

JUNE 2024



TEXAS A&M UNIVERSITY  
**Statistics**

ISSUE 6



# StatLinks

THE OFFICIAL NEWS FROM THE TEXAS A&M DEPARTMENT OF STATISTICS  
TO LINK FRIENDS AND FORMER STUDENTS

# Table of Contents

## MESSAGE FROM THE DEPARTMENT HEAD

## FACULTY

New hires, faculty promotions, faculty awards and recognitions, and emeritus faculty obituary

## STAFF

Staff promotions and new staff

## UNDERGRADUATE STUDENTS

Undergraduate students' awards and recognitions

## GRADUATE STUDENTS

Graduate students' awards and recognitions

## FORMER STUDENTS

Erin Batta and Sebastián Bravo

## EVENTS

Colloquiums, SFSN webinars, student events, faculty and staff events, Pie-A-Prof food drive, SETCASA poster competition, career night, TAMIDS sea level rise data science challenge awards, Statistical Methods for High Dimensional Complex Data Conference, Data Science Summer Camp, and other events

## OTHER ANNOUNCEMENTS

May 2024 graduates, TAMU Statistics by the numbers, 2024 Aggie Reunion and stay connected





# MESSAGE FROM THE DEPARTMENT HEAD

Dear Faculty, Staff, Students, Alumni, and Friends of the Department,

As we conclude the 2023-2024 academic year, I am thrilled to celebrate the remarkable achievements within our department. As you will see in this issue of StatLinks, this spring semester, several of our faculty members have been honored with significant research grants, awards, and recognitions. These accomplishments are a true testament to your unwavering dedication and hard work. My heartfelt congratulations go out to everyone for their outstanding contributions.

We are also delighted to welcome new members to our department. This spring, we have had the pleasure of bringing on board several new faculty and staff, each bringing unique expertise and enthusiasm. Their addition promises to

enrich our collaborative environment and propel our research and teaching efforts to new heights. I encourage everyone to extend a warm welcome and support to our new colleagues as they settle into their roles.

Finally, I want to take a moment to celebrate our May 2024 graduates. Your perseverance and commitment have culminated in this significant milestone, and we couldn't be prouder. As you embark on your next journey, whether in further education, new careers, or other endeavors, know that you carry with you the Aggie spirit and best wishes of our department. To everyone, I wish a joyful and restful summer. May this break provide you with the opportunity to recharge and prepare for the exciting year ahead.

Last, but not least, thanks to Ana Ferreira and Judith Moreno who compiled this issue of StatLinks.

Warm regards,

A handwritten signature in black ink that reads "Brani Vidakovic". The signature is fluid and cursive, with a long horizontal stroke at the end.

**BRANI VIDAKOVIC**

H.O. Hartley Chair and Department Head

# MEET OUR **NEW HIRES**

## **XIYU PENG**

Tenure-Track Assistant Professor

Research interests: interface of statistics and biomedical sciences (genomics, proteomics, and immuno-oncology).



## **ABHISHEK ROY**

Tenure-Track Assistant Professor

Research interests: non-convex optimization, uncertainty quantification, Markov Chain Monte Carlo (MCMC) sampling, multi-objective optimization, and robust learning from dependent data.



## **MARY "BETSY" LOCKHART**

Instructional Assistant Professor (APT Faculty)

Research interests: application and advancement of person-centered analytical approaches to the investigation of various STEM identities within students, and the ultimate link between the development of these identities and STEM persistence.



# FACULTY PROMOTIONS



**ALAN DABNEY**

Full Professor  
Effective 09/01/2024



**IRMA  
HERNANDEZ-MAGALLANES**

Instructional Associate Professor  
Effective 09/01/2024



**SCOTT CRAWFORD**

Instructional Full Professor  
Effective 09/01/2024



**MOHAMED ABURWEIS**

Instructional Associate Professor  
Effective 09/01/2024

# FACULTY AWARDS AND RECOGNITIONS

---



## **BANI MALLICK**

Dr. Mallick has been recognized for his 25 years of service with the Texas A&M University System. He joined the Department of Statistics as an Assistant Professor in 1998 and is now a Distinguished Professor, as well as the Director at Texas A&M NSF Research Institute for Foundations of Interdisciplinary Data Science (FIDS). Throughout his tenure, Dr. Mallick has been a model to many students and colleagues.

## **MOHAMED ABURWEIS**

Dr. Aburweis's application, "Standardizing and Enhancing the Teaching Materials Used Across All Sections of Statistical Methods Course (STAT 302)," has been selected as an awardee for the Digital Learning Faculty Support Grants totaling \$15,000.00. Also, he has received the Open Educational Resources Grant, totaling \$75,000.00.

Additionally, Dr. Aburweis has been recognized with the Department of Statistics Teaching Excellence Award.



## **RAYMOND CARROLL**

Dr. Carroll has been awarded a new grant from the National Cancer Institute (NCI) for his project on New Epidemiologic Methods for Reducing Measurement Error and Misclassification Bias in Cancer Epidemiology (2024-2028).

Research.com has ranked Dr. Carroll as the #25 mathematician internationally and #17 nationally among Best Mathematics Scientists. With over 51,000 citations and an Erdos number of 2, Dr. Carroll's work has had a profound impact on the field, solidifying his reputation as a leading figure in mathematics and statistics.



## **SHARMISTHA GUHA**

Dr. Guha was awarded a NSF-DMS (Statistics) grant of \$150,000 for her proposal “New Directions in Bayesian Heterogeneous Data Integration: Methods, Theory and Applications.”



## **TORYN SCHAFER**

Dr. Schafer has been appointed as a 2024 ConocoPhillips Data Science Faculty Fellow at Texas A&M University. She is going to be awarded \$15,000 to be used for research. This prestigious nomination comes in recognition of her outstanding contributions to the UG Certificate in Data Analytics for the Petroleum Industry.

## **ABHISHEK CHAKRABORTTY**

Dr. Chakraborty has been awarded the Teaching Excellence Award from the Department of Statistics, recognizing his outstanding commitment to the educational development of his students.





## YANG NI

Dr. Ni has been presented at the 2024 College of Arts & Sciences Award Ceremony with the Research Impact Award for Associate Professors.

This award recognizes faculty who have outstanding records of research that is transforming their discipline and/or society, and who are recognized as a leader who is advancing their research field in new directions.

## HUIYAN SANG

Dr. Sang has been honored by being named a 2024 Fellow of the American Statistical Association (ASA), a recognition of her significant contributions to the field of statistics.

As an ASA Fellow, Dr. Sang joins a select group of esteemed statisticians whose work has left a lasting mark on the discipline. The ASA Awards Ceremony, scheduled to take place during JSM, will provide an opportunity to celebrate Dr. Sang's accomplishments and further highlight her remarkable contributions to the statistical community.





---

## **PATRICIA NING**

Dr. Ning has taken on new roles as Associate Editor for two statistics journals, "Statistics and Computing" and "Journal of Computational and Graphical Statistics."

Additionally, Dr. Ning received the 2024 TAMIDS Course Development Award and secured two grants with Texas A&M University. The grants are: Individual Research Grant (2024-2025) and Team-building Grant (2024-2025).



## **RAJARSHI GUHANIYOGI**

Dr. Guhaniyogi is the recipient of the 2023 Early Career Award for Statistics and Data Sciences from the International Indian Statistical Institute.

Also, Dr. Guhaniyogi has been awarded a \$1.7 million grant as a PI from the National Institute of Health. The grant aims to support research into Primary Progressive Aphasia, a rare neurodegenerative disorder.

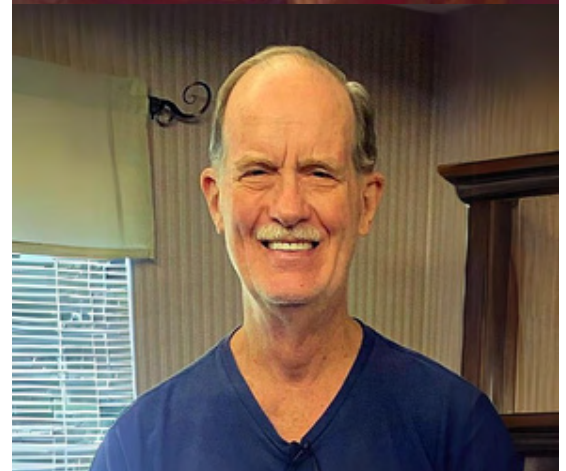
# IN LOVING MEMORY OF **FRED DAHM**

April 16, 1951 – January 18, 2024

Dr. Fred Dahm was born in Manhattan, Kansas on April 16, 1951. After earning his PhD at Iowa State University, Fred Dahm and his wife, Diane Laundry, moved to College Station, Texas. Dr. Dahm joined the Department of Statistics at Texas A&M University, where he earned tenure and dedicated 38 years of his career.

Dr. Dahm joined the department as an Assistant Professor in 1979. He then earned a promotion to Associate Professor in 1985, and later to Professor in 1993. Dr. Dahm also served as the Graduate Advisor and Program Director for 25 years. In an effort to welcome the graduate students he recruited to the Texas A&M Statistics Department, he hosted an annual Oktoberfest party at his home. Over the years this event grew in size and lore, peaking at over 100 attendees and several hundred bratwursts cooked.

In 2017, Dr. Fred Dahm retired and moved to Austin with his wife, Diane. His contributions to the department and the field of statistics are numerous and far-reaching. His mentorship has helped shape the careers of many successful statisticians and researchers.



# STAFF PROMOTIONS



## **ANDREA DAWSON**

Andrea Dawson has been promoted to Program Coordinator II. This new title more accurately reflects her comprehensive job duties, particularly in managing the graduate program. Andrea's promotion highlights her continued dedication to the university. Working in the department since March 2016, she will continue to play a crucial role in overseeing and coordinating various aspects of the graduate program.

---



## **JUDITH MORENO**

Judith Moreno has recently been promoted to the position of Senior Administrative Coordinator II within the Department of Statistics. This promotion reflects her increased responsibilities and accomplishments within the department. As Senior Administrative Coordinator II, Judith will continue to provide essential administrative support, contributing to the efficient functioning of the department.

---



## **COURTNEY SHUTTLESWORTH**

Courtney Shuttlesworth has been promoted to the position of Blocker Advising Manager, a promotion within the College of Arts & Sciences. While continuing her role as the Statistics undergraduate advisor, Courtney now manages the advising hub located in Blocker Hub Advising office, supervising advisors from multiple departments including MATH, PHYS, CHEM, and STAT. This new role highlights Courtney's dedication and expertise in the field.

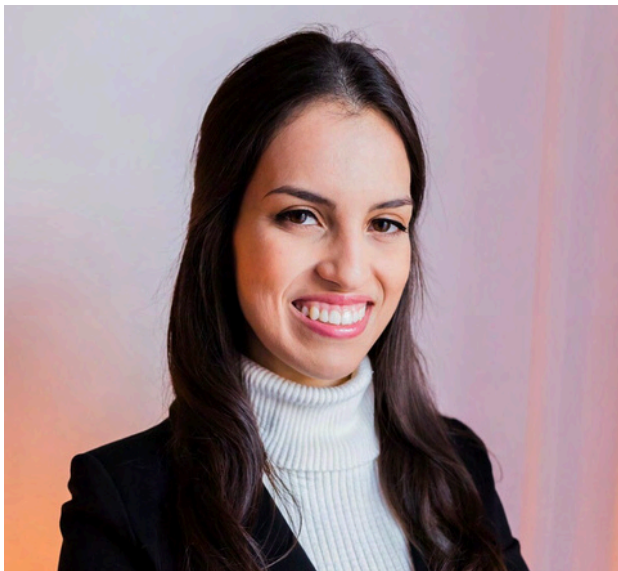


## **KIM RITCHIE**

Kim Ritchie has recently transitioned to the Dean's Office from the Department of Statistics, where she served as the Program Director for the Master of Science online program. Since 2022, she has been an integral part of the college's digital learning team, contributing her expertise and leadership. Having joined Texas A&M in 2005, Kim brings a wealth of experience to her new role, where she will lead efforts to support faculty in utilizing Canvas tools and exploring emerging educational technologies. Based in Room 226 of the Academic Building, Kim will oversee a team of two educational consultants and one multimedia specialist.

