INTERNATIONAL INDIAN STATISTICAL ASSOCIATION

2024 IISA CONFERENCE

DEC 27 - DEC 31

COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY

KOCHI, KERALA, INDIA

Scientific Program

iisaconference.org

IISA2024@intindstat.org

National Science Foundation

Johnson&Johnson Innovative Medicine

International Indian Statistical Association
Annual Conference
Cochin University of Science and Technology
Kochi, Kerala, India
27th - 31st December, 2024

Chief Patron: P. G. SANKARAN, Cochin University of Science and Technology

Scientific Programme Committee
Chair: Bertrand CLARKE, University of Nebraska-Lincoln
Co-chair: Kshitij KHARE, University of Florida, USA
Co-chair: Pallavi BASU, Indian School of Business, India

Sayan BANERJEE
Piyali BASAK
Sourish DAS
Angshuman SARKAR
Jyotiska DATTA
Soudeep DEB
Tanujit DEY
Yogita GHARDE
Asha GOPALAKRISHNAN
Kaushik JANA
Sayantee JANA
Karthik SRIRAM
Poh-Ling LOH
Steve MACEACHERN
Ranjan MAITRA
Debashis MONDAL
Somabha MUKHERJEE
Parthanil ROY
Vivekananda ROY
Sandip SINHARAY
Pragya SUR
Suresh CHENJI
Suhasini SUBBA RAO
Sandip ACHARYA
Shibasish DASGUPTA
Christian ROBERT
Dipankar BANDYOPADHYAY
Moumanti PODDER

Local Organising Committee
Chair and Local Coordinator: S. M. SUNOJ
Assistant Coordinator: M.R. IRSHAD
Assistant Coordinator: E.P. SREEDEVI

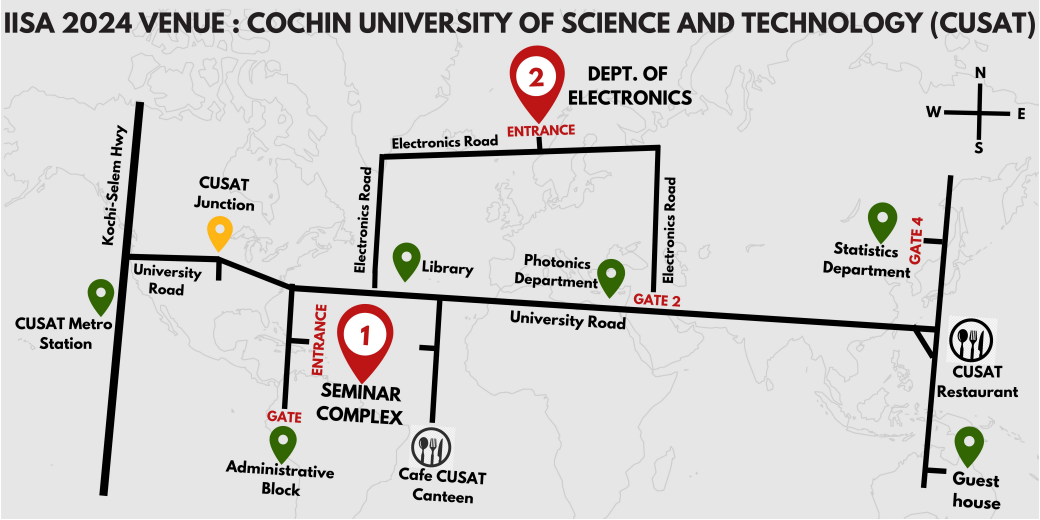
K.C. JAMES
Rajesh G.
M.R. PRINCY
Ambily JOSE
Hashmy HASSAN
N. BALAKRISHNA
K. JAYAKUMAR
C. SATHEESH KUMAR
E-I.Abdul SATHAR
Sebastian GEORGE
Joby JOSE
K. K. JOSE

International Organizing Committee:
Saonli BASU
Hiya BANERJEE
Subhashis GHOSHAL
Abhyuday MANDAL
Snigdhansu CHATTERJEE
Sanjay CHAUDHURI
Subrata KUNDU
Subhabrata MAJUMDAR
Arkaprava ROY
Debjit BISWAS

Student and Staff Volunteers:
Alfiya V. P.,Anush GEORGE, Anu Rose JOY, Linta JOSH, Maria JOSEPH, ABDUL FATHAH P. M., Nandhana DHANESH, Salma P. A., Hanan SUBAIR, Krishnapriya R S, Navya V, Avanthika S, A.K. Swamy NARAYANAN, Vishnu UNNIKRISHNAN, Kains Mathew BENOY, Anusree T R, Aparna K, AMEEN T V, Krishnanunni N. S., Thejus K J, Rahul M K, Charudarsan P, Prajul N A S, Febin Sulthana A. B., Mary Athulya DAVID, Fathima RILFA, Parvathy A. S., Chinnu C. NAIR, Sreelekha KUMAR.

Sponsorships and Endorsements:

Venue Map:



Lecture Theatres:

Seminar Complex	Room Name	Abbreviation
	Main Hall Seminar Hall Executive Hall	SC Main Hall SC Seminar Hall SC Executive Hall
Department of Electronics Building	Room Names	Abbreviation
	Auditorium Class Room 1 Class Room 2 Class Room 3 Class Room 4 Department of Computer Science Seminar Hall (2nd Floor)	DOE Auditorium DOE CL 1 DOE CL 2 DOE CL 3 DOE CL 4 DSC Seminar Hall

From The President, India Chapter

Dear Delegates,

It is my privilege to welcome you to the 2024 Annual Conference of the International Indian Statistical Association (IISA) being held at the Cochin University of Science and Technology, a state government-owned autonomous university. This year's event is in the historic coastal city of Cochin, also known as Kochi, located in the beautiful state of Kerala, India.

The IISA 2024 conference promises to be a landmark event to close out the year as the Scientific Program Committee and the Local and International Organizing Committees have put in a lot of effort over the past several months to design the agenda and plan event logistics. The conference program will consist of four plenary talks, 3 special invited talks, over 120 invited and contributed sessions, poster sessions, and a student paper competition. This year, we expect over 500 participants and about 450 talks across plenary, invited and contributed sessions.

Beyond the formal agenda, the IISA conference provides ample opportunities for networking...reconnecting with friends and colleagues, initiating collaborations, and just for catching up. I also hope that you will be able to make some time to visit places of historical interest in this port city that was an important destination on the Spice Route and was visited by many nationalities over the last 600 years, including the Chinese, Portuguese, Dutch and the British. The city's rich historical legacy can be seen at Fort Kochi, which still retains the ambience of its colonial past. Here one can see ancient temples, churches and synagogues standing in complete harmony.

Once again, welcome to Cochin/Kochi. Enjoy the event and the city!

Debjit Biswas
President, India Chapter
International Indian Statistical Association

December 21, 2024

Dear Conference Participants and Committee Members,

I bring greetings and best wishes on behalf of the Board of Directors and the membership of the American Statistical Association as you convene for the 2024 International Indian Statistical Association (IISA) Conference in Kochi, Kerala. We value our collaborations with the IISA and are delighted to support the outstanding conferences that IISA hosts.

I am impressed by the comprehensive program you have organized, featuring four distinguished plenary speakers including the R. R. Bahadur Memorial Lecture, six special invited talks, and over 100 invited sessions. The addition of two innovative short courses and the student paper and poster competition demonstrates a commitment to nurturing the next generation of statisticians. The conference, including speakers who are leaders from academia and industry, reflects the breadth and depth of our field.

Your conference's structure encourages collaboration and knowledge exchange across different sectors of our profession – academia, industry, and research institutions. With over 400 anticipated attendees from around the world, this conference will provide an excellent opportunity for fostering new connections and strengthening existing ones in our global statistical community.

Congratulations on creating a program that celebrates the importance and diversity of statistical science.

With best wishes,

M. GhoshDastidar

Madhumita Ghosh Dastidar
2024 President, American Statistical Association

Contents

Program Overview

Program

Friday December 27

Registration

Venue: Lobby

Time : **9:00 - 18:00**

Conference Inaugaration

Venue: SC Main Hall

Time : **9:30 - 10:15**

9:30 **Welcome : S. M. SUNOJ**, Cochin University of Science and Technology *Local Coordinator, IISA Conference 2024*

9:35 **Presidential Address : Saonli BASU**, University of Minnesota *President, IISA*

9:40 **About the Conference : Bertrand CLARKE**, University of Nebraska-Lincoln *Chair, Scientific Program Committee, IISA Conference 2024*

9:45 **Inaugaration : Junaid BUSHARI**, Cochin University of Science and Technology *Vice Chancellor CUSAT*

9:55 **Felicitation : Sanjay CHAUDHURI**, University of Nebraska-Lincoln *Executive Director, IISA*

10:00 **Felicitation : Sankaran PADUTHOL GODAN**, Cochin University of Science and Technology *Department of Statistics, CUSAT*

10:05 **Felicitation : Sam THOMAS**, Cochin University of Science and Technology *Director, IQAC, CUSAT*

10:10 **Vote of Thanks : Asha GOPALAKRISHNAN**, Cochin University of Science & Technology *Department of Statistics, CUSAT*

Coffee Break 10:20 - 10:40

01.M2.I1 Exploring Spatial Frontiers: Unraveling Insights Through Theory and Applications

Venue: SC Main Hall

Chair : Arvind RAO, University of Michigan

Organizer : Sagnik BHADURY, University of Michigan

10:40 **InterSpatial: Bi-directional Integration of Single Cell RNA-seq and Spatial Transcriptomic Data** [Abstract 275]

Tuhin MAJUMDER, *Duke University*
Cliburn CHAN, *Duke University*
Jichun XIE, *Duke University*

11:05 **BIPPS: Bayesian Inference for Point Patterns in Space applied to Multiplex imaging data** [Abstract 49]

Sagnik BHADURY, *University of Michigan*

11:30 **Network models for spatial transcriptomics data** [Abstract 6]

Satwik ACHARYYA, *University of Alabama at Birmingham*

01.M2.I2 Cutting-Edge Computational Methods for Diverse Scientific Challenges *Venue:* SC Seminar Hall

Chair : Arnab AUDDY, The Ohio State University
Organizer : Tanujit DEY, Brigham and Women's Hospital and Harvard Medical School

10:40 Integrative graphical modeling of multi-tissue omics data with application to modeling dependencies in the proteome of human pregnancy [Abstract 81]

Moumita CHAKRABORTY, *The University of Texas Medical Branch*

11:05 Continuous Time Reinforcement Learning using Rough Paths [Abstract 83]

Prakash CHAKRABORTY, *Pennsylvania State University*

11:30 Functional connectivity across the human subcortical auditory system using an autoregressive matrix-Gaussian copula graphical model approach with partial correlations [Abstract 88]

Noirrit Kiran CHANDRA, *The University of Texas at Dallas*
Kevin SITEK, *Northwestern University*
Bharath CHANDRASEKARAN, *Northwestern University*
Abhra SARKAR, *The University of Texas at Austin*

01.M2.I3 Causal Inference in Drug Development *Venue:* SC Executive Hall

Chair and Organizer : Suresh CHENJI, Eli Lilly

10:40 A Feasibility Study to assess the adequacy of historical clinical trial data using an external control to inform the Development of an Innovative treatment for a rare disease [Abstract 331]

Rashidkhan PATHAN, *Novartis Healthcare Pvt. Ltd.*
Gerhild ANGYALOSI , *Novartis AG Basel*
Kirsten CARTER , *Novartis AG Basel*
Rima IZEM, *Novartis AG Basel*

11:05 THE IMPACT OF TREATMENT DISCONTINUATION DUE ADVERSE EVENTS ON EFFICACY IN AN ONCOLOGY CLINICAL TRIAL [Abstract 31]

Hiya BANERJEE, *Eli Lilly*

11:30 Causal Inference 3 in Drug Development [Abstract 290]

Jit MITRA, *Eli Lilly*

01.M2.I4 Advances in predictive modelling techniques and applications *Venue:* DOE Auditorium

Chair : Snigdhansu CHATTERJEE, University of Maryland Baltimore County
Organizer : Yogita GHARDE, ICAR-Directorate of Weed Research

10:40 Cross impact analysis based models for technology forecasting in agriculture [Abstract 455]

Ramasubramanian VAIDHYANATHAN, *ICAR-National Academy of Agricultural Research Management, Hyderabad*
Umesh HUDEDAMANI, *ICAR-National Academy of Agricultural Research Management, Hyderabad*
Abin GEORGE, *ICAR-National Academy of Agricultural Research Management, Hyderabad*
Mrinmoy RAY, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

11:05 **Advances in the species distribution modelling for alien invasive weeds** [Abstract 161]

Yogita GHARDE, *ICAR-Directorate of Weed Research*

11:30 **The Impact of Data Quality on Predictive Model Performance - A Simulation Study** [Abstract 50]

Nivedita BHAKTHA, *IIT Kanpur*
Veena BANSAL,

01.M2.I5 Topics in network analysis

Venue: DOE CL 1

Chair and Organizer : Debashis MONDAL, Washington University in St Louis

10:40 **Conformal Prediction for Network-Assisted Regression** [Abstract 268]

Robert LUNDE, *Washington University in St Louis*

11:05 **Asymptotic non-identifiability of peer effects in linear-in-means models** [Abstract 267]

Keith LEVIN, *University of Wisconsin–Madison*
Alex HAYES,

11:30 **A Multiple Testing Approach for Estimating the Number of Blocks in Degree-Corrected Stochastic Block Models** [Abstract 213]

Chetkar JHA, *Ahmedabad University*
Ian BARNETT, *University of Pennsylvania*
Debashis MONDAL, *Washington University in St Louis*

01.M2.I6 Data Science in Fintech and Insurtech

Venue: DOE CL 2

Chair and Organizer : Karthik SRIRAM, IIM Ahmedabad

10:40 **Real-time claim risk assessment using ML models and technology** [Abstract 399]

Omesh SARAF, *Bajaj Allianz Life Insurance Co Ltd*

11:05 **Emotion-Aware Banking: An Artificial Intelligence Framework to Elevate Customer Interaction Experience Using Contextual Insights and Emotion Ontologies** [Abstract 470]

Dr Naveen YERI, *Wells Fargo-Rennes School of Business*

11:30 **Trade Durations and Liquidity Analysis of Government Bonds: An Econometric Study** [Abstract 227]

IRFANA K A, *federal bank*

01.M2.I7 Advances in Survey Methods and Designing of Experiments

Venue: DOE CL 3

Chair : C. Satheesh KUMAR, Department of Statistics, University of Kerala
Organizer : Cini VARGHESE, ICAR-IASRI, PUSA, Library Avenue, New Delhi - 110 012

10:40 **RANKED SET SAMPLING BASED ON OPTIMAL WEIGHTS** [Abstract 87]

Girish CHANDRA, *University of Delhi*

11:05 **On Alternative Hyper-Poisson Distribution and its Related Versions** [Abstract 252]

C. Satheesh KUMAR, *Department of Statistics, University of Kerala*

11:30 **Mappable Nearly Orthogonal Arrays using Projective Geometry** [Abstract 429]

Poonam SINGH, *Department of Statistics, University of Delhi, Delhi*

01.M2.I8 Recent advances in analyzing complex data structures

Venue: DOE CL 4

Chair : Rishikesh YADAV, Indian Institute of Technology Mandi
Organizer : Soudeep DEB, Indian Institute of Management Bangalore

10:40 A Regularized Low Tubal-Rank Model for High-dimensional Time Series Data [Abstract 368]

Samrat ROY, *Indian Institute of Management Ahmedabad*
George MICHAILIDIS, *University of California Los Angeles*

11:05 Relative Efficiency of Estimators in tensor Ising models [Abstract 302]

Somabha MUKHERJEE, *National University of Singapore*
Jaesung SON, *Columbia University*
Swarnadip GHOSH, *Radix Trading LLC*
Sourav MUKHERJEE, *Bristol Myers Squibb*

01.M2.I9 Recent Advances in Statistical Prediction Problems

Venue: DSC Seminar Hall

Chair : Saran Ishika MAITI, Visva-Bharati
Organizer : Sandip SINHARAY, Educational Testing Service

10:40 The robust desparsified lasso and the focused information criterion for high-dimensional generalized linear models [Abstract 353]

T. V. RAMANATHAN, *Savitribai Phule Pune University & Plaksha University*

11:05 A rank-based test of independence between nonparametric covariate and error in semiparametric regression with missing at random response [Abstract 272]

Saran Ishika MAITI, *Visva-Bharati*
Sthitadhi DAS,

11:30 Jacobi Prior: An Alternate Bayesian Method for Supervised Learning [Abstract 117]

Sourish DAS, *Chennai Mathematical Institute*

Lunch 12:00 - 13:30

Plenary Lecture 1 Pandurang Kulkarni

Venue: SC Main Hall

Chair : Hiya BANERJEE, Eli Lilly

13:30 Navigating a Dynamic Career in Statistics: Insights, Adaptation, and Building the Future [Abstract 250]

Pandurang KULKARNI, *Eli Lilly*

Short Break 14:30 - 14:40

01.A1.I10 High and infinite dimensional inference

Venue: SC Seminar Hall

Chair : Arnab AUDDY, The Ohio State University
Organizer : Steven MACEACHERN, The Ohio State University

14:40 **Regularized empirical likelihood for Bayesian inference: theory and applications** [Abstract 337]

Mario PERUGGIA, *The Ohio State University*
Eunseop KIM, *The Ohio State University*
Steven MACEACHERN, *The Ohio State University*

15:05 **Approximate, Leave-one-out, Cross Validation for Regression with ℓ_1 Regularizers** [Abstract 18]

Arnab AUDDY, *The Ohio State University*
Haolin ZOU, *Columbia University*
Kamiar RAHNAMA RAD, *Baruch College, CUNY*
Arian MALEKI, *Columbia University*

15:30 **SMART-MC: Sparse Matrix estimation with covaRiate-based Transitions in Markov Chain modelling via parallelized Multiple Spherically Constrained Optimization Routine (MSCOR)** [Abstract 112]

Priyam DAS, *Virginia Commonwealth University*

01.A1.I11 Statistics in Pediatric Drug Development

Venue: SC Executive Hall

Chair and Organizer : Suresh CHENJI, Eli Lilly

14:40 **Statistical Methodologies for Pediatric Drug Development: Enhancing Safety and Efficacy** [Abstract 451]

Anil Kumar THUKKAMATTATHIL, *Stat Analytics*

15:05 **Statistics 2 in Pediatric Drug Development** [Abstract 45]

Ramakrishna BATTULA, *Eli Lilly*

15:30 **Enhancing Pediatric Drug Development through Bayesian Dynamic Borrowing: A Case Study in Rare and Serious Diseases** [Abstract 228]

Lijina KAIPRATH, *GSK, India*

01.A1.I12 Recent Advances in Time Series and Network Statistics

Venue: DOE Auditorium

Chair and Organizer : Somabha MUKHERJEE, National University of Singapore

14:40 **Nonparametric Regression of Spatio-temporal Data using Infinite-dimensional Covariates** [Abstract 131]

Soudeep DEB, *Indian Institute of Management Bangalore*
Subhrajyoty ROY,
Sayar KARMAKAR,
Rishideep ROY,

15:05 **SMAUG: Sample-specific Multiomic Association networks Using Gaussian graphical models** [Abstract 388]

Enakshi SAHA, *University of South Carolina*
Katherine H. SHUTTA, *Channing Division of Network Medicine, Brigham and Women's Hospital*
John QUACKENBUSH, *Harvard T. H. Chan School of Public Health*

15:30 **Community Detection With Contextual Multilayer Networks** [Abstract 314]

Sagnik NANDY, *University of Chicago*
Zongming MA, *Yale University*

01.A1.I13 Modern statistical methods for passively collected data from wearables/smartphones

Venue: DOE CL 1

Chair and Organizer : Samprit BANERJEE, Weill Medical College of Cornell University

14:40 A semi-supervised and transfer learning support vector machine for predicting adherence to psychotherapy in mood disorders [Abstract 33]

Samprit BANERJEE, *Weill Medical College of Cornell University*

15:05 Detecting emotionally stressful periods from passive sensing data via mobile devices [Abstract 44]

Sumanta BASU, *Cornell University*
Younghoon KIM, *Cornell University*
Samprit BANERJEE, *Weill Cornell Medical College*

15:30 Multivariate Principal Component Analysis for Mixed-Type Functional Data with application to mHealth in Mood Disorders [Abstract 137]

Debangan DEY, *National Institute of Mental Health, USA*

01.A1.I14 Sustainability and Multi-Criteria Decision Making

Venue: DOE CL 2

Chair and Organizer : Leena KULKARNI, NMIMS (Deemed-to-be) University

14:40 Current Account Deficit Sustainability: The Indian Case [Abstract 330]

Savita PAREEK, *Auburn University*

15:05 Impact of Psychological Contract Breach on Innovative Behavior and Well-being amongst Academicians [Abstract 276]

Sunita MALL, *MICA,Ahmedabad*
Anushree KARANI , *Shri Jairambhai Patel Institute of Business Management and Computer Applications, Gandhinagar, India*
Sunita MALL, *MICA,Ahmedabad*
Revati DESHPANDE, *Independent Researchers, Gujarat, India*

15:30 ESG Performance and Sustainable Investment - Company's Comparative Analysis using the TOPSIS Method [Abstract 432]

Rushina SINGHI, *NMIMS (Deemed-to-be) University, Mumbai*

01.A1.I15 Recent Advances in Random Matrix Theory

Venue: DOE CL 3

Chair : Debraj DAS, Indian Institute of Technology Bombay
Organizers : *Monika BHATTACHARJEE, IIT Bombay* and *Piyali BASAK, Merck & Co., Inc*

14:40 Random Matrices with independent entries: Beyond non-crossing partitions [Abstract 414]

Priyanka SEN, *IIT Bombay*
Arup BOSE, *ISI Kolkata*
Koushik SAHA, *IIT Bombay*
Arusharka SEN, *Concordia University*

15:05 On the spectrum of random simplicial complexes in thermodynamic regime. [Abstract 7]

Kartick ADHIKARI, *IISER Bhopal*
Kiran KUMAR, *IIT Bombay*
Koushik SAHA, *IIT Bombay*

15:30 **Spectrum of random Centrosymmetric matrices** [Abstract 208]

Indrajit JANA, *IIT Bhubaneswar*
Sunita RANI, *IIT Bhubaneswar*

01.A1.I16 Advances in Machine Learning Methods

Venue: DOE CL 4

Chair and Organizer : Debashis MONDAL, Washington University in St Louis

14:40 **Learning from heterogeneous preferences** [Abstract 243]

Ramya KORLAKAI VINAYAK, *University of Wisconsin-Madison*

15:05 **TBA** [Abstract 396]

Kris SANKARAN, *University of Wisconsin-Madison*

15:30 **On imbalanced spatio-temporal data with missing values** [Abstract 93]

Snigdhansu CHATTERJEE, *University of Maryland Baltimore County*
Ben BAGOZZI, *University of Delaware*
Ujjal MUKHERJEE, *UIUC*
Vishal SUBEDI, *UMBC*

01.A1.I17 Recent Developments in Applied Probability

Venue: DSC Seminar Hall

Chair : Abdul SATHAR E I, University of Kerala

14:40 **On a Unified Class of Bivariate Distributions Characterized by a Functional Equation via a Binary Operator** [Abstract 182]

Asha GOPALAKRISHNAN, *Cochin University of Science & Technology*
Durga VASUDEVAN, *Cochin University of Science & Technology*

15:05 **Structural Aspects of Steady-state Characters in Queuing Systems** [Abstract 284]

Manoharan MAVILA, *University of Calicut*

15:30 **Properties of Extropy and Its Weighted Version for Doubly Truncated Random Variables** [Abstract 407]

Abdul SATHAR E I, *University of Kerala*

Coffee Break 16:00 - 16:30

Special Session 1 Felicitation Session Professor P. G. Sankaran

Venue: SC Main Hall

Chair : S. M. SUNOJ, Cochin University of Science and Technology

01.E1.I18 Bayesian Structure Learning

Venue: SC Seminar Hall

Chair and Organizer : Subhashis GHOSHAL, North Carolina State University

16:30 **The Nested Interpretable Neural Network Model and Its Application In Evaluating Digital Competitiveness of Manufacturing Enterprises in China** [Abstract 472]

William Weimin YOO, *Heriot-Watt University Malaysia*
Siwei LIU, *Heriot-Watt University Malaysia*
Sarat Chandra DASS, *Heriot-Watt University Malaysia*

16:55 **Coverage of Credible Sets for Regression under Variable Selection** [Abstract 180]

Subhashis GHOSHAL, *North Carolina State University*
Samhita PAL, *North Carolina State University*

17:20 **Bayesian biclustering and its application in education data analysis** [Abstract 420]

Weining SHEN, *University of California, Irvine*

01.E1.I19 Statistical Methods in Health and Related Areas: Opportunities and Challenges

Venue: SC Executive Hall

Chair : Mathur SUNIL, Houston Methodist Research Institute

Organizers : *Sunil MATHUR, Houston Methodist Neal Cancer Center* and *Himel MALLICK, Cornell University*

16:30 **Next generation statistical workflows for multimodal data integration** [Abstract 266]

Leo LAHTI, *University of Turku, Finland*

16:55 **Gene identification through a sparse correlation approach** [Abstract 412]

Ananda SEN, *University of Michigan*

17:20 **Predicting Immune Checkpoint Blockade by Stereotactic Radiation and In Situ Virus Gene Therapy in Metastatic Triple-Negative Breast Cancer** [Abstract 282]

Sunil MATHUR, *Houston Methodist Neal Cancer Center*
Ethan BURNS, *Houston Methodist NEal Cancer Center*
Shreya MATHUR, *Boston Medical Center*
Jenny CHANG, *Houston Methodist Neal Cancer Center*

01.E1.I20 Topics in high dimensional and network data

Venue: DOE Auditorium

Chair : Savita PAREEK, Auburn University

Organizers : *Debraj DAS, Indian Institute of Technology Bombay* and *Subhajit DUTTA, ISI Kolkata and IIT Kanpur*

16:30 **Statistical inference in networks under snowball type sampling** [Abstract 91]

Arindam CHATTERJEE, *Stat Math Unit, Indian Statistical Institute, Delhi*

16:55 **On Exact Feature Screening in Ultrahigh-dimensional Classification** [Abstract 153]

Subhajit DUTTA, *ISI Kolkata and IIT Kanpur*

17:20 **Central Limit Theorem in High Dimension** [Abstract 110]

Debraj DAS, *Indian Institute of Technology Bombay*
Soumendra LAHIRI,

01.E1.C1 Contributed Session 1

Venue: DOE CL 1

Chair : P M SAFWANA, University of Calicut

16:30 **Exploring Quantile-Based Cumulative Residual Extropy of Record Values: A Comprehensive Analysis.** [Abstract 460]

VEENA L VIJAYAN, *Department of Statistics, University of Kerala*
E I Abdul SATHAR, *University of Kerala*

16:45 **Estimation Based on Extended Sequential Order Statistics for Lindely Distribution** [Abstract 215]

Jaaziah P JOHN, *University of Kerala*
Manoj CHACKO, *University of Kerala*
Jaaziah P JOHN, *University of Kerala*

17:00 **Asymptotic Tests for Tree Ordered Effects in One-Way ANOVA** [Abstract 196]

Subha HALDER, *IIT Kharagpur*
Somesh KUMAR, *IIT Kharagpur*

17:15 **A tensor-based algorithm for partitioning asymmetric measure of predictability: Delta index** [Abstract 207]

Riya JAIN, *Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University Jalgaon, Maharashtra, India*
Kirtee KAMALJA, *Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India*

17:30 **On Interval Entropy Loss Functions** [Abstract 384]

P M SAFWANA, *University of Calicut*

01.E1.C2 Contributed Session 2

Venue: DOE CL 2

Chair : Anand R, University of Calicut

16:30 **On estimation of multicomponent stress-strength reliability for Bilal distribution** [Abstract 244]

Ashly Elizabeth KOSHY, *University of Kerala*
Manoj CHACKO, *University of Kerala*

16:45 **PGDUS Kumaraswamy Distribution: Properties and Reliability Analysis** [Abstract 395]

ANN SANIA, *St.Thomas College(Autonomous), Thrissur, University of Calicut, Kerala*
V M CHACKO, *Professor, St Thomas College (Autonomous), Thrissur*

17:00 **On Parameter Estimation of Gompertz Distribution under Constant Stress Accelerated Life Testing using Adaptive type-II progressive censoring** [Abstract 379]

Sreelakshmi S, *University of Kerala, Trivandrum, Kerala*
Manoj CHACKO, *University of Kerala, Trivandrum, Kerala, India*

17:15 **Nonparametric estimation of Matusita's measure for residual and past lifetimes** [Abstract 217]

Chinu JOSEPH, *Mahatma Gandhi University,Kottayam,Kerala,Imdia*
Chinu JOSEPH, *Mahatma Gandhi University, Kottayam,Kerala, India*
Angel MATHEW, *Mahatma Gandhi University, Kottayam,Kerala,India*

17:30 **Reliability properties and applications of proportional reversed hazards in reversed relevation transform** [Abstract 348]

Anand R, *University of Calicut*
Dileepkumar M, *University of Calicut*
Sankaran P.G, *Cochin Univeristy of Science and Technology*

01.E1.C3 Contributed Session 3

Venue: DOE CL 3

Chair : Thomas XAVIER, Novartis, India

16:30 **Modified Polygenic Risk Score to account for Gene Effects** [Abstract 23]

Devarpita BAG, *Duke University*
Manit PAUL, *Wharton School, University of Pennsylvania*
Indranil MUKHOPADHYAY, *UNIVERSITY of NEBRASKA–LINCOLN*

16:45 **Trends of Deep learning at single-cell level for biomarker based drug discovery and development** [Abstract 291]

Sharda MOHAK, *Glaxo Smith Kline*

17:00 **Predicting Response to Immunotherapy in Cancer Using Multi-Omic Data Integration and Machine Learning** [Abstract 148]

MAHESH DIVAKARAN, *Amity University, Lucknow*
Gunjan SINGH, *Amity University, Lucknow*
Jayadevan SREEDHARAN, *Gulf Medical University*

17:15 **Joint Modeling of binary longitudinal Response and Treatment Compliance in the presence of missing data in a Clinical Trial setup** [Abstract 409]

Sarfaraz SAYYED, *Novartis Healthcare Pvt Ltd*
Ashwini MATHUR, *Onesto Consulting Limited*
Asha KAMATH, *Manipal University*

17:30 **Selection of optimal time window in restricted mean survival time under delayed treatment effect using statistical distributions** [Abstract 466]

Thomas XAVIER, *Novartis, India*

01.E1.C4 Contributed Session 4

Venue: DOE CL 4

Chair : Debika GHOSH, Indian Institute of Management, Udaipur

16:30 **Change point in functional data** [Abstract 121]

Debanjana DATTA, *Indian Statistical Institute, Bangalore*
Rituparna SEN, *Associate Professor,Indian Statistical Institute, Bangalore*

16:45 **Change Point Detection for Functional Data Based on the Notion of Maximum Mean Discrepancy** [Abstract 77]

Sourav CHAKRABARTY, *Indian Statistical Institute*
Anirvan CHAKRABORTY, *Indian Institute of Science Education and Research Kolkata*
Shyamal Krishna DE, *Indian Statistical Institute, Kolkata*

17:00 **Evaluating treatment heterogeneity in survival outcomes using causal inference methods** [Abstract 4]

Sharon Varghese A, *Novartis Healthcare Private Limited*

17:15 **Exploring the Estimation Techniques and Stress-Strength analysis on PGDUS-powered inverse Rayleigh Distribution** [Abstract 269]

AMRUTHA M, *Research Scholar, Department of Statistics, St. Thomas College (Autonomous), Thrissur, Kerala, India*
V M CHACKO, *Professor, Department of Statistics, St. Thomas College (Autonomous), Thrissur, Kerala, India.*

17:30 **Regularized Additive Matrix Autoregressive Model** [Abstract 167]

Debika GHOSH, *Indian Institute of Management, Udaipur*
Samrat ROY, *Indian Institute of Management, Ahmedabad*
Nilanjana CHAKRABORTY, *Indian Institute of Management, Udaipur*

01.E1.C5 Contributed Session 5

Venue: DSC Seminar Hall

Chair : Ashlin VARKEY, Farook College (Autonomous), Kozhikode

16:30 On Solving Uncertain Transportation Problem [Abstract 195]

Ummey HABIBA, *Indian Institute of Information Technology Guwahati*
Masihuddin MASIHUDDIN, *Indian Institute of Information Technology Guwahati*

16:45 Enhanced Breast Cancer Diagnosis Using a Novel Hybrid Deep Learning Approach [Abstract 194]

Balakrishna GURMITKAL, *Yenepoya (Deemed to be University)*
Ismail B, *Department of Statistics, Yenepoya (Deemed to be University)*
N/A N/A, *N/A*
N/A N/A, *N/A*

17:00 Characterization and Nonparametric Estimation of Bivariate Weighted Residual Inaccuracy Measure [Abstract 73]

Anju C. THOMAS, *Department of Statistics, University of Kerala*
E.I. ABDUL SATHAR, *Department of Statistics, University of Kerala.*

17:15 A New Approach to Constructing a Family of Distributions Using Sigmoid Functions and Its Applications [Abstract 89]

Greeshma CHANDRAN, *University of Calicut*
Manoharan M, *University of Calicut*

17:30 Quantile based Analysis of Residual Incomes [Abstract 458]

Ashlin VARKEY, *Farook College (Autonomous), Kozhikode*
Haritha N HARIDAS, *Farook College (Autonomous), Kozhikode*
Ashlin VARKEY, *Farook College (Autonomous), Kozhikode*

17:45 Inferences on Lifetime Performance Index of Weibull Products under Step- Stress Setup [Abstract 304]

Sarat Sindhu MUKHOPADHYAY, *Indian Statistical Institute, Bangalore*
Viswakala K.V., *Research Associate, SQC & OR Unit, Indian Statistical Institute, Bangalore*
Gijo E.V., *Faculty member, SQC & OR Unit Indian Statistical Institute, Bangalore*

Saturday December 28

Special Invited Session 1 Bhaswar Bhattacharyya, Madhu Mazumdar *Venue:* SC Main Hall

Chair : Somabha MUKHERJEE, National University of Singapore

09:00 **Kernel and Graphical Methods for Comparing Conditional Distributions** [Abstract 55]

Bhaswar BHATTACHARYA , *University of Pennsylvania*
Anirban CHATTERJEE, *University of Pennsylvania*
Ziang NIU , *University of Pennsylvania*

09:40 **Detecting and Managing Malnutrition with AI** [Abstract 285]

Madhu MAZUMDAR, *Icahn School of Medicine at Mount Sinai*
Melanie BESCULIDES, *Institute for Healthcare Delivery Science, Icahn School of Medicine at Mount Sinai*
Sara WILSON, *Institute for Healthcare Delivery Science, Icahn School of Medicine at Mount Sinai*
Pramathamesh PARCHURI , *Institute for Healthcare Delivery Science, Icahn School of Medicine at Mount Sinai*

02.M1.I21 Random Matrices and Multivariate Distributions and their applications *Venue:* SC Seminar Hall

Chair and Organizer : Sayantee JANA, IIT Hyderabad

9:00 **Bulk Spectrum of Sample Covariance Type Matrices.** [Abstract 71]

Arup BOSE, *Indian Statistical Institute*

9:25 **High-dimensional Bernstein Von-Mises theorems for covariance and precision matrices** [Abstract 404]

Partha SARKAR, *Florida State University*
Kshitij KHARE, *University of Florida*
Malay GHOSH, *University of Florida*
Matt P. WAND, *School of Mathematical and Physical Sciences, University of Technology Sydney*

9:50 **Generalized Multivariate Analysis of Variance (GMANOVA) models for volatile data** [Abstract 211]

Sayantee JANA, *IIT Hyderabad*

02.M1.I22 Innovations in spatial studies *Venue:* SC Executive Hall

Chair : Debjoy THAKUR, Washington University in St. Louis
Organizer : Soudeep DEB, Indian Institute of Management Bangalore

9:00 **Frequency Domain Resampling for Spatial Data** [Abstract 27]

Soutir BANDYOPADHYAY, *Colorado School of Mines*
Souvick BERA, *Colorado School of Mines*
Dan NORDMAN, *Iowa State University*

9:25 **Bayesian Scalar-on-Image Regression with Spatial Interactions for Modeling Alzheimer's Disease** [Abstract 82]

