ASA Statistics in Imaging Section

SMI 2019

---- Statistical Methods in Imaging ----

June 2-4, 2019

PROGRAM BOOK

UCI Department of Statistics

Donald Bren School of Information & Computer Sciences

The **Statistical Methods in Imaging** conference is the annual meeting of the ASA Statistics in Imaging section. The Conference aims at gathering investigators working on methods and applications in imaging science.

We thank the sponsors of SMI 2019 @ UCI:

Statistics in Imaging Section of the American Statistical Association

Department of Statistics, Donald Bren School of Information and Computer Sciences
University of California Irvine

UCI Brain Initiative

UCI Center for Neurobiology of Learning & Memory

We also acknowledge NSF sponsorship through grants DMS 1622490 and SES-1659921

Local Organizing Committee

Weining Shen (chair, Statistics)
Michele Guindani (Statistics)
Bin Nan (Statistics)
Zhaoxia Yu (Statistics)
Mike Yassa (Department of Neurobiology and Behavior, Director, CNLM & UCI Brain)

,

Scientific Committee

Michele Guindani (UCI, Statistics, co-chair)
John Kornak (UCSF, Epidemiology & Biostatistics)
Nicole Lazar (University of Georgia, Statistics)
Armanda Meja (Indiana University, Statistics)
Bin Nan (UCI, Statistics, co-chair)
Hernando Ombao (KAUST, Statistics)
Russell Shinoara (University of Pennsylvania, Biostatistics)
Dana Larisa Tudorascu (University of Pittsburgh, Medicine, Biostatistics, Psychiatry and Clinical and Translational Science)

SMI 2019
Program

Sunday June 2nd, 2019

R Software Development Workshop (follow the link for details)
https://sites.uci.edu/smi2019/r-workshop/

6:30pm-9pm
Opening Mixer with Posters

Monday June 3rd, 2019

8:00 Registration desk opens

8:30 – 8:45 Opening Remarks

8:45 – 10:00 Invited Session I
Advances in the study of Brain Connectivity
Chair: *Simon Vandekar*

- ***Lexin Li***
  University of California, Berkeley

  Mixed-effect time-varying stochastic blockmodel and application in brain connectivity analysis

  Time-varying networks are fast emerging in a wide range of scientific and business applications. Most existing dynamic network models are limited to a single-subject and discrete-time setting. In this article, we propose a mixed-effect stochastic blockmodel that characterizes the continuous time-varying behavior of the network at the population level, meanwhile taking into account both the individual subject variability as well as the prior module information. We develop a multi-step optimization procedure for a constrained likelihood estimation, and derive the associated asymptotic properties. We demonstrate the effectiveness of our method through both simulations and an application to a study of brain development

in youth.
Keywords: brain connectivity analysis; stochastic blockmodel.

- ***Seeonjoo Lee***
  Columbia University

  Age related network efficiency and the role of multi-modal neural underpinning: a statistical framework for multimodal imaging analysis and a case study

  The neuroanatomical underpinnings of cognitive reserve have been investigated using either structural magnetic resonance imaging (MRI) or functional MRI. This study applied a multimodal data-fusion method to a set of structural MRI, DTI and FLAIR data acquired from 177 cognitive normal participants in order to elucidate associations between aging and brain network efficiency. We compared the performance with the results when all modalities are analyzed separately.

- ***Jaroslaw Harezlak*, Damian Brzyski, Xixi Hu***
  Indiana University, Wroclaw University of Technology, Indiana University

  Matrix-variate regression methods: SpINNEr to the rescue

  Classical regression methods use covariate vectors and estimate the corresponding vector of regression coefficients. In many brain imaging applications, however, regressors are often formed as multidimensional arrays. For example, we can be interested in identifying associations of the cortical brain regions connectivity with a scalar outcome of interest. Transforming connectivity matrices into vectors provides an unsatisfactory solution, since it destroys the inherent structure of the matrix and can be computationally challenging. In our work, we propose an alternative approach , the regularized matrix regression , where the matrix of regression coefficients is defined as a solution to a specific optimization problem. The method, called Sparsity Inducing Nuclear Norm Estimator (SpINNEr), simultaneously imposes two types of penalties on the coefficient matrix ,Äì the nuclear norm and the LASSO-type norm ,Äì to encourage the solutions with low rank and entry-wise sparsity. We use the alternating direction method of multipliers (ADMM) to obtain the solution in a computationally efficient manner. We also automatically select the tuning parameters to provide optimal trade-off between the two penalties. We tested the performance of SpINNEr via an extensive simulation study, where we showed that our approach outperforms competing methods in the estimation accuracy. Finally, SpINNEr was applied to investigate the associations between brain structural connectivity and HIV disease-related outcomes.