Kim's appointment as Director of Digital Learning underscores her dedication and contributions to advancing education at Texas A&M. Her tenure in the Department of Statistics and previous roles within the digital learning team have prepared her well for this new leadership position. With her guidance, the digital learning team is poised to make significant strides in enhancing the educational experience for both faculty and students. Congratulations to Kim on this well-deserved appointment, and we look forward to her continued success in this role.

# **NEW STAFF**



## **ANA FERREIRA**

Ana Ferreira stepped into the role of Administrative Coordinator I for the Department of Statistics, effective February 1st. With a background in sales, Ana quickly became an integral part of the organization, seamlessly handling various administrative tasks with precision and flair. Her proactive approach, attention to detail, and ability to adapt to new challenges make her a valuable asset to the team.

Join us in welcoming Ana Ferreira!

# MEET OUR LEADERSHIP TEAM



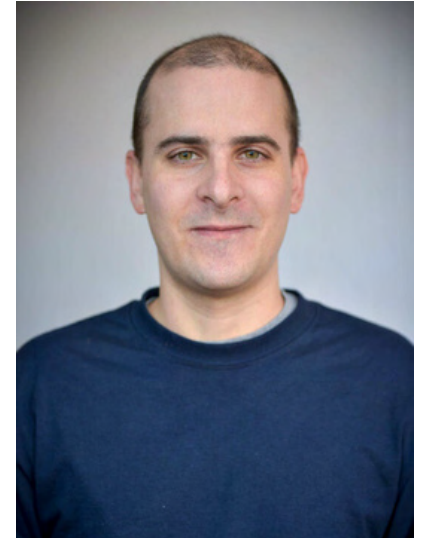
**BRANI VIDAKOVIC**

Department Head



**DAREN B.H. CLINE**

Associate Department  
Head for Operations



**ALAN DABNEY**

Associate Department Head  
for Teaching Excellence



**DEBDEEP PATI**

Director of PhD Programs



**DERYA AKLEMAN**

Director of MS Programs



**HUIYAN SANG**

Director of Undergraduate  
Programs

# UNDERGRADUATE STUDENTS

## AWARDS AND RECOGNITIONS

Nicholas Johannessen, a spring 2024 graduate from the Department of Statistics at Texas A&M University, has been awarded the 2024 Buck Weirus Spirit Award. This prestigious award, bestowed upon 65 students annually, recognizes individuals who exhibit high involvement and contribute positively to the Aggie community. Unlike other accolades, the Buck Weirus Spirit Award specifically acknowledges students who enhance student life at Texas A&M through active participation in student organizations, engagement in Aggie traditions, and involvement in university events.

Graduating with a bachelor's degree in Statistics, Johannessen's recognition for the Buck Weirus Spirit Award underscores his dedication to fostering a positive campus environment. Throughout his time at Texas A&M, Johannessen has made notable contributions to the university's community by actively participating in various student organizations and engaging in Aggie traditions. His impact has been felt across campus, reflecting his commitment to enhancing the Aggie Spirit and student life at Texas A&M University.



# GRADUATE STUDENTS AWARDS AND RECOGNITIONS



## REBECCA LEE

Rebecca Lee has recently been awarded the G.R.A.D. Aggies Advanced Professional Development Certificate and the CIRT Associate recognition for completing An Introduction to Evidence-Based Undergraduate STEM Teaching MOOC.

Additionally, she has become a Graduate Mentoring Fellow, showcasing her commitment to professional development and excellence in mentoring.

## JOSE RODRIGUEZ-ACOSTA

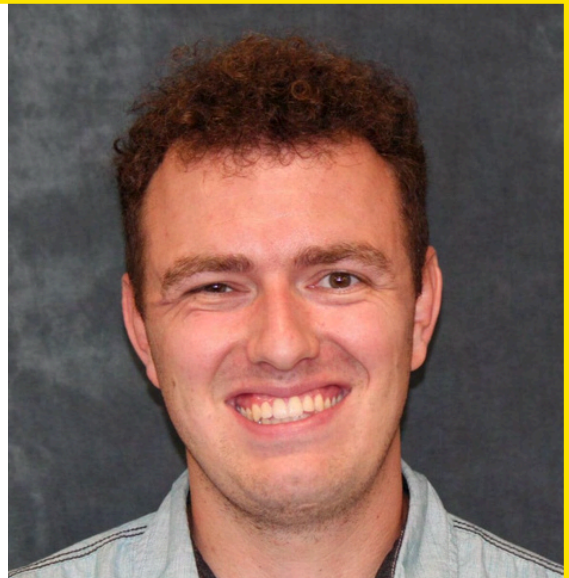
Jose Rodriguez-Acosta has been awarded the National Science Foundation Graduate Research Fellowship (GRFP).

As the oldest graduate fellowship of its kind, the GRFP has a long history of selecting recipients who achieve high levels of success in their future academic and professional careers. NSF Fellows are anticipated to become knowledge experts who can contribute significantly to research, teaching, and innovations in the field.



## JACOB ANDROS

Jacob Andros has been awarded the 2024 Ruth J. and Howard F. Newton Memorial Graduate Student Teaching Award in Statistics for his exceptional work teaching a section of STAT 302 in Fall 2023. Jacob's student evaluations were overwhelmingly positive, with many students praising his dedication, knowledge, and ability to explain complex concepts effectively.



## SAMUEL GAILLIOT

Samuel Gailliot is among the awardees of the 2024 student paper competition for the Section on Environmental Sciences (EnviBayes) of the International Society for Bayesian Analysis. He was recognized for his work on "Data Sketching and Stacking: A Confluence of Two Strategies for Predictive Inference in High-Dimensional Gaussian Process Regressions with Air Pollution Data."



## ARHIT CHAKRABARTI

Arhit Chakrabarti won the SBSS Student Paper Competition with his submission, "Graphical Dirichlet Process for Clustering Non-Exchangeable Grouped Data." This achievement highlights his knowledge and dedication to Bayesian analysis as well as in the broader field of statistics.

As a winner, Arhit will present his work at JSM 2024, where he will also receive a certificate to formally acknowledge his achievement.





# FROM NUMBERS TO IMPACT: THE POWER OF STATISTICAL ANALYSIS

**Erin Batta '24** entered Texas A&M University in 2020 with a keen interest in statistics. Although she was drawn to the university's renowned statistics program, she had yet to pinpoint a specific area of focus. However, everything changed when she received an invitation to join a research project that illuminated the potential of statistics to effect meaningful change.

During the summer before her sophomore year, Erin became involved in a research project examining the efficacy of the First Year Eats (FYE) program, part of the Texas A&M Hunger Consortium. This initiative, established in 2019 by LAUNCH, aimed to combat food insecurity on campus by providing students with access to resources like Crockpot meals, Monday meals, and the FYEats Store. Erin's role in the project was to investigate how FYE impacted students beyond their freshman year.

Under the mentorship of Dr. Alan Dabney and Dr. Sumana Datta, Erin meticulously planned and executed her research project. "She incorporated some new survey tools that allowed us to explore novel aspects of student wellbeing," Dr. Dabney said. Erin discovered compelling results by interviewing sophomores who did and did not participate in FYE during their freshman year. Students who engaged with FYE demonstrated improvements in eating habits, higher GPAs, and decreased likelihood of leaving school, highlighting the program's significant positive impact.



**"Erin's work was excellent and significantly contributed to the overall success of the FYE project, an initiative that everyone involved is proud of."**

For Erin, this research project was a turning point. Not only did it provide her with a clear career direction, but it also revealed a newfound passion for using statistics to effect positive change. As she continues her academic journey, Erin is committed to exploring new ways to apply statistics to improve lives.

"Erin's work was excellent and significantly contributed to the overall success of the FYE project, an initiative that everyone involved is proud of," Dr. Dabney said.

Erin's story exemplifies the transformative potential of statistics and the importance of practical, hands-on learning experiences in guiding one's career.



# STATISTICS

## SEBASTIÁN BRAVO: NAVIGATING THE ONLINE LANDSCAPE AS AN AGGIE

**Sebastián Bravo '23** never envisioned attending a university in Texas, especially after establishing a career at General Motors (GM) Financial. However, feeling stagnant in his role, Bravo discovered Texas A&M University's Online Statistics Master's Program, which he credits for providing him with the skills necessary for professional growth. Born in Bogota, Bravo earned his undergraduate degree in Chemical and Industrial Engineering from Universidad de Los Andes. Despite preparing for a career in engineering, he found more growth opportunities in the financial sector, leading him to work as a sales and marketing analyst for GM in 2015.

Joining GM's international rotations program and relocating to Charlotte, N.C., Bravo advanced to Senior Data Analyst, sparking a newfound interest in the data science department. To continue his career growth, he realized the need for further education. Bravo obtained certifications and technical skills but felt the need for a master's degree related to data science.

After researching various programs, he chose Texas A&M's program for its rigor and comprehensive curriculum, which included mathematics, theory, and methods simultaneously.

The program's strong reputation and high rankings solidified his decision, making it the best distance-friendly program he found.

During his two years in the program, Bravo experienced significant career growth, applying what he learned in class to his work. The program's structure and support allowed him to successfully balance work and studies, enhancing his collaborative and time-management skills. Bravo's experience at Texas A&M extended beyond education, providing him with a valuable network and skills that continue to benefit him in his career as a machine learning engineer and as a member of the Ph.D. program in Industrial Engineering at Wayne State University.

“

**Texas A&M's program was much more rigorous. We studied math, theory and methods simultaneously. Being able to master the underlying theory behind concepts is what set the program apart for me. The Department of Statistics at Texas A&M being so well-ranked also solidified my decision and made it the best distance-friendly program out of every single one I looked up.**

- Sebastián Bravo

”

S P R I N G  
2 0 2 4

**E V E N T S**

---

J A N U A R Y | F E B R U A R Y | M A R C H  
A P R I L | M A Y | J U N E



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

### ALEXANDER GIESSING

Assistant Professor, Department of Statistics,  
University of Washington

**WEDNESDAY, JANUARY 24, 2024 | 11:30 AM | BLOCKER 150**

### **Inference on High-Dimensional Data: Efficient and Robust Strategies**

How can we conduct inference and uncertainty quantification on high-dimensional data when classical asymptotic theory fails? In this talk, I discuss two complementary approaches to answer this question: First, I introduce a new framework for inference on regression functions in the presence of high-dimensional nuisance parameters. At its core is a convex program that aims to minimize the mean squared prediction error of the regression function by trading off bias and variance optimally. The resulting regression function is asymptotically normal and semiparametric efficient "in a weak sense". I will focus on a specific implementation of this framework for estimating the conditional mean when outcomes are missing at random. Second, I discuss recent advances on the high-dimensional bootstrap. I propose a new bootstrap procedure that takes the correlation structure of the data into account and, under assumption on the correlation structure, allows consistent bootstrapping of the sampling distribution of supremum-type statistics. These theoretical results generalize the existing Gaussian multiplier bootstrap and broaden its applicability for inference on high-dimensional data. They also provide new intuition for when the high-dimensional bootstrap works or does not.

### **Biography**

Dr. Giessing received a B.Sc. in Economics from the University of Bonn, Germany in 2011, a M.Sc. in Econometrics and Mathematical Economics from the London School of Economics and Political Science in 2012, and a Ph.D. in Statistics from the University of Michigan, Ann Arbor in 2018. Before joining the faculty at the University of Washington in 2021, he was a Postdoc in the Department of Operations Research and Financial Engineering at Princeton University. Dr. Giessing is broadly interested in inference on high-dimensional data; in particular, in Gaussian and bootstrap approximations and semiparametric efficient inference.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

**ABHISHEK ROY**

Postdoctoral Researcher, Halıcıoğlu Data Science  
Institute, University of California, San Diego

**FRIDAY, JANUARY 26, 2024 | 11:30 AM | BLOCKER 150**

### **Breaking Independence in Statistical Inference of Online Stochastic Algorithms**

Online optimization algorithms like Stochastic Gradient Descent (SGD) are the main workhorse behind most machine learning tasks with huge datasets. Here we will concentrate on the case of decision-dependent data where the data-distribution is strategically or adversarially modified based on the outputs (decisions) of the algorithms. In this case, an equilibrium decision is of interest which remains optimal even after the data is strategically modified. While most works focus on the prediction accuracy of algorithms in this setting, I will present our new results on online statistical inference of algorithmic estimators for such equilibrium decisions. Specifically, we will focus on two research vignettes. In the first part, I will talk about online covariance estimation for SGD to construct a confidence interval for the equilibrium point under state (decision)-dependent Markovian data. To this end, we establish the convergence rate, which matches with the rate for i.i.d data ignoring logarithmic factors, of an online overlapping batch-means covariance estimator. We apply our method to strategic classification with logistic regression, where adversaries adaptively modify features during training to affect target class classification. The second part of my talk focuses on characterizing the asymptotic randomness of an algorithmic estimator of the saddle point in a stochastic min-max optimization with state-dependent Markovian data. These optimization problems arise in multitask strategic classification, relative profit maximization in competitive markets such as electric vehicle charging and ride-sharing platforms, and robust strategic regression. We show that the averaged iterate of the Stochastic Extra-gradient (SEG) Algorithm converges almost surely to the equilibrium saddle point of a globally convex-concave and locally strongly-convex strongly-concave objective, and is asymptotically normal. To conclude, I will present a brief overview of my future research.

### **Biography**

Dr. Abhishek Roy is a postdoctoral researcher at the Halıcıoğlu Data Science Institute, University of California, San Diego. Abhishek finished his Ph.D. in June 2020 at the University of California Davis. From 2020-2022, he was an NSF-tripods postdoc in the Department of Statistics, and the Department of Computer Science at UC Davis. He completed his B.Tech (Hons.) at the Indian Institute of Technology, Kharagpur in 2013. The underlying theme of Abhishek's research has been sequential decision-making in statistical problems with multi-faceted goals such as robustness, fairness, sparsity, or even more general constraints with an emphasis on applications with decision-dependent (adaptively collected) and non-i.i.d data. Abhishek's research specifically focuses on non-convex optimization, uncertainty quantification, Markov Chain Monte Carlo (MCMC) sampling, multiobjective optimization, and robust learning from dependent data.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

**ANNA BYKHOVSKAYA**

Assistant Professor, Department of Economics,  
Duke University

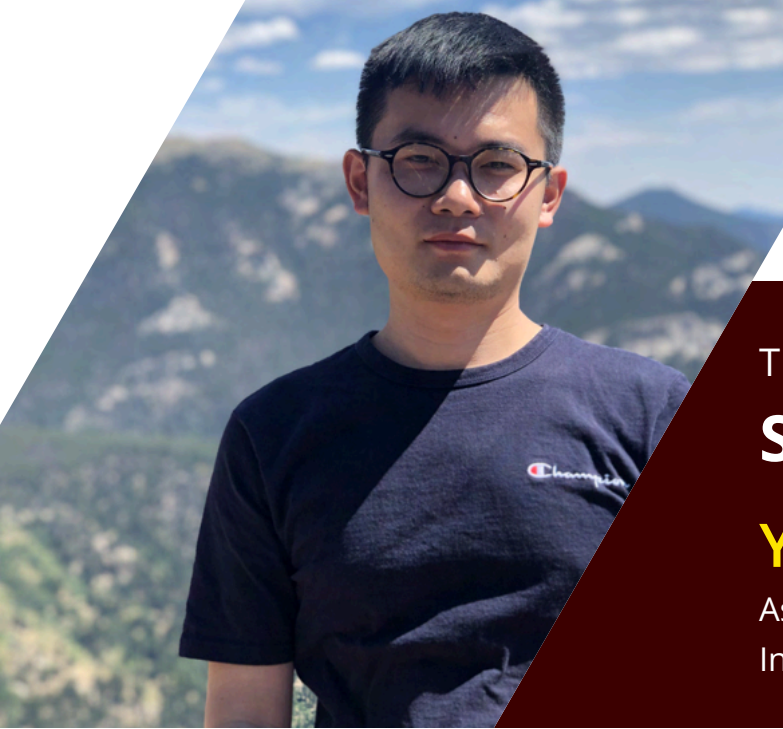
**FRIDAY, FEBRUARY 2, 2024 | 11:30 AM | BLOCKER 150**

### **Dynamic Tobit Model**

In this talk we will discuss the dynamic Tobit model, which can be used to model censored time series. We will start with characterizing its asymptotic behavior. Turns out, the asymptotics has a point of discontinuity (analogue of a unit root for linear autoregression). We will then extend local to unity asymptotics to the non-linear setting of the dynamic Tobit model, motivated by this discontinuity as well as by the application of this model to highly persistent censored time series. We will present a unit root test in the presence of censoring, and we will provide an application of our methods to the Swiss franc / euro exchange rate, during a period when this was subject to an occasionally binding lower bound.

### **Biography**

Anna Bykhovskaya is an Assistant Professor at Duke University. She received a PhD in Economics from Yale University in 2019. Prior to Yale, Anna spent 5 years at the Moscow State University, where she earned BS and MS in Mathematics. She was mostly specializing in probability theory. Anna's main research interests lie in the field of time series and panel data econometrics. In particular she focuses on the issues related to nonlinearities, non-stationarities, and high dimensionality.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

**YUZHOU CHEN**

Assistant Professor, Department of Computer and Information Sciences, Temple University

**WEDNESDAY, FEBRUARY 7, 2024 | 11:30 AM | BLOCKER 150**

### **Explainable Topology-Guided Machine Learning for Spatio-Temporal Data and Beyond**

In many real-world applications such as intelligent transportation, biosurveillance, climate science, and bioinformatics, statistical models and machine learning algorithms are applied to large-scale spatio-temporal data. Such data are typically high-dimensional and demonstrate complex nonlinear spatial and temporal dependencies. In recent years, graph machine learning (GML) has emerged as a powerful machinery to harness the rich information encoded in various spatio-temporal datasets. However, the existing models still tend to be insufficient to handle complex structural phenomena exhibited by spatio-temporal processes and do not explicitly account for time-conditioned properties of the encoded knowledge. In this talk, I will demonstrate how our innovative approaches harnessing the interdisciplinary strengths of topological data analysis, statistics, and machine learning allow us to tackle these limitations across a spectrum of spatio-temporal applications. In particular, I will focus on explainable topology-guided machine learning models for spatio-temporal processes that incorporate time-aware shape descriptors and discuss how pushing forward the performance boundary of GML models can assist in decision-making under uncertainties. I will showcase the applications of our models to challenging problems in traffic forecasting, infectious disease prediction, Ethereum blockchain price prediction, and wildland fire detection. Finally, I will conclude with an overview of future research directions at the nexus of statistics, simplicial geometry, and artificial intelligence as the driving force behind the emerging translational and use-inspired research thrusts, ranging from climate justice to open knowledge graphs to digital twins.

### **Biography**

Yuzhou Chen is an Assistant Professor in the Department of Computer and Information Sciences at Temple University. He is also a Visiting Researcher in the Department of Electrical and Computer Engineering at Princeton University. Before joining Temple, he worked as a postdoctoral scholar in the Department of Electrical and Computer Engineering at Princeton University. He received his Ph.D. in Statistics at Southern Methodist University in 2021. His research focuses on geometric machine learning, topological data analysis, knowledge discovery in graphs and spatio-temporal data, with applications to biosurveillance, energy systems, intelligent transportation, cryptocurrency, and environmental sciences. His research has appeared in the top machine learning and data mining conferences and journals, including ICML, ICLR, NeurIPS, AAAI, PNAS, etc. He won 2022 and 2021 Best Paper Awards of the Section for SDNS of ASA and the 2021 Chateaubriand Fellowship from the Embassy of France in the United States. His research is supported by NSF TIP, NSF DMS, and NASA AIST grants.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents  
**SPRING 2024 COLLOQUIUM**

**XIYU PENG**

Postdoctoral Fellow, Department of Epidemiology and  
Biostatistics, Memorial Sloan Kettering Cancer Center

**FRIDAY, FEBRUARY 9, 2024 | 11:30 AM | BLOCKER 150**

**Spatial Topic Modeling of Tumor Microenvironments with Multiplexed Imaging Data.**

Recent advancements in multiplexed tissue imaging allow for examination of tissue microenvironments in great detail. These cutting-edge technologies offer invaluable insights into cellular heterogeneity and spatial architectures, playing a crucial role in decoding mechanisms of treatment response and disease progression. However, gaining a deep understanding of complex spatial patterns remains challenging, primarily due to the lack of robust statistical and computational methods. In this talk, I will present SpaTopic, a spatial topic model designed to decipher high-level spatial architecture across multiplexed tissue images. This model integrates both cell type and spatial information using a topic model framework, originally developed in computer vision, facilitating the identification of complex spatial tissue structures. In the model, spatial information is incorporated into the flexible design of documents, which represent densely overlapped regions in images. The model is implemented using an efficient collapsed Gibbs sampling algorithm for inference. Contrasting to computationally intensive K-nearest-neighbor-based cell neighborhood analysis approaches, SpaTopic is more scalable to large datasets. Through several case studies with multiplexed imaging data from varied types of tissue, we showcase SpaTopic's capability in effectively identifying and quantifying spatial tissue structures without human intervention and highlight its utility in revealing the dynamics in spatial architecture across different disease stages.

**Biography**

Dr. Peng is a third-year postdoctoral fellow in the Department of Epidemiology and Biostatistics at Memorial Sloan Kettering Cancer Center, working with Drs. Ronglai Shen, Katherine Panageas, and Margaret Callahan. Her current research focuses on identifying biomarkers from longitudinal flow cytometry and multiplexed imaging data of cancer patients undergoing immunotherapy to predict patients' outcomes. Prior to joining the MSK, Dr. Peng obtained her Ph.D. degree in bioinformatics and statistics from Iowa State University, working with Dr. Karin Dorman. For my Ph.D., she worked on improving fundamental bioinformatics algorithms, especially in denoising and bias correction for next-generation sequencing data.





TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents  
**SPRING 2024 COLLOQUIUM**

**JIANQING FAN**

Professor of Statistics, Machine Learning, Operations Research, and Financial Engineering at Princeton University

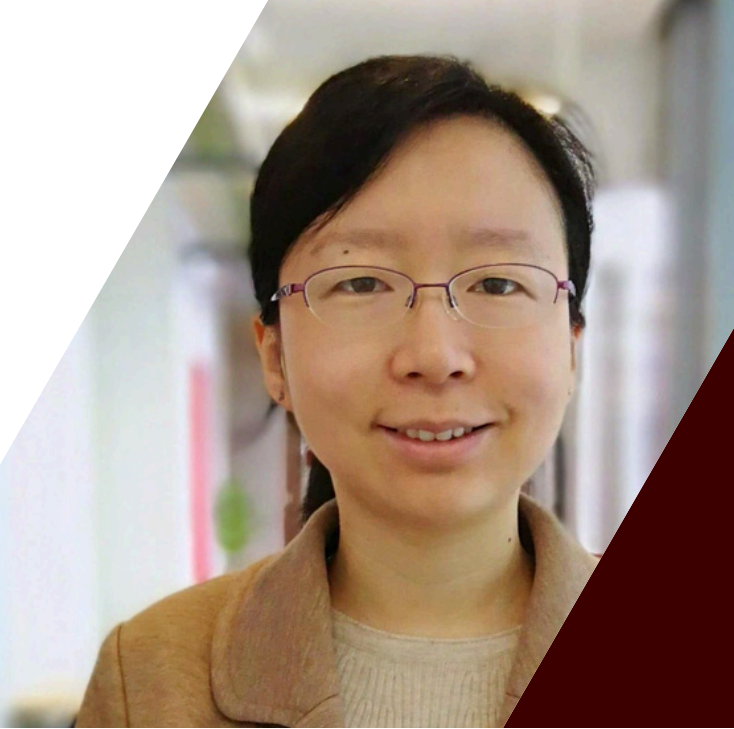
**FRIDAY, FEBRUARY 16, 2024 | 11:30 AM | BLOCKER 150**

## **Ranking Inferences Based on the Top Choice of Multiway Comparisons**

This paper considers ranking inference of  $n$  items based on the observed data on the top choice among  $M$  randomly selected items at each trial. This is a useful modification of the Plackett-Luce model for  $M$ -way ranking with only the top choice observed and is an extension of the celebrated Bradley-Terry-Luce model that corresponds to  $M = 2$ . Under a uniform sampling scheme in which any  $M$  distinguished items are selected for comparisons with probability  $p$  and the selected  $M$  items are compared  $L$  times with multinomial outcomes, we establish the statistical rates of convergence for underlying  $n$  preference scores using both  $l_2$ -norm and  $l_\infty$ -norm, with the minimum sampling complexity. In addition, we establish the asymptotic normality of the maximum likelihood estimator that allows us to construct confidence intervals for the underlying scores. Furthermore, we propose a novel inference framework for ranking items through a sophisticated maximum pairwise difference statistic whose distribution is estimated via a valid Gaussian multiplier bootstrap. The estimated distributions are then used to construct simultaneous confidence intervals for the differences in the preference scores and the ranks of individual items. They also enable us to address various inference questions on the ranks of these items. Extensive simulation studies lend further support to our theoretical results. A real data application illustrates the usefulness of the proposed methods convincingly. (Joint work with Zhipeng Lou, Weichen Wang, and Mengxin Yu)

## **Biography**

Jianqing Fan is Frederick L. Moore Professor of Finance, Professor of Operations Research and Financial Engineering, Former Chairman of Department of Operations Research and Financial Engineering and Director of Committee of Statistical Studies at Princeton University, where he directs both financial econometrics and statistics labs. After receiving his Ph.D. from the University of California at Berkeley, he has been appointed as assistant, associate, and full professor at the University of North Carolina at Chapel Hill (1989-2003), professor at the University of California at Los Angeles (1997-2000), professor and chair at Chinese University of Hong Kong, and professor at the Princeton University (2003-). He was the past president of the Institute of Mathematical Statistics and International Chinese Statistical Association. He is the joint editor of Journal of American Statistical Association and was the co-editor of The Annals of Statistics, Probability Theory and Related Fields, Econometrics Journal, Journal of Econometrics, and Journal of Business and Economics Statistics. His published work on statistics, machine learning, economics, finance, and computational biology has been recognized by The 2000 COPSS Presidents' Award, The 2007 Morningside Gold Medal of Applied Mathematics, Guggenheim Fellow in 2009, P.L. Hsu Prize in 2013, Royal Statistical Society Guy medal in silver in 2014, Noether Distinguished Scholar Award in 2018, Le Cam Award and Lecture in 2021, and election to Academician of Academia Sinica and fellow of American Associations for Advancement of Science, Institute of Mathematical Statistics, American Statistical Association, and Society of Financial Econometrics. His research interest includes high-dimensional statistics, data science, machine learning, financial econometrics, and computational biology.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

**TRACY KE**

Associate Professor, Department of Statistics,  
Harvard University

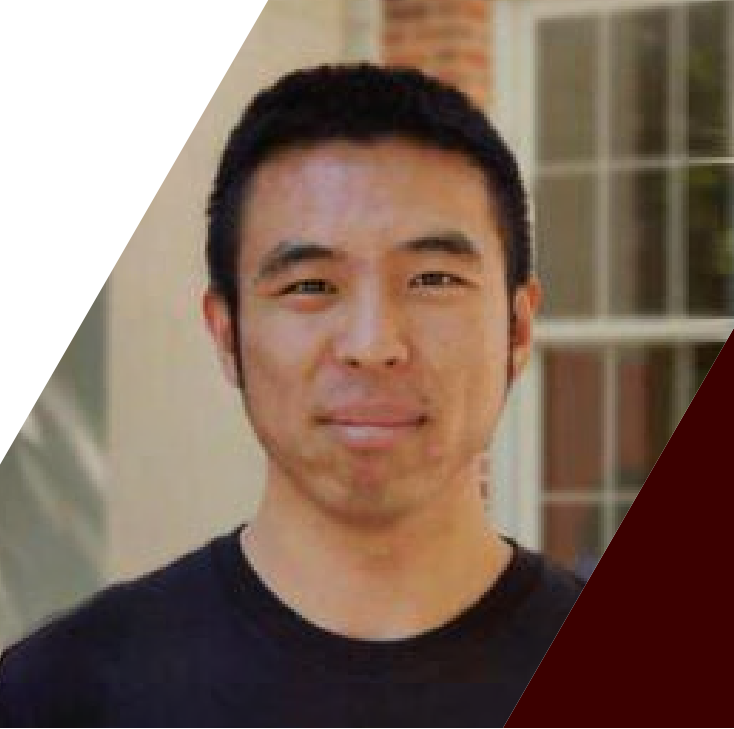
**FRIDAY, FEBRUARY 23, 2024 | 11:30 AM | BLOCKER 150**

### Recent Advances in Text Analysis

Text analysis is an interesting research area in data science and has various applications, such as in artificial intelligence, biomedical research, and engineering. In this talk, I will review popular methods for text analysis, ranging from topic modeling to the recent neural language models. In particular, I will introduce Topic-SCORE (Ke and Wang, 2022), a statistical approach to topic modeling, and discuss how to use it to analyze MADStat - a dataset on statistical publications that we collected and cleaned on our own. The application of Topic-SCORE and other methods on MADStat leads to interesting findings. For example, 11 representative topics in statistics are identified. For each journal, the evolution of topic weights over time can be visualized, and these results are used to analyze the trends in statistical research. In particular, we propose a new statistical model for ranking the citation impacts of 11 topics, and we also build a cross-topic citation graph to illustrate how research results on different topics spread to one another.

### Biography

Tracy Ke is currently Associate Professor of Statistics at Harvard University. She received her PhD from Princeton University in 2014, advised by Professor Jianqing Fan. Prior to joining Harvard, she was Assistant Professor of Statistics at The University of Chicago from 2014 to 2018. Her research interests include network data analysis, high-dimensional statistics, text mining, and machine learning. Recently, she is especially interested in developing optimal spectral methods for network and text data. She was the recipient of NSF CAREER award, ASA Gottfried E. Noether Young Scholar Award, IMS Peter Gavin Hall Early Career Prize, and is currently a Sloan Research Fellow.



The Department of Statistics presents  
**SPRING 2024 COLLOQUIUM**

**DIDONG LI**

Assistant Professor, Department of Biostatistics,  
UNC-Chapel Hill University

**FRIDAY, MARCH 1, 2024 | 11:30 AM | BLOCKER 150**

## **A Journey to Derivatives: From Historical Foundations to Spatial Transcriptomics**

Spatial data pervades numerous fields, ranging from geospatial research to genomics and image analysis, typically within a 2-dimensional plane. Gaussian Processes (GPs) have become foundational in studying these datasets. However, the conventional approach often centers solely on the function. This talk revisits the historical essence of calculus, championed by Newton and Leibniz, and extends it to the novel territory of GPs. We commence with the derivative and curvature processes derived from GPs, known as the ‘derivative process’ and the ‘curvature process’, and their pivotal role in spatial transcriptomic data—a burgeoning domain in genomics. Transitioning to non-Euclidean domains, intrinsic to studies of surfaces such as the brain, heart, and teeth, we outline the recent advancements of GPs in these manifold domains. Through a rigorous definition of the associated derivative and curvature processes, we provide a blueprint of the conditions required for their existence and derive their joint distributions for comprehensive statistical inference. This talk, thus, traces the journey of the derivative concept, bridging its historical significance with modern applications in spatial transcriptomics.

### **Biography**

Didong Li is an assistant professor in the Department of Biostatistics at UNC-Chapel Hill. Before joining UNC, he earned a PhD of Mathematics from Duke University followed by a joint postdoc in Princeton Computer Science and UCLA Biostatistics, and a Visiting Scientist at the Gladstone Institutes. His research focus is statistical methods development for robust inference of complex and high-dimensional data, specifically covering manifold learning, nonparametric Bayesian inference, information geometry, and spatial statistics. He has applied these methods to electronic healthcare record data, large-scale genetic and survey data, and RNA-sequencing data.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

**AMARJIT BUDHIRAJA**

Professor, Department of Statistics and Operations Research, University of North Carolina at Chapel Hill

**FRIDAY, MARCH 8, 2024 | 11:30 AM | BLOCKER 150**

### **Load Balancing in Parallel Queues and Rank-based Diffusions**

Consider a queuing system with  $K$  parallel queues in which the server for each queue processes jobs at rate  $n$  and the total arrival rate to the system is  $nK - an^{1/2}$  where  $a$  is positive and  $n$  is large. It is well known that the join-the-shortest-queue (JSQ) policy has many desirable load balancing properties. In particular, in comparison with uniformly at random routing, the time asymptotic total queue-length of a JSQ system, in the heavy traffic limit, is reduced by a factor of  $K$ . However this decrease in total queue-length comes at the price of a high communication cost of order  $nK^2$  since at each arrival instant, the state of the full  $K$  dimensional system needs to be queried. In view of this it is of interest to study alternative routing policies that have lower communication costs and yet have similar load balancing properties as JSQ. In this work we study a family of such rank-based routing policies, called the Marginal Size Bias Load Balancing (MSBLB) policies, in which  $O(n)$  of the incoming jobs are routed to servers with probabilities depending on their ranked queue length and the remaining jobs are routed uniformly at random. We study the heavy traffic approximation for this system and obtain a novel diffusion limit which is the constrained analogue of the well-studied Atlas model that arise from certain problems in mathematical finance. We also establish an interchange of limits (large time and large  $n$ ) result which shows that, under conditions, the steady state of the queuing system is well approximated by that of the limiting diffusion. Using the explicit formulae for the stationary distributions we identify some striking performance features of this class of load balancing schemes.

### **Biography**

Budhiraja is a Professor of Statistics and Operations Research at the University of North Carolina at Chapel Hill. He served as the chair of the department during 2014-2019. He currently serves as a Senior Associate Dean at the new School of Data Science and Society at UNC. He is the author of over 100 papers and has co-authored a book on the theory of Large Deviations with Paul Dupuis. He is a Fellow of the Institute of Mathematical Statistics.



The Department of Statistics presents  
**SPRING 2024 COLLOQUIUM**

**GONGJUN XU**

Associate Professor, Department of Statistics,  
University of Michigan

**FRIDAY, MARCH 22, 2024 | 11:30 AM | BLOCKER 150**

## **Covariate-Adjusted Generalized Factor Analysis with Application to Testing Fairness**

In the era of data explosion, psychometricians and statisticians have been developing interpretable and computationally efficient statistical methods to measure latent factors (e.g. skills, abilities, and personalities) using large-scale assessment data. In addition to understanding the latent information, the covariate effect on responses controlling for latent factors is also of great scientific interest and has wide applications, such as evaluating the fairness of educational testing, where the covariate effect reflects whether a test question is biased toward certain individual characteristics (e.g. gender and race) taking into account their latent abilities. However, the large sample size, substantial covariate dimension, and great test length pose great challenges to developing efficient methods and drawing valid inferences. Moreover, to accommodate the commonly encountered discrete types of responses, nonlinear factor models (item response theory models) are often assumed, bringing in further complexity to the problem. To address these challenges, we consider a covariate-adjusted generalized factor model and develop novel and interpretable conditions to address the identifiability issue. Based on the identifiability conditions, we propose a joint maximum likelihood estimation method and establish estimation consistency and asymptotic normality results for the covariate effects under a practical yet challenging asymptotic regime. Furthermore, we derive estimation and inference results for latent factors and the factor loadings. We illustrate the performance of this method through extensive numerical studies and an application to a large-scale educational assessment, the Programme for International Student Assessment (PISA). This is a joint work with Jing Ouyang, Chengyu Cui, and Kean Ming Tan.

## **Biography**

Dr. Gongjun Xu is an Associate Professor in the Department of Statistics with a joint appointment in the Department of Psychology at the University of Michigan. He received his Ph.D. in Statistics from Columbia University in 2013. His research interests include latent variable models, psychometrics, statistical learning and inference, and survival analysis. Dr. Xu received NSF CAREER Award (2019), International Chinese Statistical Association (ICSA) Outstanding Young Researcher Award (2019), Bernoulli Society New Researcher Award (2019), Psychometric Society Early Career Award (2023) and Committee of Presidents of Statistical Societies (COPSS) Emerging Leader Award (2023). Dr. Xu is currently serving as Co-Editor-in-Chief for the Journal of Educational and Behavioral Statistics, and Associate Editor for Journal of American Statistical Association, Psychometrika, Annals of Applied Statistics, Statistica Sinica, and Journal of Data Science.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

**ROBERT GENTLEMAN**

Executive Director, Center for Computational  
Biomedicine, Harvard Medical School

**FRIDAY, MARCH 29, 2024 | 11:30 AM | BLOCKER 150**

### Statistics in the Modern Age

The practice of statistics is changing rapidly. Datasets are becoming large, more complex, often collected for purposes other than the one they are being used for. This leads to a number of challenges, and hopefully will lead to changes in the way we teach statistics (or data science) and the ways we practice. I will give some vignettes of the various projects that my center (Center for Computational Biomedicine) at Harvard Medical School is working on. You can view our projects page at (<https://computationalbiomed.hms.harvard.edu/projects/>). I will emphasize the applications, the technologies that are needed, and give some examples of where we have had some success. Applications will cover epidemiology, molecular biology, and image analysis.

### Biography

Dr. Robert Gentleman is the founding executive director of Harvard Medical School's Center for Computational Biomedicine. An accomplished statistician and bioinformatician, Dr. Gentleman is one of the creators of the R programming language and a founder of the Bioconductor project, an open-source collaborative software tool to promote statistical analysis of biological data. He has served as vice president of 23andMe, where he helped launch their therapeutic division, and as senior director for bioinformatics and computational biology at Genentech. He was head of computational biology at the Fred Hutchinson Cancer Research Center and held academic positions at Harvard University, University of Auckland and the University of Waterloo. Dr. Gentleman's research interests are related to genomics, machine learning, data visualization, and the application of statistical and computational methods to study human disease.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

### SOUTIR BANDYOPADHYAY

Associate Professor, Department of Applied Mathematics and Statistics, Colorado School of Mines

**FRIDAY, APRIL 5, 2024 | 11:30 AM | BLOCKER 150**

### **High-Resolution Long-term Weather Data for Energy**

The siting of future solar power generation arrays relies heavily on accurate estimates of solar radiation at an hourly time scale and high spatial resolution. While global and regional climate model projections are useful for gauging future patterns of climate variables, including solar radiation, but data from these models is often too spatio-temporally coarse, and therefore often lack the necessary detail for local applications. With the photovoltaic (PV) industry moving towards longer plant lifetimes, understanding the impact of climate change on PV production becomes crucial. We propose a novel method to downscale global horizontal irradiance (GHI) data from daily averages to hourly profiles while preserving spatial correlation. Our approach employs a diurnal template that can be adjusted based on time, location, and year. We apply this method to data from the National Solar Radiation Database and present a case study over several sub-regions of CONUS. We further assess uncertainties associated with regriding regional climate model fields, disaggregating daily totals to hourly profiles, and downscaling to a 4km resolution. The resulting multi-scale predictive model enables projections of GHI into the future, facilitating informed decisions in solar facility planning amidst a changing climate.

### **Biography**

Soutir earned a doctorate in Statistics (2010) at Texas A&M University. Before joining Mines he was an Assistant Professor in the Department of Mathematics (2010-2017) at Lehigh University. He has also been a visiting scientist at the Computational and Information Systems Laboratory at the National Center for Atmospheric Research studying climate models. His core research approaches problems related to spatial and time series data. His expertise lies in developing novel inferential procedures under spatial (and/or temporal) dependence and investigating their asymptotic properties. Soutir's research has always been motivated by and applied to problems arising from practical situations in various areas such as climate science, environmental studies, finance, and biomedical studies, among others. He is an elected member of International Statistical Institute and the recipient of the International Indian Statistical Association Young Statistical Scientist Award.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

**YONG CHEN**

Professor of Biostatistics, Department of Statistics,  
University of Pennsylvania

**FRIDAY, APRIL 12, 2024 | 11:30 AM | BLOCKER 150**

### **PDA: Privacy-preserving Distributed Algorithms and statistical inference in the era of real-world data networks**

With the increasing availability of electronic health records (EHR) data, it is important to effectively integrate evidence from multiple data sources to enable reproducible scientific discovery. However, we are still facing practical challenges in data integration, such as protection of data privacy, the high dimensionality of features, and heterogeneity across different datasets. Aim to facilitate efficient multi-institutional data analysis without sharing individual patient data (IPD), we developed a toolbox of Privacy-preserving Distributed Algorithms (PDA) that conduct distributed learning and inference for various models, such as association analyses, causal inference, cluster analyses, counterfactual analyses, and beyond. Our algorithms do not require iterative communication across sites and are able to account for heterogeneity across different hospitals. The validity and efficiency of PDA are also demonstrated with real-world use cases in Observational Health Data Sciences and Informatics (OHDSI), PCORnets including PEDSnet and OneFlorida, and Penn Medicine Biobank (PMBB).

### **Biography**

Yong Chen is a tenured Professor of Biostatistics and the Founding Director of the Center for Health AI and Synthesis of Evidence (CHASE) at the University of Pennsylvania. He is an elected fellow of American Statistical Association, International Statistical Institute, Society for Research Synthesis Methodology, American College of Medical Informatics, and American Medical Informatics Association. He founded the Penn Computing, Inference and Learning (PennCIL) lab at the University of Pennsylvania, focusing on clinical evidence generation and evidence synthesis using clinical and real-world data. During pandemic, Dr. Chen is serving as biostatistics core director for a national multi-center study on Post-Acute Sequelae of SARS CoV-2 infection (PASC), involving more than 9 million pediatric patients across 40 health systems.





TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

**KAI ZHANG**

Associate Professor with Tenure, Department of  
Statistics and Operations Research, UNC Chapel Hill

**WEDNESDAY, APRIL 17, 2024 | 11:30 AM | BLOCKER 150**

### **BET and BELIEF**

We study the problem of distribution-free dependence detection and modeling through the new framework of binary expansion statistics (BEStat). The binary expansion testing (BET) avoids the problem of non-uniform consistency and improves upon a wide class of commonly used methods (a) by achieving the minimax rate in sample size requirement for reliable power and (b) by providing clear interpretations of global relationships upon rejection of independence. The binary expansion approach also connects the symmetry statistics with the current computing system to facilitate efficient bitwise implementation. Modeling with the binary expansion linear effect (BELIEF) is motivated by the fact that two linearly uncorrelated binary variables must be also independent. Inferences from BELIEF are easily interpretable because they describe the association of binary variables in the language of linear models, yielding convenient theoretical insight and striking parallels with the Gaussian world. With BELIEF, one may study generalized linear models (GLM) through transparent linear models, providing insight into how modeling is affected by the choice of link. We explore these phenomena and provide a host of related theoretical results. This is joint work with Benjamin Brown and Xiao-Li Meng.

### **Biography**

Dr. Kai Zhang is currently an associate professor with tenure at the Department of Statistics and Operations Research, UNC Chapel Hill. Dr. Zhang obtained his bachelor's degree from Peking University in 2003, his Ph.D. degree in mathematics from Temple University in 2007, and his Ph.D. degree in statistics from the Wharton School, University of Pennsylvania in 2012. He is an elected fellow of the Institute of Mathematical Statistics. His research interests include nonparametric statistics, high-dimensional statistics, and post-selection inference. His research is supported by five grants from the National Science Foundation.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents  
**SPRING 2024 COLLOQUIUM**

**DOUGLAS NYCHKA**

Professor, Department of Applied Mathematics and  
Statistics, Colorado School of Mines

**FRIDAY, APRIL 19, 2024 | 11:30 AM | BLOCKER 150**

## **Hybrid L1 and L2 Smoothing**

Spline smoothing and more generally Gaussian process smoothing have become a successful methodology for estimating a smooth trend or surface from noisy data. Similarly the LASSO and related L1 penalties have become important tools for variable selection and also admit of a Bayesian version based on the Laplace distribution. This project combines these two approaches as a method to detect discontinuous behavior in an otherwise smooth signal. Every day the Foothills Facility of Denver Water filters more than 250 million gallons of water for the metropolitan area. This process runs continuously and is monitored across an array of filters, each the size of a small swimming pool, at 5 minute intervals. It is important to be able detect anomalous behavior in a filter in a prompt manner or to process past measurements to determine trends. The anomalies take the form of discontinuities or appear as step changes in the smooth filtering cycle. This application is the motivation for a mixed smoothing approach where normal operation is captured by a smoothing spline and the anomalies by basis function coefficients determined by an L1 penalty. As part of this research a frequentist penalty method is compared against its equivalent Bayesian hierarchical model (BHM) based on Gaussian processes and a Laplace prior for the anomaly coefficients. This talk will discuss some of the challenges in implementing both models. Specifically we study how to choose penalty parameters for the frequentist model and how to formulate the BHM in a way that the Bayesian sampling algorithm mixes efficiently. Both approaches appear to classify anomalies in the filter cycles well with the spline model being much faster but the BHM providing measures of uncertainty in the detected anomalies. The similarities between these frequentist and Bayesian models relies on the correspondence between splines and Gaussian processes. This was first described by Grace Wahba and George Kimeldorf. Some background for this historical connection will be given as part of developing the Bayesian model. This is joint work with Matthew Hofkis.