Nilanjana CHAKRABORTY, *Indian Institue of Management Udaipur*
Qi LONG, *Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania*
Suprateek KUNDU, *Department of Biostatistics, Division of Basic Science Research, The University of Texas MD Anderson Cancer Center*

9:50 **Threshold Exceedance Modeling for Spatial Extreme Data Using Neural Network** [**Abstract 450**]

Debjoy THAKUR, *Washington University in St. Louis*
Soumendra N. LAHIRI, *Washington University in St. Louis*

Student Paper Competition 1 Application of Statistics and Data Sciences

Venue: DOE Auditorium

Chair : Kshitij KHARE, University of Florida

9:00 **Bayesian Variable Selection and Sparse Estimation for High-Dimensional Graphical Models** [**Abstract 86**]

Anwesha CHAKRAVARTI, *University of Illinois Urbana Champaign*
Naveen NARISHETTY,
Feng LIANG, *University of Illinois Urbana Champaign*

9:15 **Anytime-Valid Inference for Double/Debiased Machine Learning of Causal Parameters** [**Abstract 109**]

Abhinandan DALAL, *University of Pennsylvania*
Patrick BLOEBAUM, *Amazon Web Services*
Shiva KASIVISWANATHAN, *Amazon Web Services*
Aaditya RAMDAS, *Carnegie Mellon University*

9:30 **Robust Principal Component Analysis using Density Power Divergence** [**Abstract 373**]

Subhrajyoty ROY, *Indian Statistical Institute, Kolkata*
Ayanendranath BASU, *Indian Statistical Institute, Kolkata*
Abhik GHOSH, *Indian Statistical Institute, Kolkata*

9:45 **Adaptive Rate-Optimal Lack-Of-Fit Testing For Systems Of Ordinary Differential Equations** [**Abstract 367**]

Archi ROY, *Doctoral student*
Itai DATTNER, *Proffessor*
Moumanti PODDER, *Assistant Proffessor*

10:00 **Bayesian Semi-supervised Multi-category Classification under Nonparanormality** [**Abstract 174**]

Shuvrarghya GHOSH, *North Carolina State University*
Rui ZHU, *Google Inc.*
Subhashis GHOSAL, *North Carolina State University*

02.M1.I23 Modern advances in statistical learning

Venue: DOE CL 1

Chair : Subhabrata SEN, Harvard University
Organizer : Pragya SUR, Harvard University

9:00 **Statistical Analysis on In-Context Learning** [**Abstract 200**]

Masaaki IMAIZUMI, *Dept. of Statistics, University of Tokyo*

9:25 **Single Index Batched Contextual Bandits** [**Abstract 17**]

Sakshi ARYA, *Case Western Reserve University*
Hyebin SONG, *Penn State University*

9:50 **Precise generalization error of min-norm interpolants under transfer learning** [**Abstract 443**]

Pragya SUR, *Harvard University*

02.M1.I24 Advances in Cure Rate Modeling

Venue: DOE CL 2

Chair and Organizer : Suvra PAL, University of Texas at Arlington

9:00 **Destructive cure models with proportional hazards lifetimes and associated likelihood inference** [Abstract 40]

Sandip BARUI, *Indian Statistical Institute*
Narayanaswamy BALAKRISHNAN, *McMaster University*

9:25 **Novel Machine Learning-Based Cure Models** [Abstract 327]

Suvra PAL, *University of Texas at Arlington*

9:50 **Unified Competing Risks Mixture Cure Model for Cancer Survival** [Abstract 43]

Sanjib BASU, *Division of Epidemiology and Biostatistics, University of Illinois Chicago*

02.M1.I25 High-dimensional statistics, statistical signal processing, extreme value theory

Venue: DOE CL 3

Chair : Nabin Kumar JANA, National Institute of Science Education and Research, Bhubaneswar
Organizer : Moumanti PODDER, IISER, Pune

9:00 **Modeling periodic and nearly periodic signals** [Abstract 186]

Rhythm GROVER, *Indian Institute of Technology Guwahati*
Debasis KUNDU, *IIT Kanpur*

9:25 **Asymptotic methods in high dimensional regression under dependence** [Abstract 265]

Samriddha LAHIRY, *National University of Singapore*
Pragya SUR, *Harvard University*
Jean BARBIER, *International Centre for Theoretical Physics*
Daria TIEPLOVA, *International Centre for Theoretical Physics*

9:50 **Top eigenvalues and eigenvectors of inhomogeneous Erdős-Rényi random graphs** [Abstract 76]

Arijit CHAKRABARTY, *Indian Statistical Institute*
Bishakh BHATTACHARYA, *Indian Statistical Institute*
Sukrit CHAKRABORTY, *Achhruram Memorial College*
Rajat HAZRA, *University of Leiden*

Coffee Break 10:20 - 10:40

Panel Discussion 1 CWS sponsored session on professional development at IISA 2024

Venue: SC Main Hall

Organizer : Sayantee JANA, IIT Hyderabad
Sponsor: Caucas for Women in Statistics

- **Raghu SENGUPTA**, IIT Kanpur
- **Soudeep DEB**, Indian Institute of Management Bangalore
- **Sourish DAS**, Chennai Mathematical Institute
- **Sowmya R. RAO**, Boston University School of Public Health

02.M2.I26 Bayesian Methods for Survey Sampling and Macroeconomic Forecasting *Venue:* SC Seminar Hall

Chair : Partha SARKAR, Florida State University
Organizer : Malay GHOSH, University of Florida

10:40 Bayesian Pseudo Posterior Mechanism under Asymptotic Differential Privacy [Abstract 408]

Terrance SAVITSKY, *U.S. Bureau of Labor Statistics*
Matthew WILLIAMS, *RTI International*
Monika HU,

11:05 Bayesian inference in high-dimensional mixed frequency regression [Abstract 238]

Kshitij KHARE, *University of Florida*

11:30 Global Local Priors for Spatial Small Area Estimation [Abstract 169]

Malay GHOSH, *University of Florida*
Xueying TANG, *University of Arizona*

02.M2.I27 Cutting-Edge Approaches to Statistical Analysis: Insights from High-Dimensional Data, Clinical Studies, and Marketing Dynamics *Venue:* SC Executive Hall

Chair : Shibasish DASGUPTA, Pfizer
Organizer : Sourish DAS, Chennai Mathematical Institute

10:40 Statistical Machine Learning Approach to Feature Selection in High Dimensional Low Sample Size (HDLSS) Data [Abstract 119]

Shibasish DASGUPTA, *Pfizer*
Vladimir IVANOV , *Pfizer*

11:05 Applications of Dynamic Linear Models in Marketing [Abstract 352]

Balaji RAMAN, *Cogitaas*

11:30 Rhetoric and Reality: Tracing the Evolution of Policy Discourse, Agenda Setting, and Populism in Indian Parliamentary Speeches [Abstract 156]

Kausik GANGOPADHYAY, *Indian Institute of Management Kozhikode*
Swarn RAJAN, *Indian Institute of Management Kozhikode*
Anirban GHATAK, *Indian Institute of Management Kozhikode*

Student Paper Competition 2 Probability and Theory of Statistics and Data Sciences *Venue:* DOE Auditorium

Chair : Bhaswar BHATTACHARYA , University of Pennsylvania

10:40 High Dimensional Behaviour of Some Two-Sample Tests Based on Ball Divergence [Abstract 29]

BILOL BANERJEE, *INDIAN STATISTICAL INSTITUTE KOLKATA*
ANIL K. GHOSH, *INDIAN STATISTICAL INSTITUTE KOLKATA*

11:55 Network sampling based inference for subgraph counts and clustering coefficient in a Stochastic Block Model framework with some extensions to a sparse case [Abstract 279]

Anirban MANDAL, *TSMU, Indian Statistical Institute, Delhi*
Anirban MANDAL, *Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Delhi*
Arindam CHATTERJEE, *Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Delhi*

11:10 Spectra of adjacency and Laplacian matrices of Erdős-Rényi hypergraphs [**Abstract 325**]

Dipranjan PAL, *Indian Statistical Institute, Kolkata*
Soumendu Sundar MUKHERJEE, *Indian Statistical Institute, Kolkata*
Himasish TALUKDAR, *Indian Statistical Institute, Kolkata*

11:25 Estimation and Inference for Change Points in Functional Regression Time Series [**Abstract 258**]

Shivam KUMAR, *PhD Candidate*
Haotian XU, *University of Warwick*
Haeran CHO, *University of Bristol*
Daren WANG, *University of Notre Dame*

11:40 Higher-Order Graphon Theory: Fluctuations, Degeneracies and Inference. [**Abstract 90**]

Anirban CHATTERJEE, *University of Pennsylvania*
Soham DAN, *IBM Research, Yorktown Heights*
Bhaswar BHATTACHARYA, *University of Pennsylvania*

02.M2.I28 Statistical Methods for Business

Venue: DOE CL 1

Chair and Organizer : Gourab MUKHERJEE, University of Southern California

10:40 Joint modeling of playing time & purchase propensity in massively multiplayer online role-playing games [**Abstract 152**]

Shantanu DUTTA, *University of Southern California*
Gourab MUKHERJEE, *University of Southern California*
Trambak BANERJEE,

11:05 The Spillover Effect of a Minimum Wage Increase on Commission based Sales Agents at a Multi-level Marketing Firm [**Abstract 426**]

Sivaramakrishnan SIDDARTH, *Marshall School of Business, University of Southern California*
Sung Joo KIM, *Purdue University*
Wreetabrata KAR, *SUNY Buffalo*
Dinesh PURANAM, *Marshall School of Business, USC*

11:30 Stochastic Optimization Algorithms for Instrumental Variable Regression with Streaming Data [**Abstract 364**]

Abhishek ROY, *Texas A&M University*
Xuxing CHEN, *University of California Davis*
Yifan HU, *College of Management of Technology, EPFL, Department of Computer Science, ETH Zurich, Switzerland*
Krishnakumar BALASUBRAMANIAN, *University of California Davis*

02.M2.I29 Statistical learning and quantification of ocean and climate signals

Venue: DOE CL 2

Chair : Kaushik JANA, Ahmedabad University
Organizer : Radhendushka SRIVASTAVA, Indian Institute of Technology Bombay

10:40 Climatologist's perspective of emergence of climate change signals [**Abstract 204**]

Suresh IYYAPPAN, *Digital University Kerala*

11:05 Bayesian model and uncertainty quantification for Time of Emergence [Abstract 178]

Sujit GHOSH , *NC State University*
Suresh IYYAPPAN, *Digital University of Kerala*
Radhendushka SRIVASTAVA , *Indian Institute of Technology Bombay*

11:30 Statistical modelling and uncertainty quantification of temperature inversion in the Bay of Bengal [Abstract 442]

Radhendushka SRIVASTAVA, *Indian Institute of Technology Bombay*
Suresh IYYAPPAN, *Digital University Kerala*
Jovi D'SILVA, *CSIR-National Institute of Oceanography, Goa*

02.M2.I30 Methodology for the Analysis of Structured Data

Venue: DOE CL 3

Chair : Rajarshi GUHANIYOGI, Texas A&M University
Organizer : Ranjan MAITRA, Iowa State University

10:40 Bayesian approaches for Modeling Brain Network Dynamics [Abstract 187]

Sharmistha GUHA, *Texas A&M University*

11:05 Statistical Test for Diagnosing Rotational Symmetry in Rotation Data with Applications in Materials Science [Abstract 64]

Eva BISWAS, *Iowa State University*
Daniel NORDMAN, *Iowa State University*
Ulrike GENSCHEL, *Iowa State University*

11:30 TBA [Abstract 189]

Rajarshi GUHANIYOGI, *Texas A&M University*

02.M2.I31 Discrete probability, probabilistic methods, random graphs

Venue: DOE CL 4

Chair : Rhythm GROVER, Indian Institute of Technology Guwahati
Organizer : Moumanti PODDER, IISER, Pune

10:40 Community detection on multi-view networks [Abstract 415]

Subhabrata SEN, *Harvard University*
Xiaodong YANG, *Harvard University*
Buyu LIN,

11:05 Central Limit Theorem for Exponential GREM [Abstract 210]

Nabin Kumar JANA, *National Institute of Science Education and Research, Bhubaneswar*

11:30 Markov chains for spin systems from random initializations [Abstract 162]

Reza GHEISSARI, *Northwestern University*

Lunch 12:00 - 13:30

Plenary Lecture 2 Sudipto Banerjee

Venue: SC Main Hall

Chair : Saonli BASU, University of Minnesota

13:30 Artificially Intelligent Data Analysis: A Case Study in Spatial Energetics [Abstract 36]

Sudipto BANERJEE, *University of California Los Angeles*

Short Break 14:30 - 14:40

Student Poster Competition Students' Poster Competition

Venue: Lobby

Time : **14:40 - 18:00**

1. **Improving the Adaptive Gauss-Hermite Quadrature approach for Approximate Bayesian Inference** [Abstract 346]

   Priyanka PRIYANKA, *Indian Institute of Technology Hyderabad*
   Patrick BROWN, *University of Toronto*
   Sayantee JANA, *Indian Institute of Technology Hyderabad*

2. **High-dimensional Quadratic Discriminant Analysis using Random Projections** [Abstract 129]

   Annesha DEB, *Indian Institute of Technology, Kanpur*
   Minerva MUKHOPADHYAY, *Indian Statistical Institute, Kolkata*
   Subhajit DUTTA, *Indian Statistical Institute, Kolkata*

3. **Bayesian deep generative reinforcement learning** [Abstract 436]

   Shreya SINHA ROY, *University of Warwick*
   Ritabrata DUTTA, *University of Warwick*
   Richard EVERITT, *University of Warwick*
   Christian ROBERT, *Université Paris-Dauphine (France) , University of Warwick*

4. **Dynamical Survival Analysis of SIR model** [Abstract 374]

   Suchismita ROY, *Duke University*
   Jason XU, *Duke University*
   Alexander FISHER, *Duke University*

5. **Signal-to-noise-ratio aware minimax analysis of sparse linear regression** [Abstract 173]

   Shubhangi GHOSH, *Columbia University*
   Yilin GUO, *Two Sigma*
   Haolei WENG, *Michigan State University*
   Arian MALEKI, *Columbia University*

6. **PACE : Privacy Aware Collaborative Estimation for Heterogeneous GLMs** [Abstract 360]

   Bhaskar RAY, *North Carolina State University*
   Srijan SENGUPTA, *North Carolina State University*
   Aritra MITRA, *North Carolina State University*

7. **LLM based approach for binary classification in peptide bioactivity** [Abstract 358]

   Ravi , *Bioinformatics Division, ICAR- Indian Agricultural Statistics Research Institute*

8. **A note on Quasi Periodic Gaussian Process** [Abstract 317]

   Unnati NIGAM, *IIT Bombay, Mumbai, India*
   Radhendushka SRIVASTAVA, *IIT Bombay, India*
   Michael BURKE, *Monash University, Australia*
   Faezeh MARZBANRAD, *Monash University, Australia*

9. **On martingale characterizations of generalized counting process and its time-changed variants** [Abstract 2]

   Manisha , *Indian Institute of Technology Bhilai*

10. **L-Estimation in Instrumental Variables Regression for Censored Data in Presence of Endogeneity and Dependent Errors** [Abstract 425]

    SWATI SHUKLA, *Indian Institute Of Technology Kanpur*
    Subhra Sankar DHAR, *Indian Institute Of Technology Kanpur*

11. **Dir-SPGLM: A Bayesian semiparametric GLM with data-driven reference distribution** [Abstract 12]

    Entejar ALAM, *University of Texas at Austin*
    Peter MÜLLER, *University of Texas at Austin*
    Paul J. RATHOUZ, *University of Texas at Austin*

12. **High-dimensional Bayesian Compressed Mixed-Effects Models** [Abstract 406]

    Sreya SARKAR, *University of Iowa*
    Sanvesh SRIVASTAVA, *University of Iowa*
    Kshitij KHARE, *University of Florida*

13. **Approximate Bayesian inference for high-resolution spatial disaggregation using alternative data sources** [Abstract 324]

    Anis PAKRASHI, *PhD Student in Statistics, Department of Statistics, the Pennsylvania State University, USA*
    Arnab HAZRA, *Assistant Professor of Statistics, Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, India*
    Sooraj M RAVEENDRAN, *Senior Consultant - Urban Informatics Lab, Indian Institute of Human Settlements, Bengaluru, India*
    Krishnachandran BALAKRISHNAN, *Lead - Research, Indian Institute of Human Settlements, Bengaluru, India*

14. **A unified Bayesian approach to transcriptome-wide association study** [Abstract 236]

    Arnab Kumar KHAN, *Indian Statistical Institute, Kolkata*
    Tanushree HALDAR, *Institute for Human Genetics, University of California San Francisco, San Francisco, California 94143, USA*
    Arunabha MAJUMDAR, *Department of Mathematics, Indian Institute of Technology Hyderabad, Kandi, Telangana 502285, India*

15. **Robust Bayesian Model Averaging for Linear Regression Models With Heavy-Tailed Errors** [Abstract 127]

    Shamriddha DE, *The University of Iowa*
    Joyee GHOSH, *The University of Iowa*

16. **Modern Problems Require Precise Solutions: Beating Traditional RCTs by Leveraging External Clinical Data and Adaptive Enrichment** [Abstract 181]

Souradipto GHOSH DASTIDAR, *University of Minnesota Twin Cities*
Jialing LIU, *University of Minnesota Twin Cities*
Aidan NEHER, *University of Minnesota Twin Cities*
Steffen VENTZ, *University of Minnesota Twin Cities*

17. **A novel non-linear intervention approach using time series model and Horel function** [**Abstract 67**]

Subhankar BISWAS, *ICAR-Indian Agricultural Research Institute*
Amrit Kumar PAUL, *Principal Scientist and Head, Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute*
Ranjit PAUL, *Senior Scientist, Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute*
Md YEASIN, *Scientist, Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute*

18. **Exploring Block Clustering with Probabilistic Distance: Theory and Validation** [**Abstract 52**]

Shrirksihna BHAT K, *Pondicherry University*
Shrikrishna BHAT K, *Department of Statistics, Pondicherry University*
Kiruthika C, *Department of Statistics, Pondicherry University*

19. **Testing synchronization of change-points for multiple time series** [**Abstract 70**]

Soham BONNERJEE, *University of Chicago*
SAYAR KARMAKAR, *UNIVERSITY OF FLORIDA*
MAGGIE CHENG, *ILLINOIS INSTITUTE OF TECHNOLOGY*
WEI BIAO WU, *UNIVERSITY OF CHICAGO*

20. **The Curious Problem of the Inverse Mean** [**Abstract 176**]

Soham GHOSH, *University of Wisconsin, Madison*
Uttaran CHATTERJEE, *School of Industrial Engineering, Purdue University*
Jyotishka DATTA, *Department of Statistics, Virginia Tech*

21. **Asymptotic Properties of Generalized Elephant Random Walks** [**Abstract 382**]

Tamojit SADHUKHAN, *Indian Statistical Institute, Kolkata*
Krishanu MAULIK, *Indian Statistical Institute, Kolkata*
Parthanil ROY, *Indian Statistical Institute, Bangalore*

22. **Extreme learning machine ensemble model for time series forecasting boosted by particle swarm optimization** [**Abstract 113**]

Saikath DAS, *ICAR - Indian Agricultural Research Institute*
Ranjit Kumar PAUL, *Senior Scientist, Division of Statistical Genetics, ICAR-Indian Agricultural Statistical Research Institute, New Delhi*
Md YEASIN, *Scientist, Division of Statistical Genetics, ICAR-Indian Agricultural Statistical Research Institute, New Delhi*
Amrit Kumar PAUL, *Principal Scientist and Head, Division of Statistical Genetics, ICAR-Indian Agricultural Statistical Research Institute, New Delhi*

23. **Asymptotic Inference in Genetic Association Studies Leveraging Genetic Correlations** [**Abstract 398**]

Madhav SANKARANARAYANAN, *Harvard T.H. Chan School of Public Health*
Rajarshi MUKHERJEE, *Harvard T.H. Chan School of Public Health*
Tamar SOFER, *Beth Israel Deaconess Medical Center*
Yana HRYTSENKO, *Beth Israel Deaconess Medical Center*

24. **Construction of Cyclic Minimal Balanced and Cyclic Minimal Strongly Balanced Crossover Designs** [**Abstract 191**]

Ashish GUPTA, *ICAR-IASRI*
Baidya Nath MANDAL, *ICAR-IARI, Jharkhand*
Rajender PARSAD, *ICAR-IASRI*
Cini VARGHESE, *ICAR-IASRI*

02.A1.I32 Recent advances in Machine Learning for Spatial Data Analysis *Venue:* SC Main Hall

Chair and Organizer : Arkajyoti SAHA, Department of Statistics, University of California, Irvine

14:40 Random Forests for Geospatial Data [**Abstract 386**]

Arkajyoti SAHA, *Department of Statistics, University of California, Irvine*
Sumanta BASU, *Department of Statistics and Data Science, Cornell University*
Abhirup DATTA, *Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

15:05 Modeling Spatio-temporal Extremes with Conditional Variational Autoencoders [**Abstract 473**]

Likun ZHANG, *University of Missouri - Columbia*
Xiaoyu MA, *University of Missouri - Columbia*
Christopher K. WIKLE, *University of Missouri - Columbia*

15:30 Sparse Spatiotemporal Dynamic Generalized Linear Models for Inference and Prediction of Bike Counts [**Abstract 467**]

Rishikesh YADAV, *Indian Institute of Technology Mandi*

02.A1.I33 Applications of Reliability Theory *Venue:* SC Seminar Hall

Chair : Shyamal GHOSH, School of Dada Science, IISER Thiruvananthapuram
Organizer : Priyanka MAJUMDER, Indian Institute of Science Education and Research Thiruvananthapuram

14:40 A stochastic comparison study on replacement policies [**Abstract 274**]

Priyanka MAJUMDER, *Indian Institute of Science Education and Research Thiruvananthapuram*
Akhila Anna VARGHESE, *IISER Thiruvananthapuram*

15:05 Parameter Estimation for Multi-state Coherent Series and Parallel Systems having Random Degradation Rates [**Abstract 249**]

Leena KULKARNI, *NMIMS (Deemed-to-be) University*
Sanjeev SABNIS, *IIT Bombay*

15:30 Bayesian Inference for Stress-Strength Reliability Models Using Phase-Type Distributions [**Abstract 239**]

Joby K JOSE, *Kannur University*

02.A1.I34 Recent Developments in Bayesian Modeling for Biostatistics and Health Data Science *Venue:* SC Executive Hall

Chair : Sagnik BHADURY, University of Michigan
Organizer : Jyotishka DATTA, Virginia Polytechnic Institute and State University

14:40 "Bayesian clustering with feature selection." [**Abstract 316**]

Leelavati NARLIKAR, *IISER Pune*

15:05 **Bayesian compositional regression with flexible microbiome feature aggregation and selection** [Abstract 389]

Satabdi SAHA, *The University of Texas MD Anderson Cancer Center*

15:30 **Bayesian Beta Regression for Analyzing Well-Being Patterns in ZIP Codes** [Abstract 292]

Shariq MOHAMMED, *Boston University*

02.A1.I35 Advances in Statistical and Data Science Pedagogy

Venue: DOE Auditorium

Chair : Arpita MONDAL, Indian Institute of Technology Bombay
Organizer : Kaushik JANA, Ahmedabad University

14:40 **Exploring Data Science Foundations in Engineering Education: Leveraging Computer Simulations for Pedagogical Advancement** [Abstract 61]

Amiya Ranjan BHOWMICK, *Institute of Chemical Technology, Mumbai*
Riddhi BHARANI, *Vivekanand Education Society's College of Arts, Science and Commerce (Autonomous), Mumbai*

15:05 **A FUN PROBLEM FROM MATHEMATICAL STATISTICS** [Abstract 355]

Marepalli RAO, *University of Cincinnati*
Neelakshi CHATTERJEE, *University of Cincinnati*
Zhaochong YU, *University of Cincinnati*
Shesh RAI, *University of Cincinnati*

02.A1.I36 Interpretable Models for Omics Data Science and their Diverse Applications

Venue: DOE CL 1

Chair and Organizer : Himel MALLICK, Cornell University

14:40 **Generalized propensity scores to obtain causal estimates in observational studies of environment, microbiome and health** [Abstract 280]

Siddhartha MANDAL, *Center for Chronic Disease Control, New Delhi, India; Center for Health Analytics and Trends, Ashoka University, India*

15:05 **Re-purposing without Reinventing the Wheel - Ensemble Models for Differential Analysis** [Abstract 277]

Himel MALLICK, *Cornell University*

15:30 **Group Heteroscedasticity - A Silent Saboteur of Power and False Discovery in RNA-Seq Differential Expression** [Abstract 94]

Suvo CHATTERJEE, *Indiana University, Bloomington*
Himel MALLICK, *Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA*
Arindam FADIKAR, *Decision and Infrastructure Sciences Division, Argonne National Laboratory, Lemont, IL, USA*
Arunkumar GANESAN, *Department of Cell Biology and Physiology, University of New Mexico, Albuquerque, NM, USA*

02.A1.I37 Statistical Applications in Management

Venue: DOE CL 2

Chair and Organizer : Karthik SRIRAM, IIM Ahmedabad

14:40 **On Mis-specified Newsvendor Models: A Precision and Complexity Comparison** [Abstract 305]

Sujay MUKHOTI, *Indian Institute of Management Indore*
Soham GHOSH, *Indian Institute of Technology Indore*

15:05 Optimal Reliability Sampling Plans under Progressive Type-I Interval Censoring Schemes [**Abstract 372**]

Soumya ROY, *Indian Institute of Management Kozhikode*

15:30 Data-driven crime prediction with human inputss [**Abstract 439**]

Karthik SRIRAM, *IIM Ahmedabad*
Ankur SINHA, *IIM Ahmedabad*
Suvashis CHOUDHARY, *Police Department (retired)*

02.A1.I38 Recent Advances in Survival Analysis and Biostatistics

Venue: DOE CL 3

Chair and Organizer : Savitri JOSHI, Indian Institute of Information Technology Allahabad

14:40 Truncation Effect and it's Adjustment via Parametric and Non Parametric approaches: Illustration through First Birth Interval Data [**Abstract 251**]

Anup KUMAR, *Department of Biostatistics and Health Informatics Sanjay Gandhi Postgraduate Institute of Medical Sciences Lucknow, Uttar Pradesh*

15:05 Joint Modeliing of Longitudinal PSA Levels and Survival Data on Prostate Cancer with Missing Grades [**Abstract 329**]

Mahaveer Singh PANWAR, *Banaras Hindu University, Varanasi*
Vikas BARNWAL, *Department of Biostatistics, Boston University, USA*

15:30 Hazard Regression Change Point Model with Applications to Bone marrow transplant data [**Abstract 219**]

Savitri JOSHI, *Indian Institute of Information Technology Allahabad*

02.A1.I39 Statistical Methods for High-Dimensional Complex Data

Venue: DOE CL 4

Chair : Minerva MUKHOPADHYAY, IIT Kanpur, ISI Kolkata
Organizer : Vivekananda ROY, Iowa State University

14:40 Analyzing Latent Trajectory in Longitudinal Item Responses Using Gaussian Processes [**Abstract 138**]

Dipak DEY, *University of Connecticut*
Yuhao LI, *University of Connecticut*
Xiaojing WANG, *University of Connecticut*

15:05 Bayesian prior for achieving multiple goals and its relationship with the adjusted maximum likelihood method [**Abstract 264**]

Parthasarathi LAHIRI, *University of Maryland College Park*
Masayo HIROSE, *Kyushu University, Japan*

15:30 A test for the temporal changepoint detection in spatiotemporal data [**Abstract 303**]

Minerva MUKHOPADHYAY, *IIT Kanpur, ISI Kolkata*
Arnab HAZRA,

02.A1.I40 Geometry and dynamics of random networks

Venue: DSC Seminar Hall

Chair and Organizer : Sayan BANERJEE, University of North Carolina, Chapel Hill

14:40 **Ising inverse critical temperature for preferential attachment models** [Abstract 362]

Rounak RAY, *TU Eindhoven*

15:05 **Mean-field methods under interference** [Abstract 57]

Sohom BHATTACHARYA, *University of Florida*

15:30 **Exploration-driven networks** [Abstract 34]

Sayan BANERJEE, *University of North Carolina, Chapel Hill*

Coffee Break 16:00 - 16:30

Memorial Session 1 In Memorium: Professor P. K. Sen

Venue: SC Main Hall

Chair and Organizer : Malay GHOSH, University of Florida

16:30 **Professor Pranab Kumar Sen – the Advisor and the Academic Guru** [Abstract 116]

Shubhabrata DAS, *Indian Institute of Management Bangalore*

16:55 **Professor P.K. Sen-A Friend , Guide and Philosopher** [Abstract 170]

Malay GHOSH, *University of Florida*

17:20 **Limit theorems via moments and cumulants** [Abstract 72]

Arup BOSE, *Indian Statistical Institute*

02.E1.I41 Statistics and Machine Learning for Climate Extreme

Venue: SC Seminar Hall

Chair : Auroop GANGULY, Northeastern University

Organizers : *Kaushik JANA, Ahmedabad University* and *Arpita MONDAL, Indian Institute of Technology Bombay*

16:30 **Exploration of Hydrologic Persistence using Principles of Complex Systems** [Abstract 140]

Pankaj DEY, *IIT Roorkee*
Pankaj DEY,

16:55 **Is deep learning-based climate downscaling faithful to extreme events?** [Abstract 289]

Adway MITRA, *Indian Institute of Technology Kharagpur*
Sumanta Chandra MISHRA SHARMA, *IIT Kharagpur*
Deepayan CHAKRABORTY, *IIT Kharagpur*

17:20 **Extreme Value Analysis for Attribution of Floods and Heat Waves to Climate Change** [Abstract 293]

Arpita MONDAL, *Indian Institute of Technology Bombay*
Roshan JHA, *Indian Institute of Technology Bombay*
Manish Kumar DHASMANA, *Indian Institute of Technology Bombay*

02.E1.I42 Statistical Methods for Complex Health data

Venue: SC Executive Hall

Chair : Samiran SINHA, Texas A&M University
Organizer : Souparno GHOSH, University of Nebraska-Lincoln

16:30 **Analysis of spatially clustered survival data with unobserved covariates using SBART** [Abstract 433]

Debajyoti SINHA, *Florida State University*

16:55 **CMPLE: Correlation Modeling to Decode Photosynthesis Using the Minorize–Maximize Algorithm** [Abstract 435]

Samiran SINHA, *Texas A&M University*
Abhijnan CHATTOPADHYAY,
David KRAMER,
Tapabrata MAITI,

17:20 **Wavelet-Based Multiscale Analysis for Enhanced Detection of Heart Murmurs in Phonocardiographic Signals** [Abstract 461]

Horahenage VIMALAJEEWA, *University of Nebraska-Lincoln*

02.E1.C6 Contributed Session 6

Venue: DOE Auditorium

Chair : Himasish TALUKDAR, Indian Statistical Institute, Kolkata

16:30 **Adaptive Estimation of the Transition Kernel of Controlled Markov Chains** [Abstract 32]

Imon BANERJEE, *northwestern university*
Imon BANERJEE, *Northwestern University*
Harsha HONNAPPA, *Purdue University*
Vinayak RAO, *Purdue University*

16:45 **Time changed Counting processes** [Abstract 193]

Neha GUPTA, *IIT Kanpur*
Neha GUPTA, *IIT Kanpur*

17:00 **A quantile-based bivariate distribution** [Abstract 343]

Shifna P R, *Cochin University of Science and Technology*
N. Unnikrishnan NAIR, *Cochin University of Science and Technology*
S. M. SUNOJ, *Cochin University of Science and Technology*

17:15 **The Exponentiated Teissier Distribution and Its Bivariate Extension: Properties, Estimation Techniques, and Applications** [Abstract 419]

Vikas Kumar SHARMA, *Banaras Hindu University, Varanasi*
Sudhanshu V. SINGH, *IITRAM, Ahemedabad*
Komal SHEKHAWAT, *IITRAM, Ahmedabad*
Ashok Kumar PATHAK, *Central University of Panjab, Bathinda*

17:30 **Generalized inaccuracy measure of order α in order statistics** [Abstract 339]

TINCY PHILIP, *COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY*
RAJESH G, *COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY*

17:45 **Bulk Spectra of Truncated Sample Covariance Matrices** [Abstract 449]

Himasish TALUKDAR, *Indian Statistical Institute, Kolkata*
Subhroshekhar GHOSH, *National University of Singapore*
Soumendu Sundar MUKHERJEE, *Indian Statistical Institute, Kolkata*

02.E1.C7 Contributed Session 7

Venue: DOE CL 1

Chair : Jaya Naga Sri AKURATHI, University of Hyderabad

16:30 Generalized Marshall-Olkin Bivariate Exponential Distribution and Its Application in Constant-Stress Accelerated Life Test [Abstract 22]

Sneha BABU, *Cochin University of Science and Technology, Cochin, Kerala*
Princy T,

16:45 Realtime Mortality Prediction of Left Truncated Right Censored (LTRC) MIMIC-IV data Via Piecewise Linear Approximation (PLA) of Hazard Function in Presence of Competing Risks [Abstract 326]

Sandip PAL, *S&P Global Inc*
Arnab KOLEY, *Indian Institute of Management, Indore*
Debasis KUNDU, *Indian Institute of Technology, Kanpur*
Gayatri PAL, *Tesco India*

17:00 Inference on progressive-stress accelerated life testing under generalized adaptive progressive hybrid censoring [Abstract 345]

Aman PRAKASH, *NIT Surat, Gujarat (395007)*
Aman PRAKASH, *NIT Surat*
Raj Kamal MAURYA, *NIT Surat*

17:15 Optimal life testing plan in presence of hybrid censoring using compound optimal design strategy [Abstract 144]

Vaibhavbhai DHAMELIYA, *SVNIT Surat, Gujarat (395007)*
Vaibhavbhai DHAMELIYA, *SVNIT Surat, Gujarat (395007)*
Raj Kamal MAURYA, *SVNIT Surat, Gujarat (395007)*

17:30 Fiducial inference on lifetime performance index based on type-II censored samples [Abstract 47]

KUNTAL BERA, *Indian Statistical Institute, Kolkata*
M.Z. ANIS, *Indian Statistical Institute*

17:45 Analysis of Interval Censored Competing Risks Data with Covariate Measurement Error [Abstract 10]

Jaya Naga Sri AKURATHI, *University of Hyderabad*
Anjana S, *University of Hyderabad*

02.E1.C8 Contributed Session 8

Venue: DOE CL 2

Chair : Saurabh MISHRA, Indian Institute of Technology, Hyderabad

16:30 Biomarker Selection in Randomized Clinical Trials with Survival Outcomes: Addressing High-Dimensional and Low Sample Size Challenges [Abstract 311]

Vipin N, *Novartis Healthcare Private Limited*
Vipin N,

16:45 Novel and Efficient Pipeline for Metagenomics Binning [Abstract 177]

SUBHAM GHOSH, *ICAR-Indian Agricultural Research Institute*
ULAVAPPA ANGADI, *Principal Scientist, Division of Agricultural Bioinformatics,ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*
DIPRO SINHA, *Research Associate,Division of Agricultural Bioinformatics,ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*
MIR ASIF IQUEBAL, *Principal Scientist,Division of Agricultural Bioinformatics,ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*

17:00 **Estimation Of FGMBEW and FGMBGE distributions under Progressive censored data** [Abstract 255]

Saneesh KUMAR, *M G University*
Ansa Alphonsa ANTONY, *Mg University*

17:15 **A multi-region discrete time chain binomial model for infectious disease transmission** [Abstract 434]

Pallab Kumar SINHA, *Indian Institute of Technology Bombay*
Siuli MUKHOPADHYAY, *Department of Mathematics, Indian Institute of Technology Bombay*

17:30 **A New Zero Truncated Discrete Odd Lindley Half Logistic distribution with its Application to real data** [Abstract 41]

Shabana BASHEER, *Maharaja's College, Ernakulam*
Dr. Ansa ALPHONSA ANTONY, *St. Xavier's College for Women, Aluva*

17:45 **A multivariate approach to joint testing of main genetic and gene-environment interaction effects** [Abstract 288]

Saurabh MISHRA, *Indian Institute of Technology, Hyderabad*
Arunabha MAJUMDAR, *Indian Institute of Technology Hyderabad*

02.E1.C9 Contributed Session 9

Venue: DOE CL 3

Chair : Manit PAUL, Wharton School, University of Pennsylvania

16:30 **On adaptivity of nearest neighbors in non-smooth factor models** [Abstract 383]