  Keywords: matrix-variate regression; regularization; structural connectivity

10:00-10:15 Break

10:15 -11:30 Invited Session II

Bayesian models for Connectivity Networks

Chair: *Amanda Mejia*

- ***Raquel Prado***
  University of California, Santa Cruz

  Recent Bayesian Approaches for Analysis of Neuroimaging Data

  In this talk we review some recent statistical models and related computational and inferential methods for analyzing different types of neuroimaging data. We begin discussing Bayesian approaches for detectingactivation and co-activation from complex-valued fMRI data. We show how these approaches lead to more accurate detection of activation when compared to alternative methods based on magnitude-only data. We illustrate our results in extensive simulation studies and in human studies. We then present an approach for magnitude-only data that makes use of Bayesian tensor regression models for joint estimation of activation¬and connectivity. This framework combines low-rank tensor decompositions and multiway stick breaking priors for inferring activation at the voxel level. Connectivity is modeled at the region of interest level using a Gaussian graphical prior structure. These models are illustrated in the context of analyzing multi-subject fMRI data from the balloon-analog risk-taking experiment. Finally, we present new spectral, time-domain and time-frequency approaches for analyzing multi-channel electroencephalogram data.

  Keywords: fMRI, EEG, Bayesian spatio-temporal models, hierarchical models, tensor regressions, time-frequency methods

- ***Jian Kang***
  University of Michigan

  Bayesian network-on-scalar regression with application to neuroimaging data

  In neuroimaging studies, multiple subject-specific brain networks can be constructed from brain connectivity measures, which comprise the networks of the brain regions connected by anatomical tracts or by functional associations. To study how the individual brain network is associated with potential factors such as demographics and clinical symptoms, we develop a network-on-scalar regression model where the response is a network and the predictors are scalars. We propose a new step function prior for the regression coefficients that enjoy both sparsity and homogeneity, leading to efficient posterior inference on variable selection and the community detection of multiple networks. We investigate theoretical properties of the proposed model and conduct simulation studies to evaluate its performance compared with existing alternatives. We illustrate the proposed method to the analysis of resting-state fMRI data in multiple neuroimaging studies.
  Keywords: Brain network, community detection, homogeneity, sparsity

- ***Marina Vannucci***
  Rice University

  Bayesian Modeling of Multiple Structural Connectivity Networks During the Progression of Alzheimer's Disease

  Alzheimer's disease is the most common neurodegenerative disease. In this talk I will first describe a novel approach for inference of multiple networks with related edge values across groups. I will then use the method to infer structural changes in brain connectivity resulting from disease progression using cortical thickness measurements from a cohort of participants who were either healthy control, or with mild cognitive impairment, or Alzheimer's disease patients. Specifically, the method infers a Gaussian graphical model for each group within a joint framework that relies on Bayesian hierarchical priors to link the precision matrix entries across groups. Results identify key alterations in structural connectivity which may reflect disruptions to the healthy brain. I will also illustrate the proposed method through simulations, demonstrating its performance in structure learning and precision matrix estimation with respect to alternative approaches.
  Keywords: Alzheimer's disease; Gaussian graphical model; AIBL study; MRI data; Bayesian inference

11:30-12:30 Keynote I:
Charles DeCarli (UC Davis)

Statistical Issues in Neuroimaging of Human Aging and the Transition to Dementia

Chair: *Raquel Prado*

The last 20 years have witnessed a dramatic increase in imaging technologies as they are applied to human aging and dementia. Not only have technological advances included newer methods of medical imaging that more accurately describe in vivo biological processes, but similarly rapid advances in computational resources allow for novel imaging processing methods to mine further information from the raw image data. Moreover, increasing use of these methods to study ever larger populations of individuals, either clinically or in a research setting, affords the opportunity for "big science" discoveries that could not have been imagined previously. My presentation will begin by a brief review of the biology of "normal" brain aging and the impact of the two most common causes of dementia--Alzheimer's and cerebrovascular disease on the brain. This will be followed by a brief discussion of current neuroimaging methods to assess these processes and the impact neuroimaging currently has on the definition of disease. I will conclude by presenting a series of brief "vignettes" that raise important methodological and statistical questions that remain unanswered. My goal is to offer fertile soil for new statistical ideas to advance our understanding of this critical area of scientific endeavor.

12:30-13:30 Lunch

13:30-14:30 Collaborative Case Studies I

Michael A. Yassa (UCI), Nick Tustison (University of Virginia)
Dan Gillen (UCI), Andrew Holbrook (UCLA)

Longitudinal mapping of cortical thickness measurements: an ADNI-based evaluation study

Chair: *Mark Fiecas*

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a large-scale, longitudinal research study designed to facilitate the early detection and treatment of Alzheimer's disease (AD) using a variety of data sources including neuroimaging. From these neuroimaging data, specifically structural magnetic resonance imaging, researchers can quantify AD-specific, clinically relevant biomarkers, such as regional cortical thickness and its associated longitudinal changes. In this presentation, Dr. Yassa will first provide an overview of the ADNI project and, more generally, the role of neuroimaging in AD research. Dr. Tustison will then discuss the recently developed Advanced Normalization Tools (ANTs) longitudinal pipeline for extracting cortical thickness measures including pipeline design considerations for processing longitudinal imaging data. Using the first phase of the Alzheimer's Disease Neuroimaging Initiative (ADNI-1) data, comprising over 600 subjects with multiple time points from baseline to 36 months, Dr. Holbrook will conclude the presentation by describing an evaluative statistical framework for comparing the cortical thickness quantities calculated across computational platforms in terms of their scientific interpretability and clinical utility.