## **Biography**

Douglas Nychka is a statistician and data scientist whose areas of research include the theory, computation and application of curve and surface fitting with a focus on geophysical and environmental applications. Currently he is a Professor in the Department of Applied Mathematics and Statistics at the Colorado School of Mines and Senior Scientist Emeritus at the National Center for Atmospheric Research (NCAR), Boulder, Colorado. Before moving to Mines he directed the Institute for Mathematics Applied to Geosciences at NCAR. His current focus in research is the computation of spatial statistics methods for large data sets and the migration of these algorithms into easy to use R packages. He has coauthored more than 100 research articles and with an h-index of 48. He is a Fellow of the American Statistical Association, Fellow of the Institute for Mathematical Statistics and a recipient of the Jerry Sacks Award for interdisciplinary research.



TEXAS A&M UNIVERSITY

Statistics

The Department of Statistics presents

## SPRING 2024 COLLOQUIUM

### LONG NGUYEN

Professor, Department of Statistics, University of Michigan

**FRIDAY, APRIL 26, 2024 | 11:30 AM | BLOCKER 150**

### **Dendrogram of mixing measures: Hierarchical clustering and model selection using finite mixture models**

We present a new way to summarize and select mixture models via the hierarchical clustering tree (dendrogram) constructed from an overfitted latent mixing measure. Our proposed method bridges agglomerative hierarchical clustering and mixture modeling. The dendrogram's construction is derived from the theory of convergence of the mixing measures, and as a result, we can both consistently select the true number of mixing components and obtain the pointwise optimal convergence rate for parameter estimation from the tree, even when the model parameters are only weakly identifiable. In theory, it explicates the choice of the optimal number of clusters in hierarchical clustering. In practice, the dendrogram reveals more information on the hierarchy of subpopulations compared to traditional ways of summarizing mixture models. Several simulation studies are carried out to support our theory. We also illustrate the methodology with an application to single-cell RNA sequence analysis. This talk is based on joint work with Dat Do, Linh Do, Scott McKinley and Jonathan Terhorst.

### **Biography**

Long Nguyen is Professor of Statistics, and by courtesy, of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor. He studied computer science, mathematics and statistics at Pohang University of Science and Technology, Arizona State University, and University of California, Berkeley, where he received the PhD degree in computer science in 2007. He joined Michigan in 2009. Nguyen's interests include nonparametric Bayesian statistics, optimal transport and statistical inference, and ML with complex spatiotemporal models. He has served as associate editor of the Annals of Statistics, the Annals of Institute of Statistical Mathematics, Bayesian Analysis, Journal of Machine Learning Research, SIAM Journal on Mathematics of Data Science, and Journal of American Statistical Association. Nguyen is elected fellow of the IMS, the ASA, and a distinguished associate member of Vietnam Institute for Advanced Study in Mathematics.

# STATISTICS FORMER STUDENT NETWORK WEBINARS



## MANDY HERING

"MULTIVARIATE STATISTICAL PROCESS  
MONITORING IN COMPLEX WATER AND  
WASTEWATER TREATMENT SYSTEMS"

Professor of Statistical Science  
Baylor University

## DENNIS W. KING

"MY CAREER AS THE STATKING"

Founder and President  
STATKING Clinical Services (SCS)



## JOHN SCHWENCK

"AI & MACHINE LEARNING MODEL  
DEVELOPMENT: FROM ACADEMIA TO  
INDUSTRY"

Quantitative Analyst  
Wells Fargo  
Consumer AI & ML Model Development Center



# STUDENT EVENTS

The Statistics Graduate Student Association (SGSA) hosted a series of game nights during the spring semester, offering an evening filled with a variety of games and activities. Attendees enjoyed classic games such as Ludo, Jenga, Yut, Yacht, Werewolf, Catan, Carrom, and Uno, along with trivia and other online games. The game nights were open to faculty, postdocs, and students, including those in the distance master's program, with a zoom link provided for remote participants.

These game nights provided a fun and relaxed environment for members of the statistics community to socialize and unwind. Participants had the opportunity to engage in friendly competition, with attractive prizes offered to the winners. These events were a great success, bringing together individuals from various backgrounds and fostering a sense of camaraderie among attendees.



# STAT CAFE



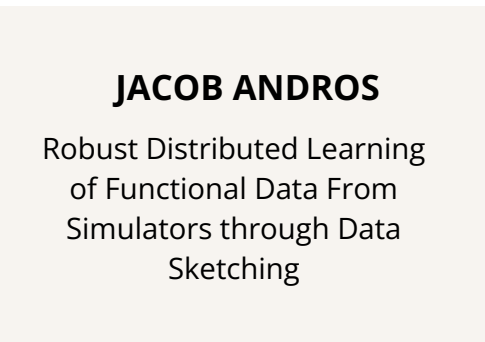
**RAYMOND CARROLL**

Mentoring Talk



**ABHISEK CHAKRABORTY**

Probabilistic  
Approaches for Fair  
Clustering



**JACOB ANDROS**

Robust Distributed Learning  
of Functional Data From  
Simulators through Data  
Sketching



**RUNMIN WANG**

High-Dimensional  
Inference and  
Beyond



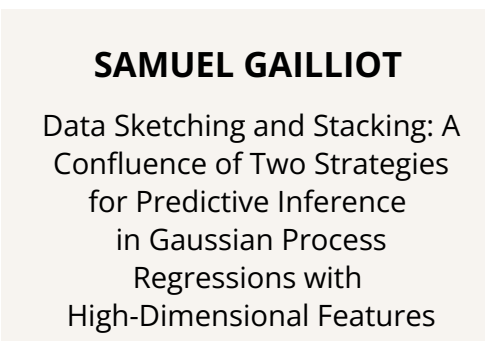
**ISAAC RAY**

GGBoost - Graph Gradient  
Boosting



**CHANGWOO LEE**

Logistic-Beta Processes  
for Modeling Dependent  
Random Probabilities  
with Beta Marginals



**SAMUEL GAILLIOT**

Data Sketching and Stacking: A  
Confluence of Two Strategies  
for Predictive Inference  
in Gaussian Process  
Regressions with  
High-Dimensional Features



**ADRIAN THURSTIN**

What is a Model Validation  
Analyst? Working in Banking  
and How to Get the Job (And  
Various Other Advice)



**ALICIA CARRIQUIRY**

Statistics and the  
Evaluation of Forensic  
Evidence



**CARSON JAMES**

Learning Joint and Individual  
Structure in Network Data  
with Covariates

# WORKFLOW WORKSHOP



**BRANI VIDAKOVIC**

NSF Funding  
Opportunities & Tips for  
Better NSF Proposal



**SHEELABHADRA DEY**

Deep Q-Learning:  
Mastering RL with  
Neural Networks



**REBECCA LEE**

Organization Tools  
for Graduate Student  
Research



**ISAAC RAY**

R Users  
Hate Him!



## WOMEN IN STATISTICS LUNCH

On April 4, 2024, the Department of Statistics' Climate Committee organized its inaugural Women in Statistics Lunch, featuring Dr. Alicia Carriquiry as the guest speaker. Dr. Carriquiry delivered a talk titled "Key to Success: Find and Develop Your Super-Powers," providing insights into navigating the field of Statistics.

The Women in Statistics Lunch provided attendees with an opportunity to engage in discussions and networking. Dr. Carriquiry's presentation emphasized the significance of recognizing and cultivating individual strengths.

## PROSPECTIVE STUDENT VISIT DAY

On March 18, 2024, the Department of Statistics organized a prospective student visit day for students accepted into the PhD program for the Fall 2024 semester. The event provided an opportunity for prospective students to familiarize themselves with the department, gain insights into the graduate program, and interact with current students and faculty. Activities included departmental tours and informational sessions designed to showcase the department's offerings and provide a glimpse into academic and research opportunities.

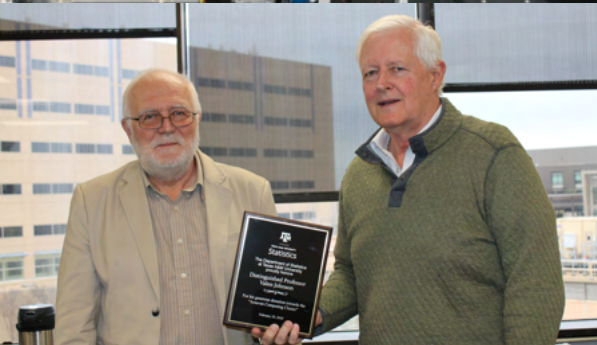
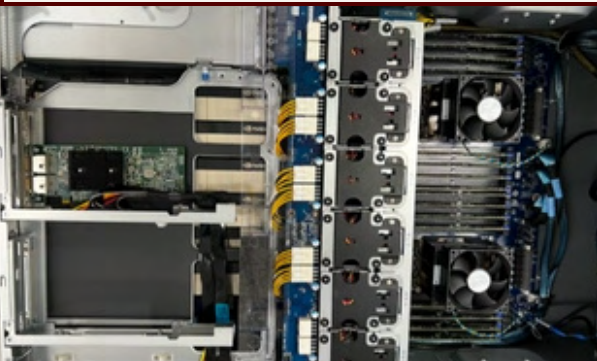




# ALICIA CARRIQUIRY'S WELCOME DINNER

The welcome dinner for Dr. Alicia Carriquiry, our new Hagler Fellow, was a delightful event held on February 10, marking the beginning of her fellowship at Texas A&M University. Faculty and staff gathered to celebrate her arrival and honor her contributions to the field of statistics and data science. The dinner provided a warm and welcoming atmosphere for networking and camaraderie, setting the stage for a fruitful tenure for Dr. Carriquiry at Texas A&M.

During the dinner, attendees had the opportunity to engage with Dr. Carriquiry, learning more about her research interests and the projects she plans to pursue as a Hagler Fellow. Her expertise and passion for statistics were evident, sparking intriguing discussions and laying the groundwork for potential collaborations. The event was not only a celebration of Dr. Carriquiry's achievements but also a testament to Texas A&M's commitment to advancing research excellence in statistics and data science.



# ARSEVEN COMPUTER CLUSTER

On February 29, a celebration took place for the arrival of the Arseven Computer Cluster, and to recognize the invaluable contributions of Dr. Ersen Arseven and Dr. Valen Johnson to the department. This computer cluster consists of 13 compute nodes and 2 storage nodes designed to tackle various research problems, including storing and analyzing massive datasets, running large-scale simulations for developing and evaluating new methodology, and implementing computationally intensive machine learning and deep learning procedures.

Each compute node is equipped with two 64-core 2.45 GHz CPUs and 1024 GB of RAM capable of running many analytical tasks simultaneously. These nodes collectively offer over 1600 cores of parallel processing power for supporting varied research activities. Three of the compute nodes come equipped with six NVIDIA A30 24GB GPUs which excel at parallel processing, making them ideal for deep learning, image analysis, and other data-hungry tasks. The cluster also includes a staggering 180 TB of storage space where researchers can temporarily park their massive datasets. Whether crunching census data or analyzing genetic sequences, this storage ensures no data gets left out in the cold.

# PIE-A-PROF FOOD DRIVE

The Statistics Student Ambassadors hosted the 3rd annual Statistics Pie-A-Prof food drive. This drive benefits the 12th Can, a local student-run food pantry that aims to serve all students, faculty, and staff in need of assistance.