Tathagata SADHUKHAN, *Cornell University*
Manit PAUL, *University of Pennsylvania*
Dwivedi RAAZ, *Cornell University*

16:45 **A study on truncated ROC curves** [Abstract 19]

Anu AUGUSTINE, *Mahatma Gandhi University kottayam*
Anu AUGUSTINE, *Mahatma Gandhi University Kottayam, Kerala*
Dr. Angel MATHEW, *Mahatma Gandhi University Kottayam, Kerala*

17:00 **Least Absolute Deviation (LAD) Regression for Cross Sectional Data with Cluster Dependence: Inference based on Averaging** [Abstract 142]

Subhodeep DEY, *Indian Statistical Institute, Kolkata*
Gopal K. BASAK, *Indian Statistical Institute, Kolkata*
Samarjit DAS, *Indian Statistical Institute, Kolkata*

17:15 **Two-dimensional stochastic volatility models for image processing** [Abstract 205]

FATHIMA JAFNA, *DEPARTMENT OF STATISTICS, UNIVERSITY OF CALICUT*
KRISHNARANI S. D., *DEPARTMENT OF STATISTICS, UNIVERSITY OF CALICUT*

17:30 **TAVIE: Tangent Approximation for Variational Inference in different Exponential Families** [Abstract 371]

Somjit ROY, *Department of Statistics, Texas A&M University*
Pritam DEY, *Department of Statistics, Texas A&M University*
Debdeep PATI, *Department of Statistics, University of Wisconsin-Madison*
Bani K. MALLICK, *Department of Statistics, Texas A&M University*

17:45 **Inference for Median and a Generalization of HulC** [Abstract 334]

Manit PAUL, *Wharton School, University of Pennsylvania*
Arun Kumar KUCHIBHOTLA, *Carnegie Mellon University*

02.E1.C10 Contributed Session 10

Venue: DOE CL 4

Chair : Anurag RAWAT, ICAR-IASRI

16:30 LSD of Large Kendall's Correlation Matrix and it's Application [Abstract 421]

Raunak SHEVADE, *IIT Bombay*
Monika BHATTACHARJEE, *IIT Bombay*
Raunak SHEVADE, *IIT Bombay*

16:45 Efficient Designs for Multivariate Crossover Trials [Abstract 318]

Shubham Sanjay NIPHADKAR, *Indian Institute of Technology Bombay*
Siuli MUKHOPADHYAY, *Indian Institute of Technology Bombay*

17:00 Statistical inference and optimal censoring scheme for generalized exponential distribution under improved adaptive type-II progressive censoring [Abstract 410]

Shilpa S DEV, *University of Kerala*
Manoj CHACKO, *University of Kerala*

17:15 Two-stage drop-the-losers design for the selection of effective treatments and estimating their worth [Abstract 232]

Yogesh KATARIYA, *Indian Institute of Technology Kanpur*

17:30 An algorithm for construction of Uniform designs [Abstract 102]

Rakesh CHHALOTRE, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*
Rajender PARSAD, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*
Baidya Nath MANDAL, *ICAR - Indian Agricultural Research Institute, Jharkhand – 825405*
Sukanta DASH, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*

17:45 Block Design for Two-level factorial experiments in block size four [Abstract 359]

Anurag RAWAT, *ICAR-IASRI*
Sukanta DASH, *Senior Scientist*

Meeting 1 IISA General Body Meeting

Venue: SC Seminar Hall

Time : **18:00 - 19:00**

Chair and Organizer : Saonli BASU, University of Minnesota

Sunday December 29

Short Course 1 Omics Data Science Workshop

Venue: DSC Seminar Hall

Time : **9:30 - 12:30**

- **Himel MALLICK**, Cornell University

Special Invited Session 2 Jennifer Clarke, Debashree Ray

Venue: SC Main Hall

Chair : Jyotishka DATTA, Virginia Polytechnic Institute and State University

09:00 **Stability as an Objective Criterion for Shrinkage Penalty Selection with Application in High Dimensional Classification** [Abstract 108]

Jennifer CLARKE, *Department of Statistics, University of Nebraska-Lincoln, USA*
Dean DUSTIN, *University of Nebraska-Lincoln*
Bertrand CLARKE, *University of Nebraska-Lincoln*
Laura KRESTY, *University of Michigan*

09:40 **TBA** [Abstract 361]

Debashree RAY, *Johns Hopkins University*

03.M1.I43 Health Data Sciences

Venue: SC Seminar Hall

Chair and Organizer : Tanujit CHAKRABORTY, Sorbonne University

9:00 **Contextual Bandits for Mobile Health Studies with Zero-inflated Count Outcomes** [Abstract 80]

Bibhas CHAKRABORTY, *National University of Singapore*

9:25 **Unveiling Cancer Complexity Through Integrative Multi-View Machine Learning** [Abstract 235]

Aparajita KHAN, *Indian Institute of Technology Roorkee*
Aparajita KHAN, *Indian Institute of Technology Roorkee*

9:50 **Epicasting: Neural Networks for Epidemic Forecasting** [Abstract 85]

Tanujit CHAKRABORTY, *Sorbonne University*

03.M1.I44 Application of empirical process theory in non-standard problems with independent and dependent data

Venue: SC Executive Hall

Chair and Organizer : Debarghya MUKHERJEE, Boston University

9:00 **Trade-off between dependence and complexity in Wasserstein distance learning** [Abstract 130]

Nabarun DEB, *University of Chicago*
Debarghya MUKHERJEE,

9:25 **A novel longitudinal rank-sum test for multiple primary endpoints in clinical trials** [Abstract 168]

Dhrubajyoti GHOSH, *Duke University*
Sheng LUO, *Duke University*
Xiaoming XU,

9:50 **Change Point Detection for Random Objects using Distance Profiles** [Abstract 149]

Paromita DUBEY, *University of Southern California*
Minxing ZHENG, *University of Southern California*

03.M1.I45 Modern regression methods with applications in various fields of science

Venue: DOE Auditorium

Chair : Prajamitra BHUYAN, Indian Institute of Management Calcutta

Organizer : Jayant JHA, Indian Statistical Institute

9:00 **Two-stage Circular-circular Regression with Zero-inflation: Application to Medical Sciences** [Abstract 214]

Jayant JHA, *Indian Statistical Institute*
Prajamitra BHUYAN , *IIM Calcutta*

9:25 **A feature-based image comparison method** [Abstract 299]

Partha Sarathi MUKHERJEE, *Indian Statistical Institute, Kolkata*
Anik ROY, *Indian Statistical Institute, Kolkata*
Partha Sarathi MUKHERJEE, *Indian Statistical Institute, Kolkata*

9:50 **A Bayesian method for estimating gene-level polygenicity under the framework of transcriptome-wide association study** [Abstract 273]

Arunabha MAJUMDAR, *IIT Hyderabad*

03.M1.I46 Reliability & Survival Analysis

Venue: DOE CL 1

Chair : Priyanka MAJUMDER, Indian Institute of Science Education and Research Thiruvananthapuram

Organizer : Sudipta DAS, RKMVERI, Belur

9:00 **Estimating scale-disparity in the proportional hazards model using martingale residuals** [Abstract 393]

Shyamsundar SAHOO, *Department of Statistics, Haldia Government College*

9:25 **Moment inequality for decreasing mean time to failure distributions with hypothesis testing application** [Abstract 175]

Shyamal GHOSH, *School of Dada Science, IISER Thiruvananthapuram*

9:50 **Current Status Data with two Competing risks and Missing failure types: A Non-Parametric Approach** [Abstract 242]

Tamalika KOLEY, *Indian Institute of Management Lucknow*

03.M1.I47 Beyond the Noise: Collaborative Statistical Learning and Regularized Multivariate Regression for High-Dimensional Data Inference

Venue: DOE CL 2

Chair : Priyam DAS, Virginia Commonwealth University

Organizer : Tanujit DEY, Brigham and Women's Hospital and Harvard Medical School

9:00 **Communication Efficient Distributed Learning with Bayesian Kernel Machine Models** [Abstract 84]

Sounak CHAKRABORTY, *University of Missouri*
Tanujit DEY, *Brigham and Women's Hospital. Harvard Medical School*
Anjishnu BANERJEE, *Medical College of Wisconsin*

9:25 **"Inference from bits and pieces of information: Collaborative statistical learning** [Abstract 28]

Anjishnu BANERJEE, *Medical College of Wisconsin*
Hengrui HU, *MCW*

9:50 **Bayesian Regularized Multivariate Regression: Unveiling Master Predictors in High-Dimensional Data** [Abstract 143]

Tanujit DEY, *Brigham and Women's Hospital and Harvard Medical School*
Sounak CHAKRABORTY, *University of Missouri*
Priyam DAS, *Virginia Commonwealth University*

03.M1.C11 Contributed Session 11

Venue: DOE CL 3

Chair : Rajesh G, Cochin University of Science and Technology

9:00 **A neural network approach to modelling count time series data** [Abstract 262]

Divya KUTTENCHALIL ANDREWS, *Cochin University of Science and Technology, Kochi, India.*
Narayana BALAKRISHNA, *Indian Institute of Technology, Tirupati, India; Cochin University of Science and Technology, Kochi, India.*

9:15 **CUSUM Control Chart for Monitoring Mean of the INAR(1)PCJ Process.** [Abstract 376]

Aswathy S, *Cochin University of Science and Technology*
Irshad M. R. ,

9:30 **Parsimonious Wavelet-Based Models for Some Periodic Financial Time Series** [Abstract 126]

Rhea DAVIS, *Cochin University of Science and Technology, Kochi*
N BALAKRISHNA, *Indian Institute of Technology, Tirupati*

9:45 **Bivariate Cumulative Residual Entropy of Equilibrium Distribution of Order n** [Abstract 155]

Rajesh G, *Cochin University of Science and Technology*
Unnikrishnan NAIR N, *Cochin University of Science and Technology*
Sajily V.S., *Cochin University of Science and Technology*

Coffee Break 10:20 - 10:40

Panel Discussion 2 Opportunities for Statisticians and other Data/AI/ML Scientists in the U.S. and India - A Panel Discussion

Venue: SC Main Hall

Organizer : Sowmya R. RAO, Boston University School of Public Health

- **Mahesh V IYER**, BMS
- **Sayantee JANA**, IIT Hyderabad
- **Madhu MAZUMDAR**, Icahn School of Medicine at Mount Sinai
- **Amarjot KAUR**, Merck Research Labs
- **Subho MAJUMDAR**, Co-founder and Head of AI, Vijil

03.M2.I48 Topics in spatial statistics

Venue: SC Seminar Hall

Chair : Sounak CHAKRABORTY, University of Missouri
Organizer : Debashis MONDAL, Washington University in St Louis

10:40 Spatial Regression: The Curious Case of Negative Spatial Dependence [Abstract 241]

Malabika KOLEY, *IIT, Kanpur*
Yu-Hsien KAO,
Anil BERA,

11:05 Robust Tests for Latent and Simultaneous Spatial Autoregressive Tobit Model with Spatial Autoregressive Disturbances [Abstract 46]

Anil BERA, *University of Illinois at Urbana–Champaign*
Chang LU,

11:30 Flexible Modeling of Nonstationary Extremal Dependence Using Spatially-Fused LASSO and Ridge Penalties [Abstract 199]

Arnab HAZRA, *Indian Institute of Technology Kanpur*
Xuanjie SHAO, *King Abdullah University of Science and Technology*
Jordan RICHARDS, *University of Edinburgh*
Raphael HUSER, *King Abdullah University of Science and Technology*

03.M2.I49 Innovative Methods for Complex Statistical Problems

Venue: SC Executive Hall

Chair : Sanjay CHAUDHURI, University of Nebraska-Lincoln
Organizer : Sandip SINHARAY, Educational Testing Service

10:40 Estimation Based on Progressively Type II Hybrid Censored Data [Abstract 74]

Manoj CHACKO, *University of Kerala*

11:05 TBD [Abstract 107]

Bertrand CLARKE, *University of Nebraska-Lincoln*

11:30 On the Properties of the Gradient Function of Log-empirical Likelihoods. [Abstract 100]

Sanjay CHAUDHURI, *University of Nebraska-Lincoln*

03.M2.I50 Recent trends in statistical inference

Venue: DOE Auditorium

Chair : Pragya SUR, Harvard University
Organizer : Rajarshi MUKHERJEE, Harvard T.H. Chan School of Public Health

10:40 Nuisance Function Tuning for Optimal Doubly Robust Estimation [Abstract 300]

Rajarshi MUKHERJEE, *Harvard T.H. Chan School of Public Health*

11:05 A general theory for robust clustering: from initialization to mislabeling minimization [Abstract 212]

Soham JANA, *University of Notre Dame*
Jianqing FAN, *Princeton University*
Sanjeev KULKARNI, *Princeton University*

11:30 Identifying arbitrary transformation between the slopes in functional regression [Abstract 145]

Subhra Sankar DHAR, *IIT Kanpur, India*
Pratim GUHA NIYOGI, *Johns Hopkins Bloomberg School of Public Health*

03.M2.I51 Advanced Bayesian multivariate methods

Venue: DOE CL 1

Chair and Organizer : Sameer DESHPANDE, University of Wisconsin-Madison

10:40 Deep Kernel Posterior Learning under Infinite Variance Prior Weights [Abstract 48]

Anindya BHADRA, *Purdue University*
Jorge LORIA, *Aalto University*

11:05 Horseshoe-type Priors for Independent Component Estimation [Abstract 122]

Jyotishka DATTA, *Virginia Polytechnic Institute and State University*
Nicholas POLSON, *University of Chicago Booth School of Business*

11:30 Scalable piece-wise smoothing in high-dimensions with BART [Abstract 133]

Sameer DESHPANDE, *University of Wisconsin-Madison*
Ryan YEE, *University of Wisconsin–Madison*
Soham GHOSH, *University of Wisconsin–Madison*

03.M2.I52 Test for Independence

Venue: DOE CL 2

Chair : Asha GOPALAKRISHNAN, Cochin University of Science & Technology
Organizer : Sudheesh KATTUMANNIL, Indian Statistical Institute, Chennai

10:40 Testing Independence of a pair of random variables - one discrete, another continuous [Abstract 135]

Isha DEWAN, *Indian Statistical Institute*

11:05 Modified semi-distance correlation and its use in testing indepndence [Abstract 233]

Sudheesh KATTUMANNIL, *Indian Statistical Institute, Chennai*

11:30 Transformed jackknife empirical likelihood inference for Bergsma's Covariance Coefficient [Abstract 51]

Deepesh BHATI, *Central University of Rajasthan*
Sudheesh KATTUMANNIL, *ISI Chennai*
Isha DEWAN, *ISI Delhi*

03.M2.I53 Inference for high-dimensional data

Venue: DOE CL 3

Chair : Sarbojit ROY, King Abdullah University of Science and Technology
Organizer : Ursula U. MÜLLER, Texas A&M University

10:40 Bayesian Semi-supervised Inference via a Debiased Modeling Approach [Abstract 78]

Abhishek CHAKRABORTTY, *Texas A&M University*

11:05 Clustering high-dimensional data [Abstract 349]

Shahina RAHMAN, *Amazon LLC*

11:30 Penalized regression with multiple loss functions and variable selection by voting [Abstract 309]

Ursula U. MÜLLER, *Texas A&M University*

Lunch 12:00 - 13:30

Bahadur Memorial Lecture Probal Chaudhuri

Venue: SC Main Hall

Chair : Subhashis GHOSHAL, North Carolina State University

13:30 On Bahadur Representation of Quantiles [Abstract 99]

Probal CHAUDHURI, *Indian Statistical Institute, Kolkata*

Short Break 14:30 - 14:40

03.A1.I54 Recent advances in time series Analysis

Venue: SC Main Hall

Chair : Anindya BHADRA, Purdue University

Organizer : Suhasini SUBBA RAO, Texas A&M

14:40 Interpretable classification of high-dimensional time series in spectral domain [Abstract 370]

Sarbojit ROY, *King Abdullah University of Science and Technology*
Malik SULTAN, *King Abdullah University of Science and Technology*
Hernando OMBAO, *King Abdullah University of Science and Technology*

15:05 A statistical framework for analyzing shape in a time series of random geometric objects [Abstract 456]

Anne VAN DELFT, *Columbia University*
Anne VAN DELFT, *Columbia University*
Andrew J. BLUMBERG, *Columbia University*

15:30 Quasi-maximum likelihood estimation for causal time series [Abstract 39]

Jean-Marc BARDET, *SAMM, Université Paris 1 Panthéon-Sorbonne, France*

03.A1.I55 Design and Learning

Venue: SC Seminar Hall

Chair : Satya SINGH, Indian Institute of Technology Kanpur

Organizer : Abhyuday MANDAL, University of Georgia

14:40 Symmetric Ordering Factorial Experiments [Abstract 120]

Anindita DATTA, *ICAR IASRI, New Delhi*

15:05 Quantiles for Control: Robust inference in Deep Networks [Abstract 390]

Snehanshu SAHA, *APPCAIR and Dept. of CSIS, BITS Pilani Goa Campus*
Aditya CHALLA,
Soma DHAVALA,

15:30 A two-stage design for A/B testing with unknown network interference [Abstract 417]

Srijan SENGUPTA, *North Carolina State University*

03.A1.I56 Statistical Methods in Child Health and Development Studies

Venue: SC Executive Hall

Chair : Asha GOPALAKRISHNAN, Cochin University of Science & Technology

Organizer : Vishal DEO, ICMR-National Institute for Research in Digital Health and Data Science, New Delhi

14:40 Mathematical Characterization of Growth of Physical Parameters in the Human Developmental Age: Growth Curve Analysis of an Indian Birth Cohort [Abstract 132]

Vishal DEO, *ICMR-National Institute for Research in Digital Health and Data Science, New Delhi*
Saumyadipta PYNE, *Department of Statistics and Applied Probability, University of California Santa Barbara, CA, USA*
Meghana RAY, *Health Analytics Network, Pittsburgh, PA, USA*

15:05 Intersectional Inequalities in Anthropometric Failure among Indian Children: Evidence from the National Family Health Survey (2015-2016) [Abstract 8]

Tulsi ADHIKARI, *Indian Council of Medical Research*
Niharika TRIPATHI, *Indraprastha College for Women University of Delhi (North Campus)*

15:30 An Analytical Framework to Combine Mortality and Morbidity to Study Determinants of Child Health [Abstract 101]

Anuradha Rajkonwar CHETIYA, *Ramjas College, University of Delhi*
Vishal DEO , *ICMR-NIRDH&DS, New Delhi*

03.A1.I57 Advancements in Applied Probability and Statistical Modeling

Venue: DOE Auditorium

Chair : Arka GHOSH, Iowa State University
Organizer : Vivekananda ROY, Iowa State University

14:40 Bernoulli Spiders [Abstract 25]

Srinivasan BALAJI, *George Washington University*
Hosam MAHMOUD, *George Washington University*

15:05 Interval-Valued Data Analysis with Applications [Abstract 263]

Arnab LAHA, *Indian Institute of Management Ahmedabad*
MAHESH, K. C.,

15:30 Heterogeneous Graphon JSQ(d) model [Abstract 165]

Arka GHOSH, *Iowa State University*
Ruoyu Wu - Iowa State University (United States) YAN-HAN CHEN - IOWA STATE UNIVERSITY (UNITED STATES),

03.A1.I58 Recent Advancements in Theoretical Statistics and Machine Learning

Venue: DOE CL 1

Chair : Pallavi BASU, Indian School of Business
Organizer : Somabha MUKHERJEE, National University of Singapore

14:40 Testing Informative Intra-Cluster Group Sizes in Clustered Data [Abstract 465]

Hasika WICKRAMA SENEVIRATHNE, *National University of Singapore*
Sandipan DUTTA, *Old Dominion University*

15:05 Stabilizing Non normal matrices and Davies Conjecture [Abstract 301]

Satyaki MUKHERJEE, *National University of Singapore*
Jess BANKS, *University of California, Berkeley*
Archit KULKARNI,
Nikhil SRIVASTAVA, *University of California, Berkeley*

15:30 **Robustness and Overparameterization** [Abstract 441]

Piyush SRIVASTAVA, *Tata Institute of Fundamental Research*
Santanu DAS, *Tata Institute of Fundamental Research*
Jatin BATRA, *Tata Institute of Fundamental Research*

03.A1.I59 Data Science in Sports

Venue: DOE CL 2

Chair : Rishideep ROY, University of Essex
Organizer : Soudeep DEB, Indian Institute of Management Bangalore

14:40 **Data Science and its future in emerging sports and markets** [Abstract 400]

Subrat SARANGI, *MICA*
Subrat SARANGI,

15:05 **Optimizing Prices of Tickets and Advertisement in Sports Series with Variable Game Counts** [Abstract 115]

Shubhabrata DAS, *Indian Institute of Management Bangalore*

15:30 **Analyzing soccer shots with spatial latent trait models** [Abstract 294]

Debashis MONDAL, *Washington University in St Louis*
Sayan DAS, *Washington University in St Louis*

03.A1.I60 New Directions in Statistical Modeling and Inference: Predictive Inference, Likelihood-Based Inference, and Nonparametric Bayes

Venue: DOE CL 3

Chair and Organizer : Jyotishka DATTA, Virginia Polytechnic Institute and State University

14:40 **Empirical Bayes Estimation with Side Information** [Abstract 298]

Gourab MUKHERJEE, *University of Southern California*
Jiajun LUO,
Trambak BANERJEE,
Wenguang SUN,

15:05 **Personalized Medicine: Adaptive Randomization in SMART Designs** [Abstract 171]

Palash GHOSH, *IIT Guwahati*

15:30 **Bivariate Exponential Distribution through Entropy Optimization** [Abstract 448]

Princy T, *Cochin University of Science and Technology*
Sneha BABU, *Cochin University of Science and Technology*

03.A1.C12 Contributed Session 12

Venue: DOE CL 4

Chair : Aritra SAHA, Indian Statistical Institute, Kolkata

14:40 **Nonparametric estimation of doubly truncated stress-strength reliability** [Abstract 245]

DEEPA K R, *MG UNIVERSITY KOTTAYAM*
ANGEL MATHEW, *MAHATMA GANDHI UNIVERSITY, KOTTAYAM*

14:55 **Predicting patient mortality in the presence of competing risks** [Abstract 60]

Amol Ajit BHAVE, *S&P Global*
Sandip Kumar PAL, *S&P Global*

15:10 Analyzing Progressive Censoring in k out of n Load Sharing Systems under Exponential Distribution [Abstract 283]

Anil MAURYA, *Sardar Vallabhbhai National Institute Technology*
Anil MAURYA, *Sardar Vallabhbhai National Institute of Technology, Surat, 395007, Gujarat, India*
Raj Kamal MAURYA, *Sardar Vallabhbhai National Institute of Technology, Surat, 395007, Gujarat, India*

15:25 Testing Exponentiality for Progressively Type II Censored Data Using Equilibrium Distributions [Abstract 464]

Sajily V S, *Cochin University of Science and Technology*
Rajesh G, *Cochin University of Science and Technology*

15:40 A Class of tests for trend change in hazard rate function with random right censored data [Abstract 385]

Aritra SAHA, *Indian Statistical Institute, Kolkata*
Md. Zafar ANIS, *Indian Statistical Institute, Kolkata, India*

Coffee Break 16:00 - 16:30

03.E1.I61 Recent Advances in Causal Inference and Observational Studies *Venue:* SC Main Hall

Chair and Organizer : Samrat ROY, Indian Institute of Management Ahmedabad

16:30 A marginal structural model for partial compliance in SMARTs [Abstract 56]

Indrabati BHATTACHARYA, *Florida State University*
William J. ARTMAN, *University of Rochester*
Ashkan ERTEFAIE, *University of Rochester*
Brent A. JOHNSON, *University of Rochester*

16:55 Inferring the Effect of a Confounded Treatment by Calibrating Resistant Population's Variance [Abstract 230]

Bikram KARMAKAR, *University of Florida*
Zikun QIN, *University of Florida*
Bikram KARMAKAR, *University of Florida*

17:20 Quasi-randomization tests for network interference [Abstract 42]

Pallavi BASU, *Indian School of Business*
Supriya TIWARI, *ISB*
Pallavi BASU, *ISB*

03.E1.I62 Analysis of Statistical Methods for Complex Data *Venue:* SC Seminar Hall

Chair and Organizer : Sharmodeep BHATTACHARYYA, Oregon State University

16:30 Analyzing Political Polarization in the Presence of Partially Observed Social Networks [Abstract 59]

Sharmodeep BHATTACHARYYA, *Oregon State University*

16:55 statistical theory of deep neural network: estimating composite functions to solving inverse problem via physics informed neural networks [Abstract 297]

Debarghya MUKHERJEE, *Boston University*

17:20 **On differentially private U statistics** [Abstract 405]

Purnamrita SARKAR, *UT Austin*
Shourya PANDEY, *UT Austin*
Po-Ling LOH, *Cambridge*
Kamalika CHAUDHURI, *UCSD/Meta*

03.E1.C13 Contributed Session 13

Venue: SC Executive Hall

Chair : Rik GHOSH, Indian Institute of Technology, Guwahati

16:30 **Batting strike rate inTwenty20 and One Day International Cricket: A comparison** [Abstract 430]

RK Renin SINGH, *Chitkara University*
Subrat SARANGI, *MICA*
Dhawal JADHAV, *MICA*

16:45 **Sea Root Clustering / RouteFinder : A Novel Approach to Estimate Vessels Route by Trajectory Clustering based on AIS Data** [Abstract 221]

Patra JYOTIRMOY, *S&P Global*
Shubham S, *S&P Global*
Sandip PAL,

17:00 **Matrix-Variate K-Means: An Extended K-Means Approach using Euclidean and Mahalanobis Distances** [Abstract 261]

Shiva Kumar KURVA, *Pondicherry University*
Kiruthika C, *Pondicherry University*

17:15 **The Reversed Aging Intensity Function- A Quantile Approach** [Abstract 342]

ANJALI P R, *Research Scholar, Department of Statistics, University of Calicut*
Dileepkumar M, *Assistant Professor, Department of Statistics, University of Calicut*

17:30 **Optimal Adaptive Strategies in SMART** [Abstract 172]

Rik GHOSH, *Indian Institute of Technology, Guwahati*
Bibhas CHAKRABORTY, *Centre for Quantitative Medicine, Duke-NUS Medical School, National University of Singapore, Singapore*
Inbal, Megan E. NAHUM-SHANI, PATRICK, *Institute for Social Research, University of Michigan*
Palash GHOSH, *Indian Institute of Technology Guwahati, Assam*

03.E1.C14 Contributed Session 14

Venue: DOE Auditorium

Chair : John Olutunji OLAOMI, University of South Africa, Department of Statistics

16:30 **On the Computational Complexity of Private High-dimensional Model Selection** [Abstract 369]

Saptarshi ROY, *University of Texas, Austin*
Zehua WANG, *University of Michigan, Ann Arbor*
Ambuj TEWARI, *University of Michigan, Ann Arbor*

16:45 **Two sample test of high-dimensional means test under missing observations** [Abstract 468]

Shiv Kumar Yadav YADAV, *Indian Institute if Bombay*

17:00 **Estimating Shannon Entropy of the Selected Population** [Abstract 281]

Masihuddin MASIHUDDIN, *Indian Institute of Information Technology Guwahati*
Neeraj MISRA, *IIT Kanpur*

17:15 **Study on fractional order entropy from a decision-making perspective** [**Abstract 335**]

Poulami PAUL, *Rajiv Gandhi Institute of Petroleum Technology, Jais*

17:30 **STOCHASTIC MODELING ON RAINFALL VARIABILITY IN NORTHERN NIGERIA** [**Abstract 319**]

John Olutunji OLAOMI, *University of South Africa, Department of Statistics*
James Babatunde EHIMONY, *Department of Statistics, University of South Africa, and Department of Statistics, Kogi State Polytechnic, Lokoja, Nigeria*

03.E1.C15 Contributed Session 15

Venue: DOE CL 1

Chair : JINTO E G, Regional Cancer Centre, Thiruvananthapuram, Kerala

16:30 **Mathematical Modeling of Malaria Transmission in Mumbai** [**Abstract 437**]

Adithya SOMARAJ, *National Disease Modeling Consortium, IIT Bombay*
Usha ANANTHAKUMAR, *Shailesh J. Mehta School of Management, IIT Bombay*

16:45 **MODELLING AND COMPARISON OF POTENTIAL EVAPOTRANSPIRATION IN TIRUCHIRAPALLI USING STATISTICAL TECHNIQUES** [**Abstract 3**]

ARCHANA A, *ICAR-IASRI, NEW DELHI*
RADHA M, *Assistant Professor, Department of Agricultural Economics, TNAU, Madurai-625104, Tamil Nadu, India*

17:00 **Beyond Single Variants: Exploring Gene-Gene Interactions and Their Impact on Disease Risk Prediction** [**Abstract 123**]

Kallol DATTA, *National Institute of Biomedical Genomics, Kalyani*
Samsiddhi BHATTACHARJEE, *National Institute of Biomedical Genomics, Kalyani*

17:15 **Inference on Overlap Index: With an Application to Cancer Data** [**Abstract 141**]

RAJU DEY, *Indian Institute of Technology Kharagpur*
Arne C. BATHKE, *Intelligent Data Analytics (IDA) Lab Salzburg, Department of Artificial Intelligence and Human Interfaces*
Somesh KUMAR, *Indian Institute of Technology Kharagpur, Department of Mathematics*

17:30 **Forecasting Lung Cancer Incidence in South Kerala, India using Bootstrap Methods** [**Abstract 154**]

JINTO E G, *Regional Cancer Centre, Thiruvananthapuram, Kerala*
Aleyamma MATHEW, *Professor and Head Division of Cancer Epidemiology & Biostatistics Regional Cancer Centre, Thiruvananthapuram*
Preethi Sara GEORGE, *Additional Professor In Biostatistics Division of Cancer Epidemiology & Biostatistics Regional Cancer Centre, Thiruvananthapuram*

03.E1.C16 Contributed Session 16

Venue: DOE CL 2

Chair : Jyoti J PRAJAPATI, Department of Mathematics, Institute of Chemical Technology, Mumbai

16:30 **On Bootstrapping Lasso in Generalized Linear Models and the Cross Validation** [**Abstract 104**]

Mayukh CHOUDHURY, *Department of Mathematics, IIT Bombay*
Debraj DAS, *Department of Mathematics, IIT Bombay*

16:45 **MCMC Importance Sampling via Moreau-Yosida Envelopes** [Abstract 423]

Apratim SHUKLA, *Indian Institute of Technology Kanpur*
Dootika VATS, *Indian Institute of Technology Kanpur*
Eric CHI, *Rice University*

17:00 **Computationally efficient Bayesian joint modeling of mixed-type multivariate spatial data** [Abstract 295]

Arghya MUKHERJEE, *Indian Institute of Technology Kanpur*
Arnab HAZRA, *Indian Institute of Technology Kanpur*
Dootika VATS, *Indian Institute of Technology Kanpur*

17:15 **Identifying key drivers of extinction for Chitala populations: Data-driven insights from an intraguild predation model using a Bayesian framework** [Abstract 286]

Dipali V. MESTRY, *Department of Mathematics, Institute of Chemical Technology, Mumbai*

17:30 **Towards Strategic Management for Controlling Invasive Alien Plants in India: Integrating Ensemble Based Species Distribution Models** [Abstract 344]

Jyoti J PRAJAPATI, *Department of Mathematics, Institute of Chemical Technology, Mumbai*

03.E1.C17 Contributed Session 17

Venue: DOE CL 3

Chair : Soumadeb PAIN, Indian Institute of Technology Kanpur

16:30 **A generalized dynamic spatio-temporal model for panel count data** [Abstract 422]

Shrinivas SHIRKE, *Department of Statistics , Savitribai Phule Pune University*

16:45 **Spatial Modeling of Residential Property Launch Prices in Bengaluru, India** [Abstract 192]

Kapil GUPTA, *IIM Bangalore*
Soudeep DEB,

17:00 **Rotation and Translation Invariant Monitoring of Shape and Size in Image Data: An Application in Satellite Imaging** [Abstract 366]

Anik ROY, *Indian Statistical Institute, Kolkata*
Partha Sarathi MUKHERJEE, *Indian Statistical Institute, Kolkata*

17:15 **Multiscale Geographically Weighted Regression Modeling Approach for Assessing the Localized Spatial Effects of Lung Cancer Incidence** [Abstract 380]

SRUTHI S, *Regional Cancer Centre, Thiruvananthapuram, Kerala*
Preethi Sara GEORGE, *Additional proferssor, Division of Cancer Epidemiology & Biostatistics, Regional Cancer Centre, Thiruvananthapuram*
Aleyamma MATHEW, *Professor and Head, Division of Cancer Epidemiology & Biostatistics, Regional Cancer Centre, Thiruvananthapuram*

17:30 **Optimal cohort Stepped Wedge design for unequal cluster size** [Abstract 323]

Soumadeb PAIN, *Indian Institute of Technology Kanpur*
Satya Prakash SINGH,

03.E1.C18 Contributed Session 18

Venue: DOE CL 4

Chair : FATHIMA K, University of Kerala

16:30 **Analysis of two-component load-sharing systems in the presence of frailty** [**Abstract 65**]

Shilpi BISWAS, *IIT Guwahati*
Debanjan MITRA, *IIM Udaipur, Rajasthan*
Ayon GANGULY, *IIT Guwahati, Assam*

16:45 **Statistical inference of a competing risks model on improved adaptive type-II progressive censored data** [**Abstract 136**]

Amlan DEY, *Scholar*

17:00 **New characterization driven test for Pareto family** [**Abstract 237**]

Sakshi KHANDELWAL, *Central University of Rajasthan*
Deepesh BHATI, *Central University of Rajasthan*

17:15 **Redundancy Allocation for Series and Parallel Systems: A Copula-based Approach** [**Abstract 253**]

Ravi KUMAR, *IIT Hyderabad*

17:30 **A Study on some new Reliability Measures in the Univariate Case** [**Abstract 223**]

FATHIMA K, *University of Kerala*
E I Abdul SATHAR, *University of Kerala*

Banquet and Awards Ceremony 19:00 - 22:00 Venue: J. W. Marriott Hotel, Kochi

Monday December 30

Short Course 2 **Interpretable and Interactive Machine Learning** *Venue:* DSC Seminar Hall

Time : **9:30 - 12:30**

- **Kris SANKARAN**, University of Wisconsin-Madison

Special Invited Session 3 **Debashis Paul, Anish Sarkar** *Venue:* SC Main Hall

Chair : Susmita DATTA, University of Florida

09:00 **Spectra of sequential sample covariance matrices with application to detecting structural changes** [**Abstract** 333]

Debashis PAUL, *Indian Statistical Institute, Kolkata*
Nina DÖRNEMANN, *Aarhus University*

09:40 **First Collision Time of simple symmetric random walks** [**Abstract** 403]

Anish SARKAR, *Indian Statistical Institute*

04.M1.I63 **Bayesian methods in genetics and privacy** *Venue:* SC Seminar Hall

Chair : Subharup GUHA, University of Florida
Organizer : Steven MACEACHERN, The Ohio State University

9:00 **Exact MCMC for Bayesian Inference for Privatized Data** [**Abstract** 356]

Vinayak RAO, *Purdue University*
Nianqiao JU, *Purdue University*
Jordan AWAN, *Purdue University*
Ruobin GONG, *Rutgers University*

9:25 **Nonparametric Bayes Differential Analysis of Multigroup DNA Methylation Data** [**Abstract** 188]

Subharup GUHA, *University of Florida*
Chiyu GU, *Bayer Crop Science*
Veerabhadran BALADANDAYUTHAPANI, *University of Michigan*
Subharup GUHA, *University of Florida*

04.M1.I64 **Theory and Computational Designs** *Venue:* SC Executive Hall

Chair : Yasmeen AKHTAR, Birla Institute of Technology and Science, Pilani-Goa Campus
Organizer : Abhyuday MANDAL, University of Georgia

9:00 **Some game theoretic ideas in the design of experiments** [**Abstract** 125]

Ori DAVIDOV, *University of Haifa*

9:25 **Combinatorial Testing using Locating Arrays** [**Abstract** 9]

Yasmeen AKHTAR, *Birla Institute of Technology and Science, Pilani-Goa Campus*

9:50 **Credible Distributions of Overall Ranking of Entities** [**Abstract** 278]

Abhyuday MANDAL, *University of Georgia*

04.M1.I65 Advances in statistical machine learning

Venue: DOE Auditorium

Chair : Ashwin PANANJADY, Georgia Institute of Technology
Organizer : Po-Ling LOH, University of Cambridge