14:30-15:45 Invited Session III

Advances in Radiomics

Chair: *Marina Vannucci*

- ***Nichole Carlson, PhD***
  University of Colorado, Anschutz Medical Campus

  Radiomics of lung CT as a tool for developing disease phenotypes in lung disease

  Radiomics of lung CT as a tool for developing disease phenotypes in lung disease Computed tomography (CT) imaging of the lung in a becoming a standard source of a data for diagnostic and clinical decision making in lung disease. Current data summarization is largely based on visual assessment of the image by trained radiologists. However, visual assessment is known to be time-consuming and suffer from poor inter and intra-rater reliability. Thus, there is a need to develop more objective and computationally efficient approaches to quantifying CT images. Radiomics is an emerging field of study in which large numbers of quantitative imaging features are extracted from medical images, including CT of the lung. Radiomics is essentially quantifying image texture and comprised of first order features of the voxels of the image (e.g. mean, SD, skewness and kurtosis) and second order features quantifying the spatial correlation of the voxels of the image. Current applications of radiomics have largely focused on lung cancer, where a small region of the lung is segmented and quantified. However, many interstitial lung diseases, such as sarcoidosis, impact the whole lung. To date, it is unclear whether radiomic characterization of CT of the whole lung is a useful tool for developing biomarkers of lung abnormality. In this talk we develop a radiomic profile of sarcoidosis from CT of the lung and use that profile to develop new phenotypes of sarcoidosis of the lung. We quantify how the radiomic biomarkers are associated with traditional visual scoring of lung CT in sarcoidosis and investigate the clinical utility of radiomic biomarkers by quantifying how well the radiomic profile predicts lung function. This talk also highlights a true clinical-biostatistical collaboration.

- ***Xiao Li, Michele Guindani, Chaan S.Ng, Brian P.Hobbs****
  Genentech, Inc., University of California, Irvine, The University of Texas MD Anderson Cancer Center, Cleveland Clinic

A Bayesian Nonparametric approach for Cancer Radiomics: elucidating textural pattern heterogeneity of solid lesions

Cancer radiomics is an emerging discipline promising to elucidate lesion phenotypes and tumor heterogeneity through patterns of enhancement, texture, morphology, and shape. The prevailing technique for image texture analysis relies on the construction and synthesis of Gray-Level Co-occurrence Matrices (GLCM). Practice currently reduces the structured count data of a GLCM to reductive and redundant summary statistics for which analysis requires variable selection and multiple comparisons for each application, thus limiting reproducibility. In this article, we develop a Bayesian multivariate probabilistic framework for the analysis and unsupervised clustering of a sample of GLCM objects. By appropriately accounting for over and under dispersion of the observed counts and simultaneously adjusting for existing spatial autocorrelation at nearby cells, the methodology facilitates estimation of texture pattern distributions within the GLCM lattice itself. The techniques are applied to cluster images of adrenal lesions obtained from CT scans with and without administration of contrast. We further assess whether the resultant subtypes are clinically oriented by investigating their correspondence with pathological diagnoses. Additionally, we compare performance to a class of machine-learning approaches currently used in cancer radiomics with simulation studies.
Keywords: Cancer Radiomics; Gray-level co-occurrence matrix; Bayesian Nonparametrics; Multivariate count data

- ***Qiwei Li*, Xinlei Wang, Faming Liang, Guanghua Xiao***
  University of Texas Southwestern Medical Center, Southern Methodist University, Purdue University, University of Texas Southwestern Medical Center

Bayesian modeling of spatial point patterns and its application on the analysis of tumor pathology images

With the advance of imaging technology, digital pathology imaging of tumor tissue slides is becoming a routine clinical procedure for cancer diagnosis. This process produces massive imaging data that capture histological details in high resolution. Recent developments in deep-learning methods have enabled us to identify and classify individual cells from digital pathology images at large scale. The randomly distributed cells can be considered from a marked point process, where each point is defined by its position and cell type. Reliable statistical approaches to model such marked spatial point patterns can provide new insight into tumor progression and shed light on the biological mechanisms of cancer. In this talk, I consider the problem of modeling spatial correlations among three commonly seen cells (i.e. lymphocyte, stromal, and tumor) observed in tumor pathology images. Two novel spatial models of marked point patterns, with interpretable underlying parameters (some of which are clinically meaningful), are proposed in a Bayesian framework. Markov chain Monte Carlo (MCMC) sampling techniques, combined with the double Metropolis-Hastings (DMH) algorithm, are used to sample from the posterior distribution with an intractable normalizing constant. A case study is conducted on the pathology images of 188 lung cancer patients from the National Lung Screening Trial. The results show that the spatial correlation between tumor and stromal cells predicts patient prognosis. This statistical methodology not only presents a new model for characterizing spatial correlations in a multi-type spatial point pattern but also provides a new perspective for understanding the role of cell-cell interactions in cancer progression.
Keywords: Potts model; Spatial point pattern; Lung cancer; Markov random field; Pathology image; Multitype point pattern; Spatial correlation

15:45-16:00 Break

16:00-17:00 Collaborative Case Studies II

Adam Staffaroni (UCSF) and John Kornack (UCSF)

Using brain atrophy measures to predict dementia onset in familial frontotemporal lobar degeneration

