Fall 2022: Hosted by Dr. Xueheng Shi

Wednesday, September 14th

Speaker:

Dr. Xiang Zhu, Department of Statistics and Huck Institutes of the Life Sciences, The Pennsylvania State University

Title:

Bayesian regression of genome-wide association summary statistics

Abstract:

Large-scale genome-wide association studies (GWAS) have markedly improved our understanding of how common variation in the human genome affects complex traits and diseases. Regression models have been widely used to analyze GWAS, but existing methods often require input data at the individual level, which are hard to obtain due to many administrative issues. Here we provide a Bayesian framework for multiple regression without the need of individual-level data. Specifically, we derive a "Regression with Summary Statistics" (RSS) likelihood function of the multiple regression coefficients based on the univariate regression summary statistics, which are easily available in GWAS. We combine the RSS likelihood with prior distributions that are specifically designed for a wide range of genetic applications, such as heritability estimation, phenotype prediction, pathway enrichment and gene prioritization. To estimate posterior distributions, we develop efficient Markov chain Monte Carlo and variational inference algorithms that scales well with millions of genetic variants. Applying RSS to a host of real-world GWAS summary statistics, we demonstrate that RSS not only achieves similar performance in settings where existing methods work, but also enables novel analyses and discoveries that existing methods cannot deliver. The software implementing RSS methods is available at https://github.com/stephenslab/rss.

Wednesday, September 21st

Speaker:

Dean Dustin, PhD Candidate, Department of Statistics at the University of Nebraska-Lincoln. Advised by Dr. Clarke.

Title:

Testing for Important Components of Posterior Predictive Variance

Abstract:

We propose a method to decompose the posterior predictive variance using the law of total variance and condition on a finite dimensional discrete random variable. This random variable summarizes various features of modeling that are used for prediction. Later, we test which terms in the decomposition are small enough to ignore. It allows to identify which of the discrete random variables are most important to prediction intervals.

The terms in the decomposition admit interpretations based on conditional means and variances and are analogous to the terms in the Cochran's Theorem (decomposition of squared error often used in analysis of variance). Therefore, the modeling features are treated as factors in completely randomized design. In cases where there is multiple decompositions, we suggest choosing the one that that gives the best predictive coverage with the smallest variance.

About the Speaker:

Dean Dustin is a PhD Candidate from Manchester, New Hampshire. Dean received his bachelor's degree in Mathematics from Plymouth State University in New Hampshire. At UNL, he has taught classes such as STAT 218, and also worked as a research assistant exploring model uncertainty and prediction. Dean plans to graduate in December 2022 and will be going into the industry to work in quantitative finance.

Wednesday, October 5th

Speaker:

Dr. Susan VanderPlas

Title:

Reproducible Science: Statistics, Forensics, and the Law

Abstract:

In science, we strive to design and create reproducible experiments that can be replicated by other researchers. We also usually require that scientists use good experimental design and approved statistical methods for results to be judged credibly. In this talk, I'll review the status quo in forensics: what studies support the use of forensic comparison methods as reliable? Are these studies reliable based on the criteria used in other disciplines? I'll discuss two major, national reports about the validity of forensic science, why statistics as a discipline is critically important to forensics - and what statisticians can do and are doing to improve forensic science.

About the Speaker:

Dr. Susan VanderPlas is an assistant professor in the Statistics Department at the University of Nebraska, Lincoln, researching the perception of statistical charts and graphs, and applying computer vision and machine learning techniques to image data. She also works with the Center for Statistical Applications in Forensic Evidence (CSAFE) at Iowa State University, developing statistical methods for examination of bullets, cartridges, and footwear.