9:00 **Optimal Iterative Algorithms for Structured PCA with Invariant Noise** [Abstract 150]

Rishabh DUDEJA, *UW Madison*
Songbin LIU, *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*
Junjie MA, *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*

9:25 **Graph Structure Learning from DeGroot Dynamics** [Abstract 118]

Gautam DASARATHY, *Arizona State University*
Vignesh TIRUKKONDA, *Arizona State University*

9:50 **Le Cam meets Turing: Computationally efficient reductions between some statistical models** [Abstract 328]

Ashwin PANANJADY, *Georgia Institute of Technology*

04.M1.I66 Pushing the Boundaries of Statistical Analysis: New Methods for Complex Data

Venue: DOE CL 1

Chair : Subarna BHATTACHARJEE, Department of Mathematics, Ravenshaw University, Cuttack
Organizer : Tanujit DEY, Brigham and Women's Hospital and Harvard Medical School

9:00 **Detecting localized dependence in bivariate datasets** [Abstract 365]

Angshuman ROY, *Indian Institute of Technology Tirupati*
Angshuman ROY, *Indian Institute of Technology*
Sagnik DAS, *Indian Institute of Science Education and Research Kolkata*

9:25 **Statistical hydrology in the age of AI** [Abstract 157]

Auroop GANGULY, *Northeastern University*

9:50 **On weighted failure rate: an alternative approach to study weighted distributions** [Abstract 54]

Subarna BHATTACHARJEE, *Department of Mathematics, Ravenshaw University, Cuttack*
S. M. SUNOJ, *Cochin University of Science and Technology*

04.M1.I67 Modeling the Untangling Threads: Unveiling Spatial, Behavioral, and Cure Rate Effects in Survival Data

Venue: DOE CL 2

Chair : Prajamitra BHUYAN, Indian Institute of Management Calcutta
Organizer : Tanujit DEY, Brigham and Women's Hospital and Harvard Medical School

9:00 **Functional proportional hazards mixture cure model** [Abstract 164]

Rahul GHOSAL, *University of South Carolina*
Marcos MATABUENA, *Harvard University*
Jiajia ZHANG, *University of South Carolina*

9:25 **Estimation of Population Size with Heterogeneous Catchability and Behavioural Dependence: Applications to Air and Water Borne Disease Surveillance** [Abstract 62]

Prajamitra BHUYAN, *Indian Institute of Management Calcutta*
Kiranmoy CHATTERJEE, *Bidhannagar College, Kolkata*

04.M1.I68 Models for Count Time series

Venue: DOE CL 3

Chair : T. V. RAMANATHAN, Savitribai Phule Pune University & Plaksha University
Organizer : N. BALAKRISHNA, Indian Institute of Technology Tirupati

9:00 **Estimation for parameter-driven count time series** [Abstract 310]

N. BALAKRISHNA, *Indian Institute of Technology Tirupati*

9:25 **A BIVARIATE INTEGER-VALUED BILINEAR AUTOREGRESSIVE MODEL WITH RANDOM COEFFICIENTS** [Abstract 224]

Jayakumar K, *Senior Professor, Department of Statistics, University of Calicut*

9:50 **Integer Autoregressive (INAR) Models and their Applications** [Abstract 21]

Manik AWALE, *Savitribai Phule Pune University*

Coffee Break 10:20 - 10:40

04.M2.I69 Statistical Inverse Problems

Venue: SC Main Hall

Chair : Anuj ABHISHEK, Case Western Reserve University
Organizer : Sakshi ARYA, Case Western Reserve University

10:40 **Stability of nonlinear filters: numerical and theoretical results** [Abstract 15]

Amit APTE, *Indian Institute of Science Education and Research IISER Pune*
Anugu SUMITH REDDY, *Rice University*
Shashank ROY, *IMT Atlantique*
Pinak MANDAL, *University of Sydney*

11:05 **Regularization Schemes for Statistical Inverse Problems** [Abstract 357]

Abhishake RASTOGI, *LUT university, Finland*

11:30 **Local reconstruction analysis of inverting the Radon transform in the plane from noisy discrete data** [Abstract 5]

Anuj ABHISHEK, *Case Western Reserve University*
Alexander KATSEVICH, *University of Central Florida*
James WEBBER, *Cleveland Clinic*

04.M2.I70 Bayesian non-parametrics

Venue: SC Seminar Hall

Chair : Siddhartha CHIB, Washington University in Saint Louis
Organizer : Christian ROBERT, Université Paris Dauphine PSL & University of Warwick

10:40 **Investigating the price determinants of the European Carbon Trading System: a non-parametric approach** [Abstract 287]

Antonietta MIRA, *Università della Svizzera italiana and Insubria University*
Cristiano SALVAGNIN, *University of Brescia*
Aldo GLIELMO, *Team, Directorate General for IT, Banca d'Italia*
Maria Elena DE GIULI, *University of Pavia*

11:05 **On On higher order approximation of Bayesian procedures through empirical Bayes** [Abstract 338]

Sonia PETRONE, *Bocconi University, Milan. Italy*
Stefano RIZZELLI, *University of Padova*
Judith ROUSSEAU, *CEREMADE, Université Paris Dauphine and Oxford University*

11:30 Robust Probabilistic Inference via a Constrained Transport Metric [Abstract 332]

Debdeep PATI, *University of Wisconsin-Madison*
Abhisek CHAKRABORTY, *Eli Lilly and Company*
Anirban BHATTACHARYA, *Texas A&M University*

04.M2.I71 Modeling Complex Biological and Medical Data

Venue: SC Executive Hall

Chair and Organizer : Shibasish DASGUPTA, Pfizer

10:40 Optimizing patient eligibility criteria in clinical trials using real world evidence (RWE) [Abstract 53]

Abhishek BHATTACHARJEE, *Pfizer Research and Development, Pfizer*

11:05 A Bayesian quantile joint modeling of multivariate longitudinal and time-to-event data [Abstract 111]

Kiranmoy DAS, *Indian Statistical Institute, Kolkata*
Damitri KUNDU, *Intuit, Bangalore, India*
Shekhar KRISHNAN, *Tata Medical Center, Kolkata*
Manash Pratim GOGOI, *Tata Medical Center, Kolkata*

11:30 Going beyond gene-based inferences: identification of cell-cell communications for complex biological systems using a Bayesian Tweedie Model [Abstract 124]

Susmita DATTA, *University of Florida*
Dongyuan WU ,
Jeremy T. GASKINS, *University of Louisville*
Michael SEKULA, *University of Louisville*

04.M2.I72 Emprical Likelihood Inference for inequality measures

Venue: DOE Auditorium

Chair and Organizer : Sudheesh KATTUMANNIL, Indian Statistical Institute, Chennai

10:40 Jackknife empirical likelihood ratio test for testing the equality of semivariance [Abstract 445]

Saparya SURESH, *Indian Institute of Management, Kozhikode*

11:05 Jackknife empirical likelihood inference of risks measures [Abstract 315]

Rupel NARGUNAM, *Madras School of Economics*

04.M2.I73 Recent Advances in Sequential Methodologies

Venue: DOE CL 1

Chair : Shyamsundar SAHOO, Department of Statistics, Haldia Government College
Organizer : Swarnali BANERJEE, Loyola University Chicago

10:40 Adaptive sampling strategies for estimating the parameter of an Inverse Maxwell Distribution [Abstract 218]

Neeraj JOSHI, *Indian Institute of Technology Delhi*

11:05 A New Formulation of Minimum Risk Fixed-Size Confidence Region (MRFSCR) Estimation Problems for a Multivariate Normal Mean with Illustrations Incorporating Covariance Structure via Positive Definite Matrix [Abstract 459]

Swathi VENKATESAN, *Fairfield University*
Nitis MUKHOPADHYAY, *University of Connecticut*

11:30 Best Treatment Selection: Application to best Digital Experience selection [Abstract 96]

Bhargab CHATTOPADHYAY, *Indian Institute of Technology Jodhpur*

04.M2.I74 Statistical Modeling

Venue: DOE CL 2

Chair : Manik AWALE, Savitribai Phule Pune University
Organizer : Santosh SUTAR, Shivaji University, Kolhapur, India

10:40 **Modelling the Predictability of categorical response variable through Non-Symmetric Correspondence Analysis** [Abstract 229]

Kirtee KAMALJA, *Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India*
Riya JAIN, *Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India*

11:05 **Modeling Marriage Rate Fluctuations Using Survival Analysis Approach** [Abstract 231]

Akanksha KASHIKAR, *Department of Statistics, Savitribai Phule Pune University*
Neela GULANIKAR, *Dept of Statistics, Savitribai Phule Pune University*

11:30 **Reliability Modeling of Unequal Load Sharing Systems Using the Accelerated Failure Time Model** [Abstract 447]

Santosh SUTAR, *Shivaji University, Kolhapur, India*
Sukumar RAJGURU, *Shivaji University, Kolhapur*

04.M2.I75 Recent Advances in Quantile Regression and Applications

Venue: DOE CL 3

Chair : Amiya Ranjan BHOWMICK, Institute of Chemical Technology, Mumbai
Organizer : Kaushik JANA, Ahmedabad University

10:40 **Statistical modeling and mitigation techniques for sensor data imputation** [Abstract 424]

Sharvari SHUKLA, *Symbiosis Statistical Institute, Symbiosis International (Deemed University)*

11:05 **Quantile regression for high-dimensional data with uniform-over-dimension convergence** [Abstract 105]

Joydeep CHOWDHURY, *King Abdullah University of Science and Technology*

11:30 **Nonparametric quantile regression for time series with replicated observations and its application to climate data** [Abstract 209]

Kaushik JANA, *Ahmedabad University*
Soudeep DEB, *Indian Institute of Management, Bangalore*

04.M2.I76 Statistical learning theory

Venue: DOE CL 4

Chair : Vidya MUTHUKUMAR, Georgia Institute of Technology
Organizer : Po-Ling LOH, University of Cambridge

10:40 **On the Statistical Complexity of Sample Amplification: Increasing Dataset Size even when Learning is Impossible** [Abstract 418]

Vatsal SHARAN, *University of Southern California*
Brian AXELROD,
Yanjun HAN,
Shivam GARG,

11:05 **Efficient and Elastic LLMs** [Abstract 206]

Prateek JAIN, *Google Research, India*

11:30 **How does the training loss function affect in-distribution and out-of-distribution learning in high dimensional linear models?** [Abstract 308]

Vidya MUTHUKUMAR, *Georgia Institute of Technology*

Lunch 12:00 - 13:30

Special Session 2 P. Unnikrishnan Nair

Venue: SC Main Hall

Chair : S. M. SUNOJ, Cochin University of Science and Technology

13:30 TBA [Abstract 313]

P. Unnikrishnan NAIR, *Cochin University of Science and Technology*

04.L1.C19 Contributed Session 19

Venue: SC Seminar Hall

Chair : Deeksha P M, ICAR-Indian Agriculture Research Institute, New Delhi

13:30 On some properties and applications of the Modi-Frechet distribution [Abstract 320]

AKHILA P, *GOVT. ARTS AND SCIENCE COLLEGE CALICUT, KOZHIKODE, KERALA, INDIA*
GIRISH BABU M, *Govt. Arts and Science College Calicut, Kozhikode, Kerala, India*

13:45 Estimation of reliability in a multicomponent stress-strength model based on power transformed Perks distribution [Abstract 216]

SUDHEEP JOSE, *MADRAS CHRISTIAN COLLEGE*
Venugopal HARIDOSS, *Madras Christian College*
Thomas XAVIER, *Novartis*

14:00 STATISTICAL ANALYSIS OF AREA, PRODUCTION AND PRODUCTIVITY OF MAJOR PULSES IN TELANGANA STATE [Abstract 259]

VANKUDOTH KUMAR, *ICAR-IASRI, NEW DELHI*
A S KAMBLE,

14:15 An autoencoder based semi-supervised approach for accurate binning of metagenomics data [Abstract 270]

Deeksha P M, *ICAR-Indian Agriculture Research Institute, New Delhi*
Shashi Bhushan LAL, *1. ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
Anu SHARMA, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
Dwijesh Chandra MISHRA, *3. ICAR-National Academy of Agricultural Research Management, Hyderabad*

04.L1.C20 Contributed Session 20

Venue: SC Executive Hall

Chair : M.R. IRSHAD, Cochin University of Science and Technology

13:30 Space filling designs based on optimal covariate designs [Abstract 375]

NEETHU RS, *ICAR IASRI*
Cini VARGHESE, *Principal Scientist*
Mohd. HARUN, *Scientist*
Anindita DATTA, *Scientist*

13:45 On a class of AR(1) model for Z valued time series [Abstract 462]

Arathi VINAYAN, *Govt. Arts and science college Calicut, Calicut 673018*
Jilesh V, *Govt . Arts and Science and College Calicut, Calicut 673018*

14:00 Power optimal designs for comparing a set of controls to a set of treatments [Abstract 428]

Arpan SINGH, *IIT Hyderabad*
Satya Prakash SINGH, *IIT Kanpur*

14:15 The Corrected Likelihood Approach for Adjusting Measurement Error in Cox's Model [Abstract 446]

Anu SUSAN GEORGE, *Cochin University of Science and Technology*
Asha GOPALAKRISHNAN, *Cochin University of Science and Technology*

Short Break 14:30 - 14:40

Panel Discussion 3 Breaking Down Barriers and Cultivating a Culture of Collaboration and Leadership in Statistical Sciences: Enhancing Synergy for Collective Growth *Venue:* SC Main Hall
Chair : Sreya SARKAR, University of Iowa
Organizer : Arinjita BHATTACHARYA,

- **Amarjot KAUR**, Merck Research Labs
- **Shanti GOMATAM**,
- **Sesh RAI**,
- **Hiya BANERJEE**, Eli Lilly
- **Asha GOPALAKRISHNAN**, Cochin University of Science & Technology
- **Deokumar SRIVASTAVA**,

04.A1.I77 Statistics in clinical research I *Venue:* SC Seminar Hall
Chair and Organizer : Isha DEWAN, Indian Statistical Institute

14:40 Clinical and public health challenges of emerging infectious disease outbreaks: Complexity of infectious disease modelling for early prediction in India. [Abstract 98]

Himanshu K. CHATURVEDI, *Disease Modelling Division, Indian Council of Medical Research, New Delhi*
Poornima Suryanath SINGH, *Amity Institute of Public Health and Hospital Administration, Amity University, NOIDA,UP, India*

15:05 TBA [Abstract 203]

Mahesh IYER, *BMS*

15:30 Sample size re-assessment in clinical trials using predictive power [Abstract 106]

Solomon CHRISTOPHER, *Emmes*
Jonathan JAEGER, *Jaeger Consulting*

04.A1.I78 Theory and Computational Designs *Venue:* SC Executive Hall
Chair : Srijan SENGUPTA, North Carolina State University
Organizer : Abhyuday MANDAL, University of Georgia

14:40 Analysis of life testing data in presence of competing risks [Abstract 240]

Arnab KOLEY, *IIM Indore*

15:05 On max–min designs for multiple comparisons for Binary outcomes [Abstract 431]

Satya SINGH, *Indian Institute of Technology Kanpur*
Ori DAVIDOV, *University of Haifa Israel*
Dharm SINGH, *Indian Institute of Technology Kanpur*

15:30 Estimation of confidence interval of Cpy under some location scale distributions using fiducial approach [Abstract 13]

M Zafar ANIS, *ISI Calcutta*

04.A1.I79 Recent Developments in Sequential Analysis and Statistical Reliability

Venue: DOE Auditorium

Chair and Organizer : Neeraj JOSHI, Indian Institute of Technology Delhi

14:40 Adaptive estimation using records data under asymmetric loss, with applications [Abstract 416]

Raghu SENGUPTA, *IIT Kanpur*
Saibal CHATTOPADHYAY, *IIM Calcutta*
Neeraj JOSHI, *IIT Delhi*

15:05 Maximum Precision Estimation for a Step-Stress Model Using Two-Stage Methodologies [Abstract 38]

Sudeep BAPAT, *Indian Institute of Technology Bombay*
Yan ZHUANG, *Connecticut College*

15:30 Quantile based control charts for the inverse Pareto distribution with applications [Abstract 97]

Aditi CHATURVEDI, *Sharda University*
Neeraj JOSHI, *Indian Institute of Technology Delhi*
Sudeep BAPAT, *Indian Institute of Technology Bombay*

04.A1.I80 Bayesian computational advances

Venue: DOE CL 1

Chair : Antonietta MIRA, Università della Svizzera italiana and Insubria University
Organizer : Christian ROBERT, Université Paris Dauphine PSL & University of Warwick

14:40 Bayesian Difference in Differences with Extensions to Multiple Periods and Staggered Treatments [Abstract 103]

Siddhartha CHIB, *Washington University in Saint Louis*
Kenichi SHIMIZU, *University of Alberta*

15:05 Bayesian Model Averaging for Implicit Generative Neural Model [Abstract 151]

Ritabrata DUTTA, *University of Warwick*
Sherman KHOO, *University of Bristol*
Shreya SINHA ROY, *University of Warwick*

15:30 Insufficient Gibbs sampling [Abstract 363]

Christian ROBERT, *Université Paris Dauphine PSL & University of Warwick*
Antoine LUCIANO, *Université Paris Dauphine*
Robin RYDER, *Université Paris Dauphine*

04.A1.I81 Frontiers of Statistical and Machine Learning

Venue: DOE CL 2

Chair : Avishek GHOSH, IIT Bombay
Organizer : Sourish DAS, Chennai Mathematical Institute

14:40 A multiplex network approach for characterizing stock market dynamics [Abstract 66]

Soumyajyoti BISWAS, *Assistant Professor, Department of Physics & Department of Computer Science and Engineering, SRM University*

15:05 Competing for space: Self-organization among a group of crowd-avoiding agents [Abstract 454]

Sasidevan V., *Assistant Professor, Dept. of Physics, Cochin University of Science and Technology, Cochin*

15:30 Provable and Efficient Algorithms for Heterogeneous and Byzantine Robust Federated Learning [Abstract 166]

Avishek GHOSH, *IIT Bombay*
Dong YIN, *Research Scientist, Apple AI*
RajKumar MAITY , *Data Scientist, Microsoft*

04.A1.I82 Semiparamertic regression analysis of lifetime data

Venue: DOE CL 3

Chair : Sudheesh KATTUMANNIL, Indian Statistical Institute, Chennai
Organizer : Sreedevi E. P., CUSAT

14:40 Semiparametric regression analysis of mixed recurrent event and panel count data with multiple causes of failure [Abstract 397]

Sankaran PADUTHOL GODAN, *Cochin University of Science and Technology*

15:05 Semiparametric regression analysis of doubly censored recurrent event data [Abstract 438]

Sreedevi E. P., *CUSAT*
Sankaran P. G., *CUSAT*
Hari S., *CUSAT*

15:30 Semiparametric analysis of competing risks data with covariate measurement error [Abstract 14]

S ANJANA, *University of Hyderabad*

04.A1.C21 Contributed Session 21

Venue: DOE CL 4

Chair : Suvadip SANA, Cornell University

14:40 Fractional cumulative entropy function in quantile framework [Abstract 411]

Iona Ann SEBASTIAN, *Department of Statistics, CUSAT*
Iona Ann SEBASTIAN, *Department of Statistics, CUSAT*
Sunoj S M, *Department of Statistics, CUSAT*

14:55 Bivariate Distribution with Singular Component and its real-life Application [Abstract 256]

Sanjay KUMAR, *PhD candidate, Department of Mathematics & Statistics, IIT Kanpur*
Debasis KUNDU, *Professor, Department of Mathematics & Statistics, IIT Kanpur*
Sharmishtha MITRA, *Professor, Department of Mathematics & Statistics, IIT Kanpur*

15:10 Jackknife empirical likelihood ratio test for Cauchy distribution with applications to financial data [Abstract 20]

Ganesh Vishnu AVHAD, *Indian Institute of Technology Tirupati*
Ganesh Vishnu AVHAD , *Department of Mathematics and Statistics, Indian Institute of Technology, Tirupati, India*
Lahiri ANANYA , *Department of Mathematics and Statistics, Indian Institute of Technology, Tirupati, India*
K. Kattumannil SUDHEESH , *Statistical Sciences Division, Indian Statistical Institute, Chennai, Tamil Nadu, India*

15:25 Inference on the Polychoric Correlation using the Density Power Divergence [Abstract 347]

Arijit PYNE, *Novartis*
Abhik GHOSH, *Indian Statistical Institute, Kolkata*
Ayanendranath BASU, *Indian Statistical Institute, Kolkata*

15:40 Cohesive response clustering via metric space embedding [Abstract 394]

Suvadip SANA, *Cornell University*

04.A1.C22 Contributed Session 22

Venue: DSC Seminar Hall

Chair : Budhaditya GOSWAMI, Sr. Manager Biostatistics, Pfizer

14:40 Notes on an Existing Power Function Distribution Process and Introducing a Novel Approach for Modelling Proportional Reverse Hazard Processes. [Abstract 381]

Sachin SACHDEVA, *PhD Research Scholar at School of Mathematics and Statistics , University of Hyderabad*
B. G. MANJUNATH, *Associate Professor, Deptt. of Statistics, Rajiv Gandhi University of Arunachal Pradesh.*
Barry C. ARNOLD,
DISTINGUISHED PROFESSOR, DEPTT. OF STATISTICS, UNIVERSITY OF CALIFORNIA, RIVERSIDE, USA.,

14:55 WEIGHTED QUASI SUJA DISTRIBUTION AND ITS PROPERTIES [Abstract 471]

Vidya YERNENI, *Symbiosis Statistical Institute, Symbiosis international (Deemed University)*
Afaq A. RATHER,

15:10 Skill Dominance Analysis of Two(Four)-player, Three(Five)-dice Variant of the Ludo Game [Abstract 37]

Tathagata BANERJEE, *Indian Institute of Technology, Kanpur*
Diganta MUKHERJEE, *Indian Statistical Institute, Kolkata*

15:25 On the two sample test for persistent homology. [Abstract 257]

Satish KUMAR, *Indian Institute of Technology Kanpur*
Subhra Sankar DHAR, *Indian Institute of Technology Kanpur*

15:40 Optimizing Drug Labelling using Gain based Graphical MCP [Abstract 183]

Budhaditya GOSWAMI, *Sr. Manager Biostatistics, Pfizer*
Pranab GHOSH, *Director Biostatistics, Pfizer*
Margaret GAMALO, *VP Biostatistics, Pfizer*
Abhishek BHATTACHARJEE, *Sr Manager Biostatistics, Pfizer*

Coffee Break 16:00 - 16:30

04.E1.I83 Topics in non-parametric statistics

Venue: SC Main Hall

Chair : Sabyasachi CHATTERJEE, University of Illinois at Urbana Champaign
Organizer : Debashis MONDAL, Washington University in St Louis

16:30 Change point detection and inference [Abstract 271]

Carlos Misael MADRID PADILLA, *Washington University in St Louis*

16:55 Robust Recovery of the Central Subspace for Regression Using the Influence Function of the Renyi Divergence [Abstract 440]

T. N. SRIRAM, *University of Georgia*
Ross IACI, *William and Mary*

17:20 A New Locally Adaptive Nonparametric Regression Method. [Abstract 92]

Sabyasachi CHATTERJEE, *University of Illinois at Urbana Champaign*
Subhajit GOSWAMI, *Tata Institute of Fundamental Research*
Soumendu MUKHERJEE, *Indian Statistical Institute*

04.E1.I84 Statistics in clinical research II

Venue: SC Seminar Hall

Chair : Mahesh IYER, BMS
Organizer : Sangita KULATHINAL, University of Helsinki

16:30 Statistical Challenges and Opportunities in Drug Development [Abstract 234]

Amarjot KAUR, *Merck Research Labs*

16:55 Optimizing Subgroup Analysis: Shrinkage Estimation Techniques Through Hierarchical Modeling [Abstract 95]

swarnendu CHATTERJEE, *GSK*
swarnendu CHATTERJEE, *GSK*

17:20 Evaluating an intervention targeting early detection of a disease in RCTs –evidence synthesis under the framework of multistate models [Abstract 248]

Sangita KULATHINAL, *University of Helsinki*
Aapeli NEVALA,

04.E1.I85 Recent developments in Design of Experiments

Venue: SC Executive Hall

Chair and Organizer : Yogita GHARDE, ICAR-Directorate of Weed Research

16:30 ANALYSIS OF NON-CRISP DATA DERIVED FROM DESIGNED EXPERIMENTS [Abstract 457]

Cini VARGHESE, *ICAR-IASRI, PUSA, Library Avenue, New Delhi - 110 012*

16:55 Product Optimization through unified DoE and ML techniques [Abstract 198]

Mohd HARUN, *ICAR-IASRI*

04.E1.C23 Contributed Session 23

Venue: DOE Auditorium

Chair : Sumangal BHATTACHARYA, Indian Institute of Technology Tirupati, India

16:30 A Machine Learning approach to predict Selective Sweep in Genomic Region [Abstract 401]

Abhik SARKAR, *ICAR- Indian Agricultural Statistics Research Institute*
Dwijesh Chandra MISHRA, *ICAR- Indian Agricultural Statistics Research Institute, New Delhi-110012*
Dipro SINHA, *ICAR- Indian Agricultural Statistics Research Institute, New Delhi-110012*
Girish Kumar JHA, *ICAR- Indian Agricultural Statistics Research Institute, New Delhi-110012*

16:45 General weighted extropy of minimum and maximum ranked set sampling with unequal samples [Abstract 260]

Santosh KUMAR CHAUDHARY, *INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR*
Nitin GUPTA, *Indian Institute of Technology Kharagpur*

17:00 A New Approach to Mean Estimation under Ranked Set Sampling for Skewed Populations [Abstract 469]

Tanushree YADAV, *University of Allahabad*
Priyanka SINGH, *University of Allahabad*
Girish CHANDRA, *Delhi University*

17:15 Predictive Inference in Linear Mixed Models [Abstract 402]

Abir SARKAR, *Senior Associate, Data Scientist, Capital One, India (DATALABS)*
Gourab MUKHERJEE, *Associate Professor, USC Marshall School of Business*
Ishan SENGUPTA, *PhD student, Rutgers University*
Keisuke YANO, *The Institute of Statistical Mathematics*

17:30 Bivariate Empirical Mode Decomposition Based Models for Agricultural Price Forecasting [Abstract 254]

Rounak KUMAR, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*
Girish Kumar JHA, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*
Rajeev Ranjan KUMAR, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*

17:45 BMW: Inlier Prone Bayesian Models for Correlated Bivariate Data [Abstract 58]

Sumangal BHATTACHARYA, *Indian Institute of Technology Tirupati, India*
Ishapathik DAS, *Indian Institute of Technology Tirupati*

04.E1.C24 Contributed Session 24

Venue: DOE CL 1

Chair : Sreedevi E. P., CUSAT

16:30 A Study On Some New Bivariate Reliability Measures. [Abstract 452]

ATHIRA T S, *UNIVERSITY OF KERALA*
E I ABDUL SATHAR, *UNIVERSITY OF KERALA*

16:45 Dynamic Survival Prediction by Landmarking Using Parametric Proportional Hazards Models [Abstract 160]

JENET GEORGE, *Cochin University of Science and Technology*
SREEDEVI E. P., *Cochin University of Science and Technology*

17:00 Inaccuracy and divergence measures based on survival extropy, their properties and applications in testing and image analysis [Abstract 322]

SARANYA P, *COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY*
SUNOJ S M, *COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY*

17:15 DEFECTIVE REGRESSION MODELS FOR CURE RATE MODELLING WITH INTERAL CENSORED COMPETING RISKS DATA [Abstract 226]

Silpa K, *Cochin University of Science and Technology*
Sreedevi E. P., *Cochin University of Science and Technology*
P. G. SANKARAN, *Cochin University of Science and Technology*

17:30 Inference Procedures on Power Generalized DUS Transformation of Inverse Weibull Distribution [Abstract 321]

GAUTHAMI P, *St. Thomas College (Autonomous), Thrissur, University of Calicut, Kerala*
CHACKO V M, *St. Thomas College (Autonomous), Thrissur, University of Calicut, Kerala*

17:45 Modelling and Analysis of Bivariate Lifetime Data using Additive Hazards Survival Model [Abstract 444]

Namitha SURESH, *Cochin University of Science and Technology*
S. M. SUNOJ, *Cochin University of Science and Technology*
N Unnikrishnan NAIR, *Cochin University of Science and Technology*

04.E1.C25 Contributed Session 25

Venue: DOE CL 2

Chair : Kripa JOSTEN, Manipal College of Health Professions, Manipal Academy of Higher Education, Manipal

16:30 Bayesian Variable Selection in Survival Regression using Gaussian Process Prior [Abstract 354]

Rakesh RANJAN, *Banaras Hindu University*
Sourabh BHATTACHARYA, *Indian Statistical Institute*

16:45 Bayesian and E-Bayesian Estimation of Weighted Power Function Distribution [Abstract 11]

Abdullah AL-MARERI, *University of Kerala*
E I ABDUL SATHAR , *University of Kerala*

17:00 Zero-One Modified Poisson INAR(1) Process [Abstract 163]

Aishwarya GHODAKE, *PhD Scholar*
Manik AWALE, *Associate Professor*

17:15 Probabilistic Guarantees on Sensitivities of Bayesian Neural Network [Abstract 387]

Diptarka SAHA, *University of Illinois, Urbana-Champaign*
Zihe LIU, *University of Illinois, Urbana-Champaign*
Feng LIANG, *University of Illinois, Urbana-Champaign*

17:30 A Bayesian Joint Modelling of Current Status and Current Count Data [Abstract 197]

Pavithra HARIHARAN, *Department of Statistics, Cochin University of Science and Technology, Cochin*
P. G. SANKARAN, *Senior Professor and Vice-Chancellor, Department of Statistics, Cochin University of Science and Technology, Cochin*

17:45 Comparison of Bayesian and Frequentist Logistic Regression: Predictors of Smoking Among Elderly in India [Abstract 220]

Kripa JOSTEN, *Manipal College of Health Professions, Manipal Academy of Higher Education, Manipal*

04.E1.C26 Contributed Session 26

Venue: DOE CL 3

Chair : Shuvayan BANERJEE, IIT Bombay

16:30 Data-Driven Insights: Chemometric Analysis of Spectroscopic Data [Abstract 147]

Shreya DHUREKAR, *Department of Statistics Savitribai Phule Pune University*

16:45 Deciding Optimal Number of Segments in DCA by Bivariate Clustering Approach [Abstract 146]

Shantaram DHUM, *Department of Statistics, Savitribai Phule Pune University, Pune*
Akanksha KASHIKAR, *Department of Statistics, Savitribai Phule Pune University, Pune*

17:00 The Max Chart for Joint Monitoring of the Two Parameters of a Zero-inflated Poisson(ZIP) Process Using the Folded Normal Distribution [Abstract 306]

Subhradeep MUNIYAN, *Indian Statistical Institute*

17:15 An autoencoder based semi-supervised approach for accurate binning of metagenomics data [Abstract 340]

Deeksha P M, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

17:30 A robust Bayesian approach for reliability prognosis of nondestructive one-shot devices under cumulative risk model [Abstract 24]

Shanya BAGHEL, *Indian Institute of Technology (ISM) Dhanbad*
Shuvashree MONDAL, *Indian Institute of Technology (ISM) Dhanbad*

17:45 Outliers and their detection in sparse regression [Abstract 35]

Shuvayan BANERJEE, *IIT Bombay*
Ajit RAJWADE, *IIT Bombay*
James SAUNDERSON, *Monash University*
Radhendushka SRIVASTAVA, *IIT Bombay*

04.E1.C27 Contributed Session 27

Venue: DOE CL 4

Chair : Naresh GARG, Indian Statistical Institute, Delhi Centre

16:30 Evaluating Randomness Assumption: A Novel Graph Theoretic Approach for Linear and Circular Data [Abstract 159]

Shriya GEHLOT, *Indian Institute of Management Ahmedabad*
Arnab Kumar LAHA,

16:45 Bivariate Kullback-Leibler Divergence [Abstract 75]

Mary Rafflesia CHACKOCHAN, *Cochin University of Science and Technology*
P. G. SANKARAN, *Cochin University of Science and Technology*

17:00 Nonparametric estimation of residual extropy function under length-biased sampling. [Abstract 336]

Vaishnavi PAVITHRADAS, *Department of Statistics, Cochin University of Science and Technology, Kochi*
Rajesh G, *Department of Statistics, Cochin University of Science and Technology, Kochi*
Richu RAJESH, *Department of Statistics, Government Victoria College, Palakkad*

17:15 Estimation of weighted extropy with a focus on its use in reliability modeling [Abstract 222]

Archana K, *Cochin University of Science and Technology, Kochi, Kerala*
Irshad M. R., *Cochin University of Science and Technology*

17:30 Varextropy of k-Record Values and its Applications [Abstract 185]

Annie GRACE, *University of Kerala*
Manoj CHACKO, *University of Kerala*

17:45 On improved estimation of the larger location parameter [Abstract 158]

Naresh GARG, *Indian Statistical Institute, Delhi Centre*
Lakshmi Kanta PATRA, *Indian Institute of Technology Bhilai, Bhilai, India*
Neeraj MISRA, *Indian Institute of Technology Kanpur, Kanpur, India*

04.E1.C28 Contributed Session 28

Venue: DSC Seminar Hall

Chair : Rohini Bhagwanrao POTE, Indian Institute of Technology Bhilai

16:30 Generalized Birth-Death Process on Finite Lattice [Abstract 463]

Pradeep VISHWAKARMA, *Indian Institute of Technology Bhilai*
Kuldeep Kumar KATARIA, *Indian Institute of Technology Bhilai*

16:45 Stochastic comparison of extreme order statistics in Archimax copula [Abstract 202]

SARIKUL ISLAM, *Indian Institute of Technology Kharagpur*
Nitin GUPTA, *Indian Institute of Technology Kharagpur*

17:00 Choice of baseline hazards in joint modelling of longitudinal markers and time-to-event data [Abstract 246]

K M Jagathnath KRISHNA, *Regional Cancer Centre, Thiruvananthapuram*
Anand HARI, *Regional Cancer Centre, Thiruvananthapuram*
Divya DENNIS, *Regional Cancer Centre, Thiruvananthapuram*

17:15 On doubly inflated count data models [Abstract 16]

Monika ARORA, *Indraprastha Institute of Information Technology Delhi*
Dr. N. Rao CHAGANTY, *Old Dominion University*

17:30 Bivariate Generalized Geometric Distribution [Abstract 378]

Dr Harisankar S, *Asian School of Business, Technocity, Trivandrum*
Dr. C. Satheesh KUMAR,

17:45 GCP Erlang Queues [Abstract 341]

Rohini Bhagwanrao POTE, *Indian Institute of Technology Bhilai*
Kuldeep Kumar KATARIA, *Indian Institute of Technology*

Tuesday December 31

05.M1.I86 Diverse Applications of Statistics in Real-World Problems

Venue: SC Seminar Hall

Chair and Organizer : Soudeep DEB, Indian Institute of Management Bangalore

9:00 **TBA** [Abstract 139]

Kushal DEY, *Memorial Sloan Kettering Cancer Center*

9:25 **Proposing Data-Driven Regulatory Framework for the Online Gaming Sector: Challenges and the Way Forward** [Abstract 1]

Aishvarya , *SPJIMR, Mumbai*

9:50 **Estimation of Spectral Risk Measure for Left-Truncated and Right-Censored Data** [Abstract 68]

Suparna BISWAS, *Indian Statistical Institute Bangalore*
Rituparna SEN, *Indian Statistical Institute Bangalore*

05.M1.I87 Recent Advances in Time Series Analysis

Venue: SC Executive Hall

Chair : Jose K.K., Schoolof Mathematics and Statistics, Mahatma Gandhi University, Kottayam, India

9:00 **A New Generalization of Thinning-Based Integer-Valued Autoregressive Models for Count Data** [Abstract 201]

M.R. IRSHAD, *Cochin University of Science and Technology*

9:25 **On Spatio-temporal Autoregressive Models and their Applications** [Abstract 377]

Krishnarani S. D., *Department of Statistics, University of Calicut*

9:50 **Autoregressive Time Series Models with Generalized Laplacian Bilateral Gamma marginal distribution and Their Applications** [Abstract 225]

Jose K.K., *Schoolof Mathematics and Statistics, Mahatma Gandhi University, Kottayam, India*

05.M1.I88 Recent Advances in Shape-Constrained Inference

Venue: DOE Auditorium

Chair and Organizer : Sivaraman BALAKRISHNAN, Carnegie Mellon University

9:00 **New Stability Bounds in Optimal Transport and their Implications** [Abstract 26]

