samsi
NSF·Duke·NCSU·UNC

Interdisciplinary Workshop for Undergraduate Students
May 28 - June 2, 2019

TITLES & ABSTRACTS

Plenary Talk

Speaker: David Banks, SAMSI & Duke University

Title: Statistical Issues in Agent-Based Models

Abstract: Agent-based models (ABMs) are computational models used to simulate the actions and interactions of agents within a system. Usually, each agent has a relatively simple set of rules for how it responds to its environment and to other agents. These models are used to gain insight into the emergent behavior of complex systems with many agents, in which the emergent behavior depends upon the micro-level behavior of the individuals. ABMs are widely used in many fields, and this talk reviews some of those applications. However, as relatively little work has been done on statistical theory for such models, this talk also points out some of those gaps and recent strategies to address them.

Project Abstracts

I. **Project-I [Tropical Cyclones]**
Leader: Pulong Ma

Title: Predictive Modeling for Tropical Cyclones with Historical Storm Data

Abstract:
Tropical cyclones are among the most destructive natural phenomena that lead to severe damage to life and property. Much of our understanding of tropical cyclones is derived from data. IBTrACS (International Best Track Archive for Climate Stewardship, https://www.ncdc.noaa.gov/ibtracs/) provides global tropical cyclone best track data with two key climate variables maximum sustained wind speed and minimum central pressure. These datasets have spatial resolution about 10km and temporal resolution of 6 hours. The main objective in this project is to use statistics and machine learning to build predictive models for the maximum sustained wind speed using the IBTrACS data. The preliminary plan is to perform exploratory data analysis and to model the maximum sustained wind speed with linear models or generalized linear models using statistical software R. As the maximum sustained wind speed is spatially and temporally correlated, a more realistic way to improve the predictability is to incorporate spatial and spatio-temporal dependence structure using time-series models, spatial/spatio-temporal process models, and machine learning techniques such as neural networks. If time permits, students can also build predictive models for minimum central pressure.

II. **Project-II [Heart Disease]**
Leader: Wenjia Wang

Title: Subgroup analysis of the Cleveland Heart Disease

Abstract:
It is described in the literature that the maximum heart rate achieved (thalach) variable is related to cardiac mortality. In addition, some categorical variables are also used to check heart problems including chest pain type, exercise-induced angina indicator, ST depression induced by exercise relative to rest, slope of the peak exercise ST segment, the number of major vessels colored by fluoroscopy, and the heart status (normal = 3; fixed defect = 6; reversible defect = 7). In this project, our interest is to conduct subgroup analysis for the thalach as the response.

III. **Project-III [Brain Imaging]**
Leader: Xinyi Li

Title: Imaging genetics studies on Alzheimer's disease (AD)

Abstract:
The Alzheimer's Disease Neuroimaging Initiative (ADNI) collects different biosignatures (neuroimaing, demographic, genetic and cognitive measures) from hundreds early-aged and aged people, which contain complementary information for the study of AD.The goal of this project is to apply modern statistical tools to explore the association between the brain-imaging measure and other measures. The main difficulties lie in the high-dimensionality and complex spatial structures in imaging data. Students with advanced knowledge on functional data analysis, nonparametric smoothing, and/or machine learning are preferred.

IV. **Project-IV [Sampling for ODEs]**
Leader: John Nardini

Title: How to sample for equation learning methods

Abstract:
While not required, an understanding of ordinary differential equations will be helpful with this project (some knowledge of partial differential equations would allow one to work on even more interesting aspects of this project!)

Differential equation models are used to model many scientific systems, including biological population growth, stock market fluctuations, and heat transfer. Typically, a model may be fit to noisy experimental data as a means to understand the system under consideration. Recent equation learning methods aim to reverse this process and instead learn the underlying differential equation(s) from noisy data. There are many unknown aspects of these methods, however, which leaves a plethora of areas for exploration. This project will focus on determining which locations of a dataset are most informative for learning the true underlying differential equation.

V. **Project-V [Affordable Care Act]**
Leader: Nikolas Bravo

Title: Forecast the impact of the implementation of the Patient Protection and Affordable Care Act and the Health Care and Education Reconciliation Act

Abstract:
In March of 2010, Congress passed the Patient Protection and Affordable Care Act and the Health Care and Education Reconciliation Act of 2010 (ACA). The ACA and the final Medicaid eligibility rules issued by the Centers for Medicare and Medicaid Services (CMS) make major changes to the medical assistance program eligibility landscape starting in 2014. The Medicaid eligibility rules require states to reorganize their existing eligibility structures in light of the new federal structure. Many studies predict that the resulting increase in the demand of care will not be adequately supported by an increase in the supply of care in terms of total number of providers available to provide care services. The aim of this project is to forecast the impact of the implementation of the ACA on the level of demand for the state of Georgia.