to an ultra-fast variable screening procedure. Our simulation studies show that the proposed approach can perform variable screening on one million predictors within seconds and achieve higher selection accuracy compared with existing methods. We also illustrate our methods on an analysis of resting state functional magnetic resonance imaging (Rs-fMRI) data from the Autism Brain Imaging Data Exchange (ABIDE) study.

13. Linglong Kong, University of Alberta, Canada

**Title:** Estimation for Bivariate Quantile Varying Coefficient Model

**Abstract:** We propose a bivariate quantile regression method for the bivariate varying coefficient model through a directional approach. The varying coefficients are approximated by the B-spline basis and an  $L_2$ -type penalty is imposed to achieve desired smoothness. We develop a multistage estimation procedure based on the Propagation-Separation (PS) approach to borrow information from nearby directions. The PS method is capable of handling the computational complexity raised by simultaneously considering multiple directions to efficiently estimate varying coefficients while guaranteeing certain smoothness along directions. We reformulate the optimization problem and solve it by the Alternating Direction Method of Multipliers (ADMM), which is implemented using R while the core is written in C to speed it up. Simulation studies are conducted to confirm the finite sample performance of our proposed method. A real data on Diffusion Tensor Imaging (DTI) properties from a clinical study on neurodevelopment is analyzed. Joint work with Haoxu Shu, Qianchuan Chad He, Giseon Heo, John Gilmore, and Hongtu Zhu.

14. Jialiang Li, National University of Singapore, Singapore

**Title:** A Functional Varying-Coefficient Single Index Model for Functional Response Data

**Abstract:** Motivated by the analysis of imaging data, we propose a novel functional varying-coefficient single index model (FVCSIM) to carry out the regression analysis of functional response data on a set of covariates of interest. FVCSIM represents a new extension of varying-coefficient single index models for scalar responses collected from cross-sectional and longitudinal studies. An efficient estimation procedure is developed to iteratively estimate varying coefficient functions, link functions, index parameter vectors, and the covariance function of individual functions. We systematically examine the asymptotic properties of all estimators including the weak convergence of the estimated varying coefficient functions, the asymptotic distribution of the estimated index parameter vectors, and the uniform convergence rate of the estimated covariance function and their spectrum. Simulation studies are carried out to assess the finite-sample performance of the proposed procedure. We apply FVCSIM to investigating the development of white matter diffusivities along the corpus callosum skeleton obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

15. Feng Liang, University of Illinois at Urbana-Champaign, USA

**Title:** Scalable Approximation Algorithms for Bayesian Variable Selection

**Abstract:** There has been an intense development on the estimation of a sparse regression/classification model in statistics, machine learning and related fields. In this talk, we focus on the Bayesian approach to this problem, where sparsity is incorporated by the so-called spike-and-slab prior on the coefficients. Instead of relying on MCMC for posterior inference, we have developed scalable algorithms that approximate the posterior distribution and can process data batch by batch without loading all the data into memory. Asymptotic analysis of our approach, as well as empirical evaluation, will be presented. This is a joint work with Xichen Huang (from UIUC) and Jin Wang (from Expedia).

16. Xi Luo, Brown University, USA

**Title:** Multilevel Causal Mediation Analysis for Big Functional MRI Data

**Abstract:** Causal mediation analysis usually requires strong assumptions, such as ignorability of the mediator, which may not hold in many social and scientific studies. In this talk, we use mediation analysis in an functional MRI experiment to quantify the effect of randomized binary stimuli passing through a brain pathway of two brain regions. We propose a two-layer Structural Equation Modeling (SEM) framework that provides valid inference even if structured unmeasured confounding for the mediator and outcome is present. In the first layer, we use a linear SEM to model the subject level data, where the continuous mediator and outcome may contain correlated additive errors. We propose a constrained optimization approach to estimate the model coefficients, and characterize the nonidentifiability issue due to the correlation parameter. To address the identifiability issue and capture the individual variation, we introduce a mixed effects SEM at the second layer. This approach leverages the power of big functional MRI data in multiple levels, with a methodological innovation to estimate the unknown correlation parameter, instead of performing sensitivity analysis. Using extensive numerical examples, we demonstrate the improvement of our approach over existing methods. This is joint work with PhD student Yi Zhao.