Sivaraman BALAKRISHNAN, *Carnegie Mellon University*

9:25 **Some non-standard inference problems** [Abstract 247]

Arun KUCHIBHOTLA, *Carnegie Mellon University*

9:50 **Totally Concave Regression** [Abstract 190]

Aditya GUNTUBOYINA, *University of California, Berkeley*
Dohyeong KI, *University of California Berkeley*

05.M1.I89 Statistics and the tech industry: some examples

Venue: DOE CL 1

Chair and Organizer : Debornshee SEN, Google

9:00 **Application of personalization in Enhancing customer engagement** [Abstract 350]

Saurabh RAJ, *Google*

9:25 **Learning representation for mixed data types with a nonlinear deep encoder-decoder framework** [Abstract 392]

Saswata SAHOO, *Google*

9:50 **Long-term impact of short-term experiments** [Abstract 413]

Deborshee SEN, *Google*

05.M1.I90 Advanced Statistical Modeling Techniques for Time Series and Cross-Sectional Data with Dependence Structures

Venue: DOE CL 2

Chair : Amiya Ranjan BHOWMICK, Institute of Chemical Technology, Mumbai
Organizer : Tanujit DEY, Brigham and Women's Hospital and Harvard Medical School

9:00 **A generalized Pegram's operator based autoregressive (GPAR) process for modelling categorical time series** [Abstract 63]

Atanu BISWAS, *Indian Statistical Institute Kolkata*

9:25 **Cross Sectional Regression with Cluster Dependence: Inference based on Averaging** [Abstract 114]

Samarjit DAS, *Indian Statistical Institute Kolkata*

05.M1.I91 Statistical Methods for High-dimensional and Complex Datasets

Venue: DOE CL 3

Chair and Organizer : Kaushik JANA, Ahmedabad University

9:00 **Novel clustering procedures based on binary splitting using max-MMD** [Abstract 79]

Anirvan CHAKRABORTY, *IISER Kolkata*
Sourav CHAKRABARTY, *ISI Kolkata*
Shyamal Krishna DE, *ISI Kolkata*

9:25 **High-dimensional Adaptive Multiple Testing Controlling FDR and FNR for Sequential Data** [Abstract 128]

Shyamal Krishna DE, *Indian Statistical Institute Kolkata*
Rahul ROY, *Indian Statistical Institute*
Subir Kumar BHANDARI, *Indian Statistical Institute*

9:50 **Angular Co-variance using intrinsic geometry of torus: Application to Non-parametric change points detection in meteorological data** [Abstract 30]

BUDDHANANDA BANERJEE, *Department Of Mathematics , IIT Kharagpur*
Surojit BISWAS, *Dept. Of Mathematics IIT Kharagpur*
Arnab Kumar LAHA, *IIM Ahmedabad*

05.M1.C29 Contributed Session 29

Venue: DOE CL 4

Chair : Asad UDDIN, Indian Institute of Information Technology Guwahati

9:00 **A Multi-objective Economic Statistical Design of the CUSUM chart: NSGA II Approach** [Abstract 312]

Sandeep , *Indian Statistical Institute, Kolkata*
Arup RANJAN MUKHOPADHYAY,

9:15 **Length Biased Weighted Ishita Distribution and Its Applications on Real Life Data Sets** [Abstract 307]

Mustafa MUSTAFA, *Department of Statistics, University of Delhi*
Mustafa MUSTAFA, *Department of Statistics, University of Delhi*

9:30 **Discriminating between Weibull and log-normal distributions in the presence of hybrid censoring** [Abstract 351]

Ojasvi RAJPUT, *IIT kanpur*
Ojasvi RAJPUT, *IIT kanpur*
Debasis KUNDU, *IIT Kanpur*
Sharmishtha MITRA, *IIT Kanpur*

9:45 **Statistical Analysis of Bivariate Left Truncated Right Censored Data** [Abstract 391]

Sourav SAHA, *IIT Guwahati*
Ayon GANGULY, *IIT Guwahati*
Debanjan MITRA, *IIM Udaipur*

10:00 **Exact Inference For The Exponential Distribution Using Type-I Censoring on Stage Life testing Experiments** [Abstract 453]

Asad UDDIN, *Indian Institute of Information Technology Guwahati*
Erhard CRAMER, *RWTH Aachen University*
Farha SULTANA, *Indian Institute of Information technology Guwahati*

05.M1.C30 Contributed Session 30

Venue: DSC Seminar Hall

Chair : AKARSH SINGH, ICAR-indian Agricultural Research Institute

9:00 **On the Poisson phase-type process and its application in shock models** [Abstract 184]

Dheeraj GOYAL, *Indian Institute of Technology Kanpur*
Dheeraj GOYAL, *Department of Mathematics and Statistics, Indian Institute of Technology Kanpur*
Nil Kamal HAZRA, *IIT Jodhpur*
Maxim FINKELSTIEN, *University of Free State, South Africa*

9:15 **Designs for Multi-Session Sensory Trials in the Presence of Assessor Constraints** [Abstract 134]

Boyina DEVI PRIYANKA, *Indian Agricultural Statistics Research Institute*
Cini VARGHESE, *Principal Scientist and Professor*
Mohd HARUN, *Scientist*
Anindita DATTA, *Scientist*

9:30 **Changepoint problem with angular data using a measure of variation based on the intrinsic geometry of torus** [Abstract 69]

SUROJIT BISWAS, *Indian Institute of Technology Kharagpur*
BUDDHANANDA BANERJEE , *Department of Mathematics, Indian Institute of Technology Kharagpur*
ARNAB KUMAR LAHA, *Operations and Decision Sciences Area, Indian Institute of Management Ahmedabad*

9:45 **Resampling-free Inference for Infinite Dimensional Parameters in Time Series via Sample Splitting and Self-Normalization** [Abstract 179]

Deep GHOSHAL, *University of Illinois urbana-Champaign*
Xiaofeng SHAO, *University of Illinois Urbana-Champaign*

10:00 **Functional Autoregressive Model for Forecasting Prices of Major Pulses in India** [Abstract 427]

AKARSH SINGH, *ICAR-indian Agricultural Research Institute*
Ranjit Kumar PAUL, *ICAR-Indian Agricultural Statistics Research Institute*
Md YEASIN, *ICAR-Indian Agricultural Statistics Research Institute*

Coffee Break 10:20 - 10:40

Plenary Lecture 3 Bhramar Mukherjee

Venue: SC Main Hall

Chair : Aditya GUNTUBOYINA, University of California, Berkeley

10:30 Unveiling Bias: A Statistician's Quest for Data Equity in Health Research [Abstract 296]

Bhramar MUKHERJEE, *Yale University*

Lunch 11:40 -

Abstracts

1. Proposing Data-Driven Regulatory Framework for the Online Gaming Sector: Challenges and the Way Forward

[05.M1.I86, (page 59)]

Aishvarya , *SPJIMR, Mumbai*

Online gaming and Daily Fantasy Sports (DFS), a booming sector, has recently gained significant popularity owing to technological advancements, increased fan engagement, relative ease of access to the internet, affordable smartphones, and increased celebrity endorsements. With the growing popularity of DFS, there has been increased legal scrutiny around this sector, which has resulted in multiple regulatory challenges. First, there has been constant tussle between the government and DFS operators on the question of whether DFS is a game of skill or a game of chance. Most countries use the 'Dominant Factor' test to identify gambling activities. According to this test, if the outcome of a gaming activity is influenced more by the user's skill than by chance elements of the game, it is called a "game of skill" and is legal under the law. However, if the effect of chance outweighs the effect of the user's skill, it is termed a "game of chance" and is considered gambling. Unfortunately, regulators struggle to use an objective approach to address this problem, which has led to numerous court cases. The absence of a quantitative approach to evaluate the effect of users' skill and chance in a DFS contest has led to inconsistencies in the law and business losses. Second, while it has been a challenge for legal bodies to adopt an appropriate measure to address the skill versus chance debate , it is also equally important for DFS operators to ensure that their platform is skill-dominant by construct and promote responsible gaming. The recent regulatory changes and lack of a data-driven, robust approach pose many important research questions that, unfortunately, have not gained enough attention. The study highlights the importance of resolving these issues and proposes a data-driven technique to i.) evaluate the effect of skill and chance in an online gaming setup and ii.) identifies parameters of a gaming setup that can ensure the game design is skill-dominant by construct. The study also outlines societal concerns and proposes solutions that can help policymakers regulate the online gaming sector.

2. On martingale characterizations of generalized counting process and its time-changed variants

[Student Poster Competition, (page 21)]

Manisha , *Indian Institute of Technology Bhilai*

Counting processes with jumps of amplitude larger than one have potential real-life applications, since they are beneficial to characterize simultaneous but independent Poisson streams. The generalized counting process (GCP) is one among the generalizations of Poisson process in the sense that it performs k kinds of jumps of amplitude $1, 2, \ldots, k$ with positive rates $\lambda_1, \lambda_2, \ldots, \lambda_k$, respectively. In our work, we extend the Watanabe's martingale characterization for Poisson process to the case of generalized counting process (GCP). It is shown that the GCP has a unique representation as the weighted sum of independent Poisson processes. Moreover, the martingale characterization for two time-changed variants of the GCP are obtained. The random time-change components considered are the inverse stable and inverse mixed stable subordinators.

3. MODELLING AND COMPARISON OF POTENTIAL EVAPOTRANSPIRATION IN TIRUCHIRAPALLI USING STATISTICAL TECHNIQUES

[03.E1.C15, (page 41)]

ARCHANA A, *ICAR-IASRI, NEW DELHI*
RADHA M, *Assistant Professor, Department of Agricultural Economics, TNAU, Madurai-625104, Tamil Nadu, India*

Potential evapotranspiration which includes both evaporation and transpiration, is a key process that determines how water is removed from soil and plant surfaces and subsequently return to the atmosphere. This study focuses on calculating the weekly potential evapotranspiration (PET) for the Tiruchirapalli district. Several models- FAO-Penman- Monteith, Modified Penman, Penman, Priestley Taylor, Blaney Criddle and Hargreaves were evaluated through Robust regression. The performance of this models was assessed using accuarcy metrics like Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Mean sbsolute Error (MAE), Mean Absolute Percentage Error and Coeeficient of determination (R2). Among the models, the Penman model proved to be the most effective, showing the least bias and the highest R2 value

4. Evaluating treatment heterogeneity in survival outcomes using causal inference methods

[01.E1.C4, (page 12)]

Sharon Varghese A, *Novartis Healthcare Private Limited*

In clinical trials, treatment heterogeneity poses significant challenges in estimating the true treatment effect, particularly in survival outcomes due to factors such as censoring and treatment switching. This study aims to quantify the variability in treatment effects on survival outcomes across different subgroups and identify the socio-demographic, clinical, and genetic factors contributing to this heterogeneity. The analysis will employ several causal inference methods, including Inverse Probability of Treatment Weighting (IPTW), Rank-Preserving Structural Failure Time Model (RPSFTM), and Inverse Probability of Censoring Weighting (IPCW), among others. The study is expected to provide a detailed understanding of how treatment effects on survival outcomes vary across different subgroups within the population.

5. Local reconstruction analysis of inverting the Radon transform in the plane from noisy discrete data

[04.M2.I69, (page 46)]

Anuj ABHISHEK, *Case Western Reserve University*
Alexander KATSEVICH, *University of Central Florida*
James WEBBER, *Cleveland Clinic*

In this talk, we investigate the reconstruction error, $N\_\epsilon^{\mathrm{rec}}(x)$, when a linear, filtered backprojection (FBP) algorithm is applied to noisy, discrete Radon transform data with sampling step size in two-dimensions. We show, under suitable assumptions on the first three moments of the noise perturbing the data, in any $O(\epsilon)$ sized neighborhood around a generic point, the reconstruction error is asymptotically a zero mean Gaussian random field. Once the limit is established, we compute explicitly its covariance. In addition, we validate our theory using numerical simulations and pseudo random noise.

6. Network models for spatial transcriptomics data

[01.M2.I1, (page 3)]

Satwik ACHARYYA, *University of Alabama at Birmingham*

Network models are powerful tools to investigate complex dependence structures in high throughput genomic datasets. They allow for holistic, systems-level view of the various biological processes, for intuitive understanding and coherent interpretations. However, most existing network or graphical models are developed under assumptions of homogeneity of samples and are not readily amenable to modeling spatial heterogeneity which often manifests in spatial genomics data. In this talk, I will discuss two spatial network models focusing on spatially varying covariance and precision matrices. (I) SpaceX (spatially dependent gene co-expression network) is a Bayesian methodology to identify both shared and cluster-specific co-expression networks across genes. (II) Spatial Graphical Regression (SGR) is a flexible approach based on graphical regression that enables spatially varying graphs over the spatial domain of the tissue. The framework incorporates multiple spatial covariates and provides a linear and non-linear functional mapping between the spatial domain and the precision matrices. All the approaches are illustrated by using case studies from cancer genomics.

7. On the spectrum of random simplicial complexes in thermodynamic regime.

[01.A1.I15, (page 8)]

Kartick ADHIKARI, *IISER Bhopal*
Kiran KUMAR, *IIT Bombay*
Koushik SAHA, *IIT Bombay*

In this talk we shall discuss the limiting spectral distribution (LSD) of the Linial Meshulam model in thermodynamics regime. This model is a generalization of the Erdos Renyi graph. The d-dimensional Linial model is a random simplicial complex with (d-1)-complete skeleton and each d-simplex included with probability p independently. Recently Knowles and Rosenthal studied the LSD of this model in dense regime and showed that its LSD is semicircular law. However, we show that the LSD of this model in thermodynamics regime is not semicircular law.

This talk will be based on a joint work with Kiran Kumar and Koushik Saha.

8. Intersectional Inequalities in Anthropometric Failure among Indian Children: Evidence from the National Family Health Survey (2015-2016)

[03.A1.I56, (page 37)]

Tulsi ADHIKARI, *Indian Council of Medical Research*
Niharika TRIPATHI, *Indraprastha College for Women University of Delhi (North Campus)*

Increasing body of health planning and policy

research focused upon unravelling the fundamental drivers of population health and nutrition inequities, such as wealth status, educational status, caste/ethnicity, gender, place of residence, and geographical context, that often interact to produce health inequalities. However, very few studies have employed intersectional framework to explicitly demonstrate how intersecting dimensions of privilege, power, and resources form the burden of anthropometric failures of children among low-and-middle income countries including India. Data on 2,15,554 sampled children below 5 years of age from the National Family Health Survey 2015-2016 were analysed. This study employed intersectional approach to examine caste group inequalities in the anthropometric failure (i.e. moderate stunting, severe stunting, moderate underweight, severe underweight, moderate wasting, severe wasting) among children in India. Descriptive statistics and multinomial logistic regression models were fitted to investigate the heterogeneities in the burden of anthropometric failure across demographic, socioeconomic and contextual factors. Interaction effects were estimated to model the joint effects of socioeconomic position (household wealth, maternal education, urban/rural residence and geographical region) and caste groups with the likelihood of anthropometric failure among children.

More than half of under-5 children suffered from anthropometric failure in India. Net of the demographic and socioeconomic characteristics, children from the disadvantageous caste groups whose mother were illiterate, belonged to economically poor households, resided in the rural areas, and coming from the central and eastern regions experienced disproportionately higher risk of anthropometric failure than their counterparts in India. Concerted policy processes must recognize the existing heterogeneities between and within population groups to improve the precision targeting of the beneficiary and enhance the efficiency of the nutritional program among under-5 children, particularly for the historically marginalized caste groups in India.

9. Combinatorial Testing using Locating Arrays

[04.M1.I64, (page 44)]

Yasmeen AKHTAR, *Birla Institute of Technology and Science, Pilani-Goa Campus*

Testing is essential for detecting failures caused by interactions between various factors in a system. Combinatorial testing, which uses a covering array, efficiently tests these interactions up to a certain strength. While covering arrays have proven effective in detecting the presence of interaction failures, they do not guarantee the ability to precisely locate the source of the failure. To address this limitation, Colbourn and McClary introduced the concept of a locating array in 2008. A locating array of strength t ensures that all t-way level combinations are tested, with each combination uniquely identifiable by its specific set of test runs. This allows for the precise identification of any t-way level combination responsible for a deterministic failure. The number of test runs required by a locating array increases logarithmically with the number of factors, making it a practical choice for screening experiments involving many categorical factors with multiple levels. This talk will explore a method for constructing and analyzing locating arrays. The computational results demonstrate that certain cyclotomic arrays can be used to produce locating arrays. Under the assumption of effect sparsity, our screening algorithm focuses on 2-way level combinations to identify the most significant factors.

10. Analysis of Interval Censored Competing Risks Data with Covariate Measurement Error

[02.E1.C7, (page 28)]

Jaya Naga Sri AKURATHI, *University of Hyderabad*
Anjana S, *University of Hyderabad*

The competing risks data are very often in survival analysis, where an individual may experience failure from one of several distinct causes of failure. In many practical situations, the covariates are measured with error due to biological variability and other sources of variation. The traditional inferential procedures give biased estimations when some covariates are measured with error. In the literature, a wide variety of inference approaches have been used to cope with the covariate measurement error for the failure time data. Most such correction methods focus on the Cox proportional hazards model without competing risks data. In this paper, we present a semiparametric linear transformation model for the interval-censored competing risks data with covariate measurement error. Propose an estimation method using semiparametric maximum likelihood estimation to handle the covariate measurement error. Derive the asymptotic properties of the estimators of the proposed method. Examine the performance of the proposed method through simulation

studies. Also, we illustrate the proposed method using a real dataset.

11. Bayesian and E-Bayesian Estimation of Weighted Power Function Distribution

[04.E1.C25, (page 56)]

Abdullah AL-MARERI, *University of Kerala*
E I ABDUL SATHAR , *University of Kerala*

This paper investigates the Bayesian and E-Bayesian estimation of the Weighted Power Function Distribution (WPFD), commonly used in evaluating the reliability of electrical ccomponents and semiconductor devices . The study compares maximum likelihooh estimaion, Bayesian estimation and E-Bayesian estimation under different loss function SELF, PLF, DLF and WLF. A simulation study evaluates the performance of these estimators in terms of bais and mean square error across different sample sizes and parametr values.

12. Dir-SPGLM: A Bayesian semiparametric GLM with data-driven reference distribution

[Student Poster Competition, (page 21)]

Entejar ALAM, *University of Texas at Austin*
Peter MÜLLER, *University of Texas at Austin*
Paul J. RATHOUZ, *University of Texas at Austin*

The recently developed semi-parametric generalized linear model (SPGLM) offers more flexibility as compared to the classical GLM by including the baseline or reference distribution of the response as an additional parameter in the model. However, some inference summaries are not easily generated under existing maximum-likelihood based inference (ML-SPGLM). This includes uncertainty in estimation for model-derived functionals such as exceedance probabilities. The latter are critical in a clinical diagnostic or decision-making setting. In this article, by placing a Dirichlet prior on the baseline distribution, we propose a Bayesian model-based approach for inference to address these important gaps. We establish consistency and asymptotic normality results for the implied canonical parameter. Simulation studies and an illustration with data from an aging research study confirm that the proposed method performs comparably or better in comparison with ML-SPGLM. The proposed Bayesian framework, Dir-SPGLM, is most attractive for inference with small sample training data or in sparse-data scenarios.

13. Estimation of confidence interval of Cpy under some location scale distributions using fiducial approach

[04.A1.I78, (page 51)]

M Zafar ANIS, *ISI Calcutta*

TBA

14. Semiparametric analysis of competing risks data with covariate measurement error

[04.A1.I82, (page 52)]

S ANJANA, *University of Hyderabad*

TBA

15. Stability of nonlinear filters: numerical and theoretical results

[04.M2.I69, (page 46)]

Amit APTE, *Indian Institute of Science Education and Research IISER Pune*
Anugu SUMITH REDDY, *Rice University*
Shashank ROY, *IMT Atlantique*
Pinak MANDAL, *University of Sydney*

The Bayesian formulation of the data assimilation problem leads to non-linear filtering. In many applications, the dynamical models are deterministic and chaotic, in which case most of the classical stability results for nonlinear filtering are not applicable because such systems do not satisfy the assumption of controllability. In this talk, I discuss our recent results proving asymptotic filter stability for deterministic, chaotic dynamics and the relation between the dynamical characteristics of such systems and the asymptotic filtering distribution. I will also discuss recent numerical results that use the Sinkhorn algorithm for computing distances between distributions represented by samples and the implications of these results for data assimilation problem in earth sciences.

16. On doubly inflated count data models

[04.E1.C28, (page 58)]

Monika ARORA, *Indraprastha Institute of Information Technology Delhi*
Dr. N. Rao CHAGANTY, *Old Dominion University*

Count data models have applications in various areas like health science, transportation, education, agriculture, and many more areas. Usually, Poisson or negative binomial models are used for such

datasets. However, there are scenarios where zero count is inflated in count datasets. The zero-inflated Poisson and zero-inflated negative binomial are the corresponding extensions in such situations. Recently, there have been examples where apart from zero there is another count value k>0 that is also inflated. The zero and k- inflated Poisson (ZkIP) models give better fit to these datasets. We propose alternative approach to get the estimates and standard errors for the models with and without covariates. We found that they provide better fit to the illustrated datasets.

17. Single Index Batched Contextual Bandits

[02.M1.I23, (page 15)]

Sakshi ARYA, *Case Western Reserve University*
Hyebin SONG, *Penn State University*

In applications such as clinical trials, treatment decisions are usually made in phases/batches, where information from the previous batch is used to determine the treatments allocated in the upcoming batch. Such scenarios can naturally be seen to fall in the batched bandits framework. While batched bandit frameworks have been studied in parametric and nonparametric regression settings, we propose a novel semi-parametric bandit approach that promotes interpretability and dimension reduction in nonparametric batched bandits. We assume that the reward-covariate relationship can be modelled in a reduced 1-dimensional central subspace based on the single-index regression framework. We adopt an adaptive binning and successive elimination algorithm and provide optimal regret guarantees for the same. We also illustrate the performance of the algorithm on simulated and real datasets.

18. Approximate, Leave-one-out, Cross Validation for Regression with ℓ_1 Regularizers

[01.A1.I10, (page 7)]

Arnab AUDDY, *The Ohio State University*
Haolin ZOU, *Columbia University*
Kamiar RAHNAMA RAD, *Baruch College, CUNY*
Arian MALEKI, *Columbia University*

The out-of-sample error (OO) is the main quantity of interest in risk estimation and model selection. Leave-one-out cross validation (LO) offers a (nearly) distribution-free yet computationally demanding method to estimate OO. Recent theoretical work showed that approximate leave-one-out cross validation (ALO) is a computationally efficient and statistically reliable estimate of LO (and OO) for generalized linear models with twice differentiable regularizers. For problems involving non-differentiable regularizers, despite significant empirical evidence, the theoretical understanding of ALO's error remains unknown. In this paper, we present a novel theory for a wide class of problems in the generalized linear model family with the non-differentiable regularizer. We bound the error in terms of intuitive metrics such as the size of leave–out perturbations in active sets, sample size , number of features and signal-to-noise ratio (SNR). As a consequence, for the regularized problems, we show that $|\mathrm{ALO} - \mathrm{LO}| \to 0$ while p/n and SNR remain bounded.

19. A study on truncated ROC curves

[02.E1.C9, (page 29)]

Anu AUGUSTINE, *Mahatma Gandhi University kottayam*
Anu AUGUSTINE, *Mahatma Gandhi University Kottayam, Kerala*
Dr. Angel MATHEW, *Mahatma Gandhi University Kottayam, Kerala*

Biomarkers play a crucial role in medical research and practice, enabling early disease detection, aiding in treatment development, and guiding clinical decision-making. The accuracy of a diagnostic test is often evaluated using receiver operating characteristic (ROC) curves. However, in many clinical scenarios, it is essential to focus on specific ranges of the biomarker that are most clinically relevant. Truncated ROC curves provide a valuable tool for this purpose, allowing for a more targeted analysis of diagnostic performance. In this paper, we introduce left-truncated and right-truncated ROC curves, which consider only the portions of the biomarker distribution that are of interest. We demonstrate the utility of these measures through various examples and discuss their characterizations and behaviour under monotone transformations. Keywords: Biomarker, ROC, left truncation, right truncation.

20. Jackknife empirical likelihood ratio test for Cauchy distribution with applications to financial data

[04.A1.C21, (page 53)]

Ganesh Vishnu AVHAD, *Indian Institute of Technology Tirupati*

Ganesh Vishnu AVHAD , *Department of Mathematics and Statistics, Indian Institute of Technology, Tirupati, India*
Lahiri ANANYA , *Department of Mathematics and Statistics, Indian Institute of Technology, Tirupati, India*
K. Kattumannil SUDHEESH , *Statistical Sciences Division, Indian Statistical Institute, Chennai, Tamil Nadu, India*

Heavy-tailed distributions, such as the Cauchy distribution, are acknowledged for providing more accurate models for financial returns, as the normal distribution is deemed insufficient for capturing the significant fluctuations observed in real-world assets. Data sets characterized by outlier sensitivity are critically important in diverse areas, including finance, economics, telecommunications, and signal processing. This article addresses a goodness-of-fit test for the Cauchy distribution based on characterization. The proposed test utilizes empirical likelihood methods, including the jackknife empirical likelihood (JEL) and adjusted jackknife empirical likelihood (AJEL). It is shown that the asymptotic distribution of both the JEL and AJEL ratio test statistics follows a chi-square distribution with one degree of freedom. Extensive simulation results show that the proposed tests have very good power across various alternatives. The approach is illustrated by analysing two real-world data sets, including an empirical application in the financial sector.

21. Integer Autoregressive (INAR) Models and their Applications

[04.M1.I68, (page 46)]

Manik AWALE, *Savitribai Phule Pune University*

This paper discusses the modelling of count time series data using integer Autoregressive model. The research in this area have shown that INAR models have become very popular these days. Various estimation procedures such as Yule-Walker (Y-W), conditional least squares (CLS) and conditional Maximum likelihood (CML) estimation methods are discussed for over dispersed count time series data. Various forecast accuracy measures are also discussed in this paper. A real life data has been analyzed using the proposed methodology.

22. Generalized Marshall-Olkin Bivariate Exponential Distribution and Its Application in Constant-Stress Accelerated Life Test

[02.E1.C7, (page 28)]

Sneha BABU, *Cochin University of Science and Technology, Cochin, Kerala*
Princy T,

It is very common for several competing failure causes to exist simultaneously during reliability life testing. This issue is referred to as the competing risks or failure model because it involves numerous failure modes, yet only the least failure time and its associated failure mode are noticed. This paper considers a constant-stress accelerated dependent competing risk model with type-II censoring. The dependent structure between competing risks is modelled by a generalized Marshall-Olkin bivariate exponential distribution. A decreasing hazard rate for both variables is a characteristic of the q-bivariate Marshall-Olkin exponential distribution. A simulation study is carried out to interpret the proposed model. An accelerated stress dataset is fitted using the proposed model and it is found to be a better fit than the conventional Marshall-Olkin bivariate exponential distribution.

23. Modified Polygenic Risk Score to account for Gene Effects

[01.E1.C3, (page 11)]

Devarpita BAG, *Duke University*
Manit PAUL, *Wharton School, University of Pennsylvania*
Indranil MUKHOPADHYAY, *UNIVERSITY of NEBRASKA–LINCOLN*

Common single-nucleotide polymorphisms (SNPs) have been demonstrated to play significant roles in determining susceptibility to prevalent diseases in recent genome-wide association studies (GWAS). There may be many underlying susceptibility SNPs for a particular disease, each showing only a weak disease association, but when combined, they may account for a sizeable proportion of the variation in disease incidence in the general population. Polygenic risk score is derived as a measure of disease risk due to one's genes. It combines the contribution of each associated SNP to the risk of developing a disease. But the SNP's are correlated with each other through linkage disequilibrium. On the other hand genes are independent of each other. Hence we have proposed a new risk score with genes as the the causal variables. We defined a new concept called gene effect size by performing LD-adjustment and aggregating the effect sizes of SNP's present in

the gene to get the true effect of the gene to the risk score. Furthermore, the current PRS suffers from ethnicity issue as the GWAS data which is used to estimate SNP effect sizes is dominated by European population. We derived new coefficients for the PRS score by shrinking GWAS estimates towards local population based estimates using James Stein shrinkage method.

24. A robust Bayesian approach for reliability prognosis of nondestructive one-shot devices under cumulative risk model

[04.E1.C26, (page 57)]

Shanya BAGHEL, *Indian Institute of Technology (ISM) Dhanbad*
Shuvashree MONDAL, *Indian Institute of Technology (ISM) Dhanbad*

The present study aims to determine the lifetime prognosis of highly durable nondestructive one-shot device units under a step-stress accelerated life testing (SSALT) experiment applying a cumulative risk model (CRM). In an SSALT experiment, CRM retains the continuity of hazard function by allowing the lag period before the effects of stress change emerge. In an analysis of such lifetime data, plentiful datasets might have outliers where conventional methods like maximum likelihood estimation or likelihood-based Bayesian estimation frequently fail. This work develops a robust estimation method based on density power divergence in classical and Bayesian frameworks. The hypothesis is tested by implementing the Bayes factor based on a robustified posterior. Further, the influence functions are examined to evaluate the robust behaviour of the estimators and the Bayes factor. Finally, the analytical development is validated through a simulation study and a real data analysis.

25. Bernoulli Spiders

[03.A1.I57, (page 37)]

Srinivasan BALAJI, *George Washington University*
Hosam MAHMOUD, *George Washington University*

In this joint work with Prof. Hosam Mahmoud from GWU, we study a tree structure called the spider. It grows in time (both discrete and continuous are considered). In the spider, only the root and the leaves recruit children. We look at the leaves and the size of the tree. Success in recruiting is based on a competition among the recruiters according to associated affinities. The growth at the root follows a probability model which adds to the affinity at the root, when it recruits, a Bernoulli (p) value. The case p = 0 is a degeneracy that is isolated and studied separately. For the number of leaves (appropriately normalized) with p (0, 1], in a discrete time spider we get a central limit theorem and in a continuous-time spider we get a gamma limit distribution. The size of a spider growing in discrete time is trivial, but the size of the continuous-time spider, when appropriately normalized, follows a Gaussian law. Association to certain Polya urns and in the continuous-time case solutions to partial differential equations are the main tools for the derivations. At the end we compute the degree-based Gini index for spiders grown in both time modes.

26. New Stability Bounds in Optimal Transport and their Implications

[05.M1.I88, (page 59)]

Sivaraman BALAKRISHNAN, *Carnegie Mellon University*

Given two probability distributions, the optimal transport (OT) map is a function which maps samples from one distribution into samples from the other, while minimizing the expected displacement of samples. We study a central question in statistical optimal transport – that of estimating the OT map when given samples from the two distributions.

Following past work of Manole et al. and Deb et al. we study plugin estimators, which are derived from the OT map between estimates of the underlying distributions. We develop novel stability bounds for optimal transport maps which generalize those in past work. Our new stability bounds allow us to reduce the problem of optimally estimating the transport map to that of optimally estimating densities in the Wasserstein distance. In contrast, past work provided a partial connection between these problems and relied on unnatural assumptions to bridge the gap. We develop several new results in the smooth setting studied in past work and we also develop consequences of our new stability bounds for the estimation of the transport maps without smoothness and boundedness assumptions on the underlying measures. As a consequence we obtain a new estimator for the OT map between two strongly log-concave distributions.

This talk will include joint work with Tudor Manole, Jon Niles-Weed and Larry Wasserman.

27. Frequency Domain Resampling for Spatial Data

[02.M1.I22, (page 14)]

Soutir BANDYOPADHYAY, *Colorado School of Mines*
Souvick BERA, *Colorado School of Mines*
Dan NORDMAN, *Iowa State University*

In the realm of frequency domain analysis for spatial data, estimators based on the periodogram often exhibit complex variance structures originating from aggregated periodogram covariances. Previous attempts to bootstrap these statistics face challenges in capturing these variances and quantifying estimation uncertainty. This difficulty arises because achieving consistency for various periodogram-based statistics requires evaluating the periodogram at an increasing number of frequencies as the sample size grows. Despite the diminishing dependence between periodogram ordinates, the decay rate balances the growing frequencies, preserving a dependence structure in the limiting distribution. Consequently, the validity of frequency domain bootstrap (FDB) approaches for spatial data is confined to a specific class of processes and statistics. To overcome this challenge, we propose cutting-edge FDB methods based on subsampling which can accurately capture uncertainty without necessitating additional stringent assumptions beyond those required for the existence of a target limit distribution, filling a gap in the theory by providing distributional approximations for spectral statistics.

28. "Inference from bits and pieces of information: Collaborative statistical learning

[03.M1.I47, (page 32)]

Anjishnu BANERJEE, *Medical College of Wisconsin*
Hengrui HU, *MCW*

In this talk, the main interest is in addressing the methodological constraints around the basic premise of information borrowing in Bayesian versions of federated and collaborative statistical learning. In general, distributed inference, where inference is made from piece-wise data, borrowing of information from related but mixed domain models, and cases when borrowing of information occurs in related but externally differentiated models (through model propagation or convolution) are considered. Specific inferential methods are discussed to incorporate pre-trained knowledge and external data. Enabling external data information borrowing allows one to gain efficiency without having to "reinvent the wheel". In contrast, hierarchical and adaptive structures allow deviations from information gleaned from external data. While focusing on Bayesian learning, the investigations considered are generalizable to other contexts. A novel methodology and theoretical considerations are presented, which enable inferential probabilistic guarantees and efficient model computation using both simulated and real examples.

29. High Dimensional Behaviour of Some Two-Sample Tests Based on Ball Divergence

[Student Paper Competition 2, (page 17)]

BILOL BANERJEE, *INDIAN STATISTICAL INSTITUTE KOLKATA*
ANIL K. GHOSH, *INDIAN STATISTICAL INSTITUTE KOLKATA*

We propose some two-sample tests based on ball divergence and investigate their high dimensional behaviour. First, we consider the High Dimension, Low Sample Size (HDLSS) setup. Under appropriate regularity conditions, we establish the consistency of these tests in the HDLSS regime, where the dimension grows to infinity while the sample sizes from the two distributions remain fixed. Next, we show that these conditions can be relaxed when the sample sizes also increase with the dimension, and in such cases, consistency can be proved even for shrinking alternatives. We use a simple example to show that even when there are no consistent tests in the HDLSS regime, the proposed tests can be consistent if the sample sizes increase with the dimension at an appropriate rate. This rate is obtained by establishing the minimax rate optimality of these tests over a certain class of alternatives. Several simulated and benchmark data sets are analyzed to compare the empirical performance of these tests with some state-of-the-art methods available for testing the equality of two high-dimensional distributions.

30. Angular Co-variance using intrinsic geometry of torus: Application to Non-parametric change points detection in meteorological data

[05.M1.I91, (page 60)]

BUDDHANANDA BANERJEE, *Department Of Mathematics , IIT Kharagpur*
Surojit BISWAS, *Dept. Of Mathematics IIT Kharagpur*
Arnab Kumar LAHA, *IIM Ahmedabad*

In many temporal datasets, the parameters of the underlying distribution may change abruptly at unknown times. Detecting these changepoints is crucial for numerous applications. While this problem has been extensively studied for linear data, there has been remarkably less research on bivariate angular data. For the first time, we address the changepoint problem for the mean direction of toroidal and spherical data, which are types of bivariate angular data. By leveraging the intrinsic geometry of a curved torus, we introduce the concept of the "square" of an angle. This leads us to define the "curved dispersion matrix" for bivariate angular random variables, analogous to the dispersion matrix for bivariate linear random variables. Using this analogous measure of the "Mahalanobis distance," we develop two new nonparametric tests to identify changes in the mean direction parameters for toroidal and spherical distributions. We derive the limiting distributions of the test statistics and evaluate their power surface and contours through extensive simulations. We also apply the proposed methods to detect changes in mean direction for hourly wind-wave direction measurements and the path of the cyclonic storm "Biporjoy," which occurred between 6th and 19th June 2023 over the Arabian Sea, western coast of India.

31 . THE IMPACT OF TREATMENT DISCONTINUATION DUE ADVERSE EVENTS ON EFFICACY IN AN ONCOLOGY CLINICAL TRIAL

[01.M2.I3, (page 4)]

Hiya BANERJEE, *Eli Lilly*

Randomized clinical trials persist as the gold standard for evaluating the efficacy and safety of new treatments. However, in clinical trials of reasonable size and duration, it is common for some patients to deviate from their assigned study treatment due to various reasons. One such deviation is treatment discontinuation resulting from adverse events. The lingering question revolves around how to estimate the impact of treatment discontinuation due to adverse events on the efficacy of a drug. Motivated by a question, the talk will present how principal stratum analysis can help in addressing this question for an oncology clinical trial.