The competition, which ran from April 12th to April 24th, generated a great deal of excitement and friendly competition among the participants. A total of 110 pounds of food was collected this year. Dr. Trevor Harris, the professor with the most points, was pied in the face by the runner-up, Dr. Jesus Arroyo, at a celebratory event on Thursday, April 25th. The event featured food and provided a fun and festive atmosphere for all attendees.

The Pie-A-Prof food drive not only raised awareness about food insecurity but also demonstrated the generosity and community spirit within the Department of Statistics at Texas A&M. By coming together to support their peers in need, students, staff, and faculty showed that small acts of kindness can make a big difference.



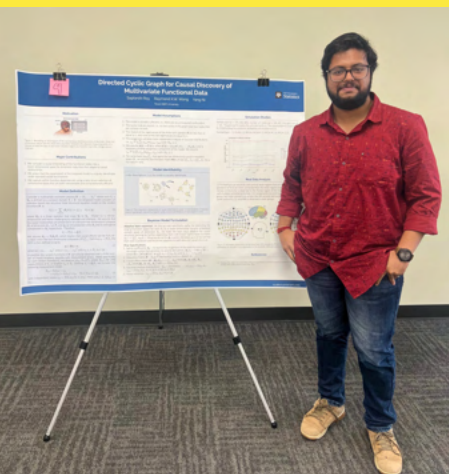
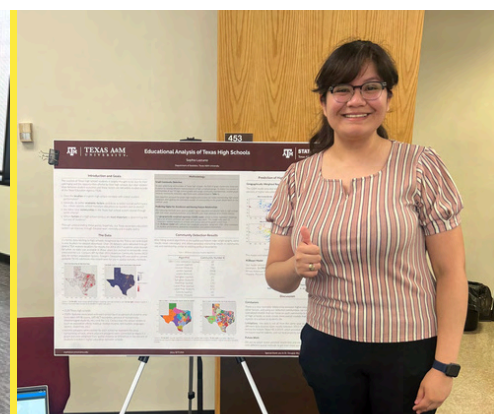
# SETCASA POSTER COMPETITION

The 2024 SETCASA Poster Competition, held on May 1st, provided a platform for junior statisticians in the region to exhibit their research and engage in professional networking. Organized annually, this event serves as a valuable opportunity for participants to share insights and discoveries in the realm of statistical science. Attendees included both statisticians and applied scientists addressing issues with a strong statistical basis, fostering interdisciplinary dialogue and collaboration.

Among the standout achievers of this year's competition were Saptarshi Roy, Niladri Kal, and Valeriya Rogovchenko, who clinched the Gold, Silver, and Bronze awards respectively in the graduate student and postdoc category. Additionally, Gözde Sert, Isaac Ray, and Anamitra Chaudhuri earned well-deserved honorable mentions, further highlighting the caliber of contributions presented at the event.

In recognition of emerging talent, the undergraduate category witnessed a remarkable performance by Sophia Lazcano, who secured the Gold award for her compelling poster presentation. As the youngest cohort of participants, undergraduate students like Sophia offer a glimpse into the promising future of statistical research, invigorating the field with fresh perspectives and innovative approaches. Their achievements underscore the inclusive ethos of SETCASA, which celebrates excellence at every stage of one's academic journey.

Through initiatives like the poster competition, SETCASA continues to cultivate a vibrant community of scholars, propelling the advancement of statistical science and its applications in addressing contemporary challenges.



# CAREER NIGHT

On March 1st, the Department of Statistics hosted an exclusive evening career fair, specifically designed for statistics students. The event provided attendees with numerous opportunities to engage with potential employers and industry professionals. The evening featured a variety of activities, including panel sessions, company presentations, and spontaneous interview opportunities, allowing students to explore different career paths and make valuable connections.

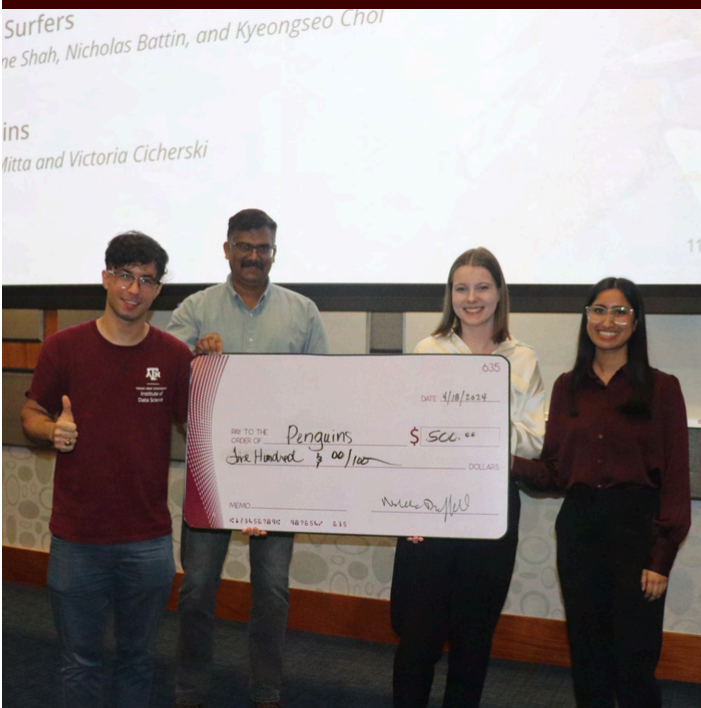
The panel sessions covered a range of relevant topics, offering insights and advice from experienced professionals. One session, titled "Women Shaping Statistics: Challenges and Triumphs," focused on the unique experiences and achievements of women in the field. Another session, "From Academia to Industry: Transitioning Successfully in Statistics," provided guidance on navigating the shift from academic studies to professional careers. Additional panels such as "Data Science and Beyond: The Versatility of a Statistical Career" and "Data Frontiers: Insights from Kickstarter Innovators in Data Science" highlighted the diverse applications of statistical skills in various industries.

Throughout the event, company presentations offered a closer look at different organizations and their roles in the statistical landscape. These presentations, combined with the opportunity for spontaneous interviews, allowed students to directly engage with employers, enhancing their understanding of the job market and potential career opportunities. Career Night proved to be a valuable event for statistics students, equipping them with the knowledge and connections necessary to succeed in their future careers.



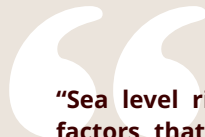


# TAMIDS SEA LEVEL RISE DATA SCIENCE CHALLENGE AWARDS



The 2024 TAMIDS Student Data Science Competition concluded on April 18th. This year's topic was the Impact of Sea Level Rise. Focused on addressing this critical issue, that threatens millions of US citizens who live in coastal regions, the competition saw the participation of eight teams tasked with providing innovative solutions to the challenges posed by sea level rise (SLR).

Sponsored by Chevron and supported by the Texas A&M University Departments of Statistics and Electrical & Computer Engineering, the event awarded a total of \$9,500 in prizes to the winning teams. The competition, organized by the Texas A&M Institute of Data Science, aimed to harness students' analytical skills to identify communities, services, or infrastructure vulnerable to SLR and provide actionable insights for decision-makers. Notably, the third-place winners, Anjana Mittal and Victoria Cicherski, both students from the Department of Statistics, received a \$500 prize for their project titled "Diving Into the Impacts of Sea Level Rise."



**"Sea level rise is influenced by an intricate interplay of factors that are unique to each part of the world. Our project seeks to address the impact of sea level rise within San Francisco and how it influences the community and environment of the Bay Area. Our report details the use of advanced machine-learning methods and predictive analytics to prevent and reduce the consequences of sea level rise on a community and environment. An efficient approach to minimizing the harmful effects of sea level rise is through prediction. To achieve this, we utilized sea level change and different types of flooding data for California from multiple sources."**

# STATISTICAL METHODS FOR HIGH DIMENSIONAL COMPLEX DATA

CONFERENCE IN HONOR OF DR. RAYMOND CARROLL



The Statistical Methods for High Dimensional Complex Data Conference, held in honor of Dr. Raymond Carroll, took place from Thursday, May 23, 2024, to Friday, May 24, 2024, at the Hilton Hotel and Conference Center in College Station. Organized by a distinguished panel including Bani Mallick from Texas A&M University, Ciprian Crainiceanu from John Hopkins Bloomberg School of Public Health, Irina Gaynanova from the University of Michigan, Mikyoung Jun from the University of Houston, Yehua Li from the University of California Riverside, Jeffrey Morris from the University of Pennsylvania, and Huiyan Sang from Texas A&M University, the conference featured a packed schedule of presentations and discussions.

The conference kicked off on Thursday with a session focusing on Measurement Error, featuring talks by Yanyuan Ma, Len Stefanski, and Anna Maria Staicu. This was followed by a session on Genomics & Epidemiology with presentations by Victor Kipnis, Xihong Lin, and Kathryn Roeder.

The day included a special Carroll Sharing session led by Jeff Morris, where former students and colleagues shared about Dr. Carroll's career and lasting impressions. The session conclude with the Carroll Young Investigator Award Lecture, delivered by the awardee Dr. Alexander Luedtke. The topic of Dr. Luedtke's talk was Simplifying Debaised Inference via Automatic Differentiation and Probabilistic Programming.



Dr. Luedtke is an Associate Professor in the Department of Statistics at the University of Washington. He also has an adjunct appointment in the Department of Biostatistics and an affiliate appointment in the Vaccine and Infectious Disease Division at the Fred Hutchinson Cancer Research Center. His methodological research focuses on developing efficient estimators in problems arising in a variety of areas, including in policy learning and infectious disease studies. He derives such estimators analytically using tools from semiparametric efficiency theory and numerically using minimax optimization schemes. He also serves as a study statistician for the HIV Vaccine Trials Network and the Covid-19 Prevention Network.



Drs. Brani Vidakovic, Alexander Luedtke and Raymond Carroll (left to right)

Friday's sessions began with a focus on Regression, featuring talks by Hans-Georg Mueller, Jane-Ling Wang, and Naisyin Wang. The day continued with a session on Robustness and Bayesian Methods, including presentations by Doug Simpson, Edward George, and Alicia Carriquiry. The conference provided a platform for young investigators to showcase their research at the Young Investigator Presentations session.

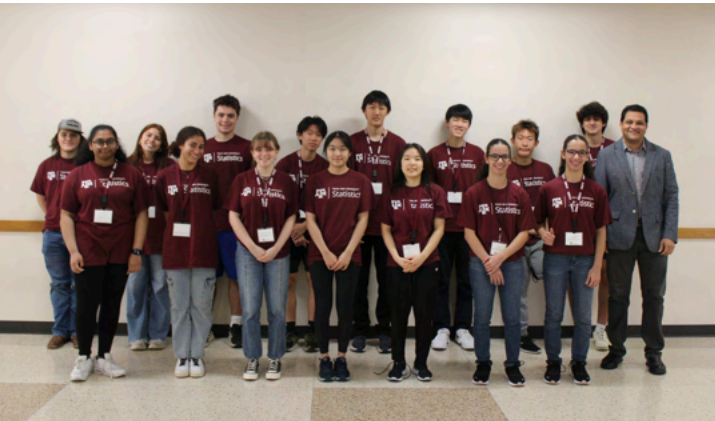
One of the key highlights of the conference was the diverse range of topics covered, which reflected the multifaceted nature of high-dimensional complex data problems. From measurement error and genomics to robust statistical methods and Bayesian approaches, the sessions underscored the breadth of Dr. Carroll's influence and the ongoing advancements in these areas. The event also facilitated meaningful networking opportunities, enabling participants to discuss potential collaborations and share their latest research findings in an engaging and supportive environment.



