The 4th International Conference on Statistical Distributions and Applications

ICOSDA 2022

October 13 – 15, 2022

Hosted by Marshall University
Department of Mathematics

Conference Venue:
The DoubleTree Hotel Huntington, WV

The organizers of ICOSDA 2022 would like to thank the Department of Mathematics and the College of Science at Marshall University for their support.

We would also like to thank the following sponsors:

Pearson

sas

SAGE Publishing

PRINCETON UNIVERSITY PRESS

O'REILLY®

Springer

siam®

ICOSDA 2022 Program Contents

A Message from the Provost

Welcome!

On behalf of the Marshall University faculty, staff, and students, I would like to extend our warmest welcome to all the participants of the 4th International Conference on Statistical Distributions and Applications (ICOSDA 2022)!

Beyond welcoming you to our beautiful campus nestled between the banks of the majestic Ohio River and the foothills of the Appalachian Mountains in "America's Best Community" - Huntington, West Virginia, I offer special congratulations to all the speakers and presenters on their scholarly accomplishments.

This conference provides an excellent venue for researchers and practitioners to share and discuss recent advances in distribution theory and its applications. Hosting the ICOSDA 2022 Conference reaffirms Marshall's shared mission to prepare society to think, learn, work, and live in an evolving world. We are extremely proud that this conference aligns with Marshall's academic excellence and national prominence as an R2 Doctoral University (High Research Activity) through the Carnegie Commission.

To plan and execute an event of this significance required the efforts of dedicated faculty and staff members. These collective efforts from the faculty and staff of our Mathematics department demonstrate our commitment to fostering research and scholarship, and I am grateful to all who have worked so diligently to make this conference a success.

Have a fantastic three days in our Jewel City, the home of the Thundering Herd!

Avinandan Mukherjee

Dr. Avinandan "Avi" Mukherjee
Provost & Senior Vice President for Academic Affairs
Marshall University

A Message from the Organizing Committee

Again, Welcome!

On behalf of the Organizing Committee, it is an honor to welcome you to the 4th International Conference on Statistical Distributions and Applications (ICOSDA 2022), hosted by the Department of Mathematics, Marshall University at the DoubleTree by Hilton, Huntington, West Virginia.

We appreciate each one of you for taking time out of your busy schedule to travel to Huntington for the conference from different parts of the globe, with all that COVID-19 has brought since the outbreak in December 2019, shortly after the last ICOSDA conference in Michigan. No wonder that one of the scopes of the conference is the statistical distributions in the era of the global pandemic! This year's conference features prominent scientists from all over the world, including four keynote and four plenary speakers. It also features topical and general invited talk sessions and a student poster session. We hope that the conference will provide an opportunity and platform for participants to share and discuss recent advances in their areas of research, as well as foster collaborative research, both on recent developments and future advances in statistical distributions and applications, and related areas.

Presenters at the conference would have the opportunity to submit their articles for consideration of publication in a special issue of the Journal of Statistical Theory and Practice. The submitted articles will go through the same rigorous peer-review process as regular submissions to the journal.

Many individuals have helped and contributed to the planning and success of the conference. We would like to thank our colleagues at Central Michigan University for allowing Marshall University to host the 4th edition of ICOSDA. Our special thanks go to our keynote speakers and plenary speakers for accepting our invitations to participate in the conference. We thank all the topic-invited session chairs and organizers, who also serve on the scientific program committee, and everyone who submitted abstracts for the general invited talks as well as poster presentations. We also want to recognize and thank members of the Organizing Committee who had worked tirelessly and enthusiastically towards a common goal of making the conference a successful, enriching, and fulfilling experience for everyone. We thank various organizations for their generous donations toward the conference. Finally, our special thanks go to the National Security Agency (NSA), the Department of Mathematics, the College of Science, and the Provost & Vice-President for Academic Affairs, Marshall University, for their financial support of the conference, and the latter for his special interest in the conference.

We are hopeful that you will find the conference, meeting your expectations and will enjoy the wonderful experience and relaxed atmosphere that our lovely city of Huntington offers.

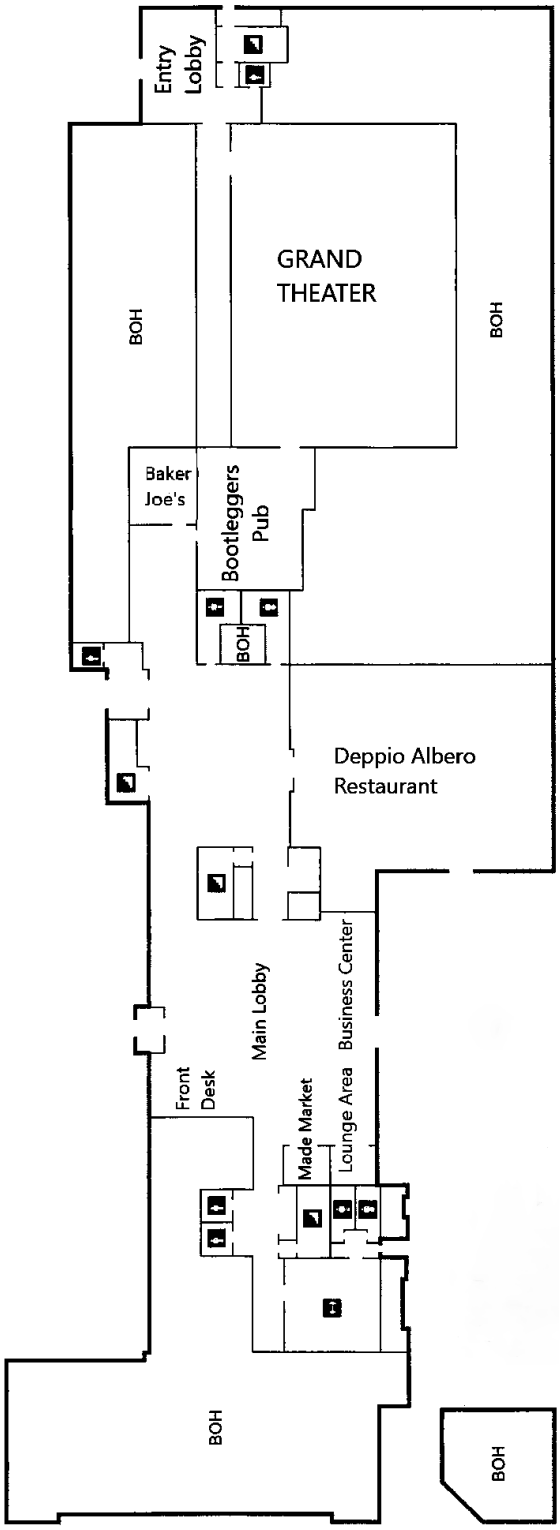
Dr. Alfred Akinsete and Dr. Avishek Mallick
Chairs, ICOSDA 2022 Organizing Committee

Organizing Committee Members

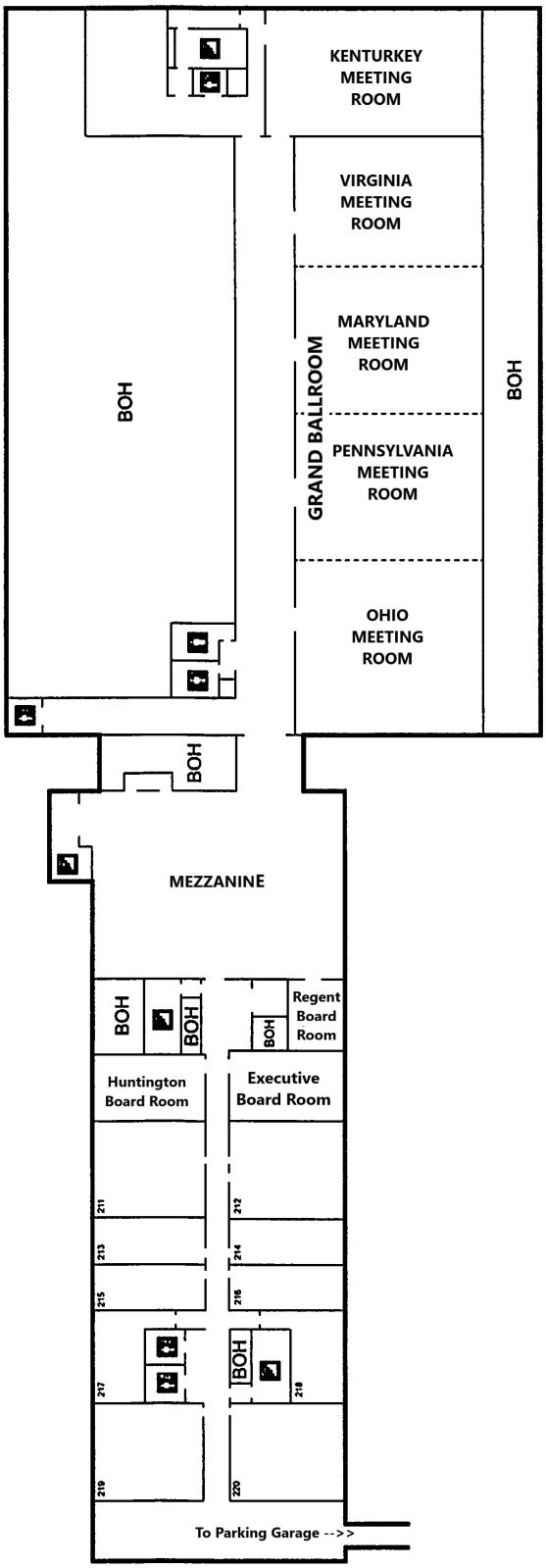
Dr. Laura Adkins
Dr. Raid Al-Aqtash
Dr. Alaa Elkadry
Dr. Carl Mummert
Dr. Michael Schroeder

Floor Maps of the DoubleTree by Hilton Huntington

First Floor



Second Floor



Friday, October 14, 2022

7:00 AM	Breakfast
8:00 AM	Opening Remarks
8:15 AM	Keynote: Barry C. Arnold **Univariate and bivariate discrete Laplace distributions and generalizations**
9:15 AM	Refreshment Break
9:30 AM	Parallel Session 1
10:50 AM	Refreshment Break
11:00 AM	Parallel Session 2
12:20 PM	Lunch
1:20 PM	Plenary: Susmitta Datta **Statistical Analysis of single cell RNA sequencing (ScRNA-seq) data**
1:55 PM	Plenary: Hon Keung Tony Ng **Degradation Models and Related Applications: From Light Intensity to Network Reliability**
2:30 PM	Refreshment Break
2:40 PM	Parallel Session 3
4:00 PM	Refreshment Break
4:10 PM	Parallel Session 4
5:30 PM	Dinner
6:30 PM	Keynote: Narayanaswamy Balakrishnan **On some forms of entropy and extropy with applications**

Saturday, October 15, 2022

7:00 AM	Breakfast
8:00 AM	Keynote: Jim Berger **Bayesian Analysis of the Covariance Matrix of a Multivariate Normal Distribution with a New Class of Prior Distributions**
9:00 AM	Refreshment Break
9:10 AM	Parallel Session 5
10:30 AM	Refreshment Break
10:40 AM	Parallel Session 6
12:00 PM	Lunch
1:00 PM	Plenary: Kimberly Sellers **Dispersed Methods for Handling Dispersed Count Data**
1:35 PM	Plenary: Carl Lee **An overview of generalized asymmetric distributions and modeling**
2:10 PM	Refreshment Break
2:20 PM	Poster Session
3:00 PM	Keynote: Thomas Mathew **Reference intervals and regions in laboratory medicine**
4:00 PM	Refreshment Break
4:15 PM	Closing and gift draw

The parallel sessions take place in the **Ohio**, **Kentucky**, **Pennsylvania**, **Maryland**, and **Virginia** rooms.

The poster session will take place in the **Ohio Room**. All other events will take place in the **Grand Theater**.

Parallel Session 1

Friday, October 14, 2022, 9:30 AM to 10:50 AM

Topic Invited Session T-1: Distributions and Applications

Chair(s): Ahmad Alzaghal, Location: Ohio Room

9:30 AM *A Versatile Family of Generalized Log-logistic Distributions*
Mahmoud Aldeni, Western Carolina University

9:50 AM *On the Gumbel Distribution and Its Generalizations*
Duha Hamed, Winthrop University

10:10 AM *A Generalized family of symmetric distributions: Properties and Applications*
Mohammad Aljarrah, Tafila Technical University

10:30 AM *A Bimodal Family Arising from the Exponentiated Exponential Distribution*
Ahmad Alzaghal, State University of New York, Farmingdale.

Topic Invited Session T-2: Statistical Distributions in Random Discrete Structures

Chair(s): Anirban DasGupta / Hosam Mahmoud, Location: Kentucky Room

9:30 AM *The Sackin index and depth of leaves in generalized Schröder trees*
Panpan Zhang, Vanderbilt University Medical Center

9:50 AM *Asymptotics of the Overflow in Urn Models*
Pawel Hitczenko, Drexel University

10:10 AM *On Hyperrecursive Trees and Their Containment Profiles*
Joshua Sparks, The George Washington University

10:30 AM *Degree profile of uniform skinny lobsters*
Hosam Mahmoud, The George Washington University

Topic Invited Session T-3: Topics in High Dimensional Inference

Chair(s): Anton Schick, Location: Pennsylvania Room

9:30 AM *Challenges in High-Dimensional Genomics - Statistical Inference with Sparse Single Cell Hi-C Data*
Shili Lin, Ohio State University

9:50 AM *Inference for High Dimensional Censored Quantile Regression*
Qi Zheng, University of Louisville

10:10 AM *Optimal False Discovery Control of Minimax Estimators*
Qifan Song, Purdue University

10:30 AM *The Bias Expansion of Parameter Estimates in High Dimensional General Estimating Equations*
Hanxiang Peng, IUPUI

Topic Invited Session T-4: Applications of Univariate Continuous and Discrete Generalized Distributions

Chair(s): Ayman Alzaatreh, Location: Maryland Room

9:30 AM *Empirical Bayes Estimation of Epidemic Reproduction Number via Branching Process and Borel-Tanner Distribution*
George Yanev, The University of Texas Rio Grande Valley & Institute ofMathematics and Informatics, Bulgarian Academy of Sciences

9:50 AM *A new class of discrete distribution arising as an analogue of gamma-Lomax distribution: Properties and Applications*
Indranil Ghosh, University of North Carolina, Wilmington, USA

10:10 AM *A Flexible Cure Model with Generalized Gamma Lifetime*
Suvra Pal, University of Texas at Arlington, USA

General Invited Session G-1

Chair(s): Alaa Elkadry, Location: Virginia Room

9:30 AM *Bayesian Analysis for Imbalanced Positive-Unlabelled Diagnosis Codes in Electronic Health Records*
Ye Liang, Oklahoma State University

9:50 AM *Jaynes-Gibbs Entropic Convex Duals & Orthogonal Polynomials*
Richard Le Blanc, Université de Sherbrooke

10:10 AM *Analysis of incomplete longitudinal data with informative drop-outs and outliers*
Sanjoy Sinha, Carleton University

10:30 AM *Bayesian Estimators of the Inverse Weibull-Weibull Composite distribution*
Nirajan Budhathoki, Central Michigan University

Parallel Session 2

Friday, October 14, 2022, 11:00 AM to 12:20 PM

Topic Invited Session T-5: Statistical Modeling and Analysis in Engineering and Biomedical Sciences

Chair(s): Suvra Pal / Tony Ng, Location: Ohio Room

11:00 AM *Semiparametric Inference in One-shot device with competing risks*
Hon Yiu So, Oakland University

11:20 AM *Statistical Modeling Approaches for the Comparison of Dissolution Profiles*
Hon Keung Tony Ng, Bentley University

11:40 AM *Using Machine Learning to Improve Predictive Accuracy of Cure*
Suvra Pal, University of Texas at Arlington

12:00 PM *Statistical learning applied to the COVID-19 pandemic study in Mexico*
Roberto Barcenas, Science Faculty, UNAM, Mexico

Topic Invited Session T-6: Interface Between Statistical Distributions and Data Analytics

Chair(s): Felix Famoye / Carl Lee, Location: Kentucky Room

11:00 AM *A Generalization of LASSO Modeling via Bayesian Interpretation of LASSO*
Gayan Warahena Liyanage, University of Dayton

11:20 AM *Framework For Generating Statistical Models With An Application*
Oluremi Abayomi, Northwood University

11:40 AM *Combining propensity score method with SMOTE for modeling imbalanced data*
Yifan Hsu, Central Michigan University

12:00 PM *Multivariate count data regression models*
Felix Famoye, Central Michigan University

Topic Invited Session T-7: Big Data: novel statistical modeling and computation for challenges in high-dimensional, spatiotemporal, and zero-inflated data

Chair(s): Hsin-Hsiung Huang, Location: Pennsylvania Room

11:00 AM *A high dimensional Cramer-von Mises test*
Mengyu Xu, University of Central Florida

11:20 AM *High-Dimensional Multivariate Time Series Forecasting for National-Level Geopolitical Events*
Hayden Hampton & Jongjin Kim, University of Central Florida

11:40 AM *Response-Aided Score-Matching Approaches for Big Data Analysis*
Keren Li, University of Alabama at Birmingham