32 . Adaptive Estimation of the Transition Kernel of Controlled Markov Chains

[02.E1.C6, (page 27)]

Imon BANERJEE, *northwestern university*
Imon BANERJEE, *Northwestern University*
Harsha HONNAPPA, *Purdue University*
Vinayak RAO, *Purdue University*

We will present estimation bounds on nonparametric estimates of the transition kernels of Controlled Markov chains (CMC's). CMC's are a natural choice for modelling various industrial and medical processes, and are also relevant to reinforcement learning (RL). Therefore, learning the transition dynamics of CMC's in a sample efficient manner is an important question. We will attempt to answer this question when the underlying state-control space is both finite and infinite. Under finiteness, we will develop a Probably Approximately Correct (PAC) bound for estimation error. We will explore the additional challenges when the state-control space is infinite, and tackle them using techniques from adaptive estimation. At the end, we will also posit some open questions.

33 . A semi-supervised and transfer learning support vector machine for predicting adherence to psychotherapy in mood disorders

[01.A1.I13, (page 8)]

Samprit BANERJEE, *Weill Medical College of Cornell University*

TBA

34. Exploration-driven networks

[02.A1.I40, (page 26)]

Sayan BANERJEE, *University of North Carolina, Chapel Hill*

TBA

35 . Outliers and their detection in sparse regression

[04.E1.C26, (page 57)]

Shuvayan BANERJEE, *IIT Bombay*
Ajit RAJWADE, *IIT Bombay*
James SAUNDERSON, *Monash University*
Radhendushka SRIVASTAVA, *IIT Bombay*

In a regression setup, outliers are extreme observations away from the data density. In a high-dimensional regression setup, the visualization of data density and detection of outliers is a challenging statistical query. Often in such scenarios, a com-

monly used model to detect the outliers is given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{X}$ is a $n \times p$ design matrix, $\boldsymbol{\beta}$ is the regression parameter and $\boldsymbol{\delta}$ is the outlier vector. Further, we assume that $\boldsymbol{\epsilon}$ is a n-dimensional Gaussian noise vector. The outlier vector $\boldsymbol{\delta}$ is sparse and its non-zero elements are the outlier observations in the data. Under the setup $n < p$ in sparse regression, LASSO is a popular estimation method for the regression parameters. However, the sampling distribution of the LASSO estimator under this setup is not available for statistical inference. In this work, we describe a new estimation strategy for the regression parameters based on LASSO estimator. We will also describe a consistent statistical test based on this estimator to detect the non-zero elements of the outlier vector $\boldsymbol{\delta}$. An illustration of the goodness of the proposed outlier detection method will be shown.

36. Artificially Intelligent Data Analysis: A Case Study in Spatial Energetics

[Plenary Lecture 2, (page 20)]

Sudipto BANERJEE, *University of California Los Angeles*

I will share my perspectives on the significant paradigm shift taking place in data analysis with the advent of AI technologies. This rapidly evolving field offers substantial intellectual space for statistical theory and methods to not only co-exist with other disciplines within computer science and machine learning, but also play a crucial role in advancing data analysis and probabilistic inference at unprecedented scales. I will elucidate three ideas that will synthesize into an artificially intelligent inferential system. The first is "amortized Bayesian inference" that considers training and calculating posterior distributions using generative AI. The second is Bayesian transfer learning for scaling Inference to massive datasets. The third is Bayesian predictive stacking that delivers exact simulation-based inference without resorting to expensive iterative methods such as Markov chain Monte Carlo. These ideas will be synthesized to offer a real time analysis of a case study in spatial energetics. Spatial energetics is broadly referred to as the study of live movement in real time and has been especially relevant in the context of mobile health data using actigraph units embedded in wearable devices. The case study is a part of the University of California Los Angeles Physical Activity and Sustainable Transportation Approaches (PASTA-LA) being conducted by the UCLA Fielding School of Public Health and is primarily concerned with learning about a subject's metabolic levels as a function of their mobility attributes and other health attributes.

37 . Skill Dominance Analysis of Two(Four)-player, Three(Five)-dice Variant of the Ludo Game

[04.A1.C22, (page 53)]

Tathagata BANERJEE, *Indian Institute of Technology, Kanpur*
Diganta MUKHERJEE, *Indian Statistical Institute, Kolkata*

This paper examines two different variants of the Ludo game, involving multiple dice and a fixed number of total turns. Within each variant, multiple game lengths (total number of turns) are considered. To compare the two variants, a set of intuitive, rule-based strategies is designed, representing different broad methods of strategic play. Game play is simulated between bots (automated software applications executing repetitive tasks over a network) following these strategies. The expected results are computed using certain game theoretic and probabilistic explanations, providing a way of interpreting the performance of the different strategies.

The different strategies are further analyzed using win percentage in a large number of gameplay simulations, and Nash Equilibrium strategies are computed for both variants for a varying number of total turns. The Nash Equilibrium strategies across different game lengths are compared. A clear distinction between performances of strategies is observed, with more sophisticated strategies outperforming the naive one. A gradual shift in optimal strategy profiles is observed with changing game length, and certain sophisticated strategies are observed to even confound each other's performance while playing against each other.

38. Maximum Precision Estimation for a Step-Stress Model Using Two-Stage Methodologies

[04.A1.I79, (page 51)]

Sudeep BAPAT, *Indian Institute of Technology Bombay*
Yan ZHUANG, *Connecticut College*

A two-stage sequential procedure to estimate the parameters of a cumulative exposure model under an

accelerated testing scenario is discussed. We focus on a step-stress model where the stress level is updated after a pre-specified number of failures occur, which is also random. This is termed as the 'random stress change time' in the literature. To obtain maximum precision, a certain variance optimality criterion is applied. A pseudo real data example from reliability studies is also analyzed to outline the performance of the proposed methodology.

39. Quasi-maximum likelihood estimation for causal time series

[03.A1.I54, (page 36)]

Jean-Marc BARDET, *SAMM, Université Paris 1 Panthéon-Sorbonne, France*

This talk will provide an overview of the latest results on the use of the quasi-maximum likelihood estimator (QMLE) for causal time series. We will start by reviewing the case of stationary processes with short memory (typically ARMA or GARCH processes), then those with infinite short memory, and finally for linear processes with long memory. Two extensions will also be proposed: the use of a non-Gaussian (Laplacian) QMLE, and the use QMLE as non-parametric estimator for locally stationary processes.

40. Destructive cure models with proportional hazards lifetimes and associated likelihood inference

[02.M1.I24, (page 16)]

Sandip BARUI, *Indian Statistical Institute*
Narayanaswamy BALAKRISHNAN, *McMaster University*

In survival analysis, cure models have gained much importance due to rapid advancements in medical sciences. More recently, a subset of cure models, called destructive cure models, have been studied extensively under competing risks scenario wherein initial competing risks undergo a destructive process. In this article, we study destructive cure models by assuming a flexible weighted Poisson distribution (exponentially weighted Poisson, length biased Poisson and negative binomial distributions) for the initial number of competing causes and lifetimes of the susceptible individuals being defined by proportional hazards. The expectation-maximization (EM) algorithm and profile likelihood approach are made use of to estimate the model parameters. An extensive simulation study is carried out under various parameter settings to examine the properties of the models, and accuracy and the robustness of the proposed estimation technique. Effects of model mis-specification on the parameter estimates are also discussed in detail. For further illustration of the proposed methodology, a real-life cutaneous melanoma data set is analyzed.

41. A New Zero Truncated Discrete Odd Lindley Half Logistic distribution with its Application to real data

[02.E1.C8, (page 29)]

Shabana BASHEER, *Maharaja's College, Ernakulam*
Dr. Ansa ALPHONSA ANTONY, *St. Xavier's College for Women, Aluva*

During the early stages of Survival Analysis, discussions were mainly confined to continuous models. While the significance of continuous models persists, the discretization of continuous distributions plays a crucial role in the study of lifetime data.

Zero-Truncated discrete distributions are important in statistical analysis because they model data that do not include zero as a possible outcome. These distributions are particularly useful in scenarios where the observed data inherently excludes the possibility of a zero count, such as modelling count data, Survival Analysis, Actuarial science, Ecology, Biology, etc. The present study proposes a zero-truncated version of discrete Odd-Lindley half Logistic distribution, exploring its vital distributional and reliability characteristics. Key functions like its Cumulative Distributive Function (CDF), Survival function, Hazard rate function, Reverse hazard rate function, Moment generating function, Residual life function, Mean residual life function and Variance residual life function have been derived. Parameters are estimated using the maximum likelihood method. Monte-Carlo simulation is carried out to check the efficiency of maximum likelihood method. Finally, distribution is applied to some real data sets to illustrate the goodness of fit of the proposed distribution. It can serve as an alternative model to other lifetime distributions in the existing statistical literature for modelling positive real data in many areas. Plots of the Probability mass functions of the distribution is also presented to illustrate the behaviour of the distribution for different values of the parameters.

42. Quasi-randomization tests for network interference

[03.E1.I61, (page 39)]
Pallavi BASU, *Indian School of Business*
Supriya TIWARI, *ISB*
Pallavi BASU, *ISB*

Many classical inferential approaches fail to hold when interference exists among the population units. This amounts to the treatment status of one unit affecting the potential outcome of other units in the population. Testing for such spillover effects in this setting makes the null hypothesis non-sharp. An interesting approach to tackling the non-sharp nature of the null hypothesis in this setup is constructing conditional randomization tests such that the null is sharp on the restricted population. In randomized experiments, conditional randomized tests hold finite sample validity. Such approaches can pose computational challenges as finding these appropriate subpopulations based on experimental design can involve solving an NP-hard problem. In this paper, we view the network amongst the population as a random variable instead of being fixed. We propose a new approach that builds a conditional quasi-randomization test. Our main idea is to build the (non-sharp) null distribution of no spillover effects using random graph null models. We show that our method is exactly valid in finite-samples under mild assumptions. Our method displays enhanced power over other methods, with substantial improvement in complex experimental designs. We highlight that the method reduces to a simple permutation test, making it easy to implement in practice. We conduct a simulation study to verify the finite-sample validity of our approach and illustrate our methodology to test for interference in a weather insurance adoption experiment run in rural China.

43. Unified Competing Risks Mixture Cure Model for Cancer Survival

[02.M1.I24, (page 16)]
Sanjib BASU, *Division of Epidemiology and Biostatistics, University of Illinois Chicago*

TBA

44. Detecting emotionally stressful periods from passive sensing data via mobile devices

[01.A1.I13, (page 8)]
Sumanta BASU, *Cornell University*
Younghoon KIM, *Cornell University*
Samprit BANERJEE, *Weill Cornell Medical College*

We develop a data-driven segmentation algorithm for detecting periods of emotional stress in mid and old-aged patients with chronic pain and depression undergoing therapy. Our method leverages the association between time series of daily self-reported levels of stress and measures of physical activities passively collected from smartwatches. This type of time series data is being collected more often in recent years, and learning such associations can potentially help intervene during future periods of emotional stress by tracking changes in passively sensed physical activity levels. These time series are typically nonstationary, and meaningful associations often exist only over short windows of time. Traditional machine learning (ML) methods, applied globally on the entire time series, often miss these time-varying local patterns. Our method first segments the passive sensing time series by detecting change points, and then looks for segment-specific associations with stress levels to identify *co-segmented periods* that exhibit distinct associations between stress and physical activity. By changing our unit of analysis from individual time points *data-driven segments of time*, and allowing for different associations in different segments, our algorithm helps detect patterns that only exist over short-time windows. We use our method to detect periods of stress in data of patients collected in ALACRITY Phase I study. We find that our data-driven segmentation algorithm can detect stress periods more accurately than traditional ML methods that do not rely on segmentation.

45. Statistics 2 in Pediatric Drug Development

[01.A1.I11, (page 7)]
Ramakrishna BATTULA, *Eli Lilly*

TBA

46. Robust Tests for Latent and Simultaneous Spatial Autoregressive Tobit Model with Spatial Autoregressive Disturbances

[03.M2.I48, (page 34)]
Anil BERA, *University of Illinois at Urbana–Champaign*
Chang LU,

The spatial autoregressive (SAR) model is one of the most important models in spatial econometrics to describe the connection between dependent and independent variables by taking account of spatial de-

pendence. While the traditional SAR model has been extensively researched in the literature, researchers in the field of spatial analysis are becoming increasingly interested in spatial models with limited dependent variables (LDV). The model specification problem for the SAR model with LDV is investigated in this paper. Our suggested tests can be viewed as a generalization of the widely used tests proposed in Anselin et. al (1996) to LDV set up. This paper proposes robust tests for SAR Tobit models with SAR disturbances (SARAR Tobit model). Two types of Tobit models are considered. One is the latent SARAR Tobit model, and the other one is the simultaneous SARAR Tobit model. The difference between these two types of SARAR Tobit model is whether a certain area's data is affected by the actual data of its neighbors or by the latent data of its neighbors. The paper includes score functions, information matrices, and proofs of asymptotic distributions for the suggested test statistics. The finite sample size and power are also investigated through an extensive simulation study. Our simulated results emonstrate that these proposed tests have good finite sample properties both in terms of size and power, even in the presence of misspecification. We also provide an empirical application to illustrate the usefulness of our tests.

47. Fiducial inference on lifetime performance index based on type-II censored samples

[02.E1.C7, (page 28)]

KUNTAL BERA, *Indian Statistical Institute, Kolkata*
M.Z. ANIS, *Indian Statistical Institute*

The lifetime performance index C_L measures the performance of a process in terms of the lifetime of produced items. In this article, we use the fiducial approach to estimate the lower confidence bound of the lifetime performance index under some location-scale distributions. We consider three important lifetime distributions: two-parameter exponential, two-parameter Rayleigh and Weibull distribution. Often lifetime data comes as censored data instead of complete data. Here we consider type-II censored samples. The fiducial quantities are estimated based on the MLE. The performance of the proposed fiducial approach is analyzed in terms of the coverage probability (CP) and average value (AV) of the estimated lower confidence bound. It is observed through the discussion that the CP of the proposed confidence intervals is very close to the nominal confidence level. Hence the proposed method estimates a very reliable confidence interval. An example is given to illustrate the application of the proposed confidence interval.

48. Deep Kernel Posterior Learning under Infinite Variance Prior Weights

[03.M2.I51, (page 35)]

Anindya BHADRA, *Purdue University*
Jorge LORIA, *Aalto University*

Neal (1996) proved that infinitely wide shallow Bayesian neural networks (BNN) converge to Gaussian processes (GP), when the network weights have bounded prior variance. Cho and Saul (2009) provided a useful recursive formula for deep kernel processes for relating the covariance kernel of each layer to the layer immediately below. Moreover, they worked out the form of the layer-wise covariance kernel in an explicit manner for several common activation functions, including the ReLU. Subsequent works have made the connection between these two works, and provided useful results on the covariance kernel of a deep GP arising as wide limits of various deep Bayesian network architectures. However, recent works, including Aitchison et al. (2021), have highlighted that the covariance kernels obtained in this manner are deterministic and hence, precludes any possibility of representation learning, which amounts to learning a non-degenerate posterior of a random kernel given the data. To address this, they proposed adding artificial noise to the kernel to retain stochasticity, and developed deep kernel Wishart and inverse Wishart processes. Nonetheless, this artificial noise injection could be critiqued in that it would not naturally emerge in a classic BNN architecture under an infinite-width limit. To address this, we show that a Bayesian deep neural network, where each layer width approaches infinity, and all network weights are elliptically distributed with infinite variance, converges to a process with α-stable marginal in each layer that has a conditionally Gaussian representation. These random covariance kernels could be recursively linked in the manner of Cho and Saul (2009), even though marginally the process exhibits stable behavior, and hence covariances are not even necessarily defined. Our results also provide useful generalizations of the recent results of Loria and Bhadra (2024) on shallow networks, to multilayer networks, and remedies the prohibitive computational burden of their approach. The computational and statistical benefits over competing ap-

proaches stand out in simulations and in a demonstration on benchmark data sets.

49. BIPPS: Bayesian Inference for Point Patterns in Space applied to Multiplex imaging data

[01.M2.I1, (page 3)]

Sagnik BHADURY, *University of Michigan*

TBA

50. The Impact of Data Quality on Predictive Model Performance - A Simulation Study

[01.M2.I4, (page 5)]

Nivedita BHAKTHA, *IIT Kanpur*
Veena BANSAL,

In survey research, low-quality data—defined as responses resulting from careless, insufficient effort, or random answering—poses a significant challenge as it fails to accurately reflect the true position of respondents on measured constructs. Recent literature has increasingly focused on detecting such low-quality responses. However, a critical gap remains in understanding how low-quality data affects predictive models, such as logistic regression, Naive Bayes classifiers, Support Vector Machines (SVM), Artificial Neural Networks (ANN), classification trees, and other supervised classification methods. This study examines the impact of low-quality data on predictive model performance through a comprehensive simulation study. We manipulated several factors: the number of features, correlation among features, the strength of the relationship between the outcome variable and features, sample size, type of low-quality data (including mid-point responding, random responding, extreme style responding, acquiescent responding, and a mixture of these types), and the percentage of low-quality responses in the sample. Various classification models were fit to the simulated datasets having the above-mentioned data generation characteristics.

Our simulation study systematically examines how these factors influence model metrics such as Area Under the Curve (AUC), accuracy, F1 score, precision, and recall. The findings indicate that even a small percentage of low-quality data can introduce significant biases into predictive models. This research underlines the importance of addressing data quality in survey-based classification modelling by providing empirical evidence on the extent to which low-quality data can affect predictive accuracy. Our results suggest that model developers and researchers need to implement robust data quality checks and develop methodologies to mitigate the adverse effects of low-quality data.

51 . Transformed jackknife empirical likelihood inference for Bergsma's Covariance Coefficient

[03.M2.I52, (page 35)]

Deepesh BHATI, *Central University of Rajasthan*
Sudheesh KATTUMANNIL, *ISI Chennai*
Isha DEWAN, *ISI Delhi*

Bergsma (2006) has recently introduced correlation coefficient having potential to detect arbitrary forms of association between two real random variables X and Y. In this work, we use jackknife empirical likelihood (JEL) and its variants to construct its confidence interval. Simulation results under various distributions and comparison of performance of these methods in terms of the coverage probability and average length is demonstrated. Finally the applicability of the proposed test is presented using real world data set.

52 . Exploring Block Clustering with Probabilistic Distance: Theory and Validation

[Student Poster Competition, (page 22)]

Shrirksihna BHAT K, *Pondicherry University*
Shrikrishna BHAT K, *Department of Statistics, Pondicherry University*
Kiruthika C, *Department of Statistics, Pondicherry University*

Probabilistic Distance (PD) clustering is a versatile method that has garnered significant attention in cluster analysis due to its probabilistic framework, which integrates distance measures and cluster membership probabilities. We propose a novel block clustering framework and algorithm based on this approach. Our framework is validated using Euclidean distance measures for continuous data. We introduce a new metric, the Average Posterior Probability of the Alternative Block (PAB) Silhouette Index, alongside the established Co-clustering Adjusted Rand Index (CARI) to assess clustering performance. This comprehensive evaluation demonstrates the effectiveness of our proposed framework in enhancing block clustering methodologies.

53. Optimizing patient eligibility criteria in clinical trials using real world evidence (RWE)

[04.M2.I71, (page 47)]

Abhishek BHATTACHARJEE, *Pfizer Research and Development, Pfizer*

TBA

54. On weighted failure rate: an alternative approach to study weighted distributions

[04.M1.I66, (page 45)]

Subarna BHATTACHARJEE, *Department of Mathematics, Ravenshaw University, Cuttack*
S. M. SUNOJ, *Cochin University of Science and Technology*

When sample observations are not equally likely, we use the weighted measures to capture the significance of their relative importance. Choosing appropriate weights, we compute various measures in a better way by giving appropriate weights based on the sample mechanism. Such biased sampling schemes are usually employed in observational studies either due to its convenience or its cost-effectiveness. Based on this, C R. Rao (1965) [Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment. In Classical and Contagious Discrete Distributions. Patil, G. P. ed. Pergamon Press and Statistical Publishing Society, Calcutta, 320-332. http://www.jstor.org/stable/25049375] identified the concept of weighted distributions in connection with the modeling statistical data, in situations where the usual practice of employing standard distributions for the purpose was not found appropriate. In this paper, we define weighted failure rate and their means from the view point of an application.

We derive some parametric and non-parametric characterization results. We discuss on the form invariance property of baseline failure rate for a specific choice of weight function. Some bounds on means of aging functions are obtained. We establish that weighted IFRA class is not closed under formation of coherent systems unlike the IFRA class. We begin by emphasizing that the formation of independent component series system having weighted failure rates with sum of weight functions being unity is same as a mixture of n distributions.

An interesting application of the present work is credited to the fact that the quantile version of means of failure rate is obtained as a special case of weighted means of failure rate. We observe that the definition of weighted concept proposed in this paper corroborates with the actual functions, namely failure rate, survival function, density function in a better way than the existing weighted concepts proposed by C. R Rao (1965).

Keywords and Phrases: weighted distribution, weighted failure rate, weighted arithmetic mean failure rate, weighted geometric mean failure rate, weighted harmonic mean failure rate

55. Kernel and Graphical Methods for Comparing Conditional Distributions

[Special Invited Session 1, (page 14)]

Bhaswar BHATTACHARYA , *University of Pennsylvania*
Anirban CHATTERJEE, *University of Pennsylvania*
Ziang NIU , *University of Pennsylvania*

In this talk we will discuss various nonparametric methods for comparing conditional distributions based on kernels and nearest-neighbor graphs. The methods can be readily applied to a broad range of problems, ranging from classical nonparametric statistics to modern machine learning. Specifically, we will discuss applications in testing model calibration, regression curve evaluation, and validation of emulators in simulation-based inference.

(Joint work with Anirban Chatterjee and Ziang Niu)

56. A marginal structural model for partial compliance in SMARTs

[03.E1.I61, (page 39)]

Indrabati BHATTACHARYA, *Florida State University*
William J. ARTMAN, *University of Rochester*
Ashkan ERTEFAIE, *University of Rochester*
Brent A. JOHNSON, *University of Rochester*

The cyclical and heterogeneous nature of many substance use disorders highlights the need to adapt the type and/or the dose of treatment to accommodate the specific and changing needs of individuals. The Adaptive Treatment for Alcohol and Cocaine Dependence study (ENGAGE) is a sequential multiple assignment randomized trial (SMART) that aimed to provide longitudinal data for constructing dynamic treatment regimes (DTRs) to improve patients' engagement in therapy. However, the high rate of noncompliance and lack of analytic tools to account for noncompliance has impeded researchers from using

the data to achieve the main goal of the trial; namely, the construction of individually tailored DTRs. We overcome this issue by defining our target parameter as the mean outcome under different DTRs for given potential compliance strata and propose a marginal structural model with principal stratification to estimate this quantity. We model the latent principal strata using a Bayesian semiparametric approach. An important feature of our work is that we consider partial rather than binary compliance strata which is more relevant in longitudinal studies. We assess the performance of our method through simulation. We illustrate its application on the ENGAGE study and demonstrate that the optimal DTRs depend on compliance strata compared with ignoring compliance information as in intention-to-treat analyses.

57. Mean-field methods under interference

[02.A1.I40, (page 26)]

Sohom BHATTACHARYA, *University of Florida*

Estimating causal effects from observational data under interference, where the interference is structured by a known network, has been a subject of significant research. In this talk, I will adopt the chain graph framework, which allows the outcome of a unit to depend on treatments assigned to distant units in the interference network. Specifically, I will focus on two types of interaction networks. For "mean-field" interaction networks, I will introduce a scalable iterative algorithm for causal effect estimation. For Gaussian-weighted networks, I will discuss an algorithm based on Approximate Message Passing. Our work establishes the consistency of parameter estimation using maximum pseudo-likelihood and demonstrates that the downstream causal effect estimators, obtained by integrating these estimated parameters into the aforementioned algorithms, are also consistent. These methods accommodate dense interactions among units, a challenge that existing techniques have struggled to address. Inspired by advances in variational inference, this talk will demonstrate the usefulness of these ideas in the context of causal effect estimation under interference. The talk is based on joint work with Subhabrata Sen.

58. BMW: Inlier Prone Bayesian Models for Correlated Bivariate Data

[04.E1.C23, (page 55)]

Sumangal BHATTACHARYA, *Indian Institute of Technology Tirupati, India*
Ishapathik DAS, *Indian Institute of Technology Tirupati*

In various fields of experiments, particularly prevalent in the lifetime of electronic components, clinical trials, rainfall data for specific locations, and insurance claim data, there are instances of zero or near-zero observations. These occurrences can be due to faulty construction, substandard quality, nonresponse to treatments, scant rain during the monsoon season, or a reduced accident rate. Typically, these observed responses are defined as instantaneous and nearly instantaneous failure observations and are known as inliers in the data. Traditional parametric distributions, such as Weibull, Pareto, exponential, gamma, and log-normal, are inaccurate for modeling life data containing inliers because such failures usually invalidate the assumption of an unimodal distribution. Many mixture distributions have been proposed in the literature to model such scenarios. Similarly, certain bivariate correlated observations, such as twin birth data, neonatal birth data, treatment responses for both eyes, and rainfall data from spatially correlated locations, exhibit a concentration of data points around a specific point (x_0, y_0) in the first quadrant, while the remaining observations follow a continuous distribution. In bivariate data, we identify these concentrated responses as inliers. In nature, correlated bivariate data can include inliers, and the region of these inliers is not restricted to a square shape. To address this, we utilize copulas to derive the Bivariate Modified Weibull (BMW) distribution, designed to capture correlated observations in the presence of inliers. This approach uses the Modified Weibull distribution as the marginal distribution, formulating it through a mixture model that combines a uniform distribution around a specific point with two parametric Weibull distributions. Moreover, we introduce a methodology for testing the presence of inliers within datasets using the supremum of Bayes factors (SBF). We support this with an extensive simulation study to assess the power of the test. The results, depicted using a confusion matrix, show that the accuracy of the test exceeds 98%. Estimation of the model parameters is carried out through a Bayesian framework, selecting suitable priors for each parameter. Given the model complexities and the dimensional breadth of the parameter space, we employ Hamiltonian Markov Chain Monte Carlo (MCMC) techniques, specifically utilizing the No-U-Turn sampler (NUTS) for efficient posterior sampling

of parameters. To demonstrate the applicability and assess the performance of our approach, we provide numerical examples and apply the method to analyze monsoon rainfall data from Assam, a popular tourist destination in India. In accordance with the prescribed testing methodology, our analysis indicates the presence of inliers within the rainfall dataset, as indicated by the SBF value exceeding 150. Our proposed model significantly fits the rainfall data better than the base model, as evidenced by the criteria of DIC and BIC.

59. Analyzing Political Polarization in the Presence of Partially Observed Social Networks

[03.E1.I62, (page 39)]

Sharmodeep BHATTACHARYYA, *Oregon State University*

We propose to investigate how political polarization and partisan identity have led individuals to engage in partisan behaviors and, in turn, have affected subsequent policy-making. By attaining a clearer understanding of the extent and nature of partisan behaviors and the underlying causal channels through which partisanship dictates human behaviors, we seek to contribute more accurate approaches to develop socio-political models in the presence of the complexities mentioned above and estimate the associated parameters in the context of public policy frameworks. Identification of underlying broader drivers of these changes that manifested in recent years would help better predict the public's reception of public policies and the efficacy of policy measures. The statistical methods would be developed based on the analysis of ANES and CCES survey data on political leanings and socio-demographic information.

60. Predicting patient mortality in the presence of competing risks

[03.A1.C12, (page 38)]

Amol Ajit BHAVE, *S&P Global*
Sandip Kumar PAL, *S&P Global*

Predicting patient mortality in the presence of competing risks is a critical task in healthcare, providing essential insights for medical decision-making and resource allocation. This study leverages the Medical Information Mart for Intensive Care (MIMIC) database to develop a robust model for mortality prediction, considering the presence of competing risks associated with different causes of death. Given the complexity of patient data, we employ a piecewise linear hazard function in conjunction with a competing risks model to effectively capture the dynamics of patient survival.

Our approach segments the hazard function into discrete intervals, allowing for varying hazard rates across different time periods, which enhances the model's adaptability to the varying risk profiles over time. The competing risks model is incorporated to account for the different causes of death (e.g., cardiovascular failure, respiratory failure, infections) and their impact on the hazard rates.

We preprocess the MIMIC data to ensure high-quality inputs, incorporating demographic, clinical, and treatment variables. Using a piecewise linear hazard model combined with a competing risks framework, we estimate the risk of mortality while accurately accounting for the presence of multiple competing causes of death. The model's performance is evaluated through extensive experiments, demonstrating its predictive accuracy and stability. Our results show that the integration of the piecewise linear hazard function with the competing risks model provides a flexible and powerful framework for understanding patient mortality, offering significant improvements over traditional methods.

The proposed model not only advances the predictive capabilities in clinical settings but also offers a comprehensive tool for healthcare professionals to better manage patient care and outcomes. Future work will focus on integrating more granular patient data and exploring the model's applicability in different clinical environments to further enhance its utility and generalizability.

61. Exploring Data Science Foundations in Engineering Education: Leveraging Computer Simulations for Pedagogical Advancement

[02.A1.I35, (page 24)]

Amiya Ranjan BHOWMICK, *Institute of Chemical Technology, Mumbai*
Riddhi BHARANI, *Vivekanand Education Society's College of Arts, Science and Commerce (Autonomous), Mumbai*

The rapid rise of data science and machine learning tools has led to increased reliance on pre-built libraries like Python and R, often at the expense of understanding fundamental statistical concepts. Many students and professionals focus on running code, overlooking core ideas such as Central Limit Theo-

rem or Weak Law of Large Numbers. Short-term courses further promote superficial learning, prioritizing speed over depth. Educators face the challenge of preparing students for an interdisciplinary landscape where statistical and mathematical foundations are crucial. Traditional textbooks, heavily focused on mathematics, often feel inaccessible, and the reduced emphasis on math in engineering curricula exacerbates the problem. This article advocates for using computer simulations to foster an interdisciplinary learning experience. Simulations offer a hands-on approach to teaching key statistical concepts, allowing students to explore sampling distributions and statistical uncertainty through experimentation. By implementing basic programming constructs, educators can bridge theory and application, encouraging deeper understanding across disciplines. This approach promotes an accessible and engaging way to teach foundation principles, preparing students to handle real-world challenges in both academic research and industry settings. Implications of such active learning strategy will be discussed in the light of the NEP 2020. Keywords: Simulation based learning, Interdisciplinary learning, Collaborative programming, Pedagogy

62. Estimation of Population Size with Heterogeneous Catchability and Behavioural Dependence: Applications to Air and Water Borne Disease Surveillance

[04.M1.I67, (page 45)]

Prajamitra BHUYAN, *Indian Institute of Management Calcutta*
Kiranmoy CHATTERJEE, *Bidhannagar College, Kolkata*

Population size estimation based on the capture-recapture experiment is an interesting problem in various fields including epidemiology, criminology, demography, etc. In many real-life scenarios, there exists inherent heterogeneity among the individuals and dependency between capture and recapture attempts. A novel trivariate Bernoulli model is considered to incorporate these features, and the Bayesian estimation of the model parameters is suggested using data augmentation. Simulation results show robustness under model misspecification and the superiority of the performance of the proposed method over existing competitors. The method is applied to analyse real case studies on epidemiological surveillance. The results provide interesting insight on the heterogeneity and dependence involved in the capture-recapture mechanism. The methodology proposed can assist in effective decision-making and policy formulation.

63. A generalized Pegram's operator based autoregressive (GPAR) process for modelling categorical time series

[05.M1.I90, (page 60)]

Atanu BISWAS, *Indian Statistical Institute Kolkata*

TBA

64. Statistical Test for Diagnosing Rotational Symmetry in Rotation Data with Applications in Materials Science

[02.M2.I30, (page 19)]

Eva BISWAS, *Iowa State University*
Daniel NORDMAN, *Iowa State University*
Ulrike GENSCHEL, *Iowa State University*

Orientation data are commonly encountered in various studies, such as human kinematics and materials science, where each observation is represented by a 3x3 rotation matrix that describes an object's position within a three-dimensional reference frame. In many cases, rotationally symmetric distributions are assumed for modeling purposes. This approach considers an orientation as a random isotropic perturbation of an underlying location parameter, effectively creating a location model for random matrices with distributionally symmetric errors. However, despite the widespread use of these models, no formal method has been available to adequately assess the validity of such symmetry assumptions. This talk introduces a valid test statistic for diagnosing rotational symmetry in orientation data. The test evaluates distributional symmetry by analyzing the joint distribution of three random variables derived from the orientation, which can be tested using empirical characteristic functions. Numerical studies demonstrate that the proposed test maintains the correct size while exhibiting high power against various deviations from symmetry. The testing method is further illustrated with orientation data collected in texture analysis from materials science.

65. Analysis of two-component load-sharing systems in the presence of frailty

[03.E1.C18, (page 43)]

Shilpi BISWAS, *IIT Guwahati*
Debanjan MITRA, *IIM Udaipur, Rajasthan*

Ayon GANGULY, *IIT Guwahati, Assam*

In this work, we consider analysis of data obtained from load-sharing systems in the presence of frailty. Frailty can be used to explain unobserved heterogeneity among systems. We used generalized Freund bivariate model, which is a family of models, to explain the stochastic behaviour of load-sharing system. The frailty is modelled using generalized gamma distribution, which is also a family of distributions. The use of families help to choose an appropriate model in the family, which gives flexibility in the analysis. We have considered likelihood inference using expectation-maximization algorithm. Confidence intervals of the model parameters are obtained using Louis missing information principle. Selection of appropriate model for a given dataset is discussed. Extensive simulation is performed to check the effectiveness of the proposed methods. The results are quite satisfactory in most of the cases. Finally, a data analysis is performed for illustrative purpose.

66. A multiplex network approach for characterizing stock market dynamics

[04.A1.I81, (page 52)]

Soumyajyoti BISWAS, *Assistant Professor, Department of Physics & Department of Computer Science and Engineering, SRM University*

TBA

67. A novel non-linear intervention approach using time series model and Horel function

[Student Poster Competition, (page 22)]

Subhankar BISWAS, *ICAR-Indian Agricultural Research Institute*
Amrit Kumar PAUL, *Principal Scientist and Head, Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute*
Ranjit PAUL, *Senior Scientist, Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute*
Md YEASIN, *Scientist, Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute*

Intervention analysis is used to investigate structural changes in data produced by external events. Analysis of time series with intervention is an important area of research in recent times. Whenever any man-made or natural event occurs which affects the behaviour of time series data, it is known as an intervention. There are several existing methods to study the effect of intervention. In the past, researchers mainly used the intervention effect as a Step function, Ramp function or as a combination of Step and Ramp functions, which are linear in nature. But in our study, we have proposed a new approach to study the effect of intervention in time series data by utilizing non-linear models. This study proposes a novel algorithm for the analysis of intervention effect using stochastic model, viz., Autoregressive Integrated Moving Average model with exogeneous variables (ARIMA-X) and machine learning model viz. Artificial Neural Networks with exogeneous variables (ANN-X). The Horel function has been introduced as non-linear intervention function. In our study, we have considered the intervention of a ban on blending in Mustard. The ban on blending in Mustard came into force in October 2020. The proposed methodology has been evaluated for different datasets and it is evident from the results that proposed methodology performs better than the existing one in terms of accuracy and performance.

68. Estimation of Spectral Risk Measure for Left-Truncated and Right-Censored Data

[05.M1.I86, (page 59)]

Suparna BISWAS, *Indian Statistical Institute Bangalore*
Rituparna SEN, *Indian Statistical Institute Bangalore*

Left-truncated and right-censored data are frequently encountered in insurance loss data due to deductibles and policy limits. Risk estimation is an important task in insurance, as it is a necessary step for determining premiums under various policy terms. Spectral risk measures are inherently coherent and have the benefit of connecting the risk measure to the user's risk aversion. In this paper, we study the estimation of spectral risk measure based on left-truncated and right-censored data.We propose a non-parametric estimator of the spectral risk measure using the product limit estimator and establish the asymptotic normality for our proposed estimator. We also develop an Edgeworth expansion of our proposed estimator. The bootstrap is employed to approximate the distribution of our proposed estimator and is shown to be second-order "accurate". Monte Carlo studies are conducted to compare the proposed spectral risk measure estimator with the existing parametric and non-parametric estimators for left-truncated and right-censored data. Based on our

simulation study, we estimate the exponential spectral risk measure for three data sets: the Norwegian fire claims data set, the Spanish automobile insurance claims data set, and the French marine losses data set.