17. Yingying Ma, Beihang University, China

**Title:** Banded Spatio-Temporal Autoregression

**Abstract:** We propose a new class of spatio-temporal models with unknown and banded autoregressive coefficient matrices. The setting represents a sparse structure for high dimensional spatial panel dynamic models when panel members represent economic (or other type) individuals at many different locations. The structure is practically meaningful when the order of panel members is arranged appropriately. Note that the implied autocovariance matrices are unlikely to be banded, and therefore, the proposal is radically different from the existing literature on the inference for high-dimensional banded covariance matrices. Due to the innate endogeneity, we apply the least squares method based on a Yule-Walker equation to estimating autoregressive matrices. A ratio-based method for determining the bandwidth of autoregressive matrices is also proposed. Some asymptotic properties of the inference methods are established. The proposed methodology is further illustrated using both simulated and real data sets.

18. Bin Nan, Presenter, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

**Title:** Classification of ADNI PET Images via Regularized 3D Functional Data Analysis

**Abstract:** We propose a penalized Haar wavelet approach for the classification of 3D brain images in the framework of functional data analysis, which treats each entire 3D brain image as a single functional input thus automatically takes into account the spatial correlations of voxel level imaging measures. We validate the proposed approach through extensive simulations and compare its classification performance with other

commonly used machine learning methods, which show that the proposed method outperforms other methods in both classification accuracy and identification of the relevant voxels. We then apply the proposed method to the practical classification problems for Alzheimer’s disease using PET images obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database to highlight the advantages of our approach.

19. Jie Peng, University of California, Davis

**Title:** Fiber Orientation Distribution Estimation by Spherical Needlets

**Abstract:** We present a novel method for estimation of the fiber orientation distribution (FOD) function based on diffusion-weighted Magnetic Resonance Imaging (D-MRI) data. We formulate the problem of FOD estimation as a regression problem through spherical deconvolution and a sparse representation of the FOD by a *spherical needlets* basis that form a multi-resolution tight frame for spherical functions. This sparse representation allows us to estimate FOD by an  $l_1$ -penalized regression under a non-negativity constraint. The resulting convex optimization problem is solved efficiently by an alternating direction method of multipliers (ADMM) algorithm. The proposed method leads to a sharp feature-preserving reconstruction of the FODs. Through extensive experiments, we demonstrate the effectiveness and favorable performance of the proposed method compared with two existing methods. Particularly, we show the ability of the proposed method in delineating fiber crossing at small angles and in automatically identifying isotropic diffusion. We also apply the proposed method to real 3T D-MRI data sets of healthy elderly individuals. The results show realistic descriptions of crossing fibers with less noise than competing methods even with a relatively small number of gradient directions.

20. Anqi Qiu, Department of Biomedical Engineering, National University of Singapore, Clinical Imaging Research Centre, National University of Singapore, Singapore Institute for Clinical Sciences, Agency for Science, Technology, and Research, Singapore

**Title:** A Set-Based Mixed Effect Model for Gene-Environment Interaction and its Application to Neuroimaging Phenotypes

**Abstract:** Imaging genetics is an emerging field for the investigation of neuro-mechanisms linked to genetic variation. Although imaging genetics has recently shown great promise in understanding biological mechanisms for brain development and psychiatric disorders, studying the link between genetic variants and neuroimaging phenotypes remains statistically challenging due to the high-dimensionality of both genetic and neuroimaging data. This becomes even more challenging when studying  $G \times E$  on neuroimaging phenotypes. In this study, we proposed a set-based mixed effect model for gene-environment interaction (MixGE) on neuroimaging phenotypes, such as structural volumes and tensor-based morphometry (TBM). This model incorporates both fixed and random effects of  $G \times E$  to investigate homogeneous and heterogeneous contributions of multiple genetic variants and their interaction with environmental risks to phenotypes. We discuss the construction of score statistics for the terms associated with fixed and random effects of  $G \times E$  to avoid direct parameter



estimation in the MixGE model, which would greatly increase computational cost. We also describe how the score statistics can be combined into a single significance value to increase statistical power. We evaluated MixGE using simulated and real Alzheimer's Disease Neuroimaging Initiative (ADNI) data, and showed statistical power superior to other burden and variance component methods. We then demonstrated the use of MixGE for exploring the voxelwise effect of  $G \times E$  on TBM, made feasible by the computational efficiency of MixGE. Through this, we discovered a potential interaction effect of gene ABCA7 and cardiovascular risk on local volume change of the right superior parietal cortex, which warrants further investigation.

