ANNUAL PETER G. HALL SYMPOSIUM: BRAIN AND DATA SCIENCE

UNIVERSITY OF CALIFORNIA, DAVIS

Department of Statistics
Room 1147, Mathematical Sciences Building

THURSDAY, MAY 03, 2018

Pre-Conference Keynote Lecture

4:00pm - 5:00pm: **Thomas Nichols** (Oxford University) 2

FRIDAY, MAY 04, 2018

08:30am - 09:00am: **Registration, Breakfast**

Welcome and Opening Remarks (Chair: **Thomas Lee**)

09:00am - 09:15am: **Elizabeth Spiller** (Dean of College of Letters and Science, UC Davis)

SESSION 1 (Chair: **Jane-Ling Wang**)

09:15am - 10:00am: **Thomas Nichols** (Oxford University) 3

10:00am - 10:30am: **Luke Bornn** (Sacramento Kings and Simon Fraser University) 3

BREAK

SESSION 2 (Chair: **Miles Lopes**)

11:00am - 11:30am: **Mark Goldman** (UC Davis) 4

11:30am - 12:00pm: **Krishna Balasubramanian** (Princeton University) 3

12:00pm - 12:30pm: **Zhou Yu** (UC Davis) 4

LUNCH

SESSION 3 (Chair: **Wolfgang Polonik**)

BREAK

SESSION 4 (Chair: **Can Le**)

Closing Remarks

5:45pm - 6:00pm: **Thomas Lee** (Chair of Department of Statistics, UC Davis)

Reception and Dinner: 6:00pm - 9:00pm

ABSTRACTS

A Journey Into Neuroimging Statistics

Thomas Nichols

Big Data Institute and Nuffield Department of Population Health, Oxford University

In this informal talk I'll introduce the speciality area that I've worked in for the past 25 years, the development of statistical methodology for human brain imaging. I'll start with a gentle introduction to the brain, how we image it and the sort of questions neuroscientists want to ask with this type of data. Then I'll discuss the type of statistical challenges neuroimaging presents. Some problems that neuroscientists face simply amount to computational ones, scaling up standard statistical models to be fit on 100,000 volume elements (voxels). Others are intrinsically statistical, like modelling neuroimaging meta-analysis data, where only 3D point patterns of locations of changes are available. Finally I'll review what I see as some of the unmet statistical challenges in brain imaging today and the ripe opportunities available for collaboration.

Challenges and Opportunities in Population Neuroimaging

Thomas Nichols
Big Data Institute and Nuffield Department of Population Health, Oxford University

Brain imaging studies have traditionally struggled to break into 3-digit sample sizes: e.g., a recent Functional Magnetic Resonance Imaging (fMRI) meta-analysis of emotion found a median sample size of n=13. However, we now have a growing collection studies with sample sizes with 4-, 5- and even 6-digits. Many of these 'population neuroimaging' studies are epidemiological in nature, trying to characterise typical variation in the population to help predict health outcomes across the life span. I will discuss some of the challenges these studies present, in terms of massive computational burden but also in ways that they expose shortcomings of existing mass univariate techniques. I will also discuss how these datasets present intriguing methodological problems heretofore absent from neuroimaging statistics. For example, the 'null hypothesis fallacy' is how H0 is never strictly true, and yet with 100,000 subjects you'll eventually find some effect even if it is meaningless. This motivates work spatial confidence sets on meaningful effect sizes (instead of thresholding test statistic images), providing intuitive measures of spatial uncertainty. I'll discuss these findings (joint with Armin Schwartzman) and other work our group had done in this area.

Possession Sketches: Mapping NBA Strategies

Luke Bornn
Strategy and Analytics, Sacramento Kings, and Department of Statistics, Simon Fraser University

We present Possession Sketches, a new machine learning method for organizing and exploring a database of basketball player-tracks. Our method organizes basketball possessions by offensive structure. We first develop a model for populating a dictionary of short, repeated, and spatially registered actions. Each action corresponds to an interpretable type of player movement. We examine statistical patterns in these actions, and show how they can be used to describe individual player behavior. Leveraging this vocabulary of actions, we develop a hierarchical model that describes interactions between players. Our approach draws on the topic-modeling literature, extending Latent Dirichlet Allocation (LDA) through a novel representation of player movement data which uses techniques common in animation and video game design. We show that our model is able to group together possessions with similar offensive structure, allowing for efficient search and exploration of the entire database of player-tracking data. We show that our model finds repeated offensive structure in teams (e.g. strategy), providing a much more sophisticated, yet interpretable lens into basketball player-tracking data. This is joint work with Andrew Miller.

Understanding Kernel-Embedding Based Goodness-of-Fit Tests

Krishna Balasubramanian
Operations Research and Financial Engineering, Princeton University

Reproducing kernel Hilbert space (RKHS) embedding of probability distributions has attracted considerable amount of attention in recent years in the machine learning community. It offers a general and flexible framework for hypothesis testing with complex non-Euclidean data structures. Two specific examples include (i) testing with graph/tree-valued data structures with applications to bioinformatics and (ii) testing with permutation-valued data structures with applications to

analyzing ranking algorithms. Despite their popularity, fairly little is known about the statistical performance of these kernel-embedding based tests. Focussing on the goodness of fit testing problem, in this talk I will first show that a vanilla version of the kernel-embedding based test is statistically suboptimal. I will then present a novel test based on a moderated embedding and show that the new approach provides minimax optimal tests for a wide range of deviations from the null and can also be made minimax adaptive over a large collection of interpolation spaces.

Inferring key network interactions underlying memory storage in a well-defined neural circuit

Mark Goldman
Department of Neurobiology, Physiology and Behavior Center for Neuroscience, UC Davis

In many neural circuits of the brain, memory over the time scale of seconds to tens of seconds has been shown to be associated with neural activity that persists in the absence of the remembered stimulus. However, despite the observation of such"persistent neural activity", relatively little is known about how actual brain circuits generate such activity in the absence of input. Working with data from a well-characterized brain region that mathematically integrates its inputs over time (in the sense of Calculus), and in the absence of input maintains a memory of the running total of recent inputs, we ask what features of the network connectivity can be directly inferred from data consisting of neural activity recordings, transient perturbations of network activity, and gross anatomical constraints. We show that this problem is challenging to address statistically because the network activity is low dimensional whereas the parameter space is much higher dimensional. Using a simple regression-based modeling framework that facilitates sensitivity analyses, we show how to formally identify which features of the network activity are, and are not, required to generate a neural circuit that integrates its inputs and maintains memory-storing activity in a manner consistent with experimental observations. This work more generally suggests a combined experimental and theoretical approach for identifying key circuit interactions underlying network computations.

Situated Multimodal Dialog Systems

Zhou Yu
Department of Computer Science, UC Davis

Communication is an intricate dance, an ensemble of coordinated individual actions. Imagine a future where machines interact with us like humans, waking us up in the morning, navigating us to work, or discussing our daily schedules in a coordinated and natural manner. Current interactive systems being developed by Apple, Google, Microsoft, and Amazon attempt to reach this goal by combining a large set of single-task systems. But products like Siri, Google Now, Cortana and Echo still follow pre-specified agendas that cannot transition between tasks smoothly and track and adapt to different users naturally. My research draws on recent developments in speech and natural language processing, human-computer interaction, and machine learning to work towards the goal of developing *situated multimodal dialog systems*. These systems can coordinate with users to achieve effective and natural interactions through understanding information from different input sources, such as vision, language and audio. I have successfully applied the proposed concepts to various tasks, such as visual dialog, social conversation, customer service, job interview training

and movie promotion. Our team recently was selected as one out of the 8 groups to compete in Amazon Alexa Prize Challenge with $250,000 (https://developer.amazon.com/alexaprize).

Identifying the multiple algorithms that contribute jointly to human learning

Anne Collins
Helen Wills Neuroscience Institute and Department of Psychology, UC Berkeley

Human learning is supported by multiple separable brain mechanisms that jointly contribute to decisions. Identifying the computations implemented by each mechanism is a difficult inference problem, because neither mechanism can be fully isolated. In this talk, I will use examples from my research to show how a combination of careful experimental design, computational modeling and measures of the brains electrical activity, can help us disentangle the multiple mechanisms involved in learning, identify their algorithmic properties, and investigate their interactions.

Network abnormalities and Interneuron Dysfunction in Alzheimers Disease

Jorge Palop
Gladstone Institute of Neurological Disease and Department of Neurology, UC San Francisco

Alzheimers disease (AD) results in deterioration of cognitive functions and abnormal patterns of neuronal network activity, but the underlying mechanisms are poorly understood. The brain relies on oscillatory rhythmic activity, generated by inhibitory interneurons, to organize information flow and precisely time the neuronal firing required for cognitive processing. We discovered that impaired inhibitory parvalbumin-expressing interneurons (PV cells) contribute to altered gamma oscillatory activity, network hyperexcitability, and memory deficits in a mouse model of AD (hAPPJ20 mice). We also identified a molecular mechanism that leads to impaired inhibitory function that involves decreased levels of the interneuron-specific and PV- cell predominant voltage-gated sodium channel Nav1.1 subunit. Reduced Nav1.1 levels were also found in AD patients. Importantly, restoring Nav1.1 levels by Nav1.1 BAC overexpression enhanced PV celldependent gamma oscillatory activity and cognitive performance in hAPPJ20 mice, revealing key functional roles for Nav1.1- and PV celldependent gamma oscillatory activity in cognition. Overall, our data support the hypothesis that altered inhibitory function critically contributes to network alterations and cognitive deficits in hAPPJ20 mice. We are now using embryonic interneuron precursors from the medial ganglionic eminence (MGE) as a source of inhibitory interneurons for cell-based therapy in mouse models of AD. During brain development, interneuron precursors are generated in the MGE and migrate into the cortex and hippocampus. MGE-derived precursors retain a remarkable capacity to migrate and integrate into neonatal and adult host brains where they mature into functional and synaptically active inhibitory interneurons. Thus, MGE-derived precursors provide a great opportunity for cell- based therapy for neurological disorders linked to impaired inhibitory function. Our data indicate that genetically modified transplants of MGE-derived interneuron precursors lead to functional restoration of brain rhythms and memory in mouse models of AD. Because alterations in inhibitory neurons and oscillatory activity are also associated with other neurological and psychiatric disorders linked to network instability, our research has implications for multiple cognitive disorders.

Neurobiological Investigation of Vocal Production Learning in the Mammalian Brain

Michael Yartsev
Helen Wills Neuroscience Institute and Department of Bioengineering, UC Berkeley

Learning a language is generally considered the crown jewel of human abilities. Yet the core question of "What is it about the human mammalian brain that allows us to learn our language?", remains unresolved. In humans, language acquisition is mediated by a process called "vocal learning". While humans are expert vocal learners, a remarkably sparse subset of mammals share this capacity and as a result, the neurobiological mechanisms of vocal learning were never studied before in the mammalian brain. To complement the remarkable research work done in the songbird and help bridge this major gap of knowledge we set out to establish the bat as a mammalian model system for studying the neurobiological mechanism of vocal learning. Here, I will present our initial efforts towards achieving this goal which included overcoming major roadblock due to the near complete absence of research efforts in this domain in mammals. These include (i) identifying the appropriate behavioral paradigms for studying the process of vocal production learning, (ii) the relevant neural circuitries which might mediate this process in the developing and adult mammalian brain and (iii) the establishments of the necessary novel technologies to support this new research direction.

Statistical Problems in Brain Functional Connectivity Analysis

Lexin Li
Division of Biostatistics, UC Berkeley

Brain functional connectivity maps the intrinsic functional architecture of the brain and reveals synchronization of brain systems through correlations in neurophysiological measures of brain activities. Accumulated evidences have suggested that it holds crucial insights of pathologies of a wide range of neurological disorders. Brain functional connectivity analysis is now at the foreground of neuroscience research, and is drawing increasing attention in the statistics field as well. A connectivity network is characterized by a graph, where nodes represent brain regions, and links represent statistical dependence that is often encoded by partial correlation. Such a graph is inferred from the matrix-valued neuroimaging data such as electroencephalography and functional magnetic resonance imaging. In this talk, we examine a number of statistical problems arising in brain connectivity analysis, including multi-graph penalized estimation, graph-based hypothesis testing, dynamic connectivity analysis, and dynamic network modeling.

Online experimental design for automated high-throughput neural circuit mapping

Shizhe Chen
Department of Statistics and Grossman Center for the Statistics of Mind, Columbia University

Recent development in neuroscience allows for the recording and manipulation of neural activities at cellular resolution in living animals. We consider the task of learning the physiological connections among neurons (i.e., synaptic connections) in vivo. However, experiments on large volumes of densely-packed neurons (e.g. cortical excitatory neurons) with two-photon optogenetics has proven challenging because of two main problems: 1) stimulation sensitivity and resolution varies across cells due to differential opsin expression and intrinsic excitability, making the precise localization of connected neurons difficult, and 2) experimental time is severely limited compared to the number of potential connections to map. We present a method which overcomes these limitations using

statistical modeling and online experimental design. To infer posterior distributions for the probability of synaptic transmission and opsin expression level of potentially connected neurons, we fit a model with three main components: a neural response model which predicts presynaptic spike rates given the power and location of stimulation targets, a connectivity model which filters presynaptic spike rates into a postsynaptic event rate, and a postsynaptic model which converts the postsynaptic event rate into electrical activity. To improve efficiency, we implement a closed-loop parallel computational system which designs batches of stimulation targets online based on fast stochastic variational inference of these posteriors. Our experimental system allows us to collect data at 20 trials per second for a large portion of experimental time while analyzing data in the background. We demonstrate the efficacy of our method in vitro and in vivo by learning connectivity in mouse cortex.

Working with the Curse of Dimensionality

Hao Chen
Department of Statistics, UC Davis

High-dimensional data and non-Euclidean data are ubiquitous nowadays. They many times exhibit counter-intuitive phenomena, making traditional approaches less effective or inapplicable. On the other hand, these counter-intuitive phenomena could lead to unique patterns, which could be helpful in developing effective approaches. In this talk, a unique pattern for high-dimensional data will be explored and discussed. This pattern is helpful in hypothesis testing and clustering for high-dimensional data.

On Robustness of Deep Neural Networks

Cho-Jui Hsieh
Department of Computer Science and Statistics, UC Davis

Recent studies have shown that deep neural networks are easily deceived by adversarial examples. In this talk, I'll discuss how to construct adversarial examples in both white-box and black-box settings, how to evaluate the robustness property of deep neural networks, and how to defend against adversarial examples.

Is Brain surveillance coming?

Fushing Hsieh
Department of Statistics, UC Davis

We develop computational techniques from domains of Combinatorial Information Theory and Machine Learning to exact patterns from 64 channels of EEG recording. The computed entropy patterns and network structural changes are able to nearly perfectly separate regimes of [eye-open vs. eye-close] X [moving vs. standing], when subjects are in a virtual landscape. Can such computing become the basis for Brain Surveillance on individuals?