Friday, October 14th (via Zoom)

Speaker:

Dr. Robert Lund

Title:

Changepoint Issues and Climate Controversies

Abstract:

This talk introduces changepoint issues in time-ordered data sequences and discusses their uses in resolving climate problems. An asymptotic description of the single mean shift changepoint case is first given. Next, a penalized likelihood method is developed for the multiple changepoint case from minimum description length information theory principles. Optimizing the objective function yields estimates of the changepoint numbers and location time(s). The audience is then walked through an example of a climate precipitation homogenization. The talk closes by addressing the climate hurricane controversy: are North Atlantic Basin hurricanes becoming more numerous and/or stronger?

About the Speaker:

Dr. Lund received his PhD degree in Statistics from UNC Chapel Hill in 1993. He is currently a full professor and the Chair of Department of Statistics at UCSC, before he was a Professor in the Department of Mathematical Sciences at Clemson University and the Department of Statistics at the University of Georgia. He is an elected Fellow of the American Statistical Association (2007) and was the 2005-2007 Chief Editor of the Journal of the American Statistical Association, Reviews Section. He served as the NSF Statistical Program Manager from 2016-2018. He has published about 100 refereed papers.

His research areas include time series, changepoint analysis, statistical climatology, applied probability, and stochastic process.

Monday, October 24th

Speaker:

Dr. Hanwen Huang, University of Georgia. UNL Bayes Candidate

Title:

Bayesian multilevel mixed-effects model for influenza dynamics

Abstract:

Influenza A viruses (IAV) are the only influenza viruses known to cause flu pandemics. Understanding the evolution of different subtypes of IAV on their natural hosts is important for preventing and controlling the virus. We propose a mechanism-based Bayesian multilevel mixed-effects model for characterizing influenza viral dynamics, described by a set of ordinarly differential equations (ODE). Both strain-specific and subject-specific random effects are included for the ODE parameters. Our models can characterize the common features in the population while taking into account the variations among individuals. The random effects selection is conducted at strain level through reparameterizing the covariance parameters of the corresponding random effect distribution. Our method does not need to solve ODE directly. We demonstrate that the posterior computation can proceed via a simple and efficient Markov chain Monte Carlo algorithm. The methods are illustrated using simulated data and real data from a study relating virus load estimates from influenza infections in ducks.

About the Speaker:

Dr. Huang received his PhD degree in Statistics from UNC Chapel Hill, and he currently works as an associate professor of Biostatistics at the University of Georgia. His research areas include statistical machine learning and data mining, high dimensional data analysis, Bayesian statistics, and dynamic modeling.

Monday, October 31st

Speaker:

Dr. Victor De Oliveira, University of Texas-San Antonio (Bayes Candidate)

Title:

Approximate Reference Priors for Gaussian Random Fields

Abstract:

When modeling spatially correlated data using Gaussian random fields, exact reference priors for the model parameters have been recommended for objective Bayesian analysis. But their use in practice is hindered by its complex formulation and the associated computational costs. In this work, we propose a new class of default prior distributions for the parameters of Gaussian random fields that approximate exact reference priors. It is based on the spectral representation of stationary random fields and their spectral density functions. These approximate reference priors maintain the major theoretical advantages of exact reference priors, but at a much lower computational cost. Unlike the situation for exact reference priors, we show that the marginal prior of the range parameter in the Matern correlation family is always proper, regardless of the mean function or degree of smoothness of the correlation function, and also establish the propriety of the joint reference posterior of the model parameters. Finally, an illustration is provided with a spatial data set of lead pollution in Galicia, Spain.

Thursday, November 3rd

This week's seminar will be held from 3-4pm in Hardin Hall 162.

Speaker:

Dr. Yoonsung Jung, Prairie View A&M University (Consulting Candidate)

Title:

Vision of Excellence in Statistical Consulting Service in the Statistics Department and IANR at University of Nebraska at Lincoln: Let's Realize Our Vision & Make It Happen, Together

Abstract:

National awareness of the Statistical Counseling Center is a sufficiently achievable goal. The director of consulting has educational and research passion and vision, motivates members to set successful goals, and provides support to achieve goals, making it possible to achieve the vision of a statistical counseling center. As a recommendation for achieving the vision, the modernized consulting website is the first to provide a regular practical training program while adding helpful information to the current consulting website to help clients. Then the visitor traffic will be increased. The second is to build a system that stores all data through statistical consultation and provides additional information to clients and general site visitors. All goal planning and dissemination are possible only with the cooperation of well-trained team members and support from college members.