21. Annie Qu, University of Illinois at Urbana-Champaign, Champaign, USA

**Title:** Weak Signal Identification and Inference in Penalized Model Selection

**Abstract:** Weak signal identification and inference are very important in the area of penalized model selection, yet they are under-developed and not well-studied. Existing inference procedures for penalized estimators are mainly focused on strong signals. In this paper, we propose an identification procedure for weak signals in finite samples, and provide a transition phase in-between noise and strong signal strengths. We also introduce a new two-step inferential method to construct better confidence intervals for the identified weak signals. Both theory and numerical studies indicate that the proposed method leads to better confidence coverage for weak signals, compared with those using asymptotic inference. In addition, the proposed method outperforms the perturbation and bootstrap resampling approaches. We illustrate our method for HIV antiretroviral drug susceptibility data to identify genetic mutations associated with HIV drug resistance.

22. Haipeng Shen, University of Hong Kong

**Title:** Clinical Research about Stroke at China National Clinical Research Center for Neurological Diseases

**Abstract:** Stroke has become the No.1 cause of death in China. Different from US and Europe where the mortality rate of stroke is going down, the situation in China is getting worse. Several large-scale medical projects about stroke have been initiated at China National Clinical Research Center for Neurological Diseases. I shall share my experience in collaborating with the medical principal investigators on such projects.

23. Peter Xuekun Song, University of Michigan, USA

**Title:** Common Data Linkage Issues: Challenges and Solutions

**Abstract:** Data enrichment has been undertaken widely in practice to increase either subject-level information by combining individual clinical data, omics data and imaging data, or cohort-level information by combining data from groups. The former leads to the so-called high-dimensional data, and the latter leads to the so-called big data. This talk presents an overview of current challenges and solutions related to linking data from multiple sources. In particular, some of the challenges such as data heterogeneity and data multimodality offer direct motivations to new developments of statistical methodologies. A set of future directions will be discussed at the end

of the talk. This talk is based on an invited talk given at the National Academies of Sciences.

24. Rui Song, North Carolina State University, USA

**Title:** Optimal Individualized Treatment Strategy with Imaging Covariates

**Abstract:** Different from the standard treatment discovery framework which is used for finding single treatments for a homogenous group of patients, personalized medicine involves finding therapies that are tailored to each individual. In this paper, we consider two approaches to estimate the optimal treatment strategy that uses both scalar and imaging covariates. The first approach is model based and takes into account the smooth nature of most imaging data. We develop an efficient penalized total variation optimization to estimate the unknown slope function and other related parameters. We also establish the error bounds for the total variation slope estimator of imaging covariates and the coefficients of scalar covariates. The second approach is built upon convolutional neural network (CNN) which exploits the correlation between adjacent pixels in the two or three dimensional imaging space. We take this opportunity to employ deep learning to approximate the contrast function and assign future patients according to the sign of the estimated contrast function. Extensive simulations demonstrate that the two proposed methods have superior performance against other possible approaches.

25. Xinyuan Song, Department of Statistics, The Chinese University of Hong Kong

**Title:** Bayesian Scalar on Image Regression with Nonignorable Nonresponses

**Abstract:** Medical imaging data have been widely applied for prognosis, screening, diagnosis, and treatment of various diseases in modern health care. In this study, we consider a scalar on image regression model that includes ultrahigh dimensional imaging data as explanatory covariates for the scalar response of interest, which is subject to nonignorable missingness. It is evident that nonignorable nonresponses distort the accuracy of statistical inference and produce misleading results. We propose the use of an efficient functional principle component analysis method to lower the dimension of the imaging observations and an imaging tilting model to handle potential effects of imaging observations on missingness probability. An instrumental variable, namely, a covariate that is associated with the response but conditionally independent of the missingness probability, is introduced to aid model identifiability. Statistical inference is conducted in a Bayesian framework with Markov chain Monte Carlo algorithms. A simulation study shows that the proposed method exhibits satisfactory finite sample performance. The methodology is illustrated through an application to Alzheimer's Disease Neuroimaging Initiative data set. The talk is mainly based on the joint works with Mr. X. N. Feng, Dr. T. F. Li and professor H. T. Zhu.

26. Niansheng Tang, Yunnan University, China

**Title:** Exponentially Tilted Likelihood Inference on Growing Dimensional Unconditional Moment Models

**Abstract:** Growing-dimensional data with likelihood unavailable are often encountered in various fields. This paper presents a penalized exponentially tilted likelihood (PETL) for variable selection and parameter estimation for growing dimensional unconditional moment models in the presence of correlation among variables and model misspecification. Under some regularity conditions, we investigate the consistent and oracle properties of the PETL estimators of parameters, and show that the constrainedly PETL ratio statistic for testing contrast hypothesis asymptotically follows the central chi-squared distribution. Theoretical results reveal that the PETL approach is robust to model misspecification. We also study high-order asymptotic properties of the proposed PETL estimators. Simulation studies are conducted to investigate the finite performance of the proposed methodologies. An example from the Boston Housing Study is illustrated.

27. Xiaofeng Wang, Cleveland Clinic Lerner Research Institute, USA

**Title:** Statistical Learning for Multi-modal Data in Lung Cancer Detection

**Abstract:** Modern Clinical research studies commonly acquire complementary multi-modal and multi-source data for each patient in order to obtain a more accurate and rigorous assessment of the disease status and likelihood of progression. Multi-modal feature learning and prediction are challenging when integrating these kind of large-scale biomedical data. Motivated from a study of lung cancer detection, we present a novel integrative learning framework for the joint analysis of multi-modal data. The method is a statistical ensemble built on several modern statistical learning techniques, including feature extraction on functional data, random forests, and supervised multidimensional scaling. We also discuss a fast unsupervised clustering method for big data using an adaptive density peak detection procedure. The proposed framework is evaluated by application to high-dimensional chemical sensor array data from the Cleveland Clinic lung cancer early detection project.

28. Yu-Ping Wang, Department of Biomedical Engineering, Biostatistics and Bioinformatics, Computer Science, and Neurosciences, Tulane University, USA

**Title:** Class Specific CCA Models for Joint MRI and Genomic Data Analysis

**Abstract:** Imaging genetics combines brain imaging and genetic information to identify the relationships between genetic variants and brain activities. When the data samples belong to different classes (e.g., disease status), the relationships may exhibit class-specific patterns that can be used to facilitate the understanding of a disease. Conventional approaches such as CCA often perform separate analysis on each class and report the differences, but ignore important shared patterns. We propose a joint sparse canonical correlation analysis (JSCCA) method, which uses a generalized fused lasso penalty to jointly estimate multiple pairs of canonical vectors with both shared and class-specific patterns. Using a data fusion approach, the method is able to discover both common and differential canonical correlations compared to conventional sparse CCA. We demonstrate the method by analyzing a real schizophrenia dataset with 208 subjects. The proposed method reveals a set of distinct SNP-voxel interaction modules for the schizophrenia patients, which are statistically and biologically significant. Finally, we extend the method to multiple classes. We apply it to detect

significant shared and group-specific interactions between DNA methylation and gene expression levels in The Cancer Genome Atlas (TCGA) datasets with 1166 samples from four cancers. This is a joint work with Drs. Jian Fang and Vince Calhoun.

29. Hulin Wu, University of Texas Health Science Center at Houston, USA

**Title:** Two Big Data Research Projects: High-Dimensional Differential Equation and Network Modeling for GEO Genomics and EHR Phenomics Data

**Abstract:** The detailed procedure and analysis pipeline for high-dimensional time course gene expression data from the GEO data repository will be presented. A series of advanced statistical methodologies and modeling techniques for high-dimensional gene regulatory networks will be discussed. In particular, we propose a novel matrix-based estimation approach for high-dimensional linear ordinary differential equation models with more than one million unknown parameters. A key idea is to use the similarity transformation of the coefficient matrix and separable least squares approach to reduce the nonlinear optimization space. Simulation studies show promising results for the new method to deal with high-dimensional systems. Two real data application examples will be used to illustrate the usefulness of the new method. I will also briefly introduce the second Big Data research project, EHR-BigData that I am involved. Both Big Data projects will be used to demonstrate a novel concept and thinking of data-driven research.