Chair: *Weining Shen*

Frontotemporal lobar dementia (FTLD) is a neurodegenerative disorder characterized by relatively early age of onset and, in many cases, rapid progression to disability and death once symptoms manifest. Some models of therapy envision starting treatment before symptoms develop. The potential to demonstrate that such treatments are effective requires accurate knowledge of when symptoms would have started without treatment. 30-40% of FTLD is considered familial (f-FTLD), and several causative mutations have been identified, the three most common being mutations in the *MAPT*, *GRN*, and *C9orf72* genes. Although all mutations are highly penetrant, such that the likelihood of symptoms is close to 100%, the age of onset can vary dramatically within and between mutation type and also within a family (e.g., onset in the 30's versus the 70's in the same family). In addition, the symptoms and brain atrophy pattern is heterogeneous both within and between mutation type. Consequently, unlike other familial disorders such as familial AD and Huntington's disease, there are no accurate predictors of when symptoms will develop in an individual at risk for FTLD. Two projects, Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects (LEFFTDS) and Advancing Research and Treatment in Frontotemporal Lobar Degeneration (ARTFL), were initiated a few years ago to identify better markers of disease in f-FTLD. In this presentation we will give background on the LEFFTDS and ARTFL studies and examine our progress toward being able to predict future disease progression for asymptomatic f-FTLD mutation carriers with imaging markers, in the face of heterogeneous brain atrophy patterns.

17:00-17:20

Briefings from the R Development Workshop

Chair: *John Muschelli and Amanda Mejia*

17:30-18:30 Special Topic Session I (15' each)

Chair: *Nichole Carlson*

1) ***Shiwei Lan***
   University of Illinois Urbana-Champaign

   Learning Temporal Evolution of Spatial Dependence in Brain Images

   Spatiotemporal processes are ubiquitous in our life and have been a trending topic in the scientific community, e.g. the dynamic brain connectivity study in neuroscience.There is usually complicated dependence among spatial locations and such relationship does not necessarily stay static over time.

Learning thetemporal evolution of spatial dependence (TESD) (among brain regions) can help shed more light in the underlying mechanism of some brain disease like Alzheimer.
Spatiotemporal Gaussian process (STGP) is a popular nonparametric method to model this type of data. However, the classic STGP has a covariance kernel with space and time separated, failed to characterize TESD. Even for some recent work on non-separable STGP, location and time are treated with no difference, which is unnecessarily inefficient. In this work we generalize STGP by introducing the time-dependence to the spatial kernel and letting its eigenvalues vary over time. A novel STGP model with the covariance kernel having a Kronecker sum structure is proposed and proved to be superior to the popular kernel with a Kronecker product structure.
With the proposed method, we will conduct a longitudinal analysis of high resolution (160x160) brain images ofAlzheimer's patients obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI). Adopting a Bayesian approach, we aim to: (i) to characterize the change of the brain structure and function over time; and (ii) to detect the spatial correlation between brain regions and describe its temporal evolution. Our proposed method demonstrates the advantage in effectively and efficiently characterizing TESD. See https://arxiv.org/abs/1901.04030

Keywords: Spatiotemporal Gaussian process (STGP), Temporal Evolution of Spatial Dependence (TESD), Alzheimer's Disease Brain Imaging Analysis

2) ***Yuan Wang***
Washington State University

Sub-Group Analysis using Second-Order Imaging Feature

Radiomics features are widely used for tissue characterization and classification. We are interested in a second-order Radiomics feature called Gray-Level Co-occurrence Matrix (GLCM) whic is a matrix-variate feature that uncovers the spatial dependency among neighboring pixels. GLCM has been used in texture analysis mostly through the summary statistics such as contrast, entropy, energy, and homogeneity. However, the summary statistics may result in unaware information loss. Important patterns for understanding the extent to disease are potentially masked by the summary statistics. Moreover, many textural features are sensitive to rotations of the image and/or GLCM. The lack of invariance impacts the robustness and reproducibility of feature-based approaches. In this paper, we are interested in ROI classification and clustering using GLCM directly. We propose a novel spatially weighted dissimilarity measure between GLCMs followed by ROI subtyping using integrated clustering methods. The method is applied to an Adrenal lesion study at MD Anderson Cancer Center and shows promising results on discovering latent sub-group structure.

Keywords: GLCM, subtping, graphical clustering

3) ***Jun Young Park*, Joerg Polzehl, Snigdhansu Chatterjee, Andrew Brechmann, Mark Fiecas***
University of Minnesota, Weierstrass Institute for Applied Analysis and Stochastics, University of Minnesota, Leibniz-Institute for Neurobiology, University of Minnesota

Semiparametric modeling of time-varying activation and connectivity in task-based fMRI data

In functional magnetic resonance imaging (fMRI), there is a rise in evidence that thetemporal change in the synchronization of brain activity, known as dynamic functional connectivity (dFC) or time-varying

connectivity, provides additional information on brain networks not captured by measures of connectivity that is static over time. While there have been many developments for statistical models for dFC when the study participants are at rest, there remains a gap in the literature on how to simultaneously model both dFC and time-varying activation when the study participants are undergoing an experimental task designed to probe at a cognitive process of interest. We propose a method to estimate the dFC between two regions of interest (ROI) in task-based fMRI where the activation effects are also allowed to vary over time. Our method uses penalized splines to model both time-varying activation effects and time-varying connectivity, and uses the bootstrap for statistical inference. We validate our approach using simulations and show that ignoring time-varying activation effects would lead to poor estimation of dFC. Our proposed model, called TVAAC (time-varying activation and connectivity), can estimate both the static and time-varying activation and functional connectivity. We give an empirical illustration of both time-varying activation and connectivity by using our proposed method to analyze two subjects in an event-related fMRI learning experiment.