Monday November 14th

This week's seminar will be held from 3-4pm. Please tune in via the Zoom link in your calendar invites.

Speaker:

Dr. Sanjay Chaudhuri, Department of Statistics and Data Science, National University of Singapore. (Bayes Candidate)

Title:

On an Empirical Likelihood-Based Solution to Approximate Bayesian Computation Problem

Abstract:

For many complex models studied in natural, engineering, and environmental sciences, it is nearly impossible to specify a likelihood for the observed data. Approximate Bayesian Computation (ABC) methods try to estimate such model parameters only by comparing the given observation and some replicates generated from the model for various input parameter values. No explicit relationship between the parameters and the data is postulated. In this article, we propose an empirical likelihood (EL) based solution to the ABC problem. By construction, our method is based on an interpretable likelihood (i.e. the EL) which is computed using estimating equations completely specified by the observed and the replicated data and a few well-chosen summary statistics. The proposed method can be justified through information projections on a specified class of densities. We further show that the posterior is consistent and discuss several of its favourable large sample and large replication properties. Illustrative examples from various real-life applications will also be presented.

This work is joint with Subhroshekhar Ghosh and Pham Thi Kim Cuc all from the National University of Singapore.

Spring 2023: Hosted by Dr. Xueheng Shi

Seminars will take place every Wednesday from 3pm-4pm in Hardin Hall 49 unless otherwise specified. You will find HARH 49 located in the basement of the north tower of Hardin Hall. The seminars are open to all UNL Students, Staff, and Faculty. If you would like to attend, but are not a part of our Statistics department, please fill out the information on the Webform below and you will receive a Zoom link via email.

Wednesday, February 1st

Speaker:

Dr. Liang Chen, Department of Earth and Atmospheric Sciences, University of Nebraska-Lincoln

Title:

Transitions in Precipitation Extremes in the US Midwest - Trends, Mechanisms, and Implications

Abstract:

Precipitation extremes present significant risks to Midwest agriculture, water resources, and natural ecosystems. Recently, there is growing attention to the transitions of precipitation extremes, or shifts between heavy precipitation and drought, due to their profound environmental and socio-economic impacts. In this presentation, I will discuss the trends, mechanisms, and implications of the flood-drought transitions based on observations and large-ensemble climate model simulations. Two Standardized Precipitation Index (SPI) based metrics, intra-annual variability and transitions, are used to quantify the magnitude, duration, and frequency of variability and transactions between wet and dry extremes. Climate projections from the Coupled Model intercomparison Project Phase 6 (CMIP6) suggest more frequent and rapid transitions over the Great Lakes region and northern Midwest. Seasonally more frequent transitions from a wet spring to a dry summer (or from a dry fall to a wet winter/spring) are projected to occur. To understand the role of circulation patterns in the projected changes in seasonal precipitation extremes, the k-means clustering approach is applied to the large ensemble experiments of Community Earth System Model version 2 (CESM2-LE) and ensemble projections of CMIP6. We identify two key atmospheric circulation patterns that are associated with the extremely wet spring and extremely dry summer in the US Midwest. The projected increase in springtime wet extremes and summertime dry extremes can be attributed to significantly more frequent occurrences of the associated atmospheric regimes. Particularly, the intensity of wet extremes is expected to increase mainly due to the enhanced moisture flux from the Gulf of Mexico. The seasonality of projected changes in precipitation extremes may pose increasing risks of flash drought through land-atmosphere interactions.