30. Ke Xu, Guanghua School of Management, Peking University, China

**Title:**  $PM_{2.5}$  Recovering by Regression with Image Information

**Abstract:** Air pollution is a severe environmental problem worldwide. This is particularly true in China for the past many years. Among all the air pollutants, the  $PM_{2.5}$  is one of the most hazardous for human health. Therefore, monitoring and reducing  $PM_{2.5}$  pollution becomes a problem of fundamental importance. Despite a comprehensive air quality monitoring system has been established by Chinese government, the air quality monitoring system in China is far less than sufficient. Then, how to develop a more comprehensive air quality monitoring system, but at an extremely low cost, becomes a problem of importance. As a promising solution, we explore in this article the possibility to recover  $PM_{2.5}$  by using imaging data.

31. Yu-Feng Zang, Center for Cognition and Brain Disorders, Institutes of Psychological Sciences, Hangzhou Normal University, China

**Title:** Concerns of Meta-analysis for Resting-state fMRI in Brain Disorders

**Abstract:** With the advantages of fairly good spatial (3 mm) and temporal resolution (0.1 - 2s), no radioactivity, and easy implementation, resting-state functional magnetic resonance imaging (RS-fMRI) is a promising functional neuroimaging technique for the diagnosis and to guide treatment of brain disorders. In addition, as compared to the experimental design and analytic methods of block- and event-related task fMRI studies, the state design and analytic methods of RS-fMRI are helpful to cognitive neuroscience studies. Voxel-based meta-analysis is crucial for precise localization of abnormal brain activity. But there are at least two concerns for RS-fMRI

meta-analysis in brain disorder studies. 1) Large variability of analytic methods among studies. There are a lot of analytic methods for RS-fMRI or state task-fMRI. But only a few of them, including amplitude of low frequency fluctuation (ALFF), regional homogeneity (ReHo), and degree centrality (DC), are sort of whole-brain voxel-wise (WBVW) analysis, and hence, are suitable for voxel-based meta-analysis. However, much less studies have used these WBVW methods than other methods, e.g, brain network methods. 2) Stringent criteria for multiple comparison correction. We usually take it for granted that stringent statistical criteria yield less false positive findings. However, we usually do not know what is true positive or the gold standard in brain disorder studies. CONCLUSION: Using very same WBVW analytic methods and reporting WBVW effect size are very important for precise localization of abnormal brain activity and further to guide focused stimulation therapy.

32. Anru Zhang, University of Wisconsin-Madison

**Title:** Cross: Efficient Low-rank Tensor Completion

**Abstract:** The completion of tensors, or high-order arrays, attracts significant attention in recent research. Current literature on tensor completion primarily focuses on recovery from a set of uniformly randomly measured entries, and the required number of measurements to achieve recovery is not guaranteed to be optimal. In addition, the implementation of some previous methods are NP-hard. In this article, we propose a framework for low-rank tensor completion via a novel tensor measurement scheme we name Cross. The proposed procedure is efficient and easy to implement. In particular, we show that a third order tensor of Tucker rank- $(r_1, r_2, r_3)$  in  $p_1$ -by- $p_2$ -by- $p_3$  dimensional space can be recovered from as few as  $r_1 r_2 r_3 + r_1(p_1 - r_1) + r_2(p_2 - r_2) + r_3(p_3 - r_3)$  noiseless measurements, which matches the sample complexity lower-bound. In the case of noisy measurements, we also develop a theoretical upper bound and the matching minimax lower bound for recovery error over certain classes of low-rank tensors for the proposed procedure. The results can be further extended to fourth or higher-order tensors. Simulation studies show that the method performs well under a variety of settings. Finally, the procedure is illustrated through a real dataset in neuroimaging.

33. Heping Zhang, Yale University, USA

**Title:** Integrating Imaging and Genetic Data for Understanding Neuropsychological Disorders

**Abstract:** Neuroimaging has the potential to better illuminate the complex genetic basis of neuropsychological disorders, which have a biological basis rooted in brain function. Because they are quantitative, neuroimaging biomarkers avoid biases arising from imprecise clinical diagnostic criteria. To tackle high dimensionality and complex spatial relationships present in neuroimaging data, we used distance correlation tests in a genomewide association study predicated on multivariate diffusion tensor imaging measurements collected by the Pediatric Imaging, Neurocognition, and Genetics study. After correcting for multiplicity, distance correlation tests of the multivariate phenotype yield significantly greater power at detecting genetic variants affecting brain structure than mass univariate analysis of individual neuroimaging biomarkers. Furthermore, the distance correlation testing framework implicitly accounts for the

covariance between neuroimaging phenotypes and is robust against stringent model specifications. These results demonstrate the potential of our approach to integrating and analyzing imaging and genomic data. This is a joint work with Canhong Wen and Chintan Mehta.

34. Tingting Zhang, University of Virginia, USA

**Title:** Bayesian Inference for the Directional Brain Network Modeled by High-dimensional Damped Harmonic Oscillators using Intracranial EEG Data

**Abstract:** We use ordinary differential equations (ODE) to model the directional interaction, also called effective connectivity, among brain regions. In contrast to existing ODE models that focus on effective connectivity among only a few brain regions and that rely on strong prior belief of the existence and strength of connections, we propose a high-dimensional ODE model motivated by statistical considerations to explore connectivity among multiple small brain regions. To characterize the brain's oscillatory activity, we extend the classical one-dimensional Harmonic oscillator equation to a set of high-dimensional oscillators, each describing one brain region's oscillation and interaction with other regions. The new ODE model also features the high-dimensional brain network in a cluster structure, which consists of modules of densely connected brain regions. We develop a unified Bayesian framework to quantify uncertainty in the assumed ODE model, identify clusters, select strongly connected brain regions, and make statistical comparison between brain networks across different experimental trials. We apply the proposed method to electrocorticography (ECoG) datasets and evaluate brain network changes across trials and different stimuli.

35. Shurong Zheng, Northeast Normal University, China

**Title:** Testing the Homogeneity of Several High Dimensional Correlation Matrices

**Abstract:** This paper tests the equality of several high dimensional correlation matrices. The proposed testing methods are valid for sparse matrices or non-sparse matrices. The fMRI data are used for analysis.

36. Hongtu Zhu, Department of Biostatistics, UNC-Chapel Hill, USA

**Title:** Population-based Structural Connectome Analysis

**Abstract:** Many challenging issues, such as statistical variability in a population, arise from the study of structural connectome maps by using diffusion MRI (dMRI) tractography data. Addressing these challenges requires the development of fast and reliable approaches for processing high-dimensional diffusion tractography data from hundreds (or even thousands) of subjects. We aim to develop a reliable Population-based Structural Connectome (PSC) Mapping framework to construct population structural connectome maps on a common space (or template), while accounting for individual variabilities. The developed PSC framework allows one to view individual structural connectome on different data level, from binary network to streamline based connectome, allowing analysis of the structural connectome at different detail levels. At the weighted network level, novel connection strength measures for a pair

of brain regions are proposed and extracted to reflect more complete description of each connection. At the streamline level, a new compression method is proposed to efficiently represent the connections. The data analysis on a test-retest data set indicates the high re-productibility of the mapped PSC at different data levels. Preliminary group-wise analysis of data extracted using PSC is demonstrated with Human Connectome Project (HCP) Dataset. This is a joint work with Zhengwu Zhang and Maxime Descoteaux.

37. Yunzhang Zhu, Department of Statistics, Ohio State University

**Title:** Maximum Likelihood Inference for a Large Precision Matrix

**Abstract:** Inference concerning Gaussian graphical models involves pairwise conditional dependencies on Gaussian random variables. In such a situation, regularization of a certain form is often employed to treat an overparameterized model, imposing challenges to inference. The common practice of inference uses either a regularized model, as in inference after model selection, or biasreduction known as de-bias. While the first ignores statistical uncertainty inherent in regularization, the second reduces the bias inbred in regularization at an expense of increased variance. In this paper, we propose a constrained maximum likelihood method for inference, with a focus of alleviating the impact of regularization on inference. Particularly, for composite hypotheses, we unregularize hypothesized parameters whereas regularizing nuisance parameters through a  $L_0$ -constraint controlling their degree of sparseness. This approach is an analogy of semiparametric likelihood inference in a high-dimensional situation. On this ground, we derive conditions under which the asymptotic distributions of the constrained likelihood ratio and the maximum likelihood estimate are established, permitting a graphs dimension increasing with the sample size. Interestingly, the corresponding distribution of the likelihood ratio is the chi-square or normal, depending on if the co-dimension of a test is finite or increases with the sample size. This goes beyond the classical Wilks phenomenon. Numerically, we demonstrate that the proposed method performs well for various types of graphs. The talk is based on joint work with Professor Xiatong Shen and Professor Wei Pan.