12:00 PM *An R Package AZIAD for Analyzing Zero-Inflated and Zero-Altered Data*
Niloufar Dousti Mousavi, University of Illinois at Chicago

General Invited Session G-2

Chair(s): Olusegun Otunuga, Location: Virginia Room

11:00 AM *Goodness-of-Fit Tests for Network Structures*
Ping-Shou Zhong, University of Illinois at Chicago

11:20 AM *Dimension reduction for random objects: a projective resampling approach with application to Covid-19 transmission in the United States*
Abdul-Nasah Soale, University of Notre Dame

11:40 AM *Generalizations with T-X Method and New Transformer with Gompertz Illustration*
Scott Smith, University of the Incarnate Word

12:00 PM *Should I substitute my logisitic regression with a machine learning model?*
Jean-Francois Plante, HEC Montréal

Parallel Session 3

Friday, October 14, 2022, 2:40 PM to 4:00 PM

Topic Invited Session T-8: Recent Developments in Inference and Modeling for High-Dimensional Data

Chair(s): Wenbo Wu, Location: Ohio Room

2:40 PM *Conditional Sufficient Variable Screening for Ultrahigh Dimensional Data with FDR Control*
Chenlu Ke, Virginia Commonwealth University

3:00 PM *Conditional Martingale Difference Divergence*
Chenglong Ye, University of Kentucky

3:20 PM *Fréchet Sufficient Variable Selection with Graph Structure Among Predictors*
Jiaying Weng, Bentley University

3:40 PM *On sufficient variable screening using log odds ratio filter*
Wenbo Wu, The University of Texas at San Antonio

Topic Invited Session T-9: Fiducial Inference

Chair(s): Kalimuthu Krishnamoorthy, Location: Kentucky Room

2:40 PM *Generalized Fiducial Inference on Differentiable Manifolds*
Jan Hannig, University of North Carolina at Chapel Hill

3:00 PM *Fiducial Inference for Hypergeometric Distributions: One- and Two-Sample Problems*
Shanshan Lv, Truman State University

3:20 PM *Recent developments of accuracy metrics in biomarker evaluation and related fiducial inferences*
Lili Tian, University at Buffalo

3:40 PM *Fiducial Inference for Location-Scale Distributions*
Kalimuthu Krishnamoorthy, University of Louisiana at Lafayette

Topic Invited Session T-10: Statistical inference based upon characterizing statistical evidence

Chair(s): Laui Al Labadi / Michael Evans, Location: Pennsylvania Room

2:40 PM *Filter Method for Variable Selection Based on Relative Belief Ratio*
Ayman Alzaatreh, American University of Sharjah, Sharjah, UAE

3:00 PM *On Robustness of the Relative Belief Ratio and the Strength of its Evidence with Respect to the Geometric Contamination Prior*
Luai Al Labadi, University of Toronto

3:20 PM *Combining Evidence*
Michael Evans, University of Toronto

General Invited Session G-3

Chair(s): Alaa Elkadry, Location: Virginia Room

2:40 PM *Independence Properties of the Truncated Multivariate Elliptical Distributions*
Michael Levine, Purdue University

3:00 PM *A family of generalized multinomial distributions*
Nobuaki Hoshino, Kanazawa University

3:20 PM *Sub-Dimensional Mardia Measures of Multivariate Skewness and Kurtosis*
Joydeep Chowdhury, King Abdullah University of Science and Technology, Saudi Arabia

3:40 PM *Multivariate Probability Distributions in Light of General Stochastic Dependence Theory*
Jerzy Filus, Oakton College

Parallel Session 4

Friday, October 14, 2022, 4:10 PM to 5:30 PM

Topic Invited Session T-11: Modeling and Application

Chair(s): Raid Al-Aqtash, Location: Ohio Room

4:10 PM *Effect of News Sentiment on Stock Price: A Deep Neural Network and Statistical Analysis*
Keshab Dahal, Truman State University

4:30 PM *The Generalized Error-Generalized Hyperbolic Secant distribution: Properties and Applications*
Hazem Al-Mofleh, Tafila Technical University, Jordan

4:50 PM *Ideas Toward Analyzing Nonstationary Gaussian Process Models*
Ahmad Hanandeh, Yarmouk University

5:10 PM *Receiver operating characteristic (ROC) analysis of combining CA125 and circulating lipid metabolites to enhance the accuracy of diagnosis and triage ovarian cancer versus benign adnexal mass*
Li Yan, Roswell Park Comprehensive Cancer Center

Topic Invited Session T-12: Statistical Distribution Theory and Applications

Chair(s): Indranil Ghosh/Tony Ng, Location: Kentucky Room

4:10 PM *On characterization of the exponential distribution via hypoexponential distributions*
George Yanev, The University of Texas Rio Grande Valley & Institute ofMathematics and Informatics, Bulgarian Academy of Sciences

4:30 PM *Truncated Family of T-X family of Distributions*
Ayman Alzaatreh, American University of Sharjah, Sharjah, UAE

4:50 PM *Testing Means of Right-Skewed Populations - a Saddlepoint Approximation.*
Kalanka Jayalath, University of Houston - Clear Lake, Houston, TX

5:10 PM *Tail Conditional Expectations Based on Kumaraswamy Dispersion Models*
Indranil Ghosh, University of North Carolina Wilmington

Topic Invited Session T-13: Statistical Methods for Count Data Analysis

Chair(s): Norou Diawara, Location: Pennsylvania Room

4:10 PM *Doubly Inflated Poisson GARCH Model*
Sumen Sen, Austin Peay State University

4:30 PM *Finite Mixtures of Mean-Parameterized Conway-Maxwell-Poisson Models*
Dongying Zhan, University of Kentucky

4:50 PM *On EM Estimation for Zero- and k-Inflated Poisson Regression Model*
Rao Chaganty, Old Dominion University

5:10 PM *Copula-Based Bivariate Zero-Inflated Poisson Time Series Models*
Dimuthu Fernando, Old Dominion University

General Invited Session G-4

Chair(s): Sher Chhetri, Location: Virginia Room

4:10 PM *Odd Pareto families of distributions for modeling loss payment Data*
Nonhle Mdziniso, Rochester Institute of Technology

4:30 PM *The Two-Sided Beta Distribution with Applications in Project Risk Analysis*
Johan van Dorp, The George Washington University

4:50 PM *Copula regression for compound distributions with endogenous covariates with applications in insurance deductible pricing*
Gee Lee, Michigan State University

5:10 PM *Parameter Estimation for a Stochastic Climate Model with Alpha-Stable L'evy Jumps*
Sher Chhetri, University of South Carolina Sumter

Parallel Session 5

Saturday, October 15, 2022, 9:10 AM to 10:30 AM

Topic Invited Session T-14: Non-Gaussian Stochastic Models

Chair(s): Tomasz Kozubowski, Location: Ohio Room

9:10 AM *Simulating entering and exiting balls by an isometric stable process*
John Nolan, American University

9:30 AM *Normal Pareto Distributions: Theoretical framework and computational issues*
Tomasz J. Kozubowski, University of Nevada

Topic Invited Session T-15: Statistical advances in high dimensional data with applications in Biostatistics and Genomics

Chair(s): Souparno Ghosh, Location: Kentucky Room

9:10 AM *Mediation Analysis of high dimensional exposures with applications in genetics*
Qi Zhang, University of New Hampshire

9:30 AM *Tumor Radiogenomics with Bayesian Layered Variable Selection*
Shariq Mohammed, Boston University School of Public Health

9:50 AM *Distance-averaged CNN ensemble learning for anti-cancer drug sensitivity prediction*
Souparno Ghosh, University of Nebraska-Lincoln

10:10 AM *Estimation of practice effects in longitudinal cohorts using Bayesian hierarchical models*
Rajesh Talluri, The University of Mississippi Medical Center

Topic Invited Session T-16: Statistical Inference and Estimation

Chair(s): Mahmoud Aldeni, Location: Pennsylvania Room

9:10 AM *Adaptive Estimation for Non-parametric Multivariate Additive Model in Random Design with Long Memory errors*
Rida Benhaddou, Ohio University

9:30 AM *On the Family of Generalized Gumbel Distributions*
Raid Al-Aqtash, Marshall University

9:50 AM *Pretest and Shrinkage Estimators for Log-normal Means: Theory and Simulation*
Mahmoud Aldeni, Western Carolina University

10:10 AM *Pretest and Shrinkage Estimators for Log-normal Means: Applications*
John Wagaman, Western Carolina University

General Invited Session G-5

Chair(s): Olusegun Otunuga, Location: Virginia Room

9:10 AM *Modeling Actuarial Data Using Iterated Trigonometric Distributions*
Shahid Mohammad, University of Wisconsin Oshkosh

9:30 AM *The expected number of distinct patterns in a random permutation*
Anant Godbole, East Tennessee State University

9:50 AM *Confidence Interval for the Mean and Upper Tolerance Limit of Zero-Inflated Gamma Data*
Derek Young, University of Kentucky

10:10 AM *Time-dependent probability density function for general stochastic logistic population model with harvesting effort*
Olusegun Otunuga, Augusta University

Parallel Session 6

Saturday, October 15, 2022, 10:40 AM to 12:00 PM

Topic Invited Session T-17: Recent developments in finite mixture modeling with applications

Chair(s): Volodymyr Melnykov, Location: Ohio Room

10:40 AM *Model-based clustering analysis of the spatio-temporal and intensity patterns of tornadoes*
Rong Zheng, Western Illinois University

11:00 AM *Conditional mixture modeling and model-based clustering*
Wang Yang, College of Charleston

11:20 AM *On model-based clustering of directional data with heavy tails and outliers*
Yingying Zhang, Western Michigan University

11:40 AM *Studying contributions of variables to classification*
Yana Melnykov, The University of Alabama

Topic Invited Session T-18: Survival Analysis in Theory and in Practice: The Analysis of Time-To-Event data

Chair(s): Drew Lazar, Location: Kentucky Room

10:40 AM *Subgroup Identification with Differential Effect of Cardio-respiratory Fitness for All-cause and Disease-specific Mortality: Survival Regression-Based Recursive Partitioning Approach*
Md Yasin Ali Parh, University of Louisville

11:00 AM *A New Robust Approach for Regression Analysis of Panel Count Data with Time-varying Covariates*
Dayu Sun, Rollins School of Public Health at Emory University

11:20 AM *Inference on Mean Quality-adjusted Lifetime Using Joint Models for Continuous Quality of Life Process and Time to Event.*
Xiaotain Gao, University of Illinois at Urbana-Champaign

11:40 AM *Neural Network Models of Survival Analysis Data*
Drew Lazar, Ball State University

General Invited Session G-6

Chair(s): Gee Lee, Location: Virginia Room

10:40 AM *Functional Regression Measures of Influence on Out-of-sample Prediction*
David Hitchcock, University of South Carolina

11:00 AM *New Generalized Extreme Value Distribution with Applications to Extreme Temperature Data*
Wilson Gyasi, Northwood University

11:20 AM *Fixed-Accuracy Confidence Interval Estimation of $P(X > c)$ for a Two-Parameter Gamma Population*
Yan Zhuang, Connecticut College

11:40 AM *Nonparametric intercept regularization*
Gee Lee, Michigan State University

Poster Presentations

Saturday, October 15, 2022, 2:20PM to 3:00PM, Ohio Room

P-1: Susan Edwards, UNC Chapel Hill, Advisor: John Preisser
Estimating Higher Order Parameters of Multivariate Bernoulli Data from Partially-Sampled Clusters

P-2: Christian Sterner, University of South Carolina Sumter, Advisor: Sher B. Chhetri
Comparing Multiple Proportions using the Marascuillo Method (An analytical study of COVID-19 data in three different care facilities)

P-3: Oluwaseun Otunuga, University of South Florida, Advisor: Dr. Lu Lu
A Case of the Pareto-G Extended Weibull Distribution

P-4: Sasanka Adikari, Old Dominion University, Advisor: Dr Norou Diawra
Utility in time description in priority best worst discrete choice models: An empirical comparison using Flynn's data.

P-5: Aye Aye Maung, University of Louisville, Advisor: Dr. Drew Lazar
Induction of Survival Trees by Quadric Splits and Dipolar Splitting Criteria

Abstracts and Biographies for Keynote and Plenary Speakers

Keynote Speaker 1: Barry C. Arnold

Professor Emeritus, University of California, Riverside, USA

Professor Barry Arnold's journey in statistics began in McMaster University from where he graduated in 1961 with a bachelor's degree in mathematics (statistics). He pursued the graduate program in statistics at Stanford University, and from there joined the faculty at Iowa State University. In 1979, Barry hung up his snow shovel, donated his winter coat to the Salvation Army, and moved to the University of California, Riverside. He became a distinguished professor in the Department of Statistics, from where he retired as Professor Emeritus.

Barry is a Fellow of the American Statistical Association, of the Institute of Mathematical Statistics and of the American Association for the Advancement of Science.

He is an elected member of the International Statistical Institute. A quick look through Barry's publications shows his broad interests, but with greater attention to Ordered Data, Distribution Theory - Univariate and Multivariate, Characterization Problems, Statistical Inference - Classical and Bayesian, Inequalities and Majorization Problems, and Multivariate Analysis. His more than 270 research papers have appeared in a broad spectrum of theoretical and applied journals. He has served or is serving as an editor or associate editor of many journals including Journal of Multivariate Analysis, Journal of the American Statistical Association, Sankhya, Communications in Statistics, Metron, and many others.

Talk KN-1: Univariate and bivariate discrete Laplace distributions and generalizations

Friday, October 14, 2022, 8:15 AM to 9:15 AM, Grand Theater

Chair: Alfred Akinsete

Abstract KN-1: Univariate discrete Laplace models are described and investigated. A more general mixture model is presented. Bivariate extensions of these models are discussed in some detail, with particular emphasis on associated parameter estimation strategies. Multivariate versions of the models are briefly introduced.

Keynote Speaker 2: Narayanaswamy Balakrishnan

Professor, McMaster University, Ontario, Canada

Professor Balakrishnan is a Distinguished University Professor at McMaster University, Hamilton, Ontario, Canada. He completed his PhD in 1981 from the Indian Institute of Technology, Kanpur, India. He is a Fellow of the American Statistical Association, Fellow of the Institute of Mathematical Statistics, Elected Member of the International Statistical Institute, Honorary Member of the Greek Statistical Institute, and many more, including the Professor C. R. Rao Lifetime Achievement Award in 2020, by the Indian Society for Probability and Statistics. His research interests include ordered data analysis, univariate and multivariate distribution theory, reliability theory, survival analysis, applied probability, stochastic orderings, nonparametric statistics, censoring methodology, and statistical inference. "Bala", as he prefers to be called, has well over 75 thousand citations of his work and a H-Index of 87! He holds editorial positions in more than 20 scholarly and professional journals, as well as over 60 distinguished visiting professorship positions in many universities around the world. He has been awarded an Honorary Doctorate Degree from the National and Kapodistrian University of Athens, Greece.

Talk KN-2: On some forms of entropy and extropy with applications

Friday, October 14, 2022, 6:30 PM to 7:30 PM, Grand Theater

Chair: Felix Famoye

Abstract KN-2: Entropy, since introduced in the pioneering work of Shannon, has found key applications in a wide range of fields. During the past two decades, different forms of such information measures have been discussed in the literature. In this talk, I will review a number of them, and provide some unifying results and then establish some key mathematical properties. Finally, I will illustrate some of their applications to classification and image processing problems.

Keynote Speaker 3: Jim Berger

Professor Emeritus, Duke University, Durham, NC, USA

Professor Berger received his Ph.D. degree in mathematics from Cornell University in 1974. He was a faculty member in the Department of Statistics at Purdue University until 1997, at which time he moved to the Department of Statistical Science at Duke University, where he is currently the Arts and Sciences Distinguished Professor Emeritus of Statistics. He was the founding director of the Statistical and Applied Mathematical Sciences Institute, serving from 2002-2010. Berger was president of the Institute of Mathematical Statistics during 1995-1996 and president of the International Society for Bayesian Analysis during 2004. His editorial activities include co-editorship of the Annals of Statistics during 1998-2000 and being a founding editor of the Journal on Uncertainty Quantification, serving from 2012-2015.

Berger has received Guggenheim and Sloan Fellowships, the COPSS President's Award in 1985, the Sigma Xi Research Award at Purdue University for contribution of the year to science in 1993, the COPSS Fisher Lecturer in 2001, the Wald Lecturer of the IMS in 2007 and the Wilks Award from the ASA in 2015. He was elected as foreign member of the Spanish Real Academia de Ciencias in 2002, elected to the USA National Academy of Sciences in 2003, was awarded an honorary Doctor of Science degree from Purdue University in 2004, and became an Honorary Professor at East China Normal University in 2011.

Berger's research has primarily been in Bayesian statistics, foundations of statistics, statistical decision theory, simulation, model selection, and various interdisciplinary areas of science and industry, including astronomy, geophysics, medicine, and validation of complex computer models. He has supervised 38 Ph.D. dissertations, published over 200 papers and has written or edited 16 books or special volumes.