Keywords: Bootstrap, dynamic functional connectivity, penalized splines, task-based fMRI, time-varying activation

4) ***Ipek Oguz*, Kathryn Ufford, Greg Fleishman, Paul Yushkevich, Simon Vandekar***
Vanderbilt University, Vanderbilt University, Janelia Research Campus, University of Pennsylvania, Vanderbilt University

Multi-atlas medical image synthesis

Image synthesis is an important problem in medical image analysis, with many applications such as cross-modality synthesis (e.g., input: T1w, desired output: T2w) and dataset harmonization (e.g., input: Siemens T1w, desired output: 'GE-style' T1w). We tackle this problem using a multi-atlas intensity fusion approach, borrowing ideas from the popular multi-atlas label fusion literature. We also propose intensity and geometric bias measures in order to quantify inaccuracies of the synthesis procedure when the target image is available. We present results from applications to noise reduction (input: noisy image, output: 'filtered' image), automated image in-painting and abnormality detection.

Keywords: image synthesis, multi-atlas label fusion

Tuesday June 4th, 2019

8:30 – 9:45 Invited Session IV

Bayesian modeling of EEG and FMRI data

Chair: *Jarek Harezlak*

- **Brian Hart, Stephen Malone, Mark Fiecas***
  University of Minnesota

  A Grouped Beta Process Model for Multivariate Resting-State EEG Microstate Analysis on Twins

  EEG microstate analysis is an investigation into the collection of distinct temporal blocks that characterize the electrical activity of the brain. Brain activity within each of these microstates is stable, but can switch rapidly between different microstates in a non-random way. We propose a Bayesian nonparametric model that concurrently estimates the number of microstates and their underlying behavior. We use a Markov switching vector autoregressive (VAR) framework, where a hidden Markov model controls the non-random state switching dynamics of the EEG activity and a VAR model defines the behavior of all time points within a given state. We analyze resting state EEG data from twin pairs collected through the Minnesota Twin Family Study, consisting of 70 epochs that correspond to 140 seconds of EEG data per participant. We fit our model at the twin pair level, sharing information within epochs from the same participant and within epochs from the same twin pair. We capture within twin pair similarity by using a Beta process Bernoulli process to consider an infinite library of microstates and allowing each participant to select a finite number of states from this library. The state spaces of highly similar twins may completely overlap while dissimilar twins could select completely distinct state spaces. In this way, our flexible Bayesian nonparametric model defines a sparse set of states which describe the EEG data. All epochs from a single participant use the same set of states and are assumed to adhere to the same state switching dynamics in the HMM model, enforcing within-participant similarity.

  Keywords: EEG Microstates, Time series, Bayesian nonparametric, switching VAR

- ***Ying Guo, Yikai Wang***
  Emory University | Emory University

  A Hierarchical Independent Component Analysis Method for Longitudinal Neuroimaging Studies

  In recent years, longitudinal neuroimaging study has become increasingly popular in neuroscience research to investigate disease-related changes in brain functions, to study neurodevelopment or to evaluate treatment effects on neural processing. One of the important goals in longitudinal imaging analysis is to study changes in brain functional networks across time and how the changes are modulated by subjects' clinical or demographic variables. In current neuroscience literature, one of the most commonly used tools to extract and characterize brain functional networks is independent component analysis (ICA), which separates multivariate signals into linear mixture of independent components. However, existing ICA methods are only applicable to cross-sectional studies and not suited for modelling repeatedly measured imaging data across visits. In this paper, we propose a novel longitudinal

independent component model (L-ICA) which provides a formal modeling framework for extending ICA to longitudinal studies. By incorporating subject-specific random effects and visit-specific covariate effects, L-ICA is able to provide more accurate estimates of changes in brain functional networks on both the population- and individual-level, borrow information across repeated scans within the same subject to increase statistical power in detecting covariate effects on the networks, and allow for model-based prediction for brain networks changes caused by disease progression, treatment or neurodevelopment. We develop a fully traceable exact EM algorithm to obtain maximum likelihood estimates of L-ICA. We further develop a subspace-based approximate EM algorithm which greatly reduce the computation time while still retaining high accuracy. Moreover, we present a statistical testing procedure for examining covariate effects on brain network changes. Simulation results demonstrate the advantages of our proposed methods. We apply L-ICA to ADNI2 study to investigate changes in brain functional networks in Alzheimer disease. Results from the L-ICA provide biologically insightful findings which are not revealed using existing methods.