About the Speaker:

Areas of expertise include: Land-atmosphere interactions, Extreme events, Impacts of climate change, Climate modeling, Land use and land cover change, and Remote sensing.

Wednesday February 15th

Speaker:

Dr. Xueheng Shi, Department of Statistics, University of Nebraska-Lincoln

Title:

Changepoint Analysis in Time Series Data: Past and Present

Abstract:

Abrupt structural changes, or changepoints, can occur in many scenarios such as mean or trend shifts in time series, and coefficient changes in regressions. Changepoint analysis is crucial in modeling and predicting time series and has wide applications in various fields, including finance, climatology, and signal processing. This talk will review notable algorithms such as Binary Segmentation, Wild Binary Segmentation, and Pruned Exact Linear Time (PELT) for detecting mean shifts in time series. However, these methods require independent and identically distributed (IID) model errors, whereas time series are often autocorrelated (serially dependent) in practice. Changepoint analysis under serial dependence is a well-known challenging problem. To address this issue, we propose a gradient-descent pruned dynamic programming algorithm for finding changepoints in time series data.

This research is a collaborative effort with Dr. Gallagher from Clemson University, Dr. Killick from Lancaster University in the UK, and Dr. Lund from UC Santa Cruz

About the Speaker:

Dr. Shi is a professor at UNL for the Department of Statistics. His research areas include: time series analysis, changepoint analysis, statistical computing, signal processing, high-dimensional statistics, machine learning, stochastic process and optimization. He is interested in developing statistical methodologies, implementing computational algorithms, and exploring statistical theories.

Monday, February 20th

Speaker:

Ying Ma, University of Michigan. Statistical Geneticist candidate for the Department of Statistics.

Title:

Statistical and Computational Methods for High-Dimensional Genomics Data

Abstract:

Spatial transcriptomics technologies have enabled gene expression profiling on complex tissues with spatial localization information. The majority of these technologies, however, effectively measure the average gene expression from a mixture of cells of potentially heterogeneous cell types on each tissue location. Here, I develop a deconvolution method, CARD, that combines cell-type-specific expression information from single-cell RNA sequencing (scRNA-seq) with correlation in cell-type composition across tissue locations. Modeling spatial correlation allows us to borrow the cell-type composition information across locations, improving accuracy of deconvolution even with a mismatched scRNA-seq reference. CARD can also impute cell-type compositions and gene expression levels at unmeasured tissue locations to enable the construction of a refined spatial tissue map with a resolution arbitrarily higher than that measured in the original study and can perform deconvolution without a scRNA-seq reference. In a real data application on the human pancreatic ductal adenocarcinoma (PDAC) dataset, CARD identified multiple cell types and molecular markers with distinct spatial localization that define the progression, heterogeneity, and compartmentalization of pancreatic cancer. In addition, if time allows, I will also discuss my other methodological work on integrative differential expression and gene set enrichment analysis in scRNA-seq studies, integrative reference-informed tissue segmentation in SRT studies, and collaborative work on polygenic risk scores for common health-related exposure traits in the Michigan Genomics Initiative (MGI) cohort.

Wednesday March 8th

Speaker:

Dr. Yisu Jia, Assistant Professor, Department of Mathematics and Statistics, The University of North Florita

Title:

Trends in Northern Hemispheric Snow Presence

Abstract:

This project develops a mathematical model and statistical methods to quantify trends in presence/absence observations of snow cover (not depths) and applies these in an analysis of Northern Hemispheric observations extracted from satellite flyovers during 1967-2021. A two-state Markov chain model with periodic dynamics is introduced to analyze changes in the data in a cell by cell fashion. Trends, converted to the number of weeks of snow cover lost/gained per century, are estimated for each study cell. Uncertainty margins for these trends are developed from the model and used to assess the significance of the trend estimates. Cells with questionable data quality are explicitly identified. Among trustworthy cells, snow presence is seen to be declining in almost twice as many cells as it is advancing. While Arctic and southern latitude snow presence is found to be rapidly receding, other locations, such as Eastern Canada, are experiencing advancing snow cover.