69. Changepoint problem with angular data using a measure of variation based on the intrinsic geometry of torus

[05.M1.C30, (page 61)]

SUROJIT BISWAS, *Indian Institute of Technology Kharagpur*
BUDDHANANDA BANERJEE , *Department of Mathematics, Indian Institute of Technology Kharagpur*
ARNAB KUMAR LAHA, *Operations and Decision Sciences Area, Indian Institute of Management Ahmedabad*

In many temporally ordered data sets, it is observed that the parameters of the underlying distribution change abruptly at unknown times. The detection of such changepoints is important for many applications. While this problem has been studied substantially in the linear data setup, not much work has been done for angular data. In this article, we utilize the intrinsic geometry of a torus to introduce the notion of the "square of an angle" and use it to propose a new measure of variation, called the "curved variance," of an angular random variable. Using the above ideas, we propose new tests for the existence of changepoint(s) in the concentration, mean direction, and/or both of these. The limiting distributions of the test statistics are derived, and their powers are obtained using extensive simulation. It is seen that the tests have better power than the corresponding existing tests. The proposed methods have been implemented on three real-life data sets, revealing interesting insights. In particular, our method, when used to detect simultaneous changes in mean direction and concentration for hourly wind direction measurements of the cyclonic storm "Amphan," identified changepoints that could be associated with important meteorological events.

70. Testing synchronization of change-points for multiple time series

[Student Poster Competition, (page 22)]

Soham BONNERJEE, *University of Chicago*
SAYAR KARMAKAR, *UNIVERSITY OF FLORIDA*
MAGGIE CHENG, *ILLINOIS INSTITUTE OF TECHNOLOGY*
WEI BIAO WU, *UNIVERSITY OF CHICAGO*

In this paper, we investigate the problem of detecting synchronization of a single change-point across components of a multivariate time series. The identification of synchronized change-points can often lead to finding a unanimous reason behind such changes whereas rejection might consequently prompt further analysis. Our proposed test statistic is simple to perceive, but its null distribution can be highly nontrivial to explicitly characterize. To overcome this, we employ a Gaussian approximation result, assisted by a clever and agnostic (to the existence of change-point) estimation of covariance matrix. Extensive simulations are provided to corroborate our theoretical results. We also provide two interesting real-world applications and discuss the implications of our findings based on the statistical tests.

71. Bulk Spectrum of Sample Covariance Type Matrices.

[02.M1.I21, (page 14)]

Arup BOSE, *Indian Statistical Institute*

We discuss some high dimensional random matrices such as the Covariance, Autocovariance and Cross covariance matrices. Some results on the behaviour of the bulk will be given. We will also discuss the joint tracial convergence of some of these matrices and present some open questions.

72. Limit theorems via moments and cumulants

[Memorial Session 1, (page 26)]

Arup BOSE, *Indian Statistical Institute*

It is well known that convergence of moments along with some mild conditions yield convergence in distribution. Analogous approach is available using cumulants and has been used extensively in the time series literature. In this talk we first discuss the one to one relation between moments and cumulants, and then show how convergence of cumulants can turn out to be an easier method than convergence of moments. In particular we show how the distributional limit of U statistics (in particular the sample mean) can be easily established using cumulants. Time permitting, we shall discuss some more examples.

73. Characterization and Nonparametric Estimation of Bivariate Weighted Residual Inaccuracy Measure

[01.E1.C5, (page 13)]

Anju C. THOMAS, *Department of Statistics, University of Kerala*
E.I. ABDUL SATHAR, *Department of Statistics, University of Kerala.*

In recent years, there has been a surge in research focused on information-theoretic measures, particularly in entropy and divergence, within the univariate setup. However, many real-world scenarios, such as reliability and life testing experiments, naturally involve bivariate or multivariate data. Also, when modeling statistical data, weighted distributions are typically considered because there are situations where standard distributions are found inappropriate. This work will delve into the bivariate extension of the weighted residual inaccuracy measure. It is extended to the definition of weighted residual inaccuracy measure to a bivariate setup and analyzed its properties, as well as extending this measure to conditionally specified models of two components, referred to as conditional residual inaccuracy measure. The properties of the measure, especially its characterization, are studied. The study also extends to nonparametric estimation of the measure based on the kernel function and explores its asymptotic properties alongside practical demonstrations through real data analysis.

74. Estimation Based on Progressively Type II Hybrid Censored Data

[03.M2.I49, (page 34)]

Manoj CHACKO, *University of Kerala*

In this paper, we consider the problem of estimation of parameters using progressively type-II hybrid censoring, which is a mixture of progressive type-II and hybrid censoring schemes. The Expectation Maximization (EM) method is employed for the parametric estimation. We illustrate this approach using life time distribution and conduct simulation studies to evaluate the performance of the estimators. Additionally, real data examples are provided for demonstration purposes.

75. Bivariate Kullback-Leibler Divergence

[04.E1.C27, (page 57)]

Mary Rafflesia CHACKOCHAN, *Cochin University of Science and Technology*
P. G. SANKARAN, *Cochin University of Science and Technology*

Kullback–Leibler (1951) introduced Kullback–Leibler divergence to measure the distance between two random variables. In the present paper, we introduce bivariate extension of Kullback–Leibler divergence and study its various properties. Bivariate Kullback–Leibler divergence is a useful tool to compute the distance between two bivariate random vectors. Some useful bounds are obtained for bivariate Kullback–Leibler divergence and some are based on entropy and inaccuracy measures. A non parametric estimator for the measure is also proposed and simulation studies are carried out for validation. Finally, the proposed measure is applied to a real life data in survival studies.

76. Top eigenvalues and eigenvectors of inhomogeneous Erdős-Rényi random graphs

[02.M1.I25, (page 16)]

Arijit CHAKRABARTY, *Indian Statistical Institute*
Bishakh BHATTACHARYA, *Indian Statistical Institute*
Sukrit CHAKRABORTY, *Achhruram Memorial College*
Rajat HAZRA, *University of Leiden*

The talk is on eigenvalues outside the spectrum of inhomogeneous Erdős-Rényi random graphs and the corresponding eigenvectors. Depending on the rank of the inhomogeneity kernel generating the random graph, the largest few eigenvalues have a much higher magnitude than that of the bulk. Assuming the rank to be finite, the second order behaviour of those few eigenvalues, after suitable centring and scaling, is shown to be multivariate Gaussian. The asymptotic behaviour of the corresponding eigenvectors is also studied.

77. Change Point Detection for Functional Data Based on the Notion of Maximum Mean Discrepancy

[01.E1.C4, (page 12)]

Sourav CHAKRABARTY, *Indian Statistical Institute*
Anirvan CHAKRABORTY, *Indian Institute of Science Education and Research Kolkata*
Shyamal Krishna DE, *Indian Statistical Institute, Kolkata*

Change point analysis is widely applicable in fields that seek to identify shifts in the underlying distribution of a sequence of time-ordered data. We propose a novel algorithm for detecting change points in functional data setup based on the notion of Maximum Mean Discrepancy (MMD). We are mainly concerned with offline change point detection for iid observations where both number of change points and change

locations are unknown to us. We have proved the consistency of estimated breakfraction in single change point and multiple change point setup. A testing procedure for checking significance of potential change point is proposed and we have proved the consistency of this testing procedure in single change point setup. Also, we have considered the scenario where the number of change points is pre-specified. We prove that if the specified number of change points and true number of change points are equal then the set of estimated breakfractions will be consistent estimator of the set of true breakfractions. On the other hand, the set of estimated breakfractions will be consistent estimator of a subset of the set of true breakfractions if the specified number of change points is less than the true number of change points. The proposed method is compared with some existing change point detection methods in simulation study to assess its performance.

78. Bayesian Semi-supervised Inference via a Debiased Modeling Approach

[03.M2.I53, (page 35)]

Abhishek CHAKRABORTTY, *Texas A&M University*

Inference in semi-supervised (SS) settings has received a great amount of attention in recent years due to increased relevance in modern big-data problems. In a typical SS setting, there is a much larger sized unlabeled data containing only observations for predictors, in addition to a moderately sized labeled data involving observations for both an outcome and a set of predictors. Such data arises naturally from settings where the outcome, unlike the predictors, is costly to obtain. One of the primary statistical objectives in SS settings is to explore whether parameter estimation can be improved by exploiting the unlabeled data. This article proposes a novel Bayesian approach to SS inference for the population mean estimation problem. The proposed approach provides improved and optimal estimators both in terms of estimation efficiency as well as inference. The method itself has several interesting artifacts. The central idea behind our method is to model certain summary statistics of the data rather than specifying a probability model for the entire raw data itself. Specifying appropriate summary statistics crucially relies on a debiased representation of the population mean in terms of nuisance parameters. Combined with careful usage of sample splitting, our debiasing approach mitigates the effect of bias due to slow rates or misspecification of the nuisance parameter from the posterior of the final parameter of interest. We establish concrete theoretical results, via Bernstein-von Mises theorems, validating all our claims and further supporting them through extensive numerical studies. To the best of our knowledge, this is the first work in Bayesian inference for SS settings. We also believe that the central idea of this article will be more broadly applicable to Bayesian semi-parametric inference.

79. Novel clustering procedures based on binary splitting using max-MMD

[05.M1.I91, (page 60)]

Anirvan CHAKRABORTY, *IISER Kolkata*
Sourav CHAKRABARTY, *ISI Kolkata*
Shyamal Krishna DE, *ISI Kolkata*

We develop novel clustering procedures for functional data in both the situations: (a) unknown cluster number and (b) specified cluster number, using the Maximum Mean Discrepancy (MMD) measure between two groups of observations. The first of the two algorithms also provides a bonafide estimate of the number of clusters. The algorithms recursively use a binary splitting strategy to partition the data set into two subgroups such that they are maximally separated in terms of the (weighted) MMD measure. This technique iteratively solves a k-population clustering problem by solving a binary clustering problem at each step. The analysis of the algorithms in an oracle scenario shows that they can achieve perfect clustering for a variety of situations. A detailed data analysis shows that near perfect clustering performance is achieved in both location and scale difference models which improves upon the state-of-the-art for functional data.

80 . Contextual Bandits for Mobile Health Studies with Zero-inflated Count Outcomes

[03.M1.I43, (page 31)]

Bibhas CHAKRABORTY, *National University of Singapore*

Mobile health (mHealth) interventions are becoming increasingly common. In this talk, we will present a cutting-edge trial design arising in mHealth, namely, the micro-randomized trial (MRT) that involves sequential, within-person randomizations. MRTs often aim to estimate the proximal effects of ”push”-type mHealth interventions, e.g., motivational text-messages to promote healthy behav-

iors. The basic MRT design can be made adaptive, thereby enabling it to learn from accumulated data as the trial progresses. This is appealing from an ethical perspective since the adaptive learning tends to make better interventions available to the trial participants. Adaptive learning in this setting can be operationalized via contextual bandit algorithms. Specifically, we will discuss the role of a particular contextual bandit algorithm, namely, Thompson sampling, in designing adaptive MRTs. While the existing literature on Thompson sampling primarily focuses on continuous outcomes, we will show how it can be modified to account for zero-inflated count data. The proposed methods will be validated via simulation studies and illustrated using real data from an MRT called Drink Less.

81 . Integrative graphical modeling of multi-tissue omics data with application to modeling dependencies in the proteome of human pregnancy

[01.M2.I2, (page 4)]

Moumita CHAKRABORTY, *The University of Texas Medical Branch*

Tissue-wide molecular heterogeneity often results in valuable information loss when omics data are analyzed independently across tissues. This loss stems from the clustering and inter-dependence of molecular profiles within biological systems, as well as the spatial orientation of these tissues. Omics data, including transcriptomics (single-cell RNA-seq) and proteomics (mass spectrometry), are inherently high-dimensional and challenging to jointly model due to the complex dependencies among genes and cell types. Recent advances, such as spatial transcriptomics (ST), have addressed some of these challenges, but network-based methods like Gaussian graphical models remain underutilized for multi-tissue omics profiles despite their wide applicability in efficiently explaining complex dependency patterns in omics studies. We propose a novel Bayesian Gaussian network-based approach designed to handle multiple high-dimensional omics data by incorporating sparsity through spike-and-slab priors on regression and precision parameters. This method can leverage similarities across cell types, providing a more nuanced understanding of protein and pathway dependencies across tissues. Using synthetic data that simulate proteomics profiles from five tissues in pre-term human pregnancy, our approach outperforms established ST methods in recovering the dependence network. Our model, applied to real data from human pre-term labor, elucidates key dependencies between proteins, maternal and fetal cell types across different gestational ages.

82 . Bayesian Scalar-on-Image Regression with Spatial Interactions for Modeling Alzheimer's Disease

[02.M1.I22, (page 14)]

Nilanjana CHAKRABORTY, *Indian Institue of Management Udaipur*
Qi LONG, *Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania*
Suprateek KUNDU, *Department of Biostatistics, Division of Basic Science Research, The University of Texas MD Anderson Cancer Center*

There has been substantial progress in predictive modeling for cognitive impairment in neurodegenerative disorders such as Alzheimer's disease (AD) based on neuroimaging biomarkers. However, existing approaches typically do not incorporate heterogeneity that may potentially arise due to interactions between the spatially varying imaging features and supplementary demographic, clinical and genetic risk factors in AD. Unfortunately, ignoring such heterogeneity may potentially result in poor prediction and biased estimation. Building on existing scalar-on-image regression framework, we address this issue by incorporating spatially varying interactions between brain image and supplementary risk factors to model cognitive impairment in AD. The proposed Bayesian method tackles spatial interactions via hierarchical representation for the functional regression coefficients depending on supplementary risk factors, which is embedded in a scalar-on-function framework involving a multi-resolution wavelet decomposition. To address the curse of dimensionality, we induce simultaneous sparsity and clustering via a spike and slab mixture prior, where the slab component is characterized by a latent class distribution. We develop an efficient Markov chain Monte Carlo algorithm for posterior computation. Extensive simulations and application to the longitudinal ADNI-1 study illustrates significantly improved prediction of cognitive impairment in AD across multiple visits by our model in comparison with alternate approaches. The proposed approach also identifies key brain regions in AD that exhibit significant association with cognitive abilities, either directly or through interactions with risk factors.

83. Continuous Time Reinforcement Learning using Rough Paths

[01.M2.I2, (page 4)]

Prakash CHAKRABORTY, *Pennsylvania State University*

TBA

84 . Communication Efficient Distributed Learning with Bayesian Kernel Machine Models

[03.M1.I47, (page 32)]

Sounak CHAKRABORTY, *University of Missouri*
Tanujit DEY, *Brigham and Women's Hospital. Harvard Medical School*
Anjishnu BANERJEE, *Medical College of Wisconsin*

In this paper we introduce Bayesian non-linear models for combining information from multiple data source platforms brought together on the same patient set. The proposed model can efficiently combine and borrow information across platforms and can provide a detailed complementary view of a specific disease. Our model can explore high dimensional covariate space with the help of reproducing kernel Hilbert space based low dimensional representation of non-linear functions based on random Fourier transformation approximation and modified Normal-Exponential-Gamma prior. In this paper we also develop an alternative approach where we can do distributed statistical inference across multiple similar data sets under Surrogate Likelihood framework. This is useful where large volume of data is accessed from multiple data sites and information of each site need to be handled and processed locally for privacy and data security issues. Our approach relies on efficient surrogate likelihood for the global likelihood and computes a quasi-posterior distribution. All Bayesian inference is based on this quasi posterior distribution improving the computational efficiency of our MCMC algorithms.

85 . Epicasting: Neural Networks for Epidemic Forecasting

[03.M1.I43, (page 31)]

Tanujit CHAKRABORTY, *Sorbonne University*

Forecasting time series data represents an emerging field of research in data science that has vast applications for the early prediction of epidemics. Numerous statistical and machine learning methods have been proposed for generating high-quality and reliable forecasts. The speaker will present a newly developed method, called Ensemble Wavelet Neural Network (EWNet), for epidemic forecasting (epicasting).

86 . Bayesian Variable Selection and Sparse Estimation for High-Dimensional Graphical Models

[Student Paper Competition 1, (page 15)]

Anwesha CHAKRAVARTI, *University of Illinois Urbana Champaign*
Naveen NARISHETTY,
Feng LIANG, *University of Illinois Urbana Champaign*

We introduce a novel Bayesian approach that can perform both covariate selection and sparse precision matrix estimation in the context of high-dimensional Gaussian graphical models involving multiple responses and covariates. Our method introduces covariate-level sparsity in the precision matrix between the multiple responses and the covariates, which induces column-wise group sparsity in our regression coefficient matrix estimates. This is achieved using a Bayesian conditional random field model with an appropriately chosen hierarchical spike and slab prior setup. A distinctive feature of our method is that it leverages the structural sparsity information gained from the presence of irrelevant covariates in the dataset to provide a sparse estimation of the three distinct sparsity structures: the conditional dependency structure among the responses, the conditional dependency structure between the responses and the covariates, and the regression coefficient matrix. This stands in contrast to existing methods, which typically concentrate on any two of these structures but seldom achieve simultaneous sparse estimation for all three. Despite the non-convex nature of the problem, we establish statistical accuracy for all points in the high posterior density region, including the maximum-a-posteriori (MAP) estimator. We also present an efficient Expectation-Maximization (EM) algorithm for computing the estimators. Through simulation experiments, we demonstrate the competitive performance of our method, particularly in scenarios where the signal strength in the precision matrices is low. Finally, we apply our method to a bikeshare data set, showcasing its predictive performance.

87 . RANKED SET SAMPLING BASED ON OPTIMAL WEIGHTS

[01.M2.I7, (page 5)]

Girish CHANDRA, *University of Delhi*

Ranked Set Sampling (RSS) is a useful technique for improving the estimator of population mean when the sampling units in a study can be easily ranked than the actual measurement. RSS performs better than simple random sampling (SRS) when the simply mean of units corresponding to each rank is used. The performance of RSS can be increased further by assigning weights to the ranked observations. In this study, simple weighted RSS procedure to estimate the population mean of positively skew distributions has been proposed. Two different cases of unequal weights are considered. It is shown that the gain the in relative precisions of the population mean for chosen distributions are uniformly higher than those based on balanced RSS. The gain in relative precisions are substantially higher. Further, the relative precisions of our estimator are slightly higher the ones based on Neyman's optimal allocation model for small sample sizes. Moreover, it is shown that, the performance of the proposed estimator increases as the skewness increases by using the example of lognormal family of distribution

88. Functional connectivity across the human subcortical auditory system using an autoregressive matrix-Gaussian copula graphical model approach with partial correlations

[01.M2.I2, (page 4)]

Noirrit Kiran CHANDRA, *The University of Texas at Dallas*
Kevin SITEK, *Northwestern University*
Bharath CHANDRASEKARAN, *Northwestern University*
Abhra SARKAR, *The University of Texas at Austin*

The auditory system comprises multiple subcortical brain structures that process and refine incoming acoustic signals along the primary auditory pathway. Due to technical limitations of imaging small structures deep inside the brain, most of our knowledge of the subcortical auditory system is based on research in animal models using invasive methodologies. Advances in ultrahigh-field functional magnetic resonance imaging (fMRI) acquisition have enabled novel noninvasive investigations of the human auditory subcortex, including fundamental features of auditory representation such as tonotopy and periodotopy. However, functional connectivity across subcortical networks is still underexplored in humans, with ongoing development of related methods. Traditionally, functional connectivity is estimated from fMRI data with full correlation matrices. However, partial correlations reveal the relationship between two regions after removing the effects of all other regions, reflecting more direct connectivity. Partial correlation analysis is particularly promising in the ascending auditory system, where sensory information is passed in an obligatory manner, from nucleus to nucleus up the primary auditory pathway, providing redundant but also increasingly abstract representations of auditory stimuli. While most existing methods for learning conditional dependency structures based on partial correlations assume independently and identically Gaussian distributed data, fMRI data exhibit significant deviations from Gaussianity as well as high-temporal autocorrelation. In this paper, we developed an autoregressive matrix-Gaussian copula graphical model (ARMGCGM) approach to estimate the partial correlations and thereby infer the functional connectivity patterns within the auditory system while appropriately accounting for autocorrelations between successive fMRI scans. Our results show strong positive partial correlations between successive structures in the primary auditory pathway on each side (left and right), including between auditory midbrain and thalamus, and between primary and associative auditory cortex. These results are highly stable when splitting the data in halves according to the acquisition schemes and computing partial correlations separately for each half of the data, as well as across cross-validation folds. In contrast, full correlation-based analysis identified a rich network of interconnectivity that was not specific to adjacent nodes along the pathway. Overall, our results demonstrate that unique functional connectivity patterns along the auditory pathway are recoverable using novel connectivity approaches and that our connectivity methods are reliable across multiple acquisitions.

89. A New Approach to Constructing a Family of Distributions Using Sigmoid Functions and Its Applications

[01.E1.C5, (page 13)]

Greeshma CHANDRAN, *University of Calicut*
Manoharan M, *University of Calicut*

This study presents a novel approach for generating a family of distributions using the Q transformation rule combined with fundamental properties of the quantile function. This method allows for creating distributions with specific characteristics based on widely used mathematical functions. For instance,

we use the sigmoid function—commonly employed in machine learning, artificial intelligence, data science, and biomedical engineering—to generate various types of distributions. To illustrate the applicability of the model, three sub-models are developed using the Exponential distribution, and their properties are thoroughly examined. The analysis reveals that the proposed model belongs to several significant classes of distributions. Maximum likelihood estimators for the model parameters are derived, and a Monte Carlo simulation is performed to assess the behavior of these estimators. The practical utility of the three proposed models is demonstrated using three real data sets.

90 . Higher-Order Graphon Theory: Fluctuations, Degeneracies and Inference.

[Student Paper Competition 2, (page 18)]

Anirban CHATTERJEE, *University of Pennsylvania*
Soham DAN, *IBM Research, Yorktown Heights*
Bhaswar BHATTACHARYA, *University of Pennsylvania*

Exchangeable random graphs, which include some of the most widely studied network models, have emerged as the mainstay of statistical network analysis in recent years. Graphons, which are the central objects in graph limit theory, provide a natural way to sample exchangeable random graphs. It is well known that network moments (motif/subgraph counts) identify a graphon (up to an isomorphism), hence, understanding the sampling distribution of subgraph counts in random graphs sampled from a graphon is pivotal for nonparametric network inference. In this paper, we derive the joint asymptotic distribution of any finite collection of network moments in random graphs sampled from a graphon, that includes both the non-degenerate case (where the distribution is Gaussian) as well as the degenerate case (where the distribution has both Gaussian or non-Gaussian components). This provides the higher-order fluctuation theory for subgraph counts in the graphon model. We also develop a novel multiplier bootstrap for graphons that consistently approximates the limiting distribution of the network moments (both in the Gaussian and non-Gaussian regimes). Using this and a procedure for testing degeneracy, we construct joint confidence sets for any finite collection of motif densities. This provides a general framework for statistical inference based on network moments in the graphon model. To illustrate the broad scope of our results we also consider the problem of detecting global structure (that is, testing whether the graphon is a constant function) based on small subgraphs. We propose a consistent test for this problem, invoking celebrated results on quasi-random graphs, and derive its limiting distribution both under the null and the alternative.

91. Statistical inference in networks under snowball type sampling

[01.E1.I20, (page 10)]

Arindam CHATTERJEE, *Stat Math Unit, Indian Statistical Institute, Delhi*

We investigate the asymptotic distribution of subgraph counts when data is collected from a population network using a snowball sampling scheme. Some inferential issues are discussed.

92. A New Locally Adaptive Nonparametric Regression Method.

[04.E1.I83, (page 54)]

Sabyasachi CHATTERJEE, *University of Illinois at Urbana Champaign*
Subhajit GOSWAMI, *Tata Institute of Fundamental Research*
Soumendu MUKHERJEE, *Indian Statistical Institute*

We propose and study a new locally adaptive nonparametric regression method. The method performs variable bandwidth local averaging/local polynomial regression. To certify its local adaptivity we show that it adapts near optimally to the local Holder smoothness exponent of the regression function at any point in the domain. Despite the vast literature on Nonparametric Regression, we only know one existing method which attains such local adaptivity proveably. This method is known as Lepski's method. There are some drawbacks to Lepski's method such as a) it is specifically tailored to a given class of functions such as Holder Smooth function class b) it is rather theoretical and impractical to implement with effectively many tuning parameters. Our proposed method seems to overcome these drawbacks. Firstly, our method is defined without any reference to any function class and secondly there is only one tuning parameter, which when set properly, adjusts all the bandwidths at all locations near optimally. Our method is practically implementable and appears to perform reasonably well in our numerical experiments.

93. On imbalanced spatio-temporal data with missing values

[01.A1.I16, (page 9)]

Snigdhansu CHATTERJEE, *University of Maryland Baltimore County*
Ben BAGOZZI, *University of Delaware*
Ujjal MUKHERJEE, *UIUC*
Vishal SUBEDI, *UMBC*

Large volumes of spatiotemporal data are collected in several application domains, such as surveys and official statistics, social media, healthcare, agriculture, transportation, and climate science. In many such domains, the data exhibits extreme class imbalance. For example, the number of individuals with rare diseases is considerably less than those with normal health conditions. Additionally, there may be extensive missing observations, owing to the sensitive nature of the data. Typical supervised learning algorithms used in artificial intelligence algorithms typically fail to consider such real-data features like large numbers of missing values, extreme class imbalance, and spatio-temporal dependency. Using data on political violence in Mexico, we illustrate the failure of several types of machine learning models in eliciting a suitable set of predictors for interpretable data science or in obtaining high predictive accuracy. We illustrate how these techniques may be improved by leveraging the high-dimensional geometry of the data.

94. Group Heteroscedasticity - A Silent Saboteur of Power and False Discovery in RNA-Seq Differential Expression

[02.A1.I36, (page 24)]

Suvo CHATTERJEE, *Indiana University, Bloomington*
Himel MALLICK, *Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA*
Arindam FADIKAR, *Decision and Infrastructure Sciences Division, Argonne National Laboratory, Lemont, IL, USA*
Arunkumar GANESAN, *Department of Cell Biology and Physiology, University of New Mexico, Albuquerque, NM, USA*

Despite the availability of several high-profile, state-of-the-art methods, analyzing bulk RNA-Seq data continues to face significant challenges. Evidence from recent studies has highlighted that popular differential expression (DE) tools, such as edgeR and DESeq2, are susceptible to an alarmingly high false discovery rate (FDR). These studies suggest that the FDR inflation observed in these models could be attributed to issues such as violations of parametric assumptions or an inability to effectively handle outliers in the data. Here, we argue that group heteroscedasticity can also contribute to this elevated FDR, a phenomenon largely overlooked by the research community. We introduce a novel statistical model, Robseq, designed for effective per-feature modeling in differential analysis, particularly when the assumption of group homoscedasticity is unmet. Robseq utilizes well-established statistical machinery from the robust statistics literature, including M-estimators to robustly estimate gene expression level changes and Huber-Cameron variance estimators to calculate robust standard errors in heteroscedastic settings. Additionally, it incorporates a degrees of freedom adjustment for the Welch t-statistic, based on Bell-McCaffrey's recommendation, for inferential purposes, effectively addressing the problem of FDR inflation in RNA-Seq differential expression. Through detailed simulations and comprehensive benchmarking, we show that Robseq successfully maintains the false discovery and type-I error rates at nominal levels while retaining high statistical power compared to well-known DE methods. Analysis of population-level RNA-Seq data further demonstrates that Robseq is capable of identifying biologically significant signals and pathways implicated in complex human diseases that otherwise cannot be revealed by published methods. The implementation of Robseq is publicly available as an R package at https://github.com/schatterjee30/Robseq.

95 . Optimizing Subgroup Analysis: Shrinkage Estimation Techniques Through Hierarchical Modeling

[04.E1.I84, (page 54)]

swarnendu CHATTERJEE, *GSK*
swarnendu CHATTERJEE, *GSK*

Exploratory subgroup analyses often yield misleading results, such as indicating a treatment effect in a subgroup that contradicts the overall effect, even when there are no true differences, or only minor differences, across subgroups. The likelihood of such misleading findings rises with the number of subgroups examined and smaller sample sizes within these subgroups. Typically, the conventional approach to subgroup analysis treats each subgroup independently, estimating treatment effects without considering data from other subgroups. However, us-

ing data from other subgroups can be particularly informative if it is believed that treatment effects are generally similar across subgroups, an assumption known as exchangeability. Unlike the conventional approach, shrinkage estimation methods integrate data from all patients to estimate subgroup-specific treatment effects. This is done by taking a weighted average of the conventional estimate and an overall treatment effect, where the weight is based on the subgroup sample size and the heterogeneity level between subgroups. This results in more precise estimates that are 'shrunk' towards the overall effect, thereby reducing the risk of misleading results. Shrinkage estimation methods can be applied to most subgroup analyses and are especially effective when there is no prior expectation of clinically significant differences between subgroup-specific treatment effects. Therefore, we advocate for the routine use of shrinkage analysis in practice unless there is a clinical, regulatory, or other justification for using the conventional estimate from a particular subgroup. This talk describes the approaches to shrinkage estimation, their assumptions, and demonstrates example applications.

96 . Best Treatment Selection: Application to best Digital Experience selection

[04.M2.I73, (page 47)]

Bhargab CHATTOPADHYAY, *Indian Institute of Technology Jodhpur*

In the current era of digital businesses, firms experiment periodically to enhance the online customer experience. This experiment tests how the impact on customer experience changes from minute tweaks to major overhauls. This pursuit of improvement requires statistical tools that can identify positive changes with respect to key performance indicators at high probability and with minimal sampling. Additionally, comparing many alternative treatments is a useful approach to the problem when the main concern is the inference of the "best" improvement. To address this task, we propose a sequential method for selecting the best digital treatment for a pre-fixed power and size level. The sequential method and its characteristics will be demonstrated using a Monte Carlo simulation.

97. Quantile based control charts for the inverse Pareto distribution with applications

[04.A1.I79, (page 51)]

Aditi CHATURVEDI, *Sharda University*
Neeraj JOSHI, *Indian Institute of Technology Delhi*
Sudeep BAPAT, *Indian Institute of Technology Bombay*

In this paper, we aim to develop the Shewhart-type control charts for monitoring the quantiles of an inverse Pareto distribution (IPD) using the complete and middle censored datasets. We obtain the maximum likelihood estimators (MLEs) of the quantile function under both cases. Furthermore, we discuss the asymptotic properties of the MLEs and use them to develop the Shewhart-type control charts. A large-scale simulation analysis is performed to examine the in-control performance of the proposed control chart schemes. This analysis is based on various criteria such as the average run length for different choices of quantiles, false-alarm rates, and sample sizes. We also study the behavior of the out-of-control performance of the proposed charts for several choices of shifts in the IPD parameter. Two real datasets are analyzed to complement the utility of the proposed control chart schemes. We establish that the proposed control charts detect the out-of-control signals very well.

98 . Clinical and public health challenges of emerging infectious disease outbreaks: Complexity of infectious disease modelling for early prediction in India.

[04.A1.I77, (page 50)]

Himanshu K. CHATURVEDI, *Disease Modelling Division, Indian Council of Medical Research, New Delhi*
Poornima Suryanath SINGH, *Amity Institute of Public Health and Hospital Administration, Amity University, NOIDA,UP, India*

Himanshu K. Chaturvedi1 and Poornima Suryanath Singh2

1. Disease Modelling Division, Indian Council of Medical Research, Delhi, India, 2. Amity Institute of Public Health and Hospital Administration, Amity University, NOIDA,UP, India

Abstract Background The global landscape of infectious diseases is ever-evolving, with emerging and re-emerging pathogens posing significant challenges to clinical and public health research. Diseases such as malaria, dengue, tuberculosis, HIV/AIDS, and recent pandemics like COVID-19 have highlighted the critical need for innovative strategies to predict, prevent, and control these infections. One of the most

promising approaches in this regard is the use of advanced disease modeling techniques and develop need based data repository system to investigate the outbreak.

Disease Modeling Complexity Disease modeling has become an indispensable tool in understanding the dynamics of infectious diseases. These models allow us to simulate the spread of infections within populations, predict potential outbreaks, and assess the impact of various intervention strategies. By incorporating data on pathogen transmission, environmental factors, and human behavior, models provide valuable insights that guide public health policies and resource allocation.

For instance, during the COVID-19 pandemic, epidemiological models played a crucial role in predicting infection trends, evaluating the effectiveness of lockdowns, and notifying vaccination strategies. Similarly, models have been used to track the spread of vector-borne diseases like malaria and dengue, enabling targeted control measures in high-risk areas.

Integrated approach Looking ahead, the success of disease modeling will depend on several factors. The integration of real-time data, including genomics, environmental monitoring, and social media trends, will enhance the accuracy and timeliness of predictions. Moreover, advances in machine learning and artificial intelligence are poised to revolutionize modeling approaches, enabling more complex simulations that account for the multifaceted nature of disease spread.

Conclusion The importance of robust and adaptable disease models cannot be overstated as we continue to face the threat of infectious diseases. These tools will be crucial in anticipating future outbreaks, optimizing public health interventions, and ultimately reducing the global burden of infectious diseases.

99 . On Bahadur Representation of Quantiles

[Bahadur Memorial Lecture, (page 36)]

Probal CHAUDHURI, *Indian Statistical Institute, Kolkata*

Raghu Raj Bahadur is one of the most outstanding mathematical statisticians of all time. In many ways, the conceptual depth and the mathematical elegance of his work is unparalleled. In 1966, in a short note published in The Annals of Mathematical Statistics, he showed how quantiles in large samples can be approximated by some form of means and consequently, asymptotic distributions of quantiles can be derived using the central limit theorem of means. This fundamental result has influenced decades of research on asymptotic properties of univariate, multivariate and infinite dimensional quantiles as well as regression quantiles in various regression problems. I shall present an overview of that research.

100. On the Properties of the Gradient Function of Log-empirical Likelihoods.

[03.M2.I49, (page 34)]

Sanjay CHAUDHURI, *University of Nebraska-Lincoln*

The support of the empirical likelihood is known to be non-convex even for very simple models. Due to this non-convexity, when applied to empirical likelihood, many standard numerical procedures often become inefficient. However, the gradient function of the log-empirical likelihood diverges at the boundary of the support and therefore provides a convenient tool to improve computational efficiency. We discuss the properties of this gradient on the whole space, including the complement of the support. Finally, we present several examples of improvement in computational efficiency that the use of the gradient function can bring.

101. An Analytical Framework to Combine Mortality and Morbidity to Study Determinants of Child Health

[03.A1.I56, (page 37)]

Anuradha Rajkonwar CHETIYA, *Ramjas College, University of Delhi*
Vishal DEO , *ICMR-NIRDH&DS, New Delhi*

Research studies on infant and child health have primarily focused on identifying determinants of the standard mortality rates like IMR (infant mortality rate), under five mortality rate and similar measures. Such studies have been useful in providing an insight into the impact of maternal health and other socioeconomic determinants of child mortality and contribute towards policy framing by the government and other agencies working in these areas. However, while mortality is the number of deaths and is absolute, morbidity is a measure of those affected by the health condition and is not absolute as there are chances that some may survive while others may not, or some may survive but with lower quality of health. Recent studies have linked health of children in their early developmental age to non-communicable risk factors at the later stage of their life. So, it be-

comes pertinent that child health status be examined not just in the terms of mortality, but also by incorporating child growth parameters as outcomes of interest, which can together present a more comprehensive picture. It will also provide key insights in managing health systems in an optimal way to ensure overall improvement of child health. This paper uses data from the latest National Family Health Survey (NFHS-5) of India, to provide an analytical framework that incorporates both mortality and morbidity as a measure of infant health to investigate their relation to maternal and other socio-economic determinants of health.

102. An algorithm for construction of Uniform designs

[02.E1.C10, (page 30)]

Rakesh CHHALOTRE, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*
Rajender PARSAD, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*
Baidya Nath MANDAL, *ICAR - Indian Agricultural Research Institute, Jharkhand – 825405*
Sukanta DASH, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*

The uniform design is an important tool used for computer experiments that allow experimental points to distributed evenly across the design space and reducing the risk of certain areas of the design space being over- or under-sampled. By adopting this design, biases are minimized and data reliability and accuracy are promoted. As per literature search, it seems that there is no computer algorithm available for constructing uniform designs specifically in the rectangular region. In the present investigation, an algorithm has been developed based on NT-nets. In the algorithm, computation of discrepancy criteria such as RMSD, AD and MD have also been added that help us to choose the best design that uniformly cover the rectangular region.