Keywords: Indepedendent Component Analysis; Longitudinal imaging studies; EM algorithm; brain network; Alzheimer disease

- ***Amanda Mejia***
Indiana University

Fast spatial Bayesian modeling of cortical surface task activation

Cortical surface functional magnetic resonance imaging (cs-fMRI) offers several advantages over volumetric fMRI, including dimension reduction, removal of extraneous tissue types, improved alignment of cortical areas across subjects, and a simplified spatial dependence structure. These make cs-fMRI attractive for use in spatial Bayesian models. However, the triangular mesh format of cs-fMRI introduces computational and modeling challenges compared with the regular lattice structure of volumetric fMRI. Here, we propose a spatial Bayesian general linear model (GLM) for cs-fMRI task activation, which to our knowledge is the first of its kind. We employ several recent advances in spatial statistics and Bayesian computation through integrated nested Laplacian approximation (INLA) to accurately and efficiently estimate latent task activation fields. To identify regions of activation, we propose a novel joint posterior probability map (PPM) method using an excursions set approach, which avoids multiple comparisons. Finally, we propose an efficient multi-subject modeling approach¬†to facilitate group inference. The proposed Bayesian GLM is validated and compared with the classical "massive univariate" GLM through simulation studies, task fMRI studies from the Human Connectome Project, and a study of Amyotrophic lateral sclerosis (ALS).

Keywords: spatial statistics; Bayesian smoothing; integrated nested Laplace approximation; stochastic partial differential equation; functional magnetic resonance imaging

9:45- 10:00 Break

10:00 – 11:00 Keynote II:
Martin Lindquist (Johns Hopkins Bloomberg School of Public Health)

How to lie with fMRI

Chair: *Lexin Li*

In this talk we deconstruct several types of fallacious arguments that are commonly made based on functional magnetic resonance imaging data. In the spirit of the classic book, Darrell Huff, "How to Lie with Statistics" we describe a number of tricks to make your results look specific, strong, and compelling, and also to make them appear precisely as your theory predicted. This will help ensure that your next paper will appear in a top journal.

11:00 – 11:45 Invited Session V
Functional analysis of PET imaging and EEG data
Chair: *Armin Schwartzman*

- ***Todd Ogden*, Denise Shieh***
  Columbia University, Columbia University

  Distance-based statistics in PET imaging

  In PET imaging for receptor mapping, the kinetic behavior of a radiotracer at any brain location is completely characterized by the "impulse response function" (IRF) which specifies the rate at which the radiotracer exits the system at each time point during the scan. Traditionally, this function has been estimated by fitting parametric kinetic models to the imaging data. An alternative to this is to take a nonparametric approach to estimating the IRF, which allows us to avoid specifying a model and making the requisite assumptions. Often it is of interest to compare these nonparametrically estimated IRFs between individuals or across regions/voxels, and for this purpose we propose some new distance metrics that are based in shape analysis. We will demonstrate how this approach can be used in analysis of PET data.
  Keywords: PET Imaging

- ***Damla Senturk*, Aaron Scheffler, Donatello Telesca, Catherine Sugar, Shafali Jeste, Abigail Dickinson, Charlotte DiStefano***
  University of California, LA

  Covariate-Adjusted Region-Referenced Generalized Functional Linear Model for EEG Data

  Electroencephalography (EEG) studies produce region-referenced functional data in the form of EEG signals recorded across electrodes on the scalp. It is of clinical interest to relate the highly structured EEG data to scalar outcomes such as diagnostic status. In our motivating study, resting state EEG is collected on both typically developing (TD) children and children with Autism Spectrum Disorder (ASD) aged two to twelve years old. The peak alpha frequency (PAF), defined as the location of a prominent peak in the alpha frequency band of the spectral density, is an important biomarker linked to neurodevelopment and

is known to shift from lower to higher frequencies as children age. To retain the most amount of information from the data, we consider the oscillations in the spectral density within the alpha band, rather than just the peak location, as a functional predictor of diagnostic status (TD vs. ASD), adjusted for chronological age. A covariate-adjusted region-referenced generalized functional linear model (CARR-GFLM) is proposed for modeling scalar outcomes from region-referenced functional predictors, which utilizes a tensor basis formed from one-dimensional discrete and continuous bases to estimate functional effects across a discrete regional domain while simultaneously adjusting for additional non-functional covariates, such as age. The proposed methodology provides novel insights into differences in neural development of TD and ASD children. The efficacy of the proposed methodology is investigated through extensive simulation studies.
Keywords: Autism spectrum disorder, electroencephalography, functional data analysis, peak alpha frequency, penalized regression

11:45 – 12:30 Invited Session VI

Advances in Hypothesis Testing of Brain Imaging data

Chair: *Zhaoxia Yu*

- ***Simon N Vandekar*, Theodore D Satterthwaite, Cedric H Xia, Azeez Adebimpe, Kosha Ruparel, Ruben C Gur, Raquel E Gur, Russell T Shinohara***
Vanderbilt University, University of Pennsylvania, University of Pennsylvania, University of Pennsylvania, University of Pennsylvania, University of Pennsylvania, University of Pennsylvania

Robust Spatial Extent Inference with a Semiparametric Bootstrap Joint Testing Procedure