About the speaker:

Dr. Yisu Jia is an Assistant Professor of Mathematics & Statistics at the University of North Florida.

Wednesday March 22

Speaker:

Dr. Xiucai Ding, Assistant Professor - Department of Statistics
University of California, Davis

Title:

Curse of dimensionality and PCA: 20 years on spiked covariance matrix model

Abstract:

This is a survey talk and mainly for random matrix non-experts and graduate students. High dimensional statistics has become one of the central topics in modern statistical theory. In this area, the dimension of the sample is usually divergent with or even larger than the size. Consequently, the classical estimation, inference and decision theory assuming fixed dimensionality usually lose their validity. The main technical reason is that the standard concentration results, like law of large number and central limit theorem usually fail without a substantial modification. To address these issues, random matrix theory has emerged as a particularly useful framework and tool. In this talk, I will explain the curse of dimensionality using principal component analysis. I will make a survey on the existing results and applications based on the simple and famous spiked model. This model was proposed by Iain Johnstone in 2000 and takes us more than 20 years to partially understand it. Open questions will also be discussed.

About the Speaker:

Dr. Ding has published several influential articles about probabilities, random matrix and nonstationary time series on the Annals of Statistics and IEEE Transactions on Information Theory. Please see his personal website: https://xcding1212.github.io/index.html

Wednesday March 29th

Speaker:

Dr. Maggie Niu, Associate Professor of Statistics and Director of the Statistical Counsulting Center at Penn State

Title:

Statistical Consulting and Collaboration: Open the Door to the World

Abstract:

In this talk, I will first motivate statistics students and faculty members as to why consulting and collaboration matter. Then I will introduce what an academic statistical consulting center looks like and the current status in North America. Finally, I will discuss my goals and vision for SC3L and how to measure success.

About the Speaker:

Dr. Niu received her PhD in Statistics from the University of Washington in 2010. Her research focuses on the development of statistical models that solve real world problems, especially with applications in health and social sciences. Niu has contributed to the professional society as the chair of the ASA Section on Statistical Consulting and JSM poster chair.

Wednesday, April 12th

Speaker:

Jiajie Kong, PhD Candidate, University of California Santa Cruz

Title:

Some Count Time Series Results

Abstract:

Count time series are widely encountered in practice. As with continuous valued data, many count series have seasonal properties. This talk examines some issues with seasonal count series. Our first topic uses a recent advance in stationary count time series to develop a general seasonal count time series modeling paradigm. The model constructed here permits any marginal distribution for the series and the most flexible autocorrelations possible, including those with negative dependence. Likelihood methods of inference are explored. Our modeling methods are discussed, which entail a discrete transformation of a Gaussian process having seasonal dynamics. Properties of this model class are then established, and particle filtering likelihood methods of parameter estimation are developed. A simulation study demonstrating the efficacy of the methods is presented and an application to the number of rainy days in successive weeks in Seattle, Washington is given. Our second topic quantifies how the Northern Hemisphere’s snow cover has recently changed by analyzing weekly zero-one snow presence/absence observations. Snow cover plays a critical role in the global energy balance due to its high albedo and insulating characteristics and is therefore a prominent indicator of climate change. Changing snow presence in Arctic regions could influence large scale releases of carbon and methane gas. Given the importance of snow cover, understanding its trends enhances our understanding of climate change.   Here, we find that cells with declining snow cover outnumber cells with increasing snow cover by about two to one.

Wednesday April 19

Speaker:

Dr. Lynette Smith, Associate Professor, Department of Biostatistics, University of Nebraska Medical Center

Title:

Repeat, Repeat - Replication Research in Practice

Abstract:

Replication and reproducibility are important components of scientific research. One reason that research fails to replicate is the misuse of statistical techniques. Internal replication research is where replication researchers use the original data collected from a published research study to attempt to reproduce the results and re-examine the article's methods and conclusions. We conducted a replication study of the influential 2012 publication, "Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial" by Sarah Baird and others. Results of this replication will be discussed including results of a pure replication, measurement and error analysis (MEA) and theory of change analysis (TCA). The pure replication found that other than a few minor discrepancies, the original study was well replicated. However, the randomization and sampling weights could not be verified due to the lack of access to raw data and a detailed sample selection plan. In the MEA it was found that the intervention effect on HIV prevalence was somewhat sensitive to model choice when comparing robust standard errors with sampling weights (original method) to weighted GLMM and weighted GEE. A TCA showed no effect of intervention on HIV awareness. In a second TCA, a wealth index for the schoolgirls' family was created, and the unconditional cash transfer intervention was found to be highly effective on HSV-2 prevalence when the wealth family is low. Inspired by the replication process, this research was extended to the classroom setting and developed into a final project for use in our Biostatistics 2 class. The students from the service course indicated gaining confidence in their analysis skills and over half the students indicated finding all aspects of the replication process

enjoyable with the exception of writing the final report. As the NIH is now requiring data sharing, replication studies will become easier to perform, with data becoming readily available.

About the Speaker:

Visit Dr. Smith's website here: https://www.unmc.edu/publichealth/departments/biostatistics/facultyandstaff/lynette-smith.html

Wednesday, May 3rd

Speaker:

Aleena Chanda, PhD Student at University of Nebraska-Lincoln & Dr. Bertrand Clarke, Department Chair and Professor in the Department of Statistics at University of Nebraska-Lincoln

Title:

Adaptive Prediction for Streaming Data

Abstract:

The main goal of this project is to make one-step ahead predictions in an M-open streaming data context. We partition the range of the data into intervals, observe the data until a specified time t, and form a histogram from which we obtain a prediction. To ensure a running time bound we reduce the data stream to a representative set at each time step before forming our predictor. Since we do not assume any distribution on the data, we introduce randomness via hash functions.

We verify that our proposed estimator satisfies an error bound and converges in probability and almost surely (a.s.) to a recognizable limit. Then we compare our method to predictors based on GPP's and DPP's. We find that when streaming K-means is used to find a representative set our method performs the best or ties for best among the three methods we have implemented so far.

About the Speaker:

Aleena Chanda is a PhD student in the Department of Statistics, University of Nebraska-Lincoln. She obtained her Masters in Statistics from the University of Calcutta, Kolkata, India. She is currently working under Dr. Bertrand Clarke. Her research focuses on prediction of streaming data.

About the Speaker:

Xiang Zhu has been an Assistant Professor of Statistics and Life Sciences at the Pennsylvania State University since 2020, and a Biostatistician (Without Compensation) at U.S. Department of Veterans Affairs Palo Alto Health Care System since 2018. He received his PhD in Statistics from The University of Chicago in 2017, and he was a Stein Fellow at Stanford University in 2017-2020. His research focuses on developing new statistical and computational methodology to mine large-scale and high-throughput genomic data collected from diverse human populations

Candidate Seminars - Messy Data

Thursday, May 11th

This week's seminar will take place at 4:00pm in HARH 162

Speaker:

Daniel Alhassan, Missouri University of Science and Technology. Messy Data candidate for the Department of Statistics.

Title:

De novo Identification of Differentially Methylated Regions

Abstract:

Identifying differentially methylated regions (DMRs) between different biological conditions is critical for developing disease biomarkers. Although methods for detecting DMRs in microarray data have been introduced, developing methods with high precision, recall, and accuracy in determining the true length of DMRs remains a challenge. In this talk we will introduce the normalized kernel-weighted model, our main approach for identifying DMRs *de novo*, which accounts for co-methylation using the relative probe distance from “nearby” CpG sites. We will compare our approach with a popular DMR detection method via simulation studies under large and small treatment effect settings. We will also discuss the susceptibility of our method in detecting the true length of the DMRs under these two settings. Next, we will demonstrate the biological usefulness of our method when combined with pathway analysis methods on oral cancer data.