Talk KN-3: Bayesian Analysis of the Covariance Matrix of a Multivariate Normal Distribution with a New Class of Prior Distributions

Saturday, October 15, 2022, 8:00 AM to 9:00 AM, Grand Theater

Chair: Alfred Akinsete

Abstract KN-3: We propose a new class of prior distributions for the covariance matrix of a multivariate normal distribution. The main motivation for the new class is to have available priors – both subjective and objective – that do not "force eigenvalues apart," which is a criticism of inverse Wishart and Jeffreys priors. Extensive comparison of these 'shrinkage priors' with inverse Wishart and Jeffreys priors is undertaken, with the new priors seeming to have considerably better performance. A number of curious facts about the new priors are also observed, such as that the posterior distribution will be proper with just three vector observations from the multivariate normal distribution – regardless of the dimension of the covariance matrix – and that useful inference about certain features of the covariance matrix can be made, even with many fewer observations than needed for identifiability. Finally, a new MCMC algorithm is developed for this class of priors and is shown to be computationally effective for matrices of up to 100 dimensions.

Keynote Speaker 4: Thomas Mathew

Professor, University of Maryland, Baltimore County, USA

Thomas Mathew is a professor of statistics at the Department of Mathematics and Statistics, University of Maryland, Baltimore County campus. He earned his Ph.D. in statistics from the Indian Statistical Institute in 1983. His research interests include methodological research in statistics, as well as novel applications of statistical methodology to other disciplines, including exposure data analysis, cost-effectiveness analysis, bioequivalence testing and the development of reference intervals and regions in laboratory medicine. He is the co-author of two books, one on hypothesis testing in mixed models, and a second book on tolerance intervals and regions, both published by Wiley.

Dr. Mathew is an elected Fellow of both the Institute of Mathematical Statistics and the American Statistical Association. In recognition of his research accomplishments, Dr. Mathew was awarded the Presidential Research Professorship at his campus in 2008.

Talk KN-4: Reference intervals and regions in laboratory medicine

Saturday, October 15, 2022, 3:00 PM to 4:00 PM, Grand Theater

Chair: Kalimuthu Krishnamoorthy

Abstract KN-4: Reference intervals are data-based intervals that are meant to capture a pre-specified large proportion of the population values of a clinical marker or analyte in a healthy population. They can be one-sided or two-sided, and they are widely used in the interpretation of results of biochemical and physiological tests of patients. A population reference range is typically expected to include 95% of the population distribution, and reference limits are often taken to be the 2.5th and 97.5th percentiles of the distribution, which is especially meaningful if normality is appropriate. Usually, the reference range is constructed based on a random sample and simply estimating the percentiles is clearly not satisfactory. This calls for the use of appropriate criteria for estimating the reference range from a random sample. When there are multiple biochemical analytes measured from each subject, a multivariate reference region is needed. Traditionally, under multivariate normality, reference regions have been constructed as ellipsoidal regions. This approach suffers from a major drawback: it cannot detect component-wise extreme observations. Thus rectangular reference regions need to be constructed based on appropriate criteria. The talk will review univariate reference intervals and multivariate reference regions, and the criteria that can be used in their construction. Both parametric and non-parametric scenarios will be addressed, and laboratory medicine examples will be used for illustration.

Plenary Speaker 1: Susmitta Datta

Professor, University of Florida

Dr. Susmita Datta is Professor at University of Florida (UF), Department of Biostatistics. She is the Co-Director of the Biostatistics, Epidemiology and Research Design Program (BERD) of UF Clinical and Translational Science Institute. Dr. Datta is widely (¿100) published in peer reviewed journals. Her work has been continuously funded by the National Science Foundation and the National Institutes of Health. She is a fellow of the American Statistical Association (ASA), an elected member of the International Statistical Institute (ISI), and fellow of the American Association for the Advancement of Science (AAAS). Her research area includes bioinformatics, genomics, proteomics, metabolomics, lipidomics, clustering and classification techniques, infectious disease modeling, statistical issues in population biology, systems biology, survival analysis, multi-state models and big data analytics. She has recently published a book on "Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry" by Springer. Professor Datta is enthusiastic in promoting women in STEM fields and has served as President of Caucus for Women in Statistics (CWS) and is presently appointed to the Committee of Women in Statistics of ASA (COWIS). She is the founding executive committee member of the Women in Statistics and Data Science conference (WSDS).

Talk PL-1: Statistical Analysis of single cell RNA sequencing (ScRNA-seq) data

Friday, October 14, 2022, 1:20 PM to 1:55 PM, Grand Theater

Chair: Carl Lee

Abstract PL-1: Transcriptomic studies such as in bulk RNA-sequencing, one can examine transcript abundance measurements averaged over bulk populations of thousands (or even millions) of cells. While these measurements have been valuable in countless studies, they often conceal cell-specific heterogeneity in expression signals that may be paramount to new biological findings. Fortunately, with single cell RNA-sequencing (scRNA-Seq), transcriptome data from individual cells are now accessible, providing opportunities to investigate functional states of cells, identify rare cell populations and uncover diverse gene expression patterns in cell populations that seemed homogeneous. Most importantly it provides an unprecedented resolution to the characterization of cellular clinical isolates. However, there are challenges analyzing such scRNA-Seq data. Amongst many challenges, the most significant are the bimodal or multimodal distribution, sparsity and tremendous heterogeneity in the data. Consequently, we will describe potential ways of statistical modeling of such data, finding differentially expressed genes and methods for constructing gene-gene interaction network using this data. Collaborators: Drs. Jeremy Gaskins and Michael Sekula (University of Louisville)

Plenary Speaker 2: Hon Keung Tony Ng

Professor, Bentley University, Waltham, Massachusetts

Hon Keung Tony Ng is a Professor in the Department of Mathematical Sciences at Bentley University (Waltham, Massachusetts). Before joining Bentley University, Dr. Ng has been a faculty member in the Department of Statistical Science, Southern Methodist University (Dallas, Texas) since 2002. He received a B.Sc. (Hon.) degree and an M.Phil. degree in statistics from The Chinese University of Hong Kong in 1997 and 1999, and an M.Sc. degree in statistics and a Ph.D. degree in mathematics from McMaster University (Hamilton, ON, Canada), in 2000, and 2002, respectively. His research interests include reliability, censoring methodology, network analysis, nonparametric methods, ordered data analysis, and statistical inference.

Dr. Ng has published more than 150 research papers in refereed journals and has co-authored/co-edited eight books. He is serving as an Associate Editor or Editorial Board Member for different scientific journals in statistics and engineering, including Communications in Statistics, Computational Statistics, IEEE Transactions on Reliability, International Journal of Reliability, Quality and Safety Engineering, Journal of Statistical Computation and Simulation, Naval Research Logistics, Sequential Analysis, and Statistics & Probability Letters. He is an elected Senior Member of the Institute of Electrical and Electronics Engineers (IEEE) (in 2008), an Elected Member of the International Statistical Institute (in 2008), and a Fellow of the American Statistical Association (in 2016).

Talk PL-2: Degradation Models and Related Applications: From Light Intensity to Network Reliability

Friday, October 14, 2022, 1:55 PM to 2:30 PM, Grand Theater

Chair: Carl Lee

Abstract PL-2: In engineering and sciences, the process that a system reduces in performance, reliability, or life span of assets gradually and irreversibly is known as a degradation process. Degradation measurements are recorded over time for prognostics and health management purposes. The gamma degradation model has been used to characterize the evolution of degradation measurements. In this talk, I will first introduce the gamma and Wiener degradation models. Then, I will present a multi-phase gamma degradation model and discuss the likelihood and Bayesian inference for this model. The gamma model and the inferential methods are applied to analyze a real data set of light-emitting diodes (LEDs). I will also discuss the application of the degradation models in power grid network reliability analysis and other applications such as network design and cybersecurity. In addition to engineering applications, an application of the degradation models in biopharmaceutical statistics will also be discussed briefly.

Plenary Speaker 3: Kimberly Sellers

Professor, Georgetown University, Washington, DC

Kimberly Sellers, Ph.D. is a Professor of Mathematics and Statistics, specializing in Statistics at Georgetown University in Washington, DC; and a Principal Researcher with the Center for Statistical Research and Methodology Division of the U.S. Census Bureau. A DC-area native, she completed her BS and MA degrees in Mathematics at the University of Maryland College Park, and then obtained her PhD in Mathematical Statistics at The George Washington University.

Prof. Sellers held previous faculty positions at Carnegie Mellon University as a Visiting Assistant Professor of Statistics, and the University of Pennsylvania School of Medicine as an Assistant Professor of Biostatistics and Senior Scholar at the Center for Clinical Epidemiology and Biostatistics (CCEB) before her return to the DC area. Her primary research interests and expertise center on statistical methods for count data that contain data dispersion with methodological interests in distribution theory, regression analysis, multivariate analysis, and stochastic processes and time series analysis.

Prof. Sellers recently became an Elected Member of the International Statistical Institute (in 2018) and a Fellow in the American Statistical Association (in 2021) in recognition of her research, as well as her active contributions to diversifying the fields of mathematical and statistical sciences with respect to gender and race/ethnicity. She was the 2017-2018 Chairperson for the American Statistical Association's Committee on Women in Statistics, and currently serves as the inaugural chairperson of the ASA's Justice, Equity, Diversity, and Inclusion (JEDI) Outreach Group, as well as an Advisory Board member for the Black Doctoral Network.

Talk PL-3: Dispersed Methods for Handling Dispersed Count Data

Saturday, October 15, 2022, 1:00 PM to 1:35 PM, Grand Theater

Chair: Souparno Ghosh

Abstract PL-3: While the Poisson distribution is a classical statistical model for count data, it hinges on the constraining equi-dispersion property (i.e. that the mean and variance are equal). This assumption, however, does not usually hold for real count data; over-dispersion (i.e. when the variance is greater than the mean) is a more common phenomenon for count data, however, data under-dispersion has also been prevalent in various settings. It would be more convenient to work with a distribution that can effectively model data (over- or under-)dispersion because it can offer more flexibility (and, thus, more appropriate inference) in the statistical methodology. This talk introduces the Conway-Maxwell-Poisson distribution along with associated statistical methods motivated by this model to better analyze count data under various scenarios (e.g. distributional theory, generalized linear modeling, control chart theory, and count processes). As time permits, this talk will likewise acquaint the audience with available associated tools for statistical computing.

Plenary Speaker 4: Carl Lee

Professor, Central Michigan University, Mt. Pleasant, MI

Professor Lee is the founding chair of the Department of Statistics, Actuarial and Data Sciences (2019-2022) at Central Michigan University. He received B.S. in Agronomy from National Taiwan University (1976) and PhD in Statistics from Iowa State University (1984). He is the cofounder of ICOSDA (with Felix Famoye). His current research interests include statistical distributions and applications, predictive modeling and applications and statistical education. He is a Dow Certified Six Sigma Black Belt. He is one of the pioneer statistics educators involving with Statistics Education Reform in early 2000, one of the founding members of the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE), and the founder of the Undergraduate Statistics Project Competition (known as USPROC).

Dr. Lee is a Fellow of the American Statistical Association and an Elected member of the International Statistical Institute. He received the 2019 Deborah & Franklin Tepper Haimo Award for Distinguished College or University Teaching of Mathematics by the Mathematical Association of America, and the 2022 Distinguished Professor Award by the Michigan Association of State University.

Talk PL-4: An overview of generalized asymmetric distributions and modeling

Saturday, October 15, 2022, 1:35 PM to 2:10 PM, Grand Theater

Chair: Souparno Ghosh

Abstract PL-4: Research in developing new and more flexible statistical distributions with more parameters for the purpose of fitting more complicated real-world phenomena has been very active during the recent decades. One critical criticism is that new distributions with many parameters may not be practically useful. Although they can fit broader types of data, these parameters may not have meaningful interpretations. Most practically useful distributions often have three parameters (location, scale and shape parameters). However, commonly used three-parameter distributions often are limited to fit a wide range of skewed data. In this talk, methods for developing three-parameter asymmetric distributions will be reviewed. A framework will be proposed to derive three-parameter generalized asymmetric distributions that are capable of fitting a wide range of skewed (left and right), highly leptokurtic or platykurtic distributions of data. Some special cases will be discussed and compared with other more complex distributions with applications.

Abstracts for
Topic Invited Sessions

Session T-1: Distributions and Applications

Friday, October 14, 2022, 9:30 AM to 10:50 AM, Ohio Room

Chair: Ahmad Alzaghal

Talk T-1-1: A Versatile Family of Generalized Log-logistic Distributions

Mahmoud Aldeni, Western Carolina University

Friday, October 14, 2022, 9:30 AM – 9:50 AM

Abstract T-1-1: In real-world applications, it is not uncommon to encounter situations in which a set of data exhibits asymmetry and bimodality. In this research, we propose a new versatile family of generalized log-logistic distributions using the method of T-R{Y} framework. The resulting flexible classes of this family includes both unimodal and bimodal distributions which can be expected to model a wide variety of data with different levels of skewness. The usefulness and goodness of-fit of some members of these classes are illustrated by means of six real data sets. In addition, a new generalized log-logistic lifetime regression model is introduced and applied to fit a right-censored data with covariates.

Talk T-1-2: On the Gumbel Distribution and Its Generalizations

Duha Hamed, Winthrop University

Friday, October 14, 2022, 9:50 AM – 10:10 AM

Abstract T-1-2: In this research, we present a review on the Gumbel distribution; the distribution that is most used in climate modeling, some of its generalizations, and some of its mathematical properties. We review many of the recent Gumbel generalizations generalized by the different generalizing method techniques. We then present new flexible generalizations for Gumbel distribution, to provide a more flexible generalization of Gumbel to include different shapes including bimodality. A real-life data application is presented to compare some of the new presented generalized Gumbel distributions with other Gumbel generalizations.

Talk T-1-3: A Generalized family of symmetric distributions: Properties and Applications

Mohammad Aljarrah, Tafila Technical University

Friday, October 14, 2022, 10:10 AM – 10:30 AM

Abstract T-1-3: Generalized distributions are valuable to applied statisticians, and some popular distributions can be extended in several ways. We introduce a family of new generalized symmetric distributions using the distribution function of a symmetric random variable. We derive some properties of this class, and a new generalized symmetric family is derived and investigated. The distributions of this family can be symmetric, left-skewed, right-skewed, and the usefulness of the generated distributions are established through applications to some data sets.

Talk T-1-4: A Bimodal Family Arising from the Exponentiated Exponential Distribution

Ahmad Alzaghal, State University of New York, Farmingdale.

Friday, October 14, 2022, 10:30 AM – 10:50 AM

Abstract T-1-4: In real-life scenarios there are several situations in which the data is bimodal, therefore unimodal distributions cannot be used to model such situations. In this research, we present a family of generalized Exponentiated-exponential distributions. The behavior of the distributions in this family can be either unimodal or bimodal, additionally, they exhibit various shape and tail properties that are consistent with data derived from several real-life phenomena. We illustrate the importance of the family by means of various applications to real data sets.

Session T-2: Statistical Distributions in Random Discrete Structures

Friday, October 14, 2022, 9:30 AM to 10:50 AM, Kentucky Room

Chair: Anirban DasGupta / Hosam Mahmoud

Talk T-2-1: The Sackin index and depth of leaves in generalized Schröder trees

Panpan Zhang, Vanderbilt University Medical Center

Friday, October 14, 2022, 9:30 AM – 9:50 AM

Abstract T-2-1: Schröoder trees are biological models of evolution, with internal nodes having two or three children. We generalize the model to grow from an arbitrary stochastic process of independent nonnegative integers (not necessarily identically distributed). We call such a process the building sequence. We study the depth of leaves and the Sackin index for some specific building sequences, such as constant additions, Bernoulli, and Poisson-like models.

Talk T-2-2: Asymptotics of the Overflow in Urn Models

Pawel Hitczenko, Drexel University

Friday, October 14, 2022, 9:50 AM – 10:10 AM

Abstract T-2-2: We consider a finite or infinite collection of urns, each with capacity r, and balls distributed among them according to a given distribution. An overflow is the number of balls that would fall to urns that already contain r balls. When r is 1 this is a well-studied quantity, namely the number of balls landing in non-empty urns. We use martingale methods to study the asymptotics of the overflow in the general situation, i.e. for arbitrary r. In particular, we provide sufficient conditions for both Poissonian and normal asymptotics for the overflow.

Talk T-2-3: On Hyperrecursive Trees and Their Containment Profiles

Joshua Sparks, The George Washington University

Friday, October 14, 2022, 10:10 AM – 10:30 AM

Abstract T-2-3: Within the area of graph theory lies an extension known as the hypergraph: a generalization of graphs to which we have vertices along with hyperedges consisting of collections of vertices. We take the well-studied structure of the recursive tree and apply its framework within the context of hypergraphs to form hyperrecursive trees, an area that shows promise in network theory and beyond. We investigate the hyperrecursive tree through its local containment profile, which observes the number of hyperedges that contains a given vertex, and its global containment profile, observing the number of vertices found within a particular containment level. We then establish an asymptotically normal distribution for the number of vertices at the smallest containment level as well as second mixed moments for the number of vertices at the smallest two containment levels. This is joint work with Hosam Mahmoud and Srinivasan Balaji (The George Washington University).

Talk T-2-4: Degree profile of uniform skinny lobsters

Hosam Mahmoud, The George Washington University

Friday, October 14, 2022, 10:30 AM – 10:50 AM

Abstract T-2-4: Deterministic lobsters—a family of trees—have been in the literature for quite some time. However, not much work has been done on random lobsters. We introduce a class of random lobsters that grows incrementally to have "skinny legs." The skinny-legs lobster has three strata, a path we call the spine, nodes in the first stratum at distance 1 and nodes in the second stratum at distance 2 from the spine. A node in the first stratum can have at most one child in the second stratum, which keeps the legs "thin." The motivation is from chemistry, as certain hydrocarbons take the shape of skinny lobsters.

We study the leaves in the lobster and find that, as the lobster grows in size, the distribution of their number converges to a fair mixture of two Poisson distributions (one of them is shifted), both parametrized by the number of nodes in the spine. We also find the asymptotic average Zagreb index of such a family of lobsters.

Session T-3: Topics in High Dimensional Inference

Friday, October 14, 2022, 9:30 AM to 10:50 AM, Pennsylvania Room

Chair: Anton Schick

Talk T-3-1: Challenges in High-Dimensional Genomics - Statistical Inference with Sparse Single Cell Hi-C Data

Shili Lin, Ohio State University

Friday, October 14, 2022, 9:30 AM – 9:50 AM

Abstract T-3-1: Single cell Hi-C (scHi-C) techniques enable one to study cell to cell variability in chromatin interactions. However, the prevalence of dropout events is a serious problem in single cell Hi-C data due to insufficient sequencing depth and coverage. Sparsity is indeed one of the major difficulties in analyzing single cell data, and it is particularly challenging for scHi-C data, as sparsity is an order of magnitude more severe compared to most of other types of single-cell data: scHi-C data are represented as two-dimensional (2D) contact matrices and its coverage (0.25-1%) is much smaller than that of single cell RNA-seq (scRNA-seq, 5-10%). Complicating things further is the fact that dropouts are confounded with structural zeros due to underlying biological mechanisms, leading to observed zeros being a mixture of both types of events. Differentiating between structural zeros and sampling zeros is important since correct inference would improve downstream analyses such as clustering of cells. Although considerable effort has been devoted to addressing the problem of structural zeros and dropouts in scRNA-seq data, little is done in scHi-C research.

In this talk, we will first discuss an adaptation of several methods from the scRNA-seq literature for inference on observed zeros in scHi-C and the evaluation of their performance. Although successful in some situations, inherent features in scHi-C 2D data matrices, including spatial correlations, cannot be easily accounted for with scRNA-seq methods, leading to deficiencies in downstream analyses. We will then describe models designed specifically for scHi-C data to infer structural zeros, impute dropouts, and improve data quality. Specifically, we will describe HiCImpute, a Bayesian hierarchy model that not only takes spatial dependencies of scHi-C 2D data structure into account, but also borrows information from similar single cells and bulk data when such are available. Through an extensive set of analyses of synthetic and real data, we will demonstrate the ability of HiCImpute for identifying structural zeros with high sensitivity, and for accurate imputation of dropout values in sampling zeros. We will also demonstrate that downstream analyses using data improved from HiCImpute can yield much more accurate clustering of cell types and discovery of subtypes compared to using observed data or data improved by several comparison methods.

Talk T-3-2: Inference for High Dimensional Censored Quantile Regression

Qi Zheng, University of Louisville

Friday, October 14, 2022, 9:50 AM – 10:10 AM

Abstract T-3-2: With the availability of high dimensional genetic biomarkers, it is of interest to identify heterogeneous effects of these predictors on patients' survival, along with proper statistical inference. Censored quantile regression has emerged as a powerful tool for detecting heterogeneous effects of covariates on survival outcomes. However, to our knowledge, few works are available to draw inference on the effects of high dimensional predictors for censored quantile regression. We propose a novel fused procedure to draw inference on all predictors within the framework of "global"" censored quantile regression, where the quantile level is over an interval, instead of several discrete values. The proposed estimator combines a sequence of low dimensional model fitting based on multi-sample splitting and variable selection. We show that, under some regularity conditions, the estimator is consistent and asymptotically follows a Gaussian process indexed by the quantile level. Simulation studies indicate that our procedure properly quantifies the uncertainty of effect estimates in high-dimensional

settings. We apply our method to analyze the heterogeneous effects of SNPs residing in the lung cancer pathways on patients' survival, using the Boston Lung Cancer Survivor Cohort, a cancer epidemiology study investigating the molecular mechanism of lung cancer.

Talk T-3-3: Optimal False Discovery Control of Minimax Estimators

Qifan Song, Purdue University

Friday, October 14, 2022, 10:10 AM – 10:30 AM

Abstract T-3-3: Two major research tasks lie at the heart of high dimensional data analysis: accurate parameter estimation and correct support recovery. The existing literature mostly aims for either the best parameter estimation or the best model selection result, however little has been done to understand the potential interaction between the estimation precision and the selection behavior. In this work, our minimax result shows that an estimator's performance of type I error control directly links with its L_2 estimation error rate, and reveals a trade-off phenomenon between the rate of convergence and the false discovery control: to achieve better accuracy, one risks yielding more false discoveries. In particular, we characterize the false discovery control behavior of rate optimal and rate suboptimal estimators under different sparsity regimes, and discover a rigid dichotomy between these two estimators under near-linear and linear sparsity settings. In addition, this work provides a rigorous explanation to the incompatibility phenomenon between selection consistency and rate minimaxity which has been frequently observed in the high dimensional literature.

Talk T-3-4: The Bias Expansion of Parameter Estimates in High Dimensional General Estimating Equations

Hanxiang Peng, IUPUI

Friday, October 14, 2022, 10:30 AM – 10:50 AM

Abstract T-3-4: In this talk, I'll present the bias expansions of the estimates for growing parameter dimension. The rates of the remainders are given under both entry-wise and global conditions on some random matrices. I'll discuss two main issues for calculating the bias. The first one is that estimates as the solutions to GEE are not well defined on the whole space, and the second is the integrability of the spectral norm of the inverse of some random matrices. My approach to the first issue is a generalization of the vector mean value theorem. For the second, I rely on the results of the modern random matrix theory. As an application, I verify the conditions for generalized linear models under the usual assumptions and derive the bias expansions with appropriate rates for the remainders. As a second application, bias-corrected estimates are constructed. Finally, I'll report a small simulation study.

Session T-4: Applications of Univariate Continuous and Discrete Generalized Distributions

Friday, October 14, 2022, 9:30 AM to 10:50 AM, Maryland Room

Chair: Ayman Alzaatreh

Talk T-4-1: Empirical Bayes Estimation of Epidemic Reproduction Number via Branching Process and Borel-Tanner Distribution

George Yanev, The University of Texas Rio Grande Valley & Institute ofMathematics and Informatics, Bulgarian Academy of Sciences

Friday, October 14, 2022, 9:30 AM – 9:50 AM

Abstract T-4-1: An epidemic outbreak can be modelled as a branching process in which the total progeny (outbreak size) follows a Borel-Tanner distribution. We construct empirical Bayes estimators for the epidemic reproduction number θ based on such a model. Simulation results indicate that the empirical estimator suffers from "jumpiness." We proceed to monotonizing it by Houwelingen's method. The monotonized estimator is much superior as having lower regret risk. Lastly, we construct an empirical Bayes estimator for θ, assuming that the initial number of infected individuals is a Poisson random variable.

Talk T-4-2: A new class of discrete distribution arising as an analogue of gamma-Lomax distribution: Properties and Applications

Indranil Ghosh, University of North Carolina, Wilmington, USA

Friday, October 14, 2022, 9:50 AM – 10:10 AM

Abstract T-4-2: This article represents how certain types of blockades in any industrial (heavy industries) production, for example, industrial strikes can be fitted with the proposed discrete probabilistic distribution. We consider the number of instances of strikes in the coal mining industry, the vehicle manufacturing industry, and the transport industry in the UK obtained from Consul (1989). We fit the data sets with the proposed discrete gamma-Lomax distribution and compare the fit with the generalized Poisson distribution (Consul, 1989). For this purpose, we explore some structural properties of the discrete gamma-Lomax distribution including hazard functions, moments, and maximum likelihood estimation under the classical setup. We briefly discuss the method of maximum likelihood estimation under the censored data set scenario. It is observed that the newly proposed model can be useful to describe strikes arising from various types of industries.

Talk T-4-3: A Flexible Cure Model with Generalized Gamma Lifetime

Suvra Pal, University of Texas at Arlington, USA

Friday, October 14, 2022, 10:10 AM – 10:30 AM

Abstract T-4-3: In this talk, I will present a two-way flexible cure rate model. The first flexibility is provided by considering a family of Box-Cox transformation cure models that includes the commonly used cure models as special cases. The second flexibility is provided by proposing the wider class of generalized gamma distributions to model the associated lifetime. The advantage of this two-way flexibility is that it allows us to carry out tests of hypotheses to select an adequate cure model

(within the family of Box-Cox transformation cure models) and a suitable lifetime distribution (within the wider class of generalized gamma distributions) that jointly provides the best fit to a given data. I will present the model fitting results as well as the bias and efficiency of the estimates of the cure rates under model mis-specification. The findings strongly suggest the importance of selecting a correct lifetime distribution and a correct cure rate model, which can be achieved through the proposed two-way flexible model. I will conclude my talk by analyzing a real breast cancer data.

Session T-5: Statistical Modeling and Analysis in Engineering and Biomedical Sciences

Friday, October 14, 2022, 11:00 AM to 12:20 PM, Ohio Room

Chair: Suvra Pal / Tony Ng

Talk T-5-1: Semiparametric Inference in One-shot device with competing risks

Hon Yiu So, Oakland University

Friday, October 14, 2022, 11:00 AM – 11:20 AM

Abstract T-5-1: One-shot devices are products that will be destroyed immediately after use. Hence, only such products' binary status, success or failure, can be observed instead of its lifetime. To avoid misspecification in lifetime distributions of the components, we proposed a proportional hazards model for analyzing the relationship between the lifetime of the components and the stress level. This study tests the one-shot devices under constant stress accelerated life-test. A link function relating to stress levels and lifetime is then applied to extrapolate the lifetimes of units from accelerated conditions to normal operating conditions. Most of the one-shot devices consist of more than one component. Malfunctioning any one of the components will result in the device's failure. The failed devices are inspected to identify the specific cause of failure. We will analyze the competing risks model with the proportional hazard assumption.

Talk T-5-2: Statistical Modeling Approaches for the Comparison of Dissolution Profiles

Hon Keung Tony Ng, Bentley University

Friday, October 14, 2022, 11:20 AM – 11:40 AM

Abstract T-5-2: Dissolution studies are a fundamental component of pharmaceutical drug development, yet many studies rely upon the f1 and f2 model-independent approach that is not capable of accounting for uncertainty in parameter estimation when comparing dissolution profiles. In this paper, we deal with the issue of uncertainty quantification by proposing several model-dependent approaches for assessing the similarity of two dissolution profiles. We take a statistical modeling approach and allow the dissolution data to be modeled using either a Dirichlet distribution, gamma process model, or Wiener process model. These parametric forms are shown to be reasonable assumptions that are capable of modeling dissolution data well. Furthermore, based on a given statistical model, we are able to use the f1 difference factor and f2 similarity factor to test the equivalency of two dissolution profiles via bootstrap confidence intervals. Illustrations highlighting the success of our methods are provided for both Monte Carlo simulation studies, as well as real dissolution data sets.

Talk T-5-3: Using Machine Learning to Improve Predictive Accuracy of Cure

Suvra Pal, University of Texas at Arlington

Friday, October 14, 2022, 11:40 AM – 12:00 PM

Abstract T-5-3: In this talk, I will present a new promotion time cure model (PCM) that uses the support vector machine (SVM) to model the incidence part. The proposed model inherits the features of the SVM and provides flexibility in capturing non-linearity in the data. Furthermore, the new model can incorporate potentially high dimensional covariates. For the estimation of model parameters, I will discuss the steps of an expectation maximization algorithm where I will make use of

the sequential minimal optimization technique together with the Platt scaling method. Next, I will present the results of a detailed simulation study and show that the proposed model outperforms the existing logistic regression-based PCM model, specifically when the true classification boundary is non-linear. I will also show that the proposed model's ability to capture complex classification boundaries can improve the estimation results related to the latency part. Finally, I will analyze a data from leukemia cancer study and show that the proposed model results in improved predictive accuracy.

Talk T-5-4: Statistical learning applied to the COVID-19 pandemic study in Mexico

Roberto Barcenas, Science Faculty, UNAM, Mexico

Friday, October 14, 2022, 12:00 PM – 12:20 PM

Abstract T-5-4: Understanding the SARS-CoV-2 infection causing COVID-19 disease among the population was crucial to determine the risk factors associated with severe cases and to prioritize their hospitalization. We conducted our study in two stages. Using a dataset from the Mexican Ministry of Health, we performed a multiclass classification for risk detection in COVID-19 patients following three statistical learning algorithms: Random Forest (89.86%), GBM (89.37%) XGBoost (89.97%). In addition, based on patient data from Mexico City, a set of classifiers weighted by the cross-entropy measure is proposed. This method enhanced the detection of cases susceptible to hospitalization, with an accuracy of 91.46%, and in a restrictive scenario, represents a preventive alert for patients. In both cases, features such as sex, age, days since symptom onset, symptoms such as dyspnea and polypnea and other comorbidities, were considered. The result was improved risk assessment to avoid progression of the disease into a severe stage requiring hospitalization.

Session T-6: Interface Between Statistical Distributions and Data Analytics

Friday, October 14, 2022, 11:00 AM to 12:20 PM, Kentucky Room

Chair: Felix Famoye / Carl Lee

Talk T-6-1: A Generalization of LASSO Modeling via Bayesian Interpretation of LASSO

Gayan Warahena Liyanage, University of Dayton

Friday, October 14, 2022, 11:00 AM – 11:20 AM

Abstract T-6-1: The purpose of this study is to introduce a generalized LASSO regression model that is derived from a generalized Laplace (GL) distribution. A family of GL distributions is derived via the T-R{Y} framework. Based on a special case of GL distributions, the Bayesian interpretation of LASSO is used to obtain some additional components to the constraint in the ordinary LASSO approach. Geometric interpretations of the additional components are presented. Some effects of the parameters of the GL distribution on the generalized LASSO regression model are examined. Lastly, the generalized LASSO regression model is demonstrated as being flexible and useful in selecting variables with better prediction capability on a real-world dataset.

Talk T-6-2: Framework For Generating Statistical Models With An Application

Oluremi Abayomi, Northwood University

Friday, October 14, 2022, 11:20 AM – 11:40 AM

Abstract T-6-2: A framework for developing generalized statistical models based on families of distributions arising from the T-R{Z} framework has been proposed. A four-parameter generalization of the Dagum distribution, the Kumaraswamy-Dagum distribution (KDD) from the T-R{Z} framework is defined. The hazard function of the KDD can be monotonic decreasing or have an upside-down bathtub shape for use in lifetime studies. A generalized Kumaraswamy-Dagum model (KDM), based on the KDD, is defined and used to model a real-life data set. Using the method of maximum likelihood estimation, the estimates of the distribution and regression parameters are obtained. Other generalized statistical models, including Weibull and exponentiated Kumaraswamy-Dagum models, are fitted to the same real-life data and their fits compared with the KDM. Using Vuong's test for non-nested models and the likelihood ratio test for nested models, the other statistical models are compared with the KDM.

Talk T-6-3: Combining propensity score method with SMOTE for modeling imbalanced data

Yifan Hsu, Central Michigan University

Friday, October 14, 2022, 11:40 AM – 12:00 PM

Abstract T-6-3: Binary classification problems often encounter data where the two classes are highly unbalanced. Various techniques have been proposed to increase the prediction accuracy of modeling imbalanced problems. The Synthetic Minority Over Sampling TEchnique (SMOTE) is a popular hybrid technique with over-sampling and under-sampling to balance imbalanced data. SMOTE technique often results in high noise level of majority class. As a consequence, the SMOTE may not provide good improvement of prediction accuracy. A new technique is developed to reduce the noise level of the majority class by integrating the propensity score method with SMOTE. The new technique is applied to several data sets with different

degrees of imbalance. The results indicate this approach performs better than the existing SMOTE technique and it's variants for modeling imbalanced data.

Talk T-6-4: Multivariate count data regression models

Felix Famoye, Central Michigan University

Friday, October 14, 2022, 12:00 PM – 12:20 PM

Abstract T-6-4: Multivariate regression models based on multivariate discrete distributions are defined and studied. Multivariate discrete distributions, including some distributions generated from the T-R{W} method are defined. Models that allow both positive and negative correlation between any pair of response variables will be considered. The model parameters will be estimated by using the method of maximum likelihood. The application of these regression models will be illustrated by one or two numerical data sets.

Session T-7: Big Data: novel statistical modeling and computation for challenges in high-dimensional, spatiotemporal, and zero-inflated data

Friday, October 14, 2022, 11:00 AM to 12:20 PM, Pennsylvania Room

Chair: Hsin-Hsiung Huang

Talk T-7-1: A high dimensional Cramer-von Mises test

Mengyu Xu, University of Central Florida

Friday, October 14, 2022, 11:00 AM – 11:20 AM

Abstract T-7-1: A Cramer-von Mises type test is developed for testing distributions of high dimensional continuous data and establish an asymptotic theory for quadratic functions of high-dimensional stochastic processes. To obtain cutoff values of our tests, we introduce two different procedures to implement high-dimensional Cramér-von Mises test in practice: a plug-in calibration method and subsampling method. Theoretical justification and numerical studies of both approaches are provided.