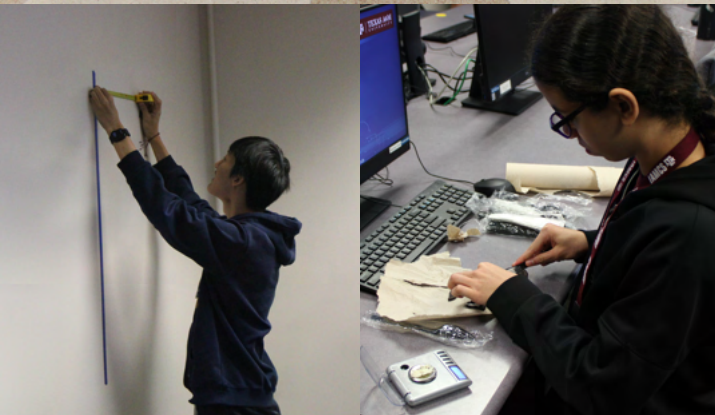
Drs. Kathryn Roeder, Bani Mallick, Raymond Carroll and Jiayi Tong (left to right)

Overall, the conference served not only as a platform for advancing the field of statistics but also as a heartwarming reunion for former students, friends, colleagues, and relatives of Dr. Raymond Carroll. The gathering celebrated both his birthday and illustrious career, bringing together a community deeply impacted by his contributions to the field. This celebratory atmosphere underscored the profound and lasting legacy of Dr. Carroll's work, making the event a memorable tribute to his enduring influence in the world of statistics.

# STATISTICS & DATA SCIENCE SUMMER CAMP



Under the leadership of Dr. Mohamed Aburweis, the 2024 Statistics and Data Sciences Summer Camp was a complete success, bringing together high school students for an immersive and educational experience from June 3-7. The camp aimed to introduce students to the exciting world of statistics and data sciences, offering a blend of theoretical knowledge and practical application.



Throughout the week-long camp, students engaged in a variety of hands-on activities and workshops designed to deepen their understanding of statistical concepts and data analysis techniques. From exploring real-world datasets to learning how to use statistical software, participants gained valuable skills that will serve them well in future academic and professional pursuits.



One of the highlights of the camp was the opportunity for students to work on a group project, where they applied their newfound knowledge to tackle a real-world problem. This not only allowed them to see the relevance of statistics and data sciences in everyday life but also encouraged teamwork and collaboration.



Thank you to our lecturers and volunteers who dedicated their time to teach different topics to our campers. Among them were Dr. Darren Homrighausen, Mr. Jacob Andros, Mrs. Courtney Shuttlesworth, Dr. Scott Bruce, Dr. Scott Crawford, Dr. Sharmistha Guha, Mr. Isaac Ray, Ms. Ana Ferreira, Mr. Wesley Fletcher, Mrs. Rebecca Lee, and Ms. Minjee Kim.

We are immensely proud of the campers for their dedication and enthusiasm throughout the program. Their curiosity and hard work have truly exemplified the spirit of learning and exploration in statistics and data sciences.



# OTHER EVENTS



## Faculty and Staff Appreciation BBQ

On Saturday, April 13th, the annual BBQ took place at Tiffany Park, organized by the graduate students as a gesture of gratitude towards the faculty and staff. The attendees enjoyed a delicious BBQ while fostering a sense of camaraderie through fellowship and fun games.



## Lunar Year Celebration

The Lunar New Year is a celebration of the arrival of spring and the beginning of a new year on the lunisolar calendar. During the 15-day festival, families and friends get together to fellowship and enjoy traditional food.



## Aggieland Saturday

On February 10, our department participated in Aggieland Saturday, a campus-wide open house event for prospective students and their families to explore the university's offerings. Our statistics students played a key role in engaging with visitors, answering questions, and creating a welcoming and informative environment for all attendees.



## Holi Festival

The Holi Festival, also known as the Festival of Colors, is a vibrant Hindu celebration that marks the arrival of spring. Participants celebrate by throwing colored powders and water at each other, symbolizing the arrival of a colorful and joyful season. The festival is also a time for reuniting with loved ones, enjoying festive foods, and sharing in music and dance.



# MAY 2024 GRADUATES

## UNDERGRADUATE STUDENTS

Sara Antonijevic | Rachel Bartolo | Johan Booc | Sloka Chandrasekaran | Tiffany Chang | Charli Cottrell | Jacob Enerio | Valerie Espinosa | Maura Flanagan | Isaac Gutierrez | Sukhyun Ha | Hanse Jeong | Silin Ji | Nicholas Johannessen | Devin Johnson | Clay Jones | Kyoungmin Jun | Alexander LeGresley | Yufan Li | Xingyu Liu | Max Moran | Caitlin O'Neil | Hannah Parker | Jackson Polston | Corina Ramont | Joshua Samuel | Grant Schweikhardt | Rahul Singh | Megan Tung | Sarah Vasicek | Mahir Vohra | Dylan Ward | Grant Westfall | Kaiqi Zhao | Mary Ziperman

## MS DISTANCE STUDENTS

Garrett Anders | Irbaz Arab | Jared Brown | Morgan Cappa | Caelli Craig | Anirban Deb | Haley Friel | Zachary Fuller | Edward Givens | Helen Goenawan | Ian Grotton | Edward Keefe | Nicholas Klein | John Lee | Skylar Liu | Luis Castaneda | Rabah Mohamed | Heather Nagy | Michael Niemerg | Luis Quilarque | Hannah Reavis | Carmen Roland | Elizabeth Salazar | Holly Schrempp | Brandon Scott | James Seymour | Manuel Sojan | Elizabeth Storm | Sadia Tasnim | Adrian Thurstin | Rohan Tiwari | Jonah Warren | Zhihua Xu | Jingwei Zhu | Ruixiao Zuo | Armel Maemble | Vanessa Rosas | Lacey Rubio

## MS ON-CAMPUS STUDENTS

Valerie Espinosa | Sarah Vasicek | Dylan Ward | Corina Ramont | Dharmenkumar Patel | Radhika Kulkarni | Riana Guha | Mark Cahill | Tiffany Chang | Silin Ji | Megan Tung | Hanse Jeong | Hussein Mansour | Udbhav Srivastava | Gunjan Joshi

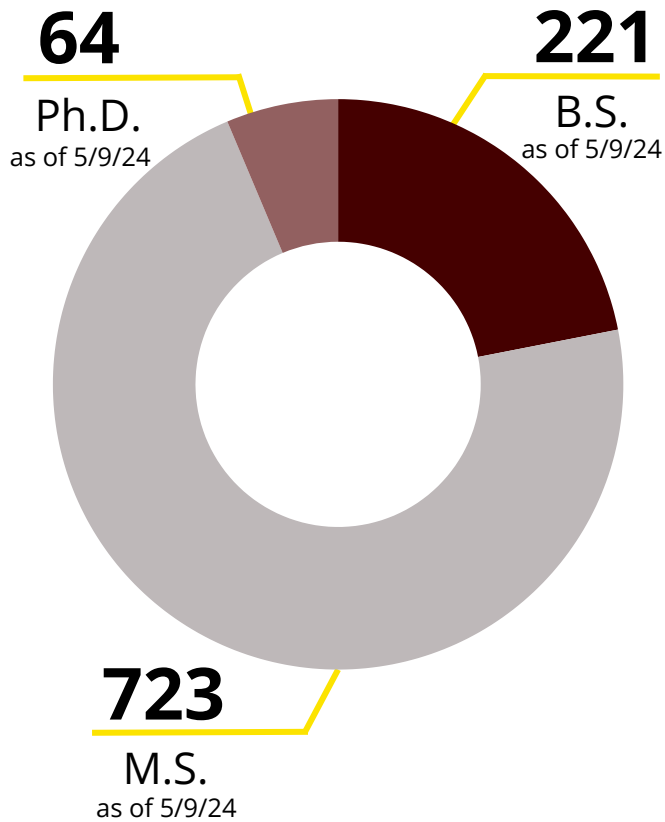
## PHD STUDENTS

Changwoo Lee | Asmita Roy

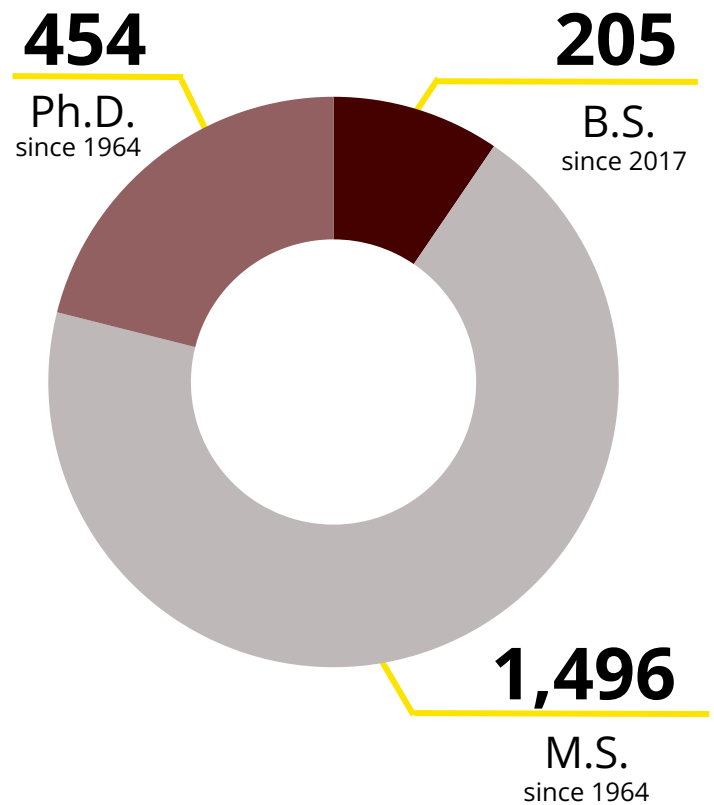


# TAMU STATISTICS BY THE NUMBERS

## TOTAL STUDENT ENROLLMENT



## TOTAL DEGREES AWARDED



**19** TENURED FACULTY

**11** TENURE TRACK FACULTY

**13** ACADEMIC PROFESSIONAL TRACK

## OUR MAJORS

**1**

UNDERGRADUATE MAJOR  
B.S. IN STATISTICS

**3**

GRADUATE MAJORS  
M.S. IN DATA SCIENCE  
M.S. IN STATISTICAL DATA SCIENCE  
Ph.D. IN STATISTICS



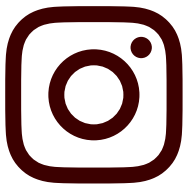
# 2024 AGGIE REUNION

Join us for the 2024 Aggie Reunion during the JSM Conference in Portland, Oregon! This exciting event will be held at the Hyatt Regency Portland in the Regency Ballroom B from 5:00 PM until 8:00 PM on Monday, August 5. We look forward to reconnecting with old friends, meeting new faces, and celebrating our shared Aggie spirit. We cannot wait to see everyone there!



**STAY CONNECTED**

# **FOLLOW US ON SOCIAL MEDIA**



@tamustats



@tamustats



Texas A&M Statistics



Texas A&M Statistics



### **IMPORTANT NOTICE**

Please help us verify your contact information and stay connected to receive future online issues of StatLinks.

Email us at [statlinks@stat.tamu.edu](mailto:statlinks@stat.tamu.edu) and let us know if your contact information has changed or if you wish to be added or removed from the StatLinks listserv.

## **CONTACT**

Department of Statistics  
3143 TAMU  
College Station, TX 77843-3143

(979) 845-3141  
[statlinks@stat.tamu.edu](mailto:statlinks@stat.tamu.edu)

[www.stat.tamu.edu](http://www.stat.tamu.edu)