About the Speaker:

Daniel Alhassan is a PhD Candidate in Mathematics with an emphasis in Statistics in the Department of Mathematics & Statistics at Missouri University of Science and Technology (Missouri S&T). His research interests include DNA methylation analysis methods for microarray data, cure survival analysis. Daniel is also engaged in interdisciplinary research. He is particularly interested in statistical applications in life & social sciences and engineering and has coauthored three peer-reviewed articles. During the first year of his PhD studies, Daniel founded the Missouri S&T ASA Student Chapter and led the chapter's first workshop. Prior to his PhD studies, he graduated with a MS in Mathematics from the University of New Orleans (UNO) and a BS in Actuarial Science from Kwame Nkrumah University of Science and Technology, Ghana. While at UNO, Daniel created a data analysis group that worked collaboratively to tackle regression analysis projects, write reports and present findings.

Monday May 22nd

This seminar will take place from 3:00-4:00pm in Hardin Hall 049. Please email the statistics department for a Zoom link.

Speaker:

Dixon Vimalajeewa, a Messy Data candidate for the Department of Statistics

Title:

Advanced Machine Learning Techniques for Processing Complex Data

Abstract:

Recent advancements in the Internet of Things (IoT), the Internet of Nano-Things (IoNT), and Information and Communication Technologies have enabled the collection of large amounts of data from previously inaccessible locations, such as the human body, at a higher sampling rate. However, the complex nature of this data presents challenges for extracting valuable insights using existing data processing techniques, including issues with scalability, interpretability, and generalizability. To overcome these obstacles, advanced machine learning techniques are required. In this talk, I will present three research schemes to address these challenges: high-frequency data analysis, distributed data processing, and mathematical modeling. I will also discuss some of the techniques that I have recently proposed. These techniques are applicable to several fields, including biomedical engineering, precision farming, and personalized healthcare. Furthermore, I will explore how these approaches can be used to unlock the full potential of the vast amounts of data being collected and extract valuable insights to drive progress in these areas.

About the Speaker:

I am currently a Postdoctoral Research Associate at Texas A&M University, working under the supervision of Professor Brani Vidakovic. I obtained my Bachelor’s degree in Mathematics and Statistics from the University of Ruhuna in Sri Lanka in 2012. In 2015, I received my Master’s degree in Computational Engineering from Lappeenranta University of Technology in Finland, under the supervision of Professors Marko Lainen and Heikki Haario. I then went on to earn my PhD in Computer Science from the School of Science at SETU, under the guidance of Professors Sasistharan Balasubramaniam and Donagh P Berry. My current research interests lie at the intersection of several data science disciplines, including wavelets, fractality and multifractality, scalable machine learning, statistical signal and image processing, and statistical inference in multiscale domains.

Thursday, May 25th

Speaker:

Bin Zhao, a candidate for the Messy Data position in the Department of Statistics

Title:

Exploring Network Analysis and Bioinformatics: Unveiling Insights into Community Structures and Genetic Variations

Abstract:

This seminar introduces cutting-edge research outcomes in network analysis, specifically focusing on community detection in censored hypergraphs. We address the challenge of missing values in network data and propose an information-theoretic approach that combines spectral analysis and a refinement step to recover community structures accurately. A one-stage semi-definite programming algorithm was also proposed, simulation of those proposed approaches was run to measure the performance. We also use real-world applications to highlight the practical utility of our method. Furthermore, we will briefly touch upon some other research areas in this seminar. We discuss insights gained from scHi-C data analysis, concretely, the distribution in scHi-C data, which enables the development of possible normalization methods. Additionally, we present our contributions to genome-wide association studies (GWAS) in Beef Cattle, investigating the interrelationship between latent factors and the related SNPs. Lastly, we showcase findings from a bioinformatics project, where we performed BLAST and annotation analyses. Our research uncovered significant differences between the studied strains, providing valuable insights into their functional variations. By emphasizing network analysis while incorporating these additional research areas, this seminar highlights the potential of statistical methods and bioinformatics approaches in addressing complex real-world problems.

About the Speaker:

Bin Zhao is a Ph.D. candidate in Statistics and a master's degree candidate in Computer Science at North Dakota State University. Bin's expertise lies in data analysis, statistical modeling, and programming. Before pursuing his Ph.D., Bin worked as a Quantitative Analyst at a fintech company, gaining valuable experience in coding and statistical analysis. Bin's research focuses on network analysis, machine learning, and computational biology, with notable contributions to community detection in censored hypergraphs and single-cell Hi-C data analysis. His work has been submitted to peer-reviewed journals and has received positive feedback. Bin's passion for research, statistical expertise, and commitment to applying statistical methods to real-world problems make him a valuable researcher in his area.

Special Seminar: Hosted by Sanjay Chaudhuri

Friday, May 19th

Speaker:

Partha Lahiri, Professor and Director, Joint Program in Survey Methodology and Professor, Department of Mathematics, University of Maryland, College Park, USA

Title:

The contribution of Joint Program in Survey Methodology (JPSM) to train graduate students in survey and data science

Abstract:

The founding of JPSM in 1993 resulted from an initiative of the United States Federal Statistical Agency heads, the head of the Office of Management and Budget's Statistical Policy Office, and the chair of the U.S. President's Council of Economic Advisors. The founders of JPSM brought together a consortium of organizations, disciplines, and researchers to provide the necessary expertise in survey methodology and official statistics. Since its inception, JPSM has grown, offering onsite Master and PhD programs, online courses, and non-credit coursework for those seeking professional development in survey methodology. We now cover three areas of specialization -- Social Science, Survey Statistics, and Data Science. The recent renaming of our Master of Science program as Master of Science in Survey and Data Science reflects the rise of nonprobability surveys and unstructured Big Data in the current state of the field, the instruction that the program provides, and the research that students conduct in their work. JPSM has increased the quality of technical staff in the U.S. Federal Statistical System and is enriching the field of survey statistics and methodology. I will end the talk by giving a brief outline of selected topics of current research interest in survey and data science.

About the Speaker:

Dr. Partha Lahiri is Professor and Director of the Joint Program in Survey Methodology (JPSM) and Professor of Department of Mathematics at the University of Maryland College Park (UMD), and an Adjunct Research Professor of the Institute of Social Research, University of Michigan, Ann Arbor. Prior to joining UMD, Dr. Lahiri was the Milton Mohr Professor of Statistics at the University of Nebraska-Lincoln. His research interests include survey statistics, Bayesian statistics, data integration, and small-area estimation. He published over 80 papers in peer-reviewed journals, delivered 17 plenary/keynote speeches and over 80 invited talks in professional meetings worldwide. Over the years, Dr. Lahiri served on the editorial board of several international journals, including the Journal of the American Statistical Association and Survey Methodology. He served on several advisory committees, including the U.S. Census Advisory committee and U.S. National Academy panel and served as consultant for international organizations such as the United Nations and the World Bank. Dr. Lahiri is a Fellow of the American Statistical Association and the Institute of Mathematical Statistics and an elected member of the International Statistical Institute. He received the 2021 SAE Award at the 63rd World Statistics Congress Satellite Meeting on Small Area Estimation in recognition of lifetime contributions to small area estimation research. More recently, Dr. Lahiri was awarded the Neyman Medal at a joint session of the 3rd Congress of Polish Statistics and 2022 International Association of Official Statistics (IAOS) held in Krakow, Poland, for outstanding contributions to the development of statistical sciences.

Homepage: https://jpsm.umd.edu/facultyprofile/lahiri/partha