Talk T-7-2: High-Dimensional Multivariate Time Series Forecasting for National-Level Geopolitical Events

Hayden Hampton & Jongjin Kim, University of Central Florida

Friday, October 14, 2022, 11:20 AM – 11:40 AM

Abstract T-7-2: Multivariate time series forecasting problems face significant challenges due to the large number of unknown parameters, nonstationary behavior, and complex interactions between predictors. This research focuses on the modeling approach taken by the team representing The University of Central Florida during the annual Pennsylvania State University's Applied Research Laboratory Algorithms for Threat Detection (ATD) 2022 Challenge. The goal of the challenge was to develop a time series forecaster capable of predicting national-level geopolitical event counts under constrained model training time. The GDELT project monitors print, broadcast, and web media to record world events and attribute them to organizations/state-actors. The data was extracted from the GDELT dataset using deep learning techniques. Due to the complex homologous relationship between language and culture, spatio-temporal clustering was performed to generate meaningful features from the dataset. Then we incorporated these features into a classical negative binomial (NB) regression model for multivariate time series prediction. Using backtesting to evaluate the performance of the NB model, we compared our model with other state-of-the-art techniques. We demonstrate the advantages to our approach, and through experimentation show why it is superior to other competing techniques.

Talk T-7-3: Response-Aided Score-Matching Approaches for Big Data Analysis

Keren Li, University of Alabama at Birmingham

Friday, October 14, 2022, 11:40 AM – 12:00 PM

Abstract T-7-3: Big data analysis has brought a lot of challenges to traditional statistical methods. We propose an efficient method called Response-Aided Score Matching Representative (RASMR) approach to facilitate the big data analysis under generalized linear models. This method utilizes representatives of natural or algorithmically generated data blocks to produce estimates of parameters, which approximate the estimate from full data set very well. In additional, a major advantage of

RASMR is that the representatives of the data set can be used for further analysis. The accurate estimation and high quality representatives enable promising application of RAMSR in a variety of statistical analysis problems, such as link function selection, variable selection and cross-validation in big data analysis.

Talk T-7-4: An R Package AZIAD for Analyzing Zero-Inflated and Zero-Altered Data

Niloufar Dousti Mousavi, University of Illinois at Chicago

Friday, October 14, 2022, 12:00 PM – 12:20 PM

Abstract T-7-4: Sparse data with a large portion of zeros arise in many scientific disciplines. Modeling sparse data is very challenging due to the skewness of the distribution. We adopt bootstrapped Monte Carlo method to estimate the p-value of the Kolmogorov-Smirnov test, as well as bootstrapped likelihood ratio tests for zero-inflated and zero-altered (or hurdle) model selection. Our new package AZIAD provides miscellaneous functions to simulate zero-inflated or zero-altered data and calculate maximum likelihood estimates of unknown parameters for a large class of discrete or continuous distributions. In addition, we calculate the Fisher information matrix and the confidence intervals of unknown parameters. Compared with other R packages available so far, our package covers many more types of zero-inflated and zero-altered distributions, provides more accurate estimates for unknown parameters, and achieves higher power for model selection. To facilitate the potential users, in this paper we provide theoretical justifications and detailed formulae for functions in AZIAD and illustrate the use of them with executable R code and real dataset.

Session T-8: Recent Developments in Inference and Modeling for High-Dimensional Data

Friday, October 14, 2022, 2:40 PM to 4:00 PM, Ohio Room

Chair: Wenbo Wu

Talk T-8-1: Conditional Sufficient Variable Screening for Ultrahigh Dimensional Data with FDR Control

Chenlu Ke, Virginia Commonwealth University

Friday, October 14, 2022, 2:40 PM – 3:00 PM

Abstract T-8-1: Variable screening has attracted extensive attention in the past decade, and conditional variable screening is a pertinent addition to the toolbox of analyzing ultrahigh dimensional data when prior information is available. In many applications, researchers know from previous investigations that certain variables are responsible for the outcomes or should be controlled for in the studies. This knowledge should be taken into account so that these control variables can assist in the selection of important predictors while being shielded in the screening procedure. However, the development of conditional variable screening methods that incorporate prior information has been less fruitful, compared to the vast literature for unconditional screening. In this paper, I propose a model-free conditional variable screening paradigm that allows high dimensional control variables and applies to either continuous or categorical responses. The contribution of each individual predictor is evaluated marginally and conditionally in the presence of other predictors and the control variables by reproducing-kernel-based R-squared and partial R-squared statistics. As a result, the proposed method enjoys the sure screening property and the rank consistency property in the notion of sufficiency, with which its superiority over existing methods is well-established. A subsequent conditional knockoff procedure is introduced to control the false discovery rate while preserving the power of retaining important predictors. The advantages of the proposed method are demonstrated by simulation studies encompassing a variety of regression and classification models as well as real data analysis.

Talk T-8-2: Conditional Martingale Difference Divergence

Chenglong Ye, University of Kentucky

Friday, October 14, 2022, 3:00 PM – 3:20 PM

Abstract T-8-2: We propose a general framework of independence measure with respect to a statistical functional of interest, which unifies some existing independent measures, such as distance covariance, Hilbert Schmidt independence criteria and martingale difference correlation. In addition, to address the potential issue of missing important variables that have no marginal utility with the response, we further propose a forward variable screening method based on a new conditional martingale difference correlation, which measures the conditional mean independence given a third variable. Under regularity conditions, it is able to select the variables that jointly but not marginally contribute to the mean or quantile of the response variable.

Talk T-8-3: Fréchet Sufficient Variable Selection with Graph Structure Among Predictors

Jiaying Weng, Bentley University

Friday, October 14, 2022, 3:20 PM – 3:40 PM

Abstract T-8-3: Fréchet regression has received considerable scholarly attention to encounter complex data that are non-Euclidean, such as images, shapes, or random densities. However, several unresolved questions remain about the development of Fréchet sufficient dimension reduction in ultra-high dimensions. We study Fréchet sufficient variable selection with graph structure among predictors and propose a penalized deference of trace loss, which avoids directly computing the inverse of a large covariance matrix. Our proposed estimation can be easily applied to high-dimensional predictors and utilizes the prior graph information of predictors to improve the accuracy and consistency. We demonstrate the superior finite-sample performance of our proposals over existing methods through comprehensive simulations and data analysis.

Talk T-8-4: On sufficient variable screening using log odds ratio filter

Wenbo Wu, The University of Texas at San Antonio

Friday, October 14, 2022, 3:40 PM – 4:00 PM

Abstract T-8-4: For ultrahigh-dimensional data, variable screening is an important step to reduce the scale of the problem, hence, to improve the estimation accuracy and efficiency. In this paper, we propose a new dependence measure which is called the log odds ratio statistic to be used under the sufficient variable screening framework. The sufficient variable screening approach ensures the sufficiency of the selected input features in modeling the regression function and is an enhancement of existing marginal screening methods. In addition, we propose an ensemble variable screening approach to combine the proposed fused log odds ratio filter with the fused Kolmogorov filter to achieve supreme performance by taking advantages of both filters. We establish the sure screening properties of the fused log odds ratio filter for both marginal variable screening and sufficient variable screening. Extensive simulations and a real data analysis are provided to demonstrate the usefulness of the proposed log odds ratio filter and the sufficient variable screening procedure.

Session T-9: Fiducial Inference

Friday, October 14, 2022, 2:40 PM to 4:00 PM, Kentucky Room

Chair: Kalimuthu Krishnamoorthy

Talk T-9-1: Generalized Fiducial Inference on Differentiable Manifolds

Jan Hannig, University of North Carolina at Chapel Hill

Friday, October 14, 2022, 2:40 PM – 3:00 PM

Abstract T-9-1: We consider the problem of defining a general fiducial density on an implicitly-defined differentiable manifold. Our proposed density extends the Generalized Fiducial Distribution (GFD) of Hannig et al. (2016). The resulting GFD formula is obtained by projecting the Jacobian differential in the ambient space onto the tangent space of the manifold. To circumvent the need for an intractable marginal integral calculation, we use two Monte Carlo algorithms that can efficiently explore a constrained parameter space and adapt them for use with the Constrained GFD. To demonstrate the new GFD formula we consider a number of simple examples. We also apply this methodology to the density estimation problem using splines and an estimation of Gaussian precision matrix with some known zero entries. Finally, we discuss how the manifold point of view could contribute to the philosophical understanding of fiducial distribution. (Joint work with A. Murph and J.P. Williams)

Talk T-9-2: Fiducial Inference for Hypergeometric Distributions: One- and Two-Sample Problems

Shanshan Lv, Truman State University

Friday, October 14, 2022, 3:00 PM – 3:20 PM

Abstract T-9-2: The problems of constructing confidence intervals (CIs) for the proportions and functions of proportions in finite populations are considered. For estimating the proportion in a finite population, we propose a CI based on the generalized fiducial method and compare it with an exact CI and score CI. We also address the construction of fiducial prediction intervals. For the two-sample problems, we consider interval estimating the difference between two proportions, the ratio of two proportions and the ratio of odds. Our solutions for the two-sample problems are based on the fiducial approach and the method of variance estimate recovery. All the CIs are evaluated on the basis of their exact coverage probabilities and expected widths. The methods are illustrated using some practical examples.

Talk T-9-3: Recent developments of accuracy metrics in biomarker evaluation and related fiducial inferences

Lili Tian, University at Buffalo

Friday, October 14, 2022, 3:20 PM – 3:40 PM

Abstract T-9-3: The development of biomarkers into diagnostic and prognostic tests can be categorized into three broad phases: discovery, performance evaluation, and impact determination when added to existing clinical measures. This talk covers some key concepts including classification types and classification metrics in performance evaluation, from a statistician's perspective. The limitations of existing classification metrics and the importance of using appropriate classification metrics are highlighted. Specifically, this talk presents some new efficient and appropriate biomarker performance measures to addresses the common pitfall caused by "naïve pooling" in biomarker evaluation, and the inefficiency and lack of flexibility in existing performance metrics in multiple ordered classification. Some recent developments in biomarker evaluation based

on fiducial inference for these newly proposed biomarker performance measures will be presented. An ovarian cancer data set from PLCO cancer study is analyzed.

Talk T-9-4: Fiducial Inference for Location-Scale Distributions

Kalimuthu Krishnamoorthy, University of Louisiana at Lafayette

Friday, October 14, 2022, 3:40 PM – 4:00 PM

Abstract T-9-4: We describe a general method of finding fiducial quantities for the parameters of location-scale (or log-location-scale) distributions based on a complete sample or type II censored sample. We then consider the inference on the parameters and function of parameters and evaluate them for their accuracy in terms of coverage probabilities and error rates. In particular, we show the applications of fiducial methods to find confidence intervals for the ratio of means or for the ratio of percentiles of two independent location-scale distributions. We compare the fiducial methods with other existing methods and illustrate them using some practical examples.

Session T-10: Statistical inference based upon characterizing statistical evidence

Friday, October 14, 2022, 2:40 PM to 4:00 PM, Pennsylvania Room

Chair: Laui Al Labadi / Michael Evans

Talk T-10-1: Filter Method for Variable Selection Based on Relative Belief Ratio

Ayman Alzaatreh, American University of Sharjah, Sharjah, UAE

Friday, October 14, 2022, 2:40 PM – 3:00 PM

Abstract T-10-1: Variable selection has become a critical step in most data mining applications to mitigate the curse of dimensionality in high-dimensional datasets. Without direct input from the target variable, filter methods evaluate the importance of features as a pre-processing operation to the learning algorithm and select the best feature subsets through some information metrics. Filters are known to be more computationally efficient than wrapper and embedded methods. In this talk, the relative belief ratio will be used to discriminate between two groups in a binary classification problem setting. The relative belief ratio is used as a filter method to rank features based on their importance in relation to a binary target variable. Several benchmark data sets are used to demonstrate the applicability of the proposed method.

Talk T-10-2: On Robustness of the Relative Belief Ratio and the Strength of its Evidence with Respect to the Geometric Contamination Prior

Luai Al Labadi, University of Toronto

Friday, October 14, 2022, 3:00 PM – 3:20 PM

Abstract T-10-2: The relative belief ratio becomes a widespread tool in many hypothesis testing problems. It measures the statistical evidence that a given statement is true based on a combination of data, model and prior. Additionally, a measure of the strength is used to calibrate its value. In this talk, robustness of the relative belief ratio and its strength to the choice of the prior is studied. Specifically, the Gateaux derivative is used to measure their sensitivity when the geometric contaminated prior is used. Examples are presented to illustrate the results.

Talk T-10-3: Combining Evidence

Michael Evans, University of Toronto

Friday, October 14, 2022, 3:20 PM – 3:40 PM

Abstract T-10-3: The problem of combining the evidence concerning an unknown, contained in each of k Bayesian inference bases, is discussed. This can be considered as a generalization of the problem of pooling k priors to determine a consensus prior. The linear opinion pool of Stone (1961) is seen to have the most appropriate properties for this role. In particular, linear pooling preserves a consensus with respect to the evidence and other rules do not. While linear pooling does not preserve prior independence, it is shown that it still behaves appropriately with respect to the expression of evidence in such a context. For the general problem of combining evidence, Jeffrey conditionalization plays a key role.

Session T-11: Modeling and Application

Friday, October 14, 2022, 4:10 PM to 5:30 PM, Ohio Room

Chair: Raid Al-Aqtash

Talk T-11-1: Effect of News Sentiment on Stock Price: A Deep Neural Network and Statistical Analysis

Keshab Dahal, Truman State University

Friday, October 14, 2022, 4:10 PM – 4:30 PM

Abstract T-11-1: Accurately predicting the price of stock markets has broad implications for investors as it helps them make informed decision when investing. While this is routinely done using the state-of-art machine learning (ML) based models, these models often do not take into account the sentiment of general public investing in it. Moreover, literature lacks studies comparing and contrasting the performance of advanced ML models for the same stock market data. In this work, we combined sentiment analysis with two deep learning approaches namely long short term memory (LSTM) and gated recurrent unit (GRU) to accurately model the behavior of stock market. Our models that combines financial news sentiments with fundamental data of stock market significantly improves upon the accuracy when compared with existing models without sentiment. The analysis is supported by the performance metrics such as root mean square error (RMSE) which decrease by 23% in LSTM and 17% in GRU when financial news sentiments are incorporated.

Talk T-11-2: The Generalized Error-Generalized Hyperbolic Secant distribution: Properties and Applications

Hazem Al-Mofleh, Tafila Technical University, Jordan

Friday, October 14, 2022, 4:30 PM – 4:50 PM

Abstract T-11-2: In this paper, a new five-parameter univariate continuous distribution called the Generalized Error-Generalized hyperbolic secant distribution (GEGHS) is defined and studied. Some general and structural distributional properties are investigated and discussed, including: central and non-central n-th moments and incomplete moments, quantile and generating functions, hazard function, Renyi and Shannon entropies, shapes: skewed right, skewed left, and symmetric. Furthermore, a special case distribution (when the parameter k=2) is defined and studied, and some mathematical probertites are investigated and discussed, including: the modality regions: unimodal and bimodal, and the maximum likelihood (MLE) estimators for the parameters. Finally, two real data sets are used to demonstrate its empirically and flexibility, and prove the strength of the new distribution.

Talk T-11-3: Ideas Toward Analyzing Nonstationary Gaussian Process Models

Ahmad Hanandeh, Yarmouk University

Friday, October 14, 2022, 4:50 PM – 5:10 PM

Abstract T-11-3: With increased use of large and complex computer models and growing interest in the field of uncertainty quantification, Gaussian process models have become a popular statistical approach for emulating expensive computer experiments, but face tremendous computational challenges for large multi-output computer models. The purpose of this article is to develop a methodology that reduces the computational complexity by combining/generalizing the merits of several popular approaches. The resulting models are called surrogate models, which we believe that they will provide

substantial computational savings, and results compare well with those from classical inferential ways to analyze Gaussian Process as well as Nonstationary Gaussian Process Models. Simulation will be used to demonstrate the performance.

Talk T-11-4: Receiver operating characteristic (ROC) analysis of combining CA125 and circulating lipid metabolites to enhance the accuracy of diagnosis and triage ovarian cancer versus benign adnexal mass

Li Yan, Roswell Park Comprehensive Cancer Center

Friday, October 14, 2022, 5:10 PM – 5:30 PM

Abstract T-11-4: Novel lipid metabolite markers have emerged as an important feature of ovarian cancer (OC), and its translational potential to aid in diagnosis and triage is under investigation. We conducted a multi-level interrogation integrating quantitative lipidomics profiling of plasma and ascites with tumor transcriptome data. We assessed 500+ plasma lipids in two cohorts of 218 women total adnexal mass. Certain lipids exhibited greater alterations in early- or late-stage cases. Lipoprotein receptor gene expression differed markedly in OC versus benign tumors. Importantly, several plasma lipid species improved the accuracy of CA125 in differentiating early-stage OC cases from benign controls, with a 15-20% increase in specificity at 90% sensitivity in multivariate models adjusted for age and BMI. This study provides novel insight into systemic and local lipid metabolic differences between OC and benign disease, and advancing plasma lipid metabolites as a complementary class of circulating biomarkers for OC diagnosis and triage.

Session T-12: Statistical Distribution Theory and Applications

Friday, October 14, 2022, 4:10 PM to 5:30 PM, Kentucky Room

Chair: Indranil Ghosh/Tony Ng

Talk T-12-1: On characterization of the exponential distribution via hypoexponential distributions

George Yanev, The University of Texas Rio Grande Valley & Institute ofMathematics and Informatics, Bulgarian Academy of Sciences

Friday, October 14, 2022, 4:10 PM – 4:30 PM

Abstract T-12-1: The sum of independent but not necessary identically distributed exponential random variables follows hypoexponential distribution. We will discuss a situation when the rate parameters of the exponential variables are not all different from each other. We obtain a representation for the Laplace transform of the hypoexponential distribution in case of two repeated parameter values. Making use of this decomposition, we prove a characterization of the exponential distribution.

Talk T-12-2: Truncated Family of T-X family of Distributions

Ayman Alzaatreh, American University of Sharjah, Sharjah, UAE

Friday, October 14, 2022, 4:30 PM – 4:50 PM

Abstract T-12-2: In this talk, we introduce right-truncated and left-truncated T-X families of distributions. These families are used to construct new generalized families of continuous distributions. Relationships between the families are investigated. Real data sets including time and cost to start a business are analyzed and the results show that the truncated families perform very well for fitting highly skewed data.

Talk T-12-3: Testing Means of Right-Skewed Populations - a Saddlepoint Approximation.

Kalanka Jayalath, University of Houston - Clear Lake, Houston, TX

Friday, October 14, 2022, 4:50 PM – 5:10 PM

Abstract T-12-3: Testing means of skewed distributions commonly rely on the central limit theorem and data transformation. Johnson's and Chen's modified t-tests that incorporate few additional sample moments may be considered powerful and robust alternatives for this purpose. This work focuses on the truncated saddlepoint (TS) approximation method that can be used to estimate the right-tail probability of the right-skewed continuous distributions using the first few sample moments. The approach is applied to obtain the p-value in one-sample mean testing. A Monte-Carlo simulation study is conducted to compare the performance of the TS approximation at four different truncation levels with the t-test and its robust alternatives. It indicates that the fifth-order truncation and Chen's method perform better for highly skewed distributions especially when the sample size is lower. A numerical example is given to exhibit the use of TS approximation and its competitors in practice.

Talk T-12-4: Tail Conditional Expectations Based on Kumaraswamy Dispersion Models

Indranil Ghosh, University of North Carolina Wilmington

Friday, October 14, 2022, 5:10 PM – 5:30 PM

Abstract T-12-4: Recently, there seems to be an increasing amount of interest in the use of the tail conditional expectation (TCE) as a useful measure of risk associated with a production process, for example, in the measurement of risk associated with stock returns corresponding to the manufacturing industry, such as the production of electric bulbs, investment in housing development, and financial institutions offering loans to small-scale industries. Companies typically face three types of risk (and associated losses from each of these sources): strategic (S); operational (O); and financial (F) (insurance companies additionally face insurance risks) and they come from multiple sources. For asymmetric and bounded losses (properly adjusted as necessary) that are continuous in nature, we conjecture that risk assessment measures via univariate/bivariate Kumaraswamy distribution will be efficient in the sense that the resulting TCE based on bivariate Kumaraswamy type copulas do not depend on the marginals. In fact, almost all classical measures of tail dependence are such, but they investigate the amount of tail dependence along the main diagonal of copulas, which has often little in common with the concentration of extremes in the copula's domain of definition. In this article, we examined the above risk measure in the case of a univariate and bivariate Kumaraswamy (KW) portfolio risk, and computed TCE based on bivariate KW type copulas. For illustrative purposes, a well-known Stock indices data set was re-analyzed by computing TCE for the bivariate KW type copulas to determine which pairs produce minimum risk in a two-component risk scenario.

Session T-13: Statistical Methods for Count Data Analysis

Friday, October 14, 2022, 4:10 PM to 5:30 PM, Pennsylvania Room

Chair: Norou Diawara

Talk T-13-1: Doubly Inflated Poisson GARCH Model

Sumen Sen, Austin Peay State University

Friday, October 14, 2022, 4:10 PM – 4:30 PM

Abstract T-13-1: A data is said to be doubly inflated when two of the data points appears disproportionately with respect to other data points. This type of double inflation is generally seen in count data, and it is important to build models acknowledging these inflation points. In this paper we proposed a doubly inflated integer GARCH (DIP-INARCH) model that accommodates the double inflation characteristic of time series data. The stationarity conditions and auto correlation functions are given. Parameter estimation was done using EM algorithm and maximum likelihood estimation process.

Talk T-13-2: Finite Mixtures of Mean-Parameterized Conway-Maxwell-Poisson Models

Dongying Zhan, University of Kentucky

Friday, October 14, 2022, 4:30 PM – 4:50 PM

Abstract T-13-2: The Conway-Maxwell-Poisson (CMP) distribution generalizes the Poisson distribution by adding the dispersion parameter to model over-dispersion and under-dispersion. The mean-parameterized CMP distribution provides and allows for simpler and more easily interpretable mean models. In this study, we propose a finite mixture of mean-parameterized CMP distributions. An EM algorithm is constructed and used to estimate the parameters in the mixture of CMP distributions. The model and estimation method are demonstrated by Monte Carlo simulations and applied on real data.

Talk T-13-3: On EM Estimation for Zero- and k-Inflated Poisson Regression Model

Rao Chaganty, Old Dominion University

Friday, October 14, 2022, 4:50 PM – 5:10 PM

Abstract T-13-3: Count data with excessive zeros are ubiquitous in healthcare, medical, and scientific studies. There are numerous articles that show how to fit Poisson and other models which account for the excessive zeros. However, in many situations, besides zero, the frequency of another count k tends to be higher in the data. The zero- and k-inflated Poisson distribution model (ZkIP) is appropriate in such situations. The ZkIP distribution essentially is a mixture distribution of Poisson and degenerate distributions at zero and k. In this talk, we will discuss the fundamental properties of this mixture distribution. Using stochastic representation, we provide details for obtaining parameter estimates of the ZkIP regression model using the Expectation–Maximization (EM) algorithm for a given data. More importantly, we derive the standard errors of the EM estimates by computing the complete, missing, and observed data information matrices. We present the analysis of two real-life data using the methods described in the talk.

Talk T-13-4: Copula-Based Bivariate Zero-Inflated Poisson Time Series Models

Dimuthu Fernando, Old Dominion University

Friday, October 14, 2022, 5:10 PM – 5:30 PM

Abstract T-13-4: Count time series data are found in multiple applications such as environmental science, biostatistics, economics, public health, and finance. Such time series counts come with inflation and in a bivariate form that captures not only serial dependence within each time series but also interdependence between the two series. To accurately study such data, one needs to account for the two types of dependence that emerge from the observed data, and the inflation. A class of bivariate integer-valued time series models is constructed via copula theory. Each time series follows Markov chains with the serial dependence is captured using copula-based transition probabilities with the Poisson and the zero-inflated Poisson (ZIP) margins. The copula theory is also used again to capture the dependence between the two series using either the bivariate Gaussian or t copula functions. Likelihood based inference is used to estimate the model parameters with the bivariate integrals of the Gaussian or t copula functions being evaluated using standard randomized Monte Carlo methods. To evaluate the proposed class of models, a comprehensive simulated study is conducted capturing both positive and negative cross correlations. Then, two real life datasets are analyzed assuming the Poisson and the ZIP marginals, respectively. The results show the promises of the proposed class of models.

Session T-14: Non-Gaussian Stochastic Models

Saturday, October 15, 2022, 9:10 AM to 10:30 AM, Ohio Room

Chair: Tomasz Kozubowski

Talk T-14-1: Simulating entering and exiting balls by an isometric stable process

John Nolan, American University

Saturday, October 15, 2022, 9:10 AM – 9:30 AM

Abstract T-14-1: Methods for simulating the entry and exit from a ball by an isotropic α-stable process are discussed. Simple rejection is possible, but can be very inefficient when starting near the boundary of a ball. Efficient methods based careful partition of the region are given.

Talk T-14-2: Normal Pareto Distributions: Theoretical framework and computational issues

Tomasz J. Kozubowski, University of Nevada

Saturday, October 15, 2022, 9:30 AM – 9:50 AM

Abstract T-14-2: We introduce a new, conditionally Gaussian, hierarchical stochastic model for heavy tailed data, which generalizes the Laplace probability distribution. We present basic properties of this model and discuss related computational issues. We also briefly consider inferential aspects of the model. This is joint work with M. Ohemeng.

Session T-15: Statistical advances in high dimensional data with applications in Biostatistics and Genomics

Saturday, October 15, 2022, 9:10 AM to 10:30 AM, Kentucky Room

Chair: Souparno Ghosh

Talk T-15-1: Mediation Analysis of high dimensional exposures with applications in genetics

Qi Zhang, University of New Hampshire

Saturday, October 15, 2022, 9:10 AM – 9:30 AM

Abstract T-15-1: We have developed a new method for estimating the mediation effect of each exposure when the exposures are high dimensional. Simulations and real genetic data analysis shows that the proposed method yields reasonable power, type I error control, and coverage probability for confidence intervals.

Talk T-15-2: Tumor Radiogenomics with Bayesian Layered Variable Selection

Shariq Mohammed, Boston University School of Public Health

Saturday, October 15, 2022, 9:30 AM – 9:50 AM

Abstract T-15-2: We propose a statistical framework to analyze radiological magnetic resonance imaging (MRI) and genomic data to identify the underlying radiogenomic associations in lower grade gliomas (LGG). We devise a novel imaging phenotype by dividing the tumor region into concentric spherical layers that mimics the tumor evolution process. MRI data within each layer is represented by voxel–intensity-based probability density functions which capture the complete information about tumor heterogeneity. Under a Riemannian-geometric framework these densities are mapped to a vector of principal component scores which act as imaging phenotypes. Subsequently, we build Bayesian variable selection models for each layer with the imaging phenotypes as the response and the genomic markers as predictors. Our novel hierarchical prior formulation incorporates the interior-to-exterior structure of the layers, and the correlation between the genomic markers. We employ a computationally-efficient Expectation–Maximization-based strategy for estimation. With a focus on the cancer driver genes in LGG, we discuss some biologically relevant findings.

Talk T-15-3: Distance-averaged CNN ensemble learning for anti-cancer drug sensitivity prediction

Souparno Ghosh, University of Nebraska-Lincoln

Saturday, October 15, 2022, 9:50 AM – 10:10 AM

Abstract T-15-3: Anti-cancer drug sensitivity prediction using deep learning models for individual cell line is a significant challenge in personalized medicine. Recently developed REFINED (REpresentation of Features as Images with NEighborhood Dependencies) -Convolutional Neural Network- based models have shown promising results in improving drug sensitivity prediction. The primary idea behind REFINED-CNN is representing high dimensional vectors as compact images with spatial correlations that can benefit from CNN architectures. However, the mapping from a high dimensional vector to a compact 2D image depends on the a priori choice of the distance metric and projection scheme with limited empirical procedures guiding these choices. We consider an ensemble of REFINED-CNN built under different choices of distance

metrics and/or projection schemes that can improve upon a single projection based REFINED-CNN model. We also develop the theoretical framework for combining different distance metrics to arrive at a single 2D mapping.

Talk T-15-4: Estimation of practice effects in longitudinal cohorts using Bayesian hierarchical models

Rajesh Talluri, The University of Mississippi Medical Center

Saturday, October 15, 2022, 10:10 AM – 10:30 AM

Abstract T-15-4: In dementia research, cognitive function and cognitive decline measures are necessary to evaluate novel treatments to improve cognitive function. It is therefore of paramount importance that cognitive measures reflect an individual's true cognitive ability. As participants are exposed to the same cognitive test repeatedly, an individual may learn and adjust to the test(practice effect). Ignoring practice effects may lead to underestimation of cognitive decline and overestimation of baseline cognition function of an individual which may bias the results of the study. We propose a statistical framework to model the effect of practice effects on the cognitive measurement in a longitudinal cohort. The dependence of practice effects on time dependent covariates and the time elapsed between the tests was modeled. Simulations show that the proposed model is identifiable, and the model parameter estimates are unbiased. Finally, data from the National Alzheimer's Coordinating Center cohort was analyzed using the proposed model to estimate the practice effects. The effect of practice effects on downstream analyses was evaluated using further simulations

Session T-16: Statistical Inference and Estimation

Saturday, October 15, 2022, 9:10 AM to 10:30 AM, Pennsylvania Room

Chair: Mahmoud Aldeni

Talk T-16-1: Adaptive Estimation for Non-parametric Multivariate Additive Model in Random Design with Long Memory errors

Rida Benhaddou, Ohio University

Saturday, October 15, 2022, 9:10 AM – 9:30 AM

Abstract T-16-1: We investigate the non-parametric bivariate additive regression estimation in the random design and long-memory errors and construct adaptive thresholding estimators based on wavelet series. The proposed approach achieves asymptotically near-optimal convergence rates when the unknown function and its univariate additive components belong to Besov space. We consider the problem under two noise structures; (1) homoskedastic Gaussian long memory errors and (2) heteroskedastic Gaussian long memory errors. In the homoskedastic long-memory error case, the estimator is completely adaptive with respect to the long-memory parameter. In the heteroskedastic long-memory case, the estimator may not be adaptive with respect to the long-memory parameter unless the heteroskedasticity is of polynomial form. In either case, the convergence rates depend on the long-memory parameter only when long-memory is strong enough, otherwise, the rates are identical to those under i.i.d. errors. A simulations study reveals the finite sample properties of the estimator under several heteroskedasticity scenarios.

Talk T-16-2: On the Family of Generalized Gumbel Distributions

Raid Al-Aqtash, Marshall University

Saturday, October 15, 2022, 9:30 AM – 9:50 AM

Abstract T-16-2: In this research, we investigate a family member of the T-R{Y} frame, using a generalized Gumbel distribution. Three members of this family are utilized and their statistical properties are investigated. Real data are used to illustrate the usefulness of using these distributions in application.

Talk T-16-3: Pretest and Shrinkage Estimators for Log-normal Means: Theory and Simulation

Mahmoud Aldeni, Western Carolina University

Saturday, October 15, 2022, 9:50 AM – 10:10 AM

Abstract T-16-3: We consider the problem of pooling means from multiple random samples from log-normal populations. Under the homogeneity assumption of means that all mean values are equal, we propose improved large sample asymptotic methods for estimating p log-normal population means when multiple samples are available, and suggest estimators based on linear shrinkage, pretest, and Stein-type methodology. In this talk, we present some useful asymptotic results of the proposed estimators accompanied by expressions of the asymptotic distributional bias and quadratic risk. We conduct a simulation study to examine the asymptotic results and efficiencies of the estimators in finite samples.

Talk T-16-4: Pretest and Shrinkage Estimators for Log-normal Means: Applications

John Wagaman, Western Carolina University

Saturday, October 15, 2022, 10:10 AM – 10:30 AM

Abstract T-16-4: We consider the problem of pooling means from multiple random samples from log-normal populations. Under the homogeneity assumption of means that all mean values are equal, we propose improved large sample asymptotic methods for estimating p log-normal population means when multiple samples are available, and suggest estimators based on linear shrinkage, pretest, and Stein-type methodology.

This talk focuses on two applications using historical data from the Dow Jones Industrial Average for 10 years from 2010-2019 and monthly precipitation data from four locations in Washington State. We compare the performance of the estimators for each of these data sets which provide varied conclusions with respect to the homogeneity assumption of means.

This is joint work with Drs. Mahmoud Aldeni, Mohamed Amezziane and S. Ejaz Ahmed.

Session T-17: Recent developments in finite mixture modeling with applications

Saturday, October 15, 2022, 10:40 AM to 12:00 PM, Ohio Room

Chair: Volodymyr Melnykov

Talk T-17-1: Model-based clustering analysis of the spatio-temporal and intensity patterns of tornadoes

Rong Zheng, Western Illinois University

Saturday, October 15, 2022, 10:40 AM – 11:00 AM

Abstract T-17-1: Tornadoes are one of nature's most violent windstorms that can occur all over the entire world except Antarctica. Previous scientific efforts were spent on studying this nature hazard from facets such as: genesis, dynamics, detection, forecasting, warning, measuring, and assessing. We model the tornado datasets by using modern statistical and computational techniques. The objective of the project is to develop novel models and perform clustering analysis of the spatial-temporal and intensity patterns in tornadoes. A summary of insights about tornado nature that can be useful in risk forecasting and assessing is provided.

Talk T-17-2: Conditional mixture modeling and model-based clustering

Wang Yang, College of Charleston

Saturday, October 15, 2022, 11:00 AM – 11:20 AM

Abstract T-17-2: Due to a potentially high number of parameters, finite mixture models are often at the risk of overparameterization even for a moderate number of components. This can lead to over-fitting individual components and result in mixture order underestimation. One of the most popular approaches to address this issue is to reduce the number of parameters by considering parsimonious models. The vast majority of techniques in this direction focus on the reparameterization of covariance matrices associated with mixture components. We propose an alternative approach that emphasizes modeling cluster locations. The developed procedure enjoys remarkable modeling flexibility, especially noticeable in the presence of non-compact clusters. Due to an attractive closed-form formulation, speedy parameter estimation is available by means of the EM algorithm. The utility of the proposed method is illustrated on synthetic and well-known classification data sets.

Talk T-17-3: On model-based clustering of directional data with heavy tails and outliers

Yingying Zhang, Western Michigan University

Saturday, October 15, 2022, 11:20 AM – 11:40 AM

Abstract T-17-3: Directional statistics deals with data that can be naturally expressed in the form of vector directions. Von Mises-Fisher distribution is one of the most fundamental parametric models to describe directional data. Mixtures of Von Mises-Fisher distributions represent a popular approach to handling heterogeneous populations. However, such models can be affected by the presence of noise, outliers, and heavy tails. To relax these model limitations, a mixture of contaminated Von Mises-Fisher distributions is proposed. The performance of the proposed methodology is tested on synthetic data and applied to the data containing abstracts from the Joint Statistical Meetings held in Denver in 2008. The obtained results demonstrate the importance of the proposed procedure and its superiority over the traditional mixture of Von Mises-Fisher distributions in the cases of heavy tails or scatter.

Talk T-17-4: Studying contributions of variables to classification

Yana Melnykov, The University of Alabama

Saturday, October 15, 2022, 11:40 AM – 12:00 PM

Abstract T-17-4: A problem of finding variables responsible for classifying a particular observation as well as detecting those that interfered with the assignment made is considered. We address this problem by providing a formal argument supported by intuitive geometric interpretation. The proposed methodology is supported by simulation studies as well as applications to well-known classification datasets. A novel graphical display is developed for visualizing obtained results.

Session T-18: Survival Analysis in Theory and in Practice: The Analysis of Time-To-Event data

Saturday, October 15, 2022, 10:40 AM to 12:00 PM, Kentucky Room

Chair: Drew Lazar

Talk T-18-1: Subgroup Identification with Differential Effect of Cardio-respiratory Fitness for All-cause and Disease-specific Mortality: Survival Regression-Based Recursive Partitioning Approach

Md Yasin Ali Parh, University of Louisville

Saturday, October 15, 2022, 10:40 AM – 11:00 AM

Abstract T-18-1: In cardio-respiratory fitness (CRF) studies, existing research is primarily on estimating the overall average effect of CRF on the risk of death. However, the effect of CRF on mortality may vary based on an individual's characteristics. A model-based recursive partitioning approach is applied to data from the BALL ST Adult Fitness Longitudinal Lifestyle Study, with CRF treated as a treatment variable. Age is selected as the partitioning variable in each case to identify the subgroup of individuals. For all-cause, CVD, and cancer specific mortality, CRF has a vital impact on the survival experiences. Our study suggests that the survival experiences of middle-aged individuals with CVD will boost if their CRF increases. High CRF also plays an important role in increasing the survival experience for cancer patients who are older than 49 years.

Talk T-18-2: A New Robust Approach for Regression Analysis of Panel Count Data with Time-varying Covariates

Dayu Sun, Rollins School of Public Health at Emory University

Saturday, October 15, 2022, 11:00 AM – 11:20 AM

Abstract T-18-2: We consider the robust inference for panel count data with time-varying covariates, which often occur in real-life applications such as medical studies and reliability experiments. Several mean model-based methods that are robust to within-subject correlation structures have been proposed and widely used. However, the robust mean model-based approach usually requires the mean function to be monotone, which may not be realistic when covariate values fluctuate over time. To address these issues, we propose a robust rate model-based procedure with rigorous theoretical justification. Under the rate model, another challenge is to develop variance estimators because the asymptotic variance usually has no closed forms. To our knowledge, the existing literature with similar problems all resorted to computationally intensive numerical methods that may not be robust. We develop novel computationally efficient robust variance estimators with closed forms based on Expectation-Maximization (EM) algorithm. We rigorously show that the variance estimators are consistent regardless of the underlying distribution assumption. An extensive simulation study is performed to demonstrate the superiority of the proposed approach and a real-life study of sexually transmitted infections is used to demonstrate the applicability of this newly proposed approach.

Talk T-18-3: Inference on Mean Quality-adjusted Lifetime Using Joint Models for Continuous Quality of Life Process and Time to Event.

Xiaotain Gao, University of Illinois at Urbana-Champaign

Saturday, October 15, 2022, 11:20 AM – 11:40 AM

Abstract T-18-3: Quality-adjusted lifetime (QAL) has been considered as an objective measurement that summarizes the quantitative and qualitative health aspects in a unitary and meaningful way. The idea is to account for each individual's health experience in adjusting the overall survival, with death at one extreme and perfect health at the other extreme. In existing literature, the health states are defined to be discrete and the number of states is taken to be finite. In this paper we propose an estimator of the mean QAL when the quality-of-life process is assumed to be continuous and observed with error over time at fixed time points via joint modeling of quality of life and time to event outcomes. We prove the asymptotic properties of the proposed estimator and study its finite sample performance through simulation. We illustrate its application with data from a hepatitis C clinical trial study.

Talk T-18-4: Neural Network Models of Survival Analysis Data

Drew Lazar, Ball State University

Saturday, October 15, 2022, 11:40 AM – 12:00 PM

Abstract T-18-4: In this talk, we consider neural network models of survival analysis data. Two multi-layer feed forward networks which assume proportional hazards, the Faraggi-Simon network from 1995 and its update as the DeepSurv neural network model from 2018, are explained and considered. Borrowing from these models the idea of predicting risk at a single final node and using the likelihood to measure loss, we construct Weibull and Log-Logistic neural network survival models. In this way, right censoring is accounted for through the loss function. These models are extended to account for frailty and recurrent events and are applied to real and simulated data sets. Comparisons are made with traditional semi-parametric and parametric survival models on simulated and real data sets in Pytorch. The flexibility of the neural network models helps them account for higher-order effects and interaction without specifying or testing for these effects in advance of model fitting.

Abstracts for
General Invited Sessions

General Invited Session G-1

Friday, October 14, 2022, 9:30 AM to 10:50 AM, Virginia Room

Chair: Alaa Elkadry

Talk G-1-1: Bayesian Analysis for Imbalanced Positive-Unlabelled Diagnosis Codes in Electronic Health Records

Ye Liang, Oklahoma State University

Friday, October 14, 2022, 9:30 AM – 9:50 AM

Abstract G-1-1: With the increasing availability of electronic health records (EHR), significant progress has been made on developing predictive inference and algorithms by health data analysts and researchers. However, the EHR data are notoriously noisy due to missing and inaccurate inputs despite the information is abundant. One serious problem is that only a small portion of patients in the database has confirmatory diagnoses while many other patients remain undiagnosed because they did not comply with the recommended examinations. The phenomenon leads to a so-called positive-unlabelled situation and the labels are extremely imbalanced. In this paper, we propose a model-based approach to classify the unlabelled patients by using a Bayesian finite mixture model. We also discuss the label switching issue for the imbalanced data and propose a consensus Monte Carlo approach to address the imbalance issue and improve computational efficiency simultaneously. Simulation studies show that our proposed model-based approach outperforms existing positive-unlabelled learning algorithms. The proposed method is applied on the Cerner EHR for detecting diabetic retinopathy (DR) patients using laboratory measurements. With only 3% confirmatory diagnoses in the EHR database, we estimate the actual DR prevalence to be 25% which coincides with reported findings in the medical literature.

Talk G-1-2: Jaynes-Gibbs Entropic Convex Duals & Orthogonal Polynomials

Richard Le Blanc, Université de Sherbrooke

Friday, October 14, 2022, 9:50 AM – 10:10 AM

Abstract G-1-2: The univariate noncentral distributions can be derived by multiplying their central distributions with translation factors. When constructed in terms of translated uniform hyperspherical distributions, these translation factors become generating functions for classical families of orthogonal polynomials. The ultraspherical noncentral t, normal, F and chi square distributions are thus found to be associated with the Gegenbauer, Hermite, Jacobi and Laguerre polynomial families, respectively, with the corresponding central distributions standing for the polynomial family-defining weights. Obtained through unconstrained minimization of the Gibbs potential, Jaynes maximal entropy priors are formally expressed in terms of empirical densities' entropic convex duals. Expanding these duals on orthogonal polynomial bases allows for expedient determination of the Jaynes-Gibbs priors. Invoking the moment problem and the duality principle, modelization can be reduced to direct determination of prior moments in parametric space in terms of Bayes factor orthogonal polynomial expansion coefficients in random variable space. Genomics and geophysics examples are provided.

Talk G-1-3: Analysis of incomplete longitudinal data with informative drop-outs and outliers

Sanjoy Sinha, Carleton University

Friday, October 14, 2022, 10:10 AM – 10:30 AM

Abstract G-1-3: In this talk, I will present a robust method for analyzing longitudinal data with informative drop-outs. The robust method, developed in the framework of the maximum likelihood, is used to bound the influence of potential outliers

in the data when estimating the model parameters. The asymptotic properties of the robust estimators will be discussed. The empirical properties will be studied using results from a simulation study. An application will be presented using longitudinal measurements on a biomarker of sepsis being studied for understanding the progression of pneumonia to severe sepsis in a group of patients.

Talk G-1-4: Bayesian Estimators of the Inverse Weibull-Weibull Composite distribution

Nirajan Budhathoki, Central Michigan University

Friday, October 14, 2022, 10:30 AM – 10:50 AM

Abstract G-1-4: In this study, Bayesian estimates with both Jeffreys and Gamma priors are compared with the least squares estimate for the parameters of Inverse Weibull-Weibull Composite (IWWC) distribution. The IWWC family of distribution is a two-parameter smooth and continuous natural composition that has an Inverse Weibull density up to a threshold and Weibull density for the remainder. The random walk Metropolis-Hastings algorithm is used to generate random samples from the posterior distributions of the parameters for inference. The goodness of fit test including Kolmogorov-Smirnov and Anderson-Darling, and the precision of estimates are considered for comparison on estimates. The methods are illustrated by a repair-times dataset and simulated datasets.

General Invited Session G-2

Friday, October 14, 2022, 11:00 AM to 12:20 PM, Virginia Room

Chair: Olusegun Otunuga

Talk G-2-1: Goodness-of-Fit Tests for Network Structures

Ping-Shou Zhong, University of Illinois at Chicago

Friday, October 14, 2022, 11:00 AM – 11:20 AM

Abstract G-2-1: Network structures are commonly used for describing the conditional dependence between random variables. The dependence information provided by network structures is important for improving statistical estimation precision. In practice, the underlying network structure of a given data set is often unknown and need to be estimated or assumed as a priory. However, whether a pre-specified network structure or an estimated network structure is consistent with the data? We tackle this problem by proposing a goodness-of-fit test for a pre-specified network structure in the high-dimensional setting with the dimension of nodes more than the number of sample size. We develop a test statistic for assessing the goodness-of-fit of a pre-specified network structure to the observed data. The asymptotic distribution is derived under mild assumptions on the data dimension and spatial dependence. Under the null hypothesis, it is shown that asymptotic distribution of the proposed test statistic to follow a Gumbel distribution. It is interesting to find that the location parameter of the limiting Gumbel distribution depends on the network structure specified under the null hypothesis. Extensive simulation studies demonstrate that the proposed test procedure is able to control the type I error under the nominal level for various settings, and it has the ability to detect the network structures that are incorrectly specified. Finally, we apply the proposed test statistic to analyze a U.S. COVID-19 data set.

Talk G-2-2: Dimension reduction for random objects: a projective resampling approach with application to Covid-19 transmission in the United States

Abdul-Nasah Soale, University of Notre Dame

Friday, October 14, 2022, 11:20 AM – 11:40 AM

Abstract G-2-2: Advanced technology led to the collection of novel data objects which may not lie in the Euclidean space. Utilizing the conditional Fréchet regression model of Petersen & Müller (2019), we propose a sufficient dimension reduction (SDR) method for metric space-valued responses. We first map the complex random response object to a real-valued distance matrix. Then, we project the distances along random directions to obtain random univariate responses. We estimate the dimension reduction subspace for each projected response using existing Euclidean-based SDR techniques, which we aggregate to estimate the Fréchet central space. Our method avoids kernels and relies on fewer tuning parameters while preserving the joint distribution of the response object. We apply our proposal to analyze the distribution of Covid-19 cases in the United States based on socio-economic and demographic characteristics. We also provide theoretical justifications and finite sample performance of our method for response objects in the Wasserstein space.

Talk G-2-3: Generalizations with T-X Method and New Transformer with Gompertz Illustration

Scott Smith, University of the Incarnate Word

Friday, October 14, 2022, 11:40 AM – 12:00 PM

Abstract G-2-3: Since the seminal paper of Alzaatreh et al (2013), many new distribution families have been explored using the T-X framework. However, many of these subsequent papers have focused on T-X constructions using the odds or logit links. In this presentation, we explore the Weibull-X construction using the log link, and we compare the use of cumulative and survival densities as transformers. We show both constructions to be proper density generalizations, and we also extend the resulting families by introducing a new link function with the logarithm as a special case.

Talk G-2-4: Should I substitute my logisitic regression with a machine learning model?

Jean-Francois Plante, HEC Montréal

Friday, October 14, 2022, 12:00 PM – 12:20 PM

Abstract G-2-4: With the rising popularity of artificial intelligence, machine learning algorithms are being considered for an increasing number of problems. When it comes to binary classification, most algorithms can provide an estimate of the probability that an event will occur, however there may be no mathematical results to guarantee their consistency. After reviewing convergence results for classification trees and random forests in the literature, we consider a number of different settings in which the estimated probabilities are utilized and demonstrate that some could be negatively impacted by biased estimates. We then run an extensive Monte Carlo simulation inspired from nine real data sets to assess numerically the ability of those machine learning algorithms to provide appropriate estimates of the probabilities despite the lack of theoretical consistency results.

General Invited Session G-3

Friday, October 14, 2022, 2:40 PM to 4:00 PM, Virginia Room

Chair: Alaa Elkadry

Talk G-3-1: Independence Properties of the Truncated Multivariate Elliptical Distributions

Michael Levine, Purdue University

Friday, October 14, 2022, 2:40 PM – 3:00 PM

Abstract G-3-1: The truncated multivariate normal distribution has been applied extensively in statistical practice over the years. Some of its numerous application areas are simultaneous equations modelling and multivariate regression, economics, econometric models for auction theory, and statistical modeling in education. Recently, mixed models using truncated multivariate normal distribution have been introduced to analyze censored data.

For practical purposes, it is very helpful be able to provide an exact characterization of independence of any two subvectors of the truncated multivariate normal distribution. Such a characterization can, for example, provide the foundation of an independence test for these two subvectors. In our work, we provide such a characterization based on the blocks of the variance-covariance matrix of the underlying multivariate normal distribution. This result is later generalized to truncated multivariate elliptical distributions. An application to educational data analysis finishes our discussion.

Talk G-3-2: A family of generalized multinomial distributions

Nobuaki Hoshino, Kanazawa University

Friday, October 14, 2022, 3:00 PM – 3:20 PM

Abstract G-3-2: We construct a multiplicative probability measure over the same support as that of the multinomial distribution. The resulting probability mass function is characterized by the Bell polynomial, and we call the generalized multinomial distribution the Bell polynomial distribution (BPD). The BPD is equivalent to the conditional distribution of independent compound Poisson variates given the total value, if the total value is finite. Correspondingly, the BPD is shown to be a mixed multinomial distribution with the normalized infinitely divisible (NID) distribution, and we observe that the BPD inherits many good properties from the multinomial distribution.

Talk G-3-3: Sub-Dimensional Mardia Measures of Multivariate Skewness and Kurtosis

Joydeep Chowdhury, King Abdullah University of Science and Technology, Saudi Arabia

Friday, October 14, 2022, 3:20 PM – 3:40 PM

Abstract G-3-3: The Mardia measures of multivariate skewness and kurtosis summarize the respective characteristics of a multivariate distribution with only two numbers and cannot reflect the sub-dimensional features of the distribution. Consequently, testing procedures based on these measures may fail to detect skewness or kurtosis present in a sub-dimension of the multivariate distribution. We introduce sub-dimensional Mardia measures of multivariate skewness and kurtosis, whose respective maxima over all the sub-dimensions reflect the maximum skewness and kurtosis present in the distribution, and allow us to identify the sub-dimension bearing the highest skewness and kurtosis. Asymptotic distributions of the vectors of sub-dimensional Mardia measures are derived, based on which testing procedures for the presence of skewness

and of deviation from Gaussian kurtosis are developed. Through empirical study, It is found that these tests often outperform the existing tests of Gaussianity when the non-Gaussian features are supported on a small sub-dimension of the whole distribution.

Talk G-3-4: Multivariate Probability Distributions in Light of General Stochastic Dependence Theory

Jerzy Filus, Oakton College

Friday, October 14, 2022, 3:40 PM – 4:00 PM

Abstract G-3-4: An attempt to develop a general theory of stochastic dependence is rendered. As the key notion applied for this aim is relatively new concept of the joiner as occurring in several different probabilistic frameworks. The most general approach where the joiner is applied is description of conjunction of any random events. However, practically, the full power of this concept emerges when constructing bi or multidimensional survival functions given all the univariate marginal or baseline distributions. Thus, the joiners play the role of 'dependence functions', while their usefulness in stochastic modeling is competitive to the copula methodology. Examples of the (often quite new) multivariate probability distributions obtained by the joiner methodology are provided. Also, some very well known classical multivariate distributions are analyzed within the framework of the joiners. Comparison of joiners and copulas methods of construction will be provided as well.

General Invited Session G-4

Friday, October 14, 2022, 4:10 PM to 5:30 PM, Virginia Room

Chair: Sher Chhetri

Talk G-4-1: Odd Pareto families of distributions for modeling loss payment Data

Nonhle Mdziniso, Rochester Institute of Technology

Friday, October 14, 2022, 4:10 PM – 4:30 PM

Abstract G-4-1: A three-parameter generalization of the Pareto distribution is presented with density function having a flexible upper tail in modeling loss payment data. This generalized Pareto distribution will be referred to as the Odd Pareto distribution since it is derived by considering the distributions of the odds of the Pareto and inverse Pareto distributions. Basic properties of the Odd Pareto distribution (OP) are studied. Model parameters are estimated using both modified and regular maximum likelihood methods. Simulation studies are conducted to compare the OP with the exponentiated Pareto, Burr, and Kumaraswamy distributions using two different test statistics based on the ml method. Furthermore, two examples from the Norwegian fire insurance claims dataset are provided to illustrate the upper tail flexibility of the distribution. Extensions of the Odd Pareto distribution are also considered to improve the fitting of data.

Talk G-4-2: The Two-Sided Beta Distribution with Applications in Project Risk Analysis

Johan van Dorp, The George Washington University

Friday, October 14, 2022, 4:30 PM – 4:50 PM

Abstract G-4-2: A two-sided (TS) framework is developed for the construction of asymmetric continuous distributions with both bounded and unbounded support. Firstly, it shall be demonstrated that this TS framework provides for a direct link between the unbounded Asymmetric Laplace and Gaussian distributions and the bounded Generalized Two-Sided Power (GTSP) distribution. In that development process, a novel asymmetric Gaussian distribution emerges. Secondly, a three-parameter bounded distribution termed the Two-Sided Beta (TSB) distribution is derived, characterized by a mode (or anti-mode), and two branch power parameters. Two separate algorithms to indirectly elicit these power parameters of the TSB distribution shall be presented. One ensures matching of an elicited most likely value, as well as the PERT mean and PERT variance, popular in project management. The second algorithm solves for the power parameters from a lower and upper quantile constraint. The application of the TSB distribution shall be demonstrated in an illustrative PERT example.

Talk G-4-3: Copula regression for compound distributions with endogenous covariates with applications in insurance deductible pricing

Gee Lee, Michigan State University

Friday, October 14, 2022, 4:50 PM – 5:10 PM

Abstract G-4-3: This talk will present a statistical model for insurance policies with deductibles, where a copula is used to jointly model the endogenous deductibles, the claim frequencies, and severities. From an actuarial standpoint, an insurer may be interested in learning about the effect of the contract deductible on a policyholder's aggregate loss. The aggregate loss may be determined by a compound distribution, where a random number of individual claims is summed to determine the total amount of loss. Policyholders may choose the deductible level based on hidden information regarding the risk they

insure, which makes deductibles endogenous in the regression for both claim frequency and claim severity. To address this in the regression model for the compound aggregate loss, a novel approach is proposed using a pair copula construct to jointly model the policyholder's deductible, the number of claims, and individual claim amounts in the context of compound distributions. The proposed method provides insurers an empirical tool to uncover the underlying risk distribution of the potential customers. An insurance portfolio from a Wisconsin property insurance program that provides property coverage for building and contents of local government entities is used for an empirical study. Empirical evidence of the endogeneity of the deductible is found from the data, and the resulting model is used for actuarial applications.

Talk G-4-4: Parameter Estimation for a Stochastic Climate Model with Alpha-Stable L'evy Jumps

Sher Chhetri, University of South Carolina Sumter

Friday, October 14, 2022, 5:10 PM – 5:30 PM

Abstract G-4-4: Alpha-stable distributions are important in many fields, including finance, economics, and climate change. In this work, we use the sample characteristic function method and least-squares estimation method to estimate all parameters of a L'evy -driven stochastic climate model first proposed by Ditlevson in 2000.

General Invited Session G-5

Saturday, October 15, 2022, 9:10 AM to 10:30 AM, Virginia Room

Chair: Olusegun Otunuga

Talk G-5-1: Modeling Actuarial Data Using Iterated Trigonometric Distributions

Shahid Mohammad, University of Wisconsin Oshkosh

Saturday, October 15, 2022, 9:10 AM – 9:30 AM

Abstract G-5-1: A new class of trigonometric distribution (TD) is developed iteratively by replacing the random variable of the TD function (or its survival function) with another TD function (or its survival function). The resulting families of distributions produce various density and hazard shapes and a wide variety of tail shapes for the parametric modeling of complex data arising in actuarial science. The basic properties of these distributions are studied and presented. The new TDs successfully model the AON Re Belgian and Danish fire insurance data sets. Risk measures such as value at risk (VaR) and conditional tail expectation (CTE) are computed and compared with their empirical values and other distributions presented in actuarial literature. Also, limited expected values (LEV) for various policy limits are computed to compare with empirical counterparts to emphasize the new TDs' validity for actuarial predictions.

Talk G-5-2: The expected number of distinct patterns in a random permutation

Anant Godbole, East Tennessee State University

Saturday, October 15, 2022, 9:30 AM – 9:50 AM

Abstract G-5-2: Let π_n be a uniformly chosen random permutation on $[n]$. Using an analysis of the probability that two overlapping consecutive k-permutations are order isomorphic, the authors of Allen et al showed that the expected number of distinct *consecutive* patterns of all lengths $k \in \{1, 2, \ldots, n\}$ in π_n is $\frac{n^2}{2}(1 - o(1))$ as $n \to \infty$. This exhibits the fact that random permutations pack consecutive patterns near-perfectly. In this talk we will examine the non-consecutive case. (Note that beginning with a question asked by Herb Wilf at the inaugural Permutation Patterns conference in 2003, several authors have studied the *maximum* number of distinct non-consecutive patterns in a permutation π_n on $[n]$.) We study the random case.

Talk G-5-3: Confidence Interval for the Mean and Upper Tolerance Limit of Zero-Inflated Gamma Data

Derek Young, University of Kentucky

Saturday, October 15, 2022, 9:50 AM – 10:10 AM

Abstract G-5-3: In practice, it is not uncommon to observe count data that possess excessive zeros (i.e., zero-inflation) relative to the assumed discrete distribution. When data are semicontinuous, the log-normal and gamma distributions are often considered for modeling the positive part of the model. The problems of constructing a confidence interval for the mean and calculating an upper tolerance limit of a zero-inflated gamma population are considered using generalized fiducial inference. Our simulation studies indicate that the proposed method is very satisfactory in terms of coverage properties and precision. The proposed intervals are calculated for lipid profiles from a study on non-small cell lung cancer. This is a joint work with Dr. Yixuan Zou.

Talk G-5-4: Time-dependent probability density function for general stochastic logistic population model with harvesting effort

Olusegun Otunuga, Augusta University

Saturday, October 15, 2022, 10:10 AM – 10:30 AM

Abstract G-5-4: We derive and analyze the time-dependent probability density function for the number of individuals in a population at a given time in a general logistic population model with harvesting effort using the Fokker-Planck equation. The time-dependent probability density function (obtained as the unique principal solution of the Fokker-Planck equation corresponding to certain initial value and boundary conditions) is used to describe how the distribution of the population process changes with time. We assume the environment is randomly varying and the population is subject to a continuous spectrum of disturbances, with fluctuations in the intrinsic growth rate and the harvesting effort. The randomness is expressed as independent white noise processes. The effect of changes in the intrinsic growth rate, harvesting effort, and noise intensities on the distribution is investigated. In addition, conditions for the existence of optimal harvesting policy are obtained using properties of the time-dependent probability density function. The results obtained in this work are validated using population and published parameters

General Invited Session G-6

Saturday, October 15, 2022, 10:40 AM to 12:00 PM, Virginia Room

Chair: Gee Lee

Talk G-6-1: Functional Regression Measures of Influence on Out-of-sample Prediction

David Hitchcock, University of South Carolina

Saturday, October 15, 2022, 10:40 AM – 11:00 AM

Abstract G-6-1: Traditional influence measures in regression (DFFITS, Cook's distance) assess an observation's influence on the prediction of one or all observations in the sample. In the context of the fully functional linear model, we propose two measures of the influence of a functional observation on the prediction of the response curve of an out-of-sample "target observation" that we wish to predict. One measure uses the L2 distance between the response curves predicted with models fit with and without the potentially influential observation. Another measure assesses whether this influence is local within the function's domain or spread across the domain. We use a novel weighted bootstrap procedure to approximate null distributions for the measures and thereby formally assess significance. Simulations show our methods have excellent power and p-value behavior. We apply our methods to functional regressions on two functional data sets, involving river stages and air and water temperatures.

Talk G-6-2: New Generalized Extreme Value Distribution with Applications to Extreme Temperature Data

Wilson Gyasi, Northwood University

Saturday, October 15, 2022, 11:00 AM – 11:20 AM

Abstract G-6-2: A four-parameter generalization of the generalized extreme value (GEV) distribution is presented. This generalized GEV distribution is referred to as the Odd generalized extreme value (OGEV) distribution. The OGEV distribution and its extensions are enlightened by not only having all six families of extreme value distributions: Gumbel, Fréchet, Weibull, reverse-Gumbel, reverse-Fréchet, and reverse-Weibull as submodels but also convenient for modeling bimodal extreme value data. Some basic properties of the OGEV distribution are studied. Quantile-based aliases of the OGEV distribution are illustrated. Distribution parameters are estimated by employing the mle method. The adequacy of the OGEV distribution and its extensions is illustrated using well-known goodness-of-fit measures. The Value-at-risk (VaR) and Expected Shortfall (ES) are discussed to check the distributions' tail fitness. Two simulations are conducted to assess the appropriateness of the goodness-of-fit measures, and to validate the estimated risk measures. Two temperature data sets are analyzed to illustrate the OGEV distribution's bimodality, flexibility, and overall fitness.

Talk G-6-3: Fixed-Accuracy Confidence Interval Estimation of $P(X > c)$ for a Two-Parameter Gamma Population

Yan Zhuang, Connecticut College

Saturday, October 15, 2022, 11:20 AM – 11:40 AM

Abstract G-6-3: Gamma distribution is a flexible right-skewed distribution that is often used for modeling lifetime data. It is of great importance to know the probability of lifetime data exceeding a given value of interest. In this paper, we construct a fixed-accuracy confidence interval for $P(X > c)$ with X being a Gamma(α, β) random variable and c being a preassigned positive constant through 1) a purely sequential procedure with known shape parameter and unknown rate parameter;

2) a non-parametric purely sequential procedure with both shape and rate parameters unknown. Both procedures enjoy appealing asymptotic first-order efficiency and asymptotic consistency properties. Extensive simulations double validate the theoretical findings. Three real-life data examples are included to illustrate both procedures.

Talk G-6-4: Nonparametric intercept regularization

Gee Lee, Michigan State University

Saturday, October 15, 2022, 11:40 AM – 12:00 PM

Abstract G-6-4: In this paper, we illustrate how the ADMM approach for subgroup analysis can be modified so that it can be more easily incorporated into an insurance claims analysis. In a subgroup analysis for an actuarial problem, the investigator would be interested in determining unique groups of intercepts of the observed policyholders, assuming there are repeated observations of the policyholder's claims experience over discrete time points. We present an approach to penalize adjacent coefficients only, and show how the algorithm can be implemented for fast estimation of the parameters. We present three different cases of the model, depending on the level of dependence among the different coverage groups within the data. We provide an interpretation of the credibility problem using both random effects and fixed effects, where the fixed effects approach corresponds to the ADMM approach to subgroup analysis, while the random effects approach represents the classic Bayesian approach. Our empirical study demonstrates how the approach can be applied to real data using the Wisconsin Local Government Property Insurance Fund (LGPIF) data.

Abstracts for Posters

Poster Session

Saturday, October 15, 2022, 2:20PM to 3:00PM, Ohio Room

Poster P-1: Estimating Higher Order Parameters of Multivariate Bernoulli Data from Partially-Sampled Clusters

Susan Edwards, UNC Chapel Hill, Advisor: John Preisser

Abstract P-1: When cluster sizes (n) are large, sampling methods may be used to reduce costs and respondent burden resulting in partially sampled clusters. Partial-mouth recording methods have been utilized to gather correlated binary outcomes of periodontal disease at the tooth level. In this context, we discuss estimation of higher order parameters of multivariate Bernoulli distributions based upon cross-classification probabilities. We consider two definitions of disease prevalence at the complete cluster (mouth) level: the probability of (1) one or more or (2) two or more teeth with disease. As clusters are incomplete, three statistical distributions assuming within-cluster exchangeability are evaluated for fit and used to estimate disease prevalence. When teeth have varying probabilities for disease and complex structures for pairwise correlations, potentially varying by location and anatomical characteristics, alternative prevalence estimators are derived under a conditional linear family of multivariate Bernoulli distributions. Simulation studies are presented.

Poster P-2: Comparing Multiple Proportions using the Marascuillo Method (An analytical study of COVID-19 data in three different care facilities)

Christian Sterner, University of South Carolina Sumter, Advisor: Sher B. Chhetri

Abstract P-2: We use publicly reported data sets by the South Carolina Department of Health and Environmental Control (SCDHEC). Data collected prior to August 2020 was used to compare the different proportions to determine statistical significance between COVID-19 confirmed cases and deaths in three facilities: Residential Care Facilities, Nursing Homes, and Intermediate Care Facilities for Individuals with Intellectual Disabilities. The spread of respiratory illnesses like COVID-19 is common in these types of facilities, and the residents who live there are at high risk for developing complications or death from COVID-19 infection. We compare confirmed cases to residents, confirmed cases to staff, deaths of residents, and staff deaths. To compare the statistical significance of the proportions, the Marascuilo method was used. There is sufficient evidence to reject the null hypothesis based on the results from the Marascuilo method that indicates a statistical significance among the three care facilities regarding staff & resident infection rates, and resident mortality. We do not have enough evidence to reject the null hypothesis for the staff mortality rates. Nursing homes had the most infections and deaths from COVID-19, and the difference is statistically significant from the other two facilities studied.

Poster P-3: A Case of the Pareto-G Extended Weibull Distribution

Oluwaseun Otunuga, University of South Florida, Advisor: Dr. Lu Lu

Abstract P-3: In this work, we introduce and discuss a special case of the family of the Pareto (II) Generalized Extended Weibull Distribution, otherwise referred to as the Lomax Extended Weibull Distribution (LEWD). The parameters of LEWD are the parameters of the extended Weibull and that of the Pareto distributions. The properties of the LEW distribution, such as the hazard rate function, the survival function, moments, skewness, kurtosis, mean deviation, and entropy are investigated and discussed. The maximum likelihood estimation of the parameters and corresponding confidence intervals of the distribution are also discussed. We demonstrate the applications and versatility of the distribution over some existing distributions by analyzing real-life datasets.

Poster P-4: Utility in time description in priority best worst discrete choice models: An empirical comparison using Flynn's data.

Sasanka Adikari, Old Dominion University, Advisor: Dr Norou Diawra

Abstract P-4: Discrete choice models (DCMs) are applied in many fields and in the statistical modelling of consumer behavior. This paper focuses on a form of choice experiment, best-worst scaling in discrete choice experiments (DCEs) and the transition probability of a choice of a consumer over time. Analysis is done by using simulated data (choice pairs) based on data from Flynn (2007) 'Quality of life Experiment.' Most of the traditional approaches assume the choice alternatives are mutually exclusive over time, which is a questionable assumption. We introduced a new copula-based model (CO-CUB) for the transition probability, which not only can handle the dependent structure of best-worst choices while applying a very practical constraint. We suggest that CO-CUB transition probability algorithm is a novel way to analyze and predict choices in future time points by expressing human choice behavior.

Poster P-5: Induction of Survival Trees by Quadric Splits and Dipolar Splitting Criteria

Aye Aye Maung, University of Louisville, Advisor: Dr. Drew Lazar

Abstract P-5: Random forests in survival analysis involve splitting data at the nodes in underlying decision trees, while accounting for censoring in the survival information. In simple cases univariate splits are used to partition nodes across a single covariate at a time. Optimality of node splits is often gauged using a ranking of different possible splits according to some criterion that measures survival time homogeneity and variability within and across the split child nodes. In this work, we propose methods to split survival data at nodes using quadric surfaces, i.e. surfaces that solve degree 2 polynomials in the covariates. Data augmentation and optimization of special piecewise-linear functions are used to find optimal quadric splits. Optimality of the quadric splits are gauged using a recently proposed non-parametric splitting criterion called the survival dipolar criterion. We compare our methods against those using univariate and oblique hyperplane splits.

List of Presenters