Spatial extent inference (SEI) is widely used across neuroimaging modalities to adjust for multiple comparisons when studying brain-phenotype associations that inform our understanding of disease. Recent studies have shown that Gaussian random field (GRF) based tools can have inflated family-wise error rates (FWERs). This has led to substantial controversy as to which processing choices are necessary to control the FWER using GRF-based SEI. The failure of GRF-based methods is due to unrealistic assumptions about the spatial covariance function of the imaging data. The permutation procedure is the most robust SEI tool because it estimates the spatial covariance function from the imaging data. However, the permutation procedure can fail because its assumption of exchangeability is violated in many imaging modalities. Here, we propose the (semi-) parametric bootstrap joint (PBJ; sPBJ) testing procedures that are designed for SEI of multilevel imaging data. The sPBJ procedure uses a robust estimate of the spatial covariance function, which yields consistent estimates of standard errors, even if the covariance model is misspecified. We use the methods to study the association between performance and executive functioning in a working memory fMRI study. The sPBJ has equal or superior power to the PBJ and permutation procedures while maintaining the nominal type 1 error rate in reasonable sample sizes.

We provide an R package https://github.com/simonvandekar/pbj to perform inference using the PBJ and sPBJ procedures.
Keywords: Spatial extent inference, FWER, semiparametric inference, bootstrap, Neuroimaging

- ***Armin Schwartzman***
  University of California, San Diego

  Do not test for activation in fMRI but estimate the regions of activation

  Null hypothesis testing lies at the foundation of human brain mapping as the core method for fMRI inference. However, recent studies have shown that under optimal conditions the null hypothesis is never true, and brain activity related to a task can be found everywhere in the brain. Rather than testing for significance, we propose to directly estimate the spatial extent of interesting brain activity, defined as excursion sets of the percentage BOLD signal change above a pre-defined threshold. The uncertainty in the estimates is then captured by a nested pair of spatial confidence regions (CRs) called inner and outer sets. These spatial CRs are defined in such a way that the true excursion sets include the inner set and are included in the outer set with a given confidence. Asymptotic coverage probabilities may be determined using the Gaussian kinematic formula or via a multiplier bootstrap. The method is illustrated in task fMRI data from the Human Connectome Project.
  Keywords: fMRI, Activation

12:30 – 13:30 Lunch

14:00-15:00 Collaborative Case Studies III

Chun Fan (UCSD) and Wes Thompson (UCSD)

Imaging Genetics:

Emerging Results from the Adolescent Brain and Cognitive Development (ABCD) Study

Chair: *Seonjoo Lee*

Genome-wide association studies (GWAS) have undergone a rapid evolution over the last 5-10 years. Complex traits shown to be heritable from familial studies (e.g., cognition, psychopathology, brain structure and function) have been discovered to be highly polygenic, with each genetic locus explaining only a tiny amount of variance. In fact, the "ominigenics" hypothesis states that almost every assayed locus in GWAS is correlated with a causal SNP for complex traits like schizophrenia diagnosis. These results has spurred 1) ever larger GWAS sample sizes, and 2) the development of novel statistical methods that can estimate and partition the total effects from millions of genetic loci simultaneously. In this talk, we draw parallels between the current state of GWAS research and population neuroscience studies, which are also moving towards large sample sizes (e.g., the UK Biobank and the ABCD Study). We describe some novel approaches developed in GWAS and how these are relevant to brain imaging data from population neuroscience. We apply some of these methods to integrate genetic and imaging data from the ABCD Study.

15:00-15:15 Break

15:15-16:30 Special Topic Session II (15' each)

Chair: *Mark Fiecas*

1) ***Suprateek Kundu***
Emory University

Bayesian Network Manifold Regression for Neuroimaging Applications

Network valued data commonly arise in areas such as neuroimaging, genetics and social sciences. Although networks contain rich information, there have been limited advances in regression approaches involving network valued covariate. The high dimensionality of the networks often results in models with inflated number of parameters leading to computational burden and inaccurate estimation. Alternative approaches seek to reduce network dimension and then use the low-rank structure in prediction. This class of methods often lack interpretability and have reduced exploratory value, and are not equipped for handling network-valued covariates with complex dependent structures. In this work, we develop a novel two stage Bayesian framework to find a node-specific low-rank representation for the network-valued covariates and then use a flexible non-linear regression framework for prediction. The approach, which is related to Gaussian process latent variable models, results in a dramatic reduction in the number of regression parameters and is able to maintain interpretability at the node level. It also enables feature selection at the level of the nodes. The computation is realized through efficient an EM algorithm and a Gibbs Sampler. We evaluate our performance in prediction and inference via extensive simulation studies. Our motivating application involves data from trauma-exposed individuals from a PTSD study, where the goal is to predict clinical PTSD phenotypes based on the brain networks and their complex interactions with environmental exposures and demographic factors. The proposed approach shows considerable improvements over linear approaches as well as alternative non-linear approaches with or without dimension reduction.

Keywords: Neuroimaging; manifold regression; Gaussian process latent variable models

2) ***Yuansi Chen***
University of California, Berkeley

The DeepTune framework for deep models interpretation in visual cortex V4

Deep neural network models have recently been shown to be effective in predicting single neuron responses. Despite their high predictive accuracy, these models are generally difficult to interpret. This limits their applicability in characterizing V4 neuron function. Here, we propose the DeepTune framework as a way to elicit interpretations of deep neural network-based models of single neurons in area V4. V4 is a mid-tier visual cortical area in the ventral visual pathway. Its functional role is not yet well understood.

First, I will describe how we reliably infer about the functional properties of neurons in visual cortex via the stability-driven DeepTune modeling framework. Given high performing predictive models with various architectures, I will discuss questions such as: What can we learn from these predictive models to infer properties of neurons? How much shall we trust the model-based interpretations?
Keywords: V4, visual cortex, interpretation, CNN, deep learning, stability

3) ***Anass El Yaagoubi Bourakna***
Biostatistics Group, KAUST

Low dimensional stationary subspace representation of high dimensional time series with applications to brain signals

Multivariate brain signals appear often as realizations of nonstationary processes that can be modeled as a linear mixture of latent processes. These latent processes tend to include both stationary and nonstationary components. Stationary subspace analysis (SSA) aims at factorizing the observed multivariate signal into stationary and nonstationary components.
Existing SSA methods involve a non-convex optimization problem that becomes computationally intractable in higher dimensions. In this talk, we discuss computational strategies to overcome this issue. First, we investigate a community structure based SSA strategy that breaks down the problem into smaller dimensions. Second, we propose a momentum based gradient descent approach to speed up the convergence of the optimization process. Third, to deal with multiple solutions that may arise from multiple epochs we consider clustering techniques involving canonical angles between subspaces, which can be used as a measure of distance between the recovered stationary space and the true stationary space. We apply these techniques to analyze local field potentials of rats to assess the impact of an induced stroke on the brain.

This is joint work with Ron Frostig (UCI, Neuroscience), Raanju Sundararajan (Biostatistics Group, KAUST) and Hernando Ombao (Biostatistics Group, KAUST).

Keywords: Stationary Subspace Analysis, Frequency-SSA, Grassmann manifold optimization, Momentum gradient descent on Grassmann manifolds

4) ***Hyun Hung,Yanling Liu,Curtis Lisle****
Frederick National Lab for Cancer Research,Frederick National Lab for Cancer Research, KnowledgeVis,LLC

Training Deep Neural Networks With Noisy Pixel-Level Annotations Using A Modified Co-Teaching Method

In medical image analysis, ground truth annotations are often expensive and time-consuming to generate and frequently can be noisy and inconsistent. Training a deep neural network using noisy pixel-level annotations is challenging, as the randomness in the ground truth annotations tends to limit the training accuracy a deep neural network can achieve. In this talk, we show how we addressed this limitation by extending a previously reported ,Co-teaching, paradigm, where two networks work together, to robustly train deep neural networks under noisy pixel-level annotations for segmenting lung carcinoma in whole-slide images. Previously, the Co-teaching technique has been applied on image-level annotations, but not to the pixel-level segmentation task. A Fully convolutional DenseNet is used as a backbone network structure and two deep neural networks are trained simultaneously to communicate with each other to refine the noisy annotations. Our approach was implemented and tested for segmenting lung carcinoma in 50 stained whole-slide images and resulted in a 0.7552 dice coefficient, higher than we could achieve using only a single deep network. This result earned our team a top ten finish in the 2019 IEEE ISBI grand challenge. As part of this talk, we are hoping to engage the statistics community as research partners to help us perform model diagnostics and better understand the results of these deep learning approaches.
Keywords: Noisy label, deep learning, modified Co-teaching, segmentation, lung cancer

5) ***Robin Yancey***
University of California, Davis

Machine Learning Methods for Image Tampering Detection

In order to protect political photos and maintain research integrity or reproducibility, image manipulation detection is a highly necessary tool [1]. With technological advances leading to an increase in mechanisms of image tampering, our fraud detection methods must continue to be upgraded to match their sophistication. Most current methods require prior knowledge of the method of forgery in order to determine which features to extract from the image to localize the region of interest.For example, detection of copy-paste fraud, added WGN (White Gaussian Noise), or color enhancements each require different filtering algorithms (eg. PCA, DWT, or DCT-based) which must also be applied at different sized bounding boxes depending on the size of the tampered region [2]. Additionally, the method needed often depends on the image type and extension such as JPEG, PNG, or .TIFF [3]. When a machine learning algorithm is used to learn different types tampering from a large set of various image types, (with a big enough database) we can easily determine which images are tampered by training on the entire image feature map for each image [4]. We can also show that these current methods may be parallelized to improve speed and combined to improve accuracy, but we still are left with the question of which features to train on, and how to localize the manipulation. Object detection networks such as Faster RCNN [5], which combine an RPN (Region Proposal Network) with a CNN have recently been adapted to fraud detection utilizing their ability to propose bounding boxes for objects of interest to localize the tampering artifacts. By making use of the computational powers of today,Äôs GPUs it also achieves a fast run-time and higher accuracy than the top current methods such as ELA (Error Level Analysis), NA (Noise Analysis), or CFA (Color Filter Array) [6]. Further, a bilinear version of this model which uses two streams, with the second one having an input of the image noise or compression level mask will be shown to provide even higher accuracy by adding training features from the segmented image map.

Keywords: Image Tampering, Fraud Detection, Machine Learning, Faster RCNN

16:30 Concluding Remarks