103. Bayesian Difference in Differences with Extensions to Multiple Periods and Staggered Treatments

[04.A1.I80, (page 51)]

Siddhartha CHIB, *Washington University in Saint Louis*
Kenichi SHIMIZU, *University of Alberta*

Starting with the natural experiment setting in which all subjects in the first period are unexposed, but a sub-sample of subjects in the second period are treated, we provide parametric and semiparametric Bayesian alternatives to the classic difference in difference estimator. The approach is based on heterogeneous subject-level models of potential outcomes in the two periods, restricted to satisfy the parallel trend assumption. The approach is generalized to settings where observations on the treated and control subjects are available for multiple additional time periods, and to settings where some subjects, unexposed in the first two periods, are subsequently treated. Marginal likelihoods, computed by the method of Chib (1995), are used to compare alternative modeling assumptions, and large-sample consistency results of causal effects and model comparisons are derived. The methods are illustrated with important examples drawn from the literature.

104. On Bootstrapping Lasso in Generalized Linear Models and the Cross Validation

[03.E1.C16, (page 41)]

Mayukh CHOUDHURY, *Department of Mathematics, IIT Bombay*
Debraj DAS, *Department of Mathematics, IIT Bombay*

Generalized linear models or GLM constitutes an important set of models which generalizes the ordinary linear regression by connecting the response variable with the covariates through arbitrary link functions. On the other hand, Lasso is a popular and easy to implement penalization method in regression when all the covariates are not relevant. However, Lasso generally has non-tractable asymptotic distribution and hence development of an alternative method of distributional approximation is required for the purpose of statistical inference. Here, we develop a Bootstrap method which works as an approximation of the distribution of the Lasso estimator for all the sub-models of GLM. To connect the distributional approximation theory based on the proposed Bootstrap method with the practical implementation of Lasso, we explore the asymptotic properties of K-fold cross validation-based penalty parameter. The results established essentially justifies drawing valid statistical inference regarding the unknown parameters based on the proposed Bootstrap method for any sub model of GLM after selecting the penalty parameter using K-fold cross validation. Good finite sample properties are also shown through a moderately large simulation study. The method is also implemented

on a real data set.

105 . Quantile regression for high-dimensional data with uniform-over-dimension convergence

[04.M2.I75, (page 48)]

Joydeep CHOWDHURY, *King Abdullah University of Science and Technology*

Conditional median and quantiles yield valuable information about aspects of the conditional distribution of a response given the covariate, and are widely used in hypothesis testing methods, which are usually more robust compared to conditional mean-based testing procedures. However, like other multivariate methods, the hypothesis testing procedures based on conditional median and quantiles are affected by the dimension of the response, necessitating development of separate procedures for multivariate and high-dimensional responses. The procedure for multivariate response is applied when the dimension of the response is small compared to the sample size, and that for a high-dimensional response is prescribed for very large dimensional response. In practical scenarios, it is hardly possible to determine whether to apply the multivariate procedure or the high-dimensional procedure when the dimension of the response is neither too small nor too large. In this work, a method of hypothesis testing involving conditional quantiles and median are proposed in a regression setup, which can be applied without being concerned about the dimension of the response. The performance of the proposed method is demonstrated in simulated and real data.

106. Sample size re-assessment in clinical trials using predictive power

[04.A1.I77, (page 50)]

Solomon CHRISTOPHER, *Emmes*
Jonathan JAEGER, *Jaeger Consulting*

Uncertainty on parameters relevant to sample size calculation may not be fully controlled at the stage of designing a clinical trial. Adaptive designs that allow for sample size re-assessment offer flexibility to update the sample size for ongoing trials while accumulating relevant information at an interim time-point.

Predictive power, under a frequentist-Bayesian hybrid methodology, facilitates combining prior information (from previous studies, expert opinion or prior beliefs) and the data collected up to the interim analysis in the ongoing trial. It utilizes all available information on the parameters while reducing the uncertainty in the parameters at sample size re-assessment. Using Monte Carlo methods, computing predictive power is intuitive. It is based on the frequentist evaluation of the samples from posterior predictive distributions of the unknown quantities, which allows trial design considerations to be directly incorporated in the computation.

An application of predictive power for recurrent event endpoints will be discussed using a case study. This application will illustrate handling design aspects such as incorporating covariate effects, strategy to handle multiple comparison and strategies to prevent unblinding of treatment effects in the sample size re-assessment.

107. TBD

[03.M2.I49, (page 34)]

Bertrand CLARKE, *University of Nebraska-Lincoln*

TBA

108. Stability as an Objective Criterion for Shrinkage Penalty Selection with Application in High Dimensional Classification

[Special Invited Session 2, (page 31)]

Jennifer CLARKE, *Department of Statistics, University of Nebraska-Lincoln, USA*
Dean DUSTIN, *University of Nebraska-Lincoln*
Bertrand CLARKE, *University of Nebraska-Lincoln*
Laura KRESTY, *University of Michigan*

In regression contexts where p » n it is common to invoke a shrinkage (or regularization) technique that determines the appropriate model by minimizing a penalized empirical risk criterion. Some shrinkage methods have the oracle property, i.e., the method is consistent in both parameter estimation and variable selection. We explore the stability of the predictive error as a metric for comparing and selecting a regularization method (with or without the oracle property) under different scenarios of sample size, number of predictors, and sparsity. We also propose a method to find an optimal penalty using evolutionary computing, as described in Dustin et al. (2023). The performance of an optimal penalty logistic regression is demonstrated in biomarker identification for esophageal adenocarcinoma, and challenges in applications with high dimensional data are discussed.

109 . Anytime-Valid Inference for Double/Debiased Machine Learning of Causal Parameters

[Student Paper Competition 1, (page 15)]

Abhinandan DALAL, *University of Pennsylvania*
Patrick BLOEBAUM, *Amazon Web Services*
Shiva KASIVISWANATHAN, *Amazon Web Services*
Aaditya RAMDAS, *Carnegie Mellon University*

Double (debiased) machine learning (DML) has seen widespread use in recent years for learning causal/structural parameters, in part due to its flexibility and adaptability to high-dimensional nuisance functions as well as its ability to avoid bias from regularization or overfitting. However, the classic double-debiased framework is only valid asymptotically for a predetermined sample size, thus lacking the flexibility of collecting more data if sharper inference is needed, or stopping data collection early if useful inferences can be made earlier than expected. This can be of particular concern in large scale experimental studies with huge financial costs or human lives at stake, as well as in observational studies where the length of confidence of intervals do not shrink to zero even with increasing sample size due to partial identifiability of a structural parameter. In this paper, we present time-uniform counterparts to the asymptotic DML results, enabling valid inference and confidence intervals for structural parameters to be constructed at any arbitrary (possibly data-dependent) stopping time. We provide conditions which are only slightly stronger than the standard DML conditions, but offer the stronger guarantee for anytime-valid inference. This facilitates the transformation of any existing DML method to provide anytime-valid guarantees with minimal modifications, making it highly adaptable and easy to use. We illustrate our procedure using two instances: a) complied average treatment effect in online experiments with non-compliance, and b) partial identification of average treatment effect in observational studies with potential unmeasured confounding.

110. Central Limit Theorem in High Dimension

[01.E1.I20, (page 10)]

Debraj DAS, *Indian Institute of Technology Bombay*
Soumendra LAHIRI,

Central Limit Theorem (CLT) is one of the oldest as well as remarkable results of classical probability theory. In most simplest words, CLT is a statement about the convergence of properly centered and scaled sample mean of a sequence of random vectors to the Gaussian random vector in distribution. The recent interest lies in establishing CLT when the dimension also grows with the sample size. I will shed light on recent developments as well as describe some results related to critical growth rate of dimension. If time permits, I will state a characterization of the Gaussian distribution in terms of high dimensional CLT and will briefly describe the benefits of self-normalization/studentization in reducing the requirement of existence of exponential moments to some polynomial moments.

111. A Bayesian quantile joint modeling of multivariate longitudinal and time-to-event data

[04.M2.I71, (page 47)]

Kiranmoy DAS, *Indian Statistical Institute, Kolkata*
Damitri KUNDU, *Intuit, Bangalore, India*
Shekhar KRISHNAN, *Tata Medical Center, Kolkata*
Manash Pratim GOGOI, *Tata Medical Center, Kolkata*

Linear mixed models are traditionally used for jointly modeling (multivariate) longitudinal outcomes and event-time(s). However, when the outcomes are non-Gaussian a quantile regression model is more appropriate. In addition, in the presence of some time-varying covariates, it might be of interest to see how the effects of different covariates vary from one quantile level (of outcomes) to the other, and consequently how the event-time changes across different quantiles. For such analyses linear quantile mixed models can be used, and an efficient computational algorithm can be developed. We analyze a dataset from the Acute Lymphocytic Leukemia (ALL) maintenance study conducted by Tata Medical Center, Kolkata. In this study, the patients suffering from ALL were treated with two standard drugs (6MP and MTx) for the first two years, and three biomarkers (e.g. lymphocyte count, neutrophil count and platelet count) were longitudinally measured. After treatment the patients were followed nearly for the next three years, and the relapse-time (if any) for each patient was recorded. For this dataset we develop a Bayesian quantile joint model for the three longitudinal biomarkers and time-to-relapse. We consider an Asymmetric Laplace Distribution (ALD) for each outcome, and exploit the mixture representation of the ALD for developing a Gibbs sampler algorithm to estimate the regression coefficients. Our proposed model allows different quantile levels for dif-

ferent biomarkers, but still simultaneously estimates the regression coefficients corresponding to a particular quantile combination. We infer that a higher lymphocyte count accelerates the chance of a relapse while a higher neutrophil count and a higher platelet count (jointly) reduce it. Also, we infer that across (almost) all quantiles 6MP reduces the lymphocyte count, while MTx increases the neutrophil count. Simulation studies are performed to assess the effectiveness of the proposed approach.

112. SMART-MC: Sparse Matrix estimation with covaRiate-based Transitions in Markov Chain modelling via parallelized Multiple Spherically Constrained Optimization Routine (MSCOR)

[01.A1.I10, (page 7)]

Priyam DAS, *Virginia Commonwealth University*

A Markov model is a widely used tool for modeling sequences of events from a finite state-space and hence can be employed to identify the transition probabilities across treatments based on treatment sequence data. To understand how subject-level covariates impact these treatment transitions, the transition probabilities are modeled as a function of subject covariates. This approach allows us to visualize how subject-level covariates affect the treatment transitions across subject visits. The proposed method automatically estimates the entries of the transition matrix with smaller numbers of empirical transitions as constant; where the user can set their desired cutoff of the number of empirical transition counts required for a particular transition probability to be estimated as a function of covariates. Firstly, this strategy automatically enforces the final estimated transition matrix to contain zeros at the locations corresponding to zero empirical transition counts, avoiding further complicated model constructs to handle sparsity in a smart fashion. Secondly, it restricts estimation of transition probabilities as a function of covariates, when the number of empirical transitions is particularly small, thus avoiding the identifiability which might arise due to the p>n scenario. To optimize the (possibly) multimodal likelihood, a parallelized global optimization routine is also developed. The proposed method is applied to understand how the transitions across disease modifying treatments (DMTs) in Multiple Sclerosis (MS) subjects are influenced by subject-level demographic and clinical phenotypes.

113. Extreme learning machine ensemble model for time series forecasting boosted by particle swarm optimization

[Student Poster Competition, (page 22)]

Saikath DAS, *ICAR - Indian Agricultural Research Institute*
Ranjit Kumar PAUL, *Senior Scientist, Division of Statistical Genetics, ICAR-Indian Agricultural Statistical Research Institute, New Delhi*
Md YEASIN, *Scientist, Division of Statistical Genetics, ICAR-Indian Agricultural Statistical Research Institute, New Delhi*
Amrit Kumar PAUL, *Principal Scientist and Head, Division of Statistical Genetics, ICAR-Indian Agricultural Statistical Research Institute, New Delhi*

Extreme learning machine is an algorithm for training single layer feedforward neural networks which skips the estimation of the input weights and biases and instead fixes them randomly, thus being much faster than gradient based backpropagation techniques used in artificial neural networks. Another concept which has gained wide popularity in recent years is that of ensemble modeling, which consists of combining forecasts from various individual models and are known to enhance the overall performance of the final model. Application of these techniques in the context of time series forecasting has been described mathematically. A weighted ensemble model has been built by combining multiple extreme learning machines, each having a different network topology. To tune the weights of the ensemble, particle swarm optimization has been employed. The proposed methodology has been validated by using two agricultural price datasets. It has been demonstrated that the employed optimization algorithm is suitable for the ensemble and that the proposed strategy is able to obtain good results for time series problems.

114. Cross Sectional Regression with Cluster Dependence: Inference based on Averaging

[05.M1.I90, (page 60)]

Samarjit DAS, *Indian Statistical Institute Kolkata*

TBA

115. Optimizing Prices of Tickets and Advertisement in Sports Series with Variable Game Counts

[03.A1.I59, (page 38)]
Shubhabrata DAS, *Indian Institute of Management Bangalore*

In play-off matches in the NBA, NHL, World Series in baseball, and bilateral series in various other sports, the winner between two competing teams or players is determined through a best-of-n games format. This work evaluates the importance of individual games within such series, enabling ticket prices to be set in proportion to the calculated significance of each game. The proposed measure accounts for both the likelihood of a game being played and its potential to be the series decider at the time of ticket sales. We explore and compare multiple pricing frameworks, offering insights into revenue management strategies for such sporting events.

116. Professor Pranab Kumar Sen – the Advisor and the Academic Guru

[Memorial Session 1, (page 26)]
Shubhabrata DAS, *Indian Institute of Management Bangalore*

In his illustrious academic career Professor Pranab Kumar Sen mentored over eighty doctoral students as their dissertation advisor in statistics as well as biostatistics department of the University of North Carolina at Chapel Hill. In addition to this, he contributed to the academic growth of many other statisticians who had the privilege of working with Prof. Sen. In this memorial session, we would look back at some of his exceptional qualities as an academic guru and guide as well as an empathetic human being that helped to shape many young careers. We plan to summarize narratives and anecdotes from many of his students which can inspire the future generations of academics.

117 . Jacobi Prior: An Alternate Bayesian Method for Supervised Learning

[01.M2.I9, (page 6)]
Sourish DAS, *Chennai Mathematical Institute*

TBA

118. Graph Structure Learning from DeGroot Dynamics

[04.M1.I65, (page 45)]
Gautam DASARATHY, *Arizona State University*
Vignesh TIRUKKONDA, *Arizona State University*

In this talk, we consider the problem of learning the structure of graphs from data generated by Gaussian DeGroot dynamics. DeGroot dynamics is a classical model of information aggregation where each node (agent) in a graph updates its value as a weighted average of its neighbors' values, capturing the evolution of opinions, beliefs, or other continuous states over a network. This work generalizes the problem of structure learning from independent and identically distributed (iid) data to scenarios where the data is highly dependent and evolves according to these dynamics on the graph.

Unlike the traditional approaches that rely on iid samples, our focus is on settings where observations are drawn from a temporally evolving process governed by the Gaussian DeGroot model. We establish that the problem of reconstructing the underlying graph structure from such dependent data is computationally feasible. Specifically, we provide a provable algorithm that efficiently recovers the graph structure in time that scales polynomially with the number of nodes and logarithmically with the inverse of the sampling interval. This extends the landscape of structure learning in graphical models and opens new avenues for understanding complex dependencies in dynamically evolving systems.

119. Statistical Machine Learning Approach to Feature Selection in High Dimensional Low Sample Size (HDLSS) Data

[02.M2.I27, (page 17)]
Shibasish DASGUPTA, *Pfizer*
Vladimir IVANOV , *Pfizer*

The recent advancement in data collection and data processing techniques generated huge amounts of high-dimensional data. Datasets originating from different disciplines, and especially biomedical domain, can contain thousands of features and still have a small sample size. This High Dimensional Low Sample Size (HDLSS) data possesses a great challenge for statistical machine learning techniques based on likelihood estimation. Feature selection (FS) helps to overcome the curse of dimensionality, risk of over-fitting and have better interpretability and precision of the final model by drastically reducing the number of input variables when building a predictive model. In our work, we explored various filter and wrapper techniques, and proposed a general feature selection framework for robust variable selection incorporating

the stability idea specifically for the HDLSS data. Particularly, we evaluated Minimum-Redundancy-Maximum-Relevance (mRMR) and Support Vector Machine Recursive Feature Elimination (SVM-RFE) methods. To compare the explored FS methods, we evaluated their predictive performance and stability on Pfizer Clinical Trial data.

120. Symmetric Ordering Factorial Experiments

[03.A1.I55, (page 36)]

Anindita DATTA, *ICAR IASRI, New Delhi*

Factorial Experiments are experiments that investigate the effects of two or more factors each at two or more levels on the output response of a process. It is assumed in factorial experiments that the response is only affected by the levels of factors. However, in some experiments, not only the level combinations of factors but also the addition orders will influence the responses. Experiments that deal with such sequential order of adding components are termed as order-of-addition (OofA) experiments. In literature, experiments considering both the levels of factors and addition orders of components have been studied as ordering factorial experiments (Yang et al., 2023). Here some new series of symmetrical ordering factorial designs have been obtained. The designs are D optimal.

121. Change point in functional data

[01.E1.C4, (page 12)]

Debanjana DATTA, *Indian Statistical Institute, Bangalore*
Rituparna SEN, *Associate Professor,Indian Statistical Institute, Bangalore*

We have devised a procedure for detecting a change point in the Functional Time Series under the state space representation where the latent process is a Functional Autoregressive (FAR) process with Gaussian innovations. The state space formulation naturally allows for sparse observations with noise. Like any other testing procedure, our approach is not based on smoothing and functional principal components which incur a huge loss of information across grids. Bayesian methodologies have been employed to test the existence and determine the location of these points.

122. Horseshoe-type Priors for Independent Component Estimation

[03.M2.I51, (page 35)]

Jyotishka DATTA, *Virginia Polytechnic Institute and State University*
Nicholas POLSON, *University of Chicago Booth School of Business*

Independent Component Estimation (ICE) has many applications in modern day machine learning as a feature engineering extraction method. Horseshoe-type priors are used to provide scalable algorithms that enables both point estimates via expectation-maximization (EM) and full posterior sampling via Markov Chain Monte Carlo (MCMC) algorithms. Our methodology also applies to flow-based methods for nonlinear feature extraction and deep learning. We also discuss how to implement conditional posteriors and envelope-based methods for optimization. Through this hierarchy representation, we unify a number of hitherto disparate estimation procedures. We illustrate our methodology and algorithms on a numerical example. Finally, we conclude with directions for future research.

123. Beyond Single Variants: Exploring Gene-Gene Interactions and Their Impact on Disease Risk Prediction

[03.E1.C15, (page 41)]

Kallol DATTA, *National Institute of Biomedical Genomics, Kalyani*
Samsiddhi BHATTACHARJEE, *National Institute of Biomedical Genomics, Kalyani*

Genome-wide Association Studies have identified several genomic variants associated with complex diseases. However, these variants do not act in isolation, rather they interact with one another to perturb disease risk. Although several candidate gene-gene (G-G) interactions have been identified, there is no evidence of large-scale G-G interactions at the genome-wide level. Moreover, interactions are frequently investigated using only the multiplicative scale, leaving out the additive scale. We address this crucial gap using UK Biobank and other datasets for 9 diseases, assessing overall enrichment using QQ plots and deriving p-values. We further investigate how patterns of interaction vary with allele frequencies, effect sizes of variants, and network proximity of mapping genes.

Next, we investigate how risk prediction can be improved with machine learning (ML) models which incorporate interaction terms as compared to simple linear models. We were able to derive more interpretable decision tree models by incorporating inter-

actions among biological pathways without compromising on risk prediction.

124. Going beyond gene-based inferences: identification of cell-cell communications for complex biological systems using a Bayesian Tweedie Model

[04.M2.I71, (page 47)]

Susmita DATTA, *University of Florida*
Dongyuan WU ,
Jeremy T. GASKINS, *University of Louisville*
Michael SEKULA, *University of Louisville*

Understanding the cellular communication through biochemical signaling process plays important role in any biological mechanism such as disease progression. In recent years, with the technology of single-cell sequencing one can acquire molecular data at a single cell resolution. However, cell signaling activities are spatially constrained, and single-cell alone data cannot provide spatial information for each cell. Spatially resolved transcriptomics data can provide that information. Most of the computational measures using such data focus on providing some adhoc estimate of the intercellular communication instead of relying on a statistical model. It is undeniable that descriptive statistics are straightforward and more accessible, but a suitable statistical model can provide more reliable and accurate inference. In this work, we propose a generalized linear regression model to infer cellular communications from spatially resolved transcriptomics data, especially spot-based data. We provide a Bayesian Tweedie modeling of the communication score estimates between cell types with the consideration of corresponding distances between them.

125. Some game theoretic ideas in the design of experiments

[04.M1.I64, (page 44)]

Ori DAVIDOV, *University of Haifa*

TBA

126. Parsimonious Wavelet-Based Models for Some Periodic Financial Time Series

[03.M1.C11, (page 33)]

Rhea DAVIS, *Cochin University of Science and Technology, Kochi*
N BALAKRISHNA, *Indian Institute of Technology, Tirupati*

Many financial series exhibit seasonal patterns in their characteristic features like conditional heteroskedasticity and conditional duration. In consideration of this, the periodic autoregressive conditional heteroskedasticity (PARCH) model, the periodic generalized autoregressive conditional heteroskedasticity (PGARCH) model, and the periodic autoregressive conditional duration model (PACD) were developed and are available in the literature. However, PARCH, PGARCH, and PACD models are plagued with a large number of parameters. Hence, it is crucial to develop the parsimonious versions of these periodic financial models. In the present study, we propose a wavelet-based method to reduce the number of parameters by identifying significant wavelet coefficients via hypothesis testing. The simulation studies of the proposed wavelet-PARCH and wavelet-PACD models yield promising results, highlighting the proposed method's potential for real-world applications.

127. Robust Bayesian Model Averaging for Linear Regression Models With Heavy-Tailed Errors

[Student Poster Competition, (page 21)]

Shamriddha DE, *The University of Iowa*
Joyee GHOSH, *The University of Iowa*

We aim to develop a Bayesian model averaging technique in linear regression models to accommodate heavier tailed error densities than the normal distribution. Motivated by the use of the Huber loss function in presence of outliers, the Bayesian Huberized lasso with hyperbolic errors has been proposed and recently implemented in the literature. Since the Huberized lasso cannot enforce regression coefficients to be exactly zero, we propose a fully Bayesian variable selection approach with spike and slab priors to address sparsity more effectively. Furthermore, the hyperbolic distribution has heavier tails than a normal distribution but thinner tails than a Cauchy distribution. Thus, we propose a novel regression model with an error distribution encompassing both hyperbolic and Student-t distributions. Our model aims to capture the benefit of using Huber loss, while adapting to heavier tails and unknown levels of sparsity, as entailed by the data. We develop an efficient Gibbs sampler with Metropolis Hastings steps for posterior computation. Through simulation studies and analyses of real datasets, we observe a superior performance of our method over various state-of-the-art methods.

128. High-dimensional Adaptive Multiple Testing Controlling FDR and FNR for Sequential Data

[05.M1.I91, (page 60)]

Shyamal Krishna DE, *Indian Statistical Institute Kolkata*
Rahul ROY, *Indian Statistical Institute*
Subir Kumar BHANDARI, *Indian Statistical Institute*

In this talk, we shall discuss multiple testing procedures with simultaneous control of false discovery and nondiscovery rates when data vectors are observed sequentially or in groups. Existing multiple testing methods for sequential data uses fixed stopping boundaries that do not depend on the sample size, and hence, are conservative when the number of hypotheses m is large. We propose sequential tests based on adaptive stopping boundaries that shrink the continue sampling region as the sample size increases. Under a two-group mixture model assumption, we propose an oracle test based on the local false discovery rate (Lfdr) statistic and prove exact control of FDR and FNR at some prefixed levels. Further, we develop a data-driven test which guarantees simultaneous control of FDR and FNR asymptotically as m tends to infinity. Both the oracle and the data-driven stopping times are shown to converge to a finite constant as m grows to infinity. Extensive analysis of simulated datasets illustrate the superiority of the proposed tests over some existing methods.

129. High-dimensional Quadratic Discriminant Analysis using Random Projections

[Student Poster Competition, (page 20)]

Annesha DEB, *Indian Institute of Technology, Kanpur*
Minerva MUKHOPADHYAY, *Indian Statistical Institute, Kolkata*
Subhajit DUTTA, *Indian Statistical Institute, Kolkata*

In binary classification , two widely used methods, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) are based on the assumption that populations follow a Gaussian distribution. LDA assumes identical covariance matrices for the two populations, while QDA allows both the means and covariance matrices to differ. These methods are effective when the number of variables (p) is fixed, and the sample size (n) increases. Notably, LDA remains asymptotically optimal even when p grows at a slower rate than $\sqrt{n}$. However, challenges arise when p exceeds $\sqrt{n}$, particularly in high-dimensional settings where $p > n$ or $p = \mathrm{O}(e^n)$.

To tackle high-dimensionality, a random projection (RP) approach has been proposed for dimensionality reduction in various statistical analyses. While RP has been applied to LDA with a derived probabilistic upper bound on misclassification probability, its theory and application to QDA remains underexplored.

In this poster, we implement RP in QDA under the assumptions of spiked covariance matrices and additional parameter constraints. We present several asymptotic results for misclassification probabilities and explore the performance of RP-QDA through numerical simulations in high-dimensional settings.

130 . Trade-off between dependence and complexity in Wasserstein distance learning

[03.M1.I44, (page 31)]

Nabarun DEB, *University of Chicago*
Debarghya MUKHERJEE,

The Wasserstein distance is a powerful tool in modern machine learning to metrize the space of probability distributions in a way that takes into account the geometry of the domain. Therefore, a lot of attention has been devoted in the literature to understanding rates of convergence for Wasserstein distances based on iid data. However, often in machine learning applications, especially in reinforcement learning, object tracking, performative prediction, and other online learning problems, observations are received sequentially, rendering some inherent temporal dependence. Motivated by this observation, we attempt to understand the problem of estimating Wasserstein distances using the natural plug-in estimator based on stationary beta-mixing sequences, a widely used assumption in the study of dependent processes. Our rates of convergence results are applicable under both short and long-range dependence. As expected, under short-range dependence, the rates match those observed in the iid. case. Interestingly, however, even under long-range dependence, we can show that the rates can match those in the iid case provided the (intrinsic) dimension is large enough.

131 . Nonparametric Regression of Spatio-temporal Data using Infinite-dimensional Covariates

[01.A1.I12, (page 7)]
Soudeep DEB, *Indian Institute of Management Bangalore*
Subhrajyoty ROY,
Sayar KARMAKAR,
Rishideep ROY,

In spatio-temporal analysis, we often record data at specific time intervals but with varying spatial locations between these timepoints. We propose a conditional model to analyze such spatio-temporal data that accommodates the dependencies alongside stationary temporal-level explanatory variables, which may be infinite-dimensional. Because of the absence of a mixing-type dependence condition in this case, which is usually required by the existing studies, we considered a weaker geometric moment contraction (GMC) condition on the covariates. In this paper, we obtain nonparametric point estimates of the mean and covariate functions of such a regression model which are statistically consistent. We also obtain a simultaneous confidence interval of the mean function using a central limit theorem for the proposed estimator. A few simulation studies and real-data analysis have been demonstrated to corroborate the findings.

132. Mathematical Characterization of Growth of Physical Parameters in the Human Developmental Age: Growth Curve Analysis of an Indian Birth Cohort

[03.A1.I56, (page 37)]
Vishal DEO, *ICMR-National Institute for Research in Digital Health and Data Science, New Delhi*
Saumyadipta PYNE, *Department of Statistics and Applied Probability, University of California Santa Barbara, CA, USA*
Meghana RAY, *Health Analytics Network, Pittsburgh, PA, USA*

Although NCDs are preventable, its burden has been on a constant rise across the world, and is expected to keep increasing in the near future. As per the study report "India: Health of the Nation's States" of the India State-Level Disease Burden Initiative by Indian Council of Medical Research (ICMR) published in 2017, it is estimated that the proportion of deaths due to NCDs in India have increased from 37.9% in 1990 to 61.8% in 2016. Based on some evidences, it has been argued that risk factors of NCDs are impacted by the sequential effects of developmental parameters and their trajectories since the early stage of life. Thus, there is a need to characterize the growth mechanism of developmental parameters, like anthropometry, body composition, etc., and their functional association with risk outcomes. For this study, our interest lies primarily in analysing the growth patterns of anthropometric and body composition variables of children measured at different stages of their childhood and early adolescence since birth, and to assess the association of early age growth patterns with the development of cardiovascular risk factors. The present study uses secondary analysis of the Mysore Parthenon Birth Cohort data which includes repeated measures on anthropometric parameters since birth, and body composition indicator since 5-year of follow-up , apart from measures on cardiovascular risk factors at 5, 9.5 and 13.5 years of follow-up. C.R. Rao in his 1958 paper published in the Biometrics journal, presented a conceptual framework for studying growth of an organism during a certain period of time and to compare the characteristics of growth under different conditions such as diet, environment, etc., using a mathematical representation known as growth curves. Further, in his paper published in 1959, C. R. Rao proposed a multivariate framework with covariance structure in addition to a growth curve structure to analyse growth curve data. These methods and their extensions have been used towards the objective of this paper.

133. Scalable piece-wise smoothing in high-dimensions with BART

[03.M2.I51, (page 35)]
Sameer DESHPANDE, *University of Wisconsin-Madison*
Ryan YEE, *University of Wisconsin–Madison*
Soham GHOSH, *University of Wisconsin–Madison*

Bayesian Additive Regression Trees (BART) is an easy-to-use and highly effective nonparametric regression model that approximates unknown functions with a sum of binary regression trees (i.e., piecewise-constant step functions). Consequently, BART is fundamentally limited in its ability to estimate smooth functions. Initial attempts to overcome this limitation replaced the constant output in each leaf of a tree with a realization of a Gaussian Process (GP). While these elaborations are conceptually elegant, most implementations thereof are computationally prohibitive, displaying a nearly cubic per-iteration complexity. We propose a version of

BART built with trees that output linear combinations of ridge functions; that is, our trees return linear combinations of compositions between an affine transformation of the inputs and a non-linear activation function. We develop a new MCMC sampler that updates trees in linear time. Our proposed model includes a random Fourier feature-inspired approximation to treed GPs as a special case. More generally, our proposed model can be viewed as an ensemble of local neural networks, which combines the representational flexibility of neural networks with the uncertainty quantification and computational tractability of BART.

134. Designs for Multi-Session Sensory Trials in the Presence of Assessor Constraints

[05.M1.C30, (page 61)]

Boyina DEVI PRIYANKA, *Indian Agricultural Statistics Research Institute*
Cini VARGHESE, *Principal Scientist and Professor*
Mohd HARUN, *Scientist*
Anindita DATTA, *Scientist*

Sensory trials form an integral part of food and nutrition experiments. It is important to recognize and control all factors that may influence or interfere with the result hindering valid conclusions. In addition to the sources associated with the preparation of the test products, there is variability due to measurements or assessment process, residual effects and assessor fatigue. In sensory trials, the assessor constriant sets a maximum number of products that an assessor can evaluate within a session before onset of sensory fatigue, and preparation constriant limits the number of products that can be prepared for a given session without loss of experimental control. Therefore, it is many times necessary to split sensory evaluations into multi-sessions. Here, general methods to construct incomplete block designs for multi-session sensory trials with the residual effects are developed.

135. Testing Independence of a pair of random variables - one discrete, another continuous

[03.M2.I52, (page 35)]

Isha DEWAN, *Indian Statistical Institute*

In competing risks one observes the failure time and the cause of failure . The failure time is a continuous random variable whereas cause of failure is a discrete random variable taking finitely many values. We look at nonparametric tests for testing the independence between these two.

136. Statistical inference of a competing risks model on improved adaptive type-II progressive censored data

[03.E1.C18, (page 43)]

Amlan DEY, *Scholar*

In this paper, a competing risks model is analyzed based on improved adaptive type-II progressive censored sample (IAT-II PCS). Two independent competing causes of failures are considered. It is assumed that lifetimes of the competing causes of failure follow exponential distributions with different means. Maximum likelihood estimators (MLEs) for the unknown model parameters are obtained. Using asymptotic normality property of MLE, the asymptotic confidence intervals are constructed. Existence and uniqueness properties of the MLEs are studied. Further, bootstrap confidence intervals are computed. The Bayes estimators are obtained under symmetric and asymmetric loss functions with non-informative and informative priors. For informative priors, independent gamma distributions are considered. Highest posterior density (HPD) credible intervals are obtained. A Monte Carlo simulation study is carried out to compare performance of the established estimates. Furthermore, three different optimality criteria are proposed to obtain the optimal censoring plan. Finally, a real-life data set is considered for illustrative purposes.

137. Multivariate Principal Component Analysis for Mixed-Type Functional Data with application to mHealth in Mood Disorders

[01.A1.I13, (page 8)]

Debangan DEY, *National Institute of Mental Health, USA*

TBA

138. Analyzing Latent Trajectory in Longitudinal Item Responses Using Gaussian Processes

[02.A1.I39, (page 25)]

Dipak DEY, *University of Connecticut*
Yuhao LI, *University of Connecticut*
Xiaojing WANG, *University of Connecticut*

Item Response Theory (IRT) models play a key

role in measurement testing. In the computerized testing scenarios, we observe a time series of item responses for an individual throughout the entire study period. Integrating time as a factor offers insights into learning how the individual's ability is changing over the time. Leveraging the flexibility of Gaussian process and its simplicity of parametric computation, we put forward a nonparametric IRT model to learn the growth of one's ability. The complexity of the proposed model makes Bayesian approach ideal in its analysis. To facilitate Bayesian computation, we propose a modified slice sampler to draw hyper-parameters in the joint posterior distribution, resulting in a much stable sampler in practice. In addition, we introduce the concepts of Bayesian surrogate residuals to access the goodness of our model fit, which can be used to assess the performance of IRT model in general. In this presentation, we have conducted several simulation studies to validate the efficacy of our approaches and applied the proposed model into a real testing dataset.

139. TBA

[05.M1.I86, (page 59)]

Kushal DEY, *Memorial Sloan Kettering Cancer Center*

TBA

140. Exploration of Hydrologic Persistence using Principles of Complex Systems

[02.E1.I41, (page 26)]

Pankaj DEY, *IIT Roorkee*
Pankaj DEY,

Complex system comprises of numerous interacting elements which give rise to a global emergent pattern due to the interaction of different elements and the surroundings. In this talk, catchments are perceived as complex systems and persistence (memory) of streamflow is modelled as an emergent phenomenon.The principles of complex systems such as emergence, self-organization and complexity are conceptualized to understand the physical mechanism of persistence in streamflow. The interactions among different elements of water resources systems generate a causal hierarchy and, therefore, to model and predict the state of the system, it is essential to demystify the causal structure for better managing the water resources. Several causal models are tested for detecting causality under the influence of persistence, and the sources of the limitation of these causal models will be discussed. This talk presents an alternative perspective in understanding memory of hydrologic systems, and linking it to catchment attributes.

141. Inference on Overlap Index: With an Application to Cancer Data

[03.E1.C15, (page 41)]

RAJU DEY, *Indian Institute of Technology Kharagpur*
Arne C. BATHKE, *Intelligent Data Analytics (IDA) Lab Salzburg, Department of Artificial Intelligence and Human Interfaces*
Somesh KUMAR, *Indian Institute of Technology Kharagpur, Department of Mathematics*

The quantification of overlap between two distributions has applications in various fields of medical, genetic, ecological, and geological research. In this article, new overlap and containment indices are considered for quantifying the niche overlap between two species/populations. Some new properties of these indices are established and the problem of estimation is studied, when the two distributions are exponential with different scale parameters. We propose several estimators and compare their relative performance with respect to different loss functions. The asymptotic normality of the maximum likelihood estimators of these indices is proved under certain conditions. We also obtain confidence intervals of the indices based on three different approaches and compare their average lengths and coverage probabilities. The point and confidence interval procedures developed here are applied on a breast cancer data set to analyze the similarity between the survival times of patients undergoing two different types of surgery.

142. Least Absolute Deviation (LAD) Regression for Cross Sectional Data with Cluster Dependence: Inference based on Averaging

[02.E1.C9, (page 29)]

Subhodeep DEY, *Indian Statistical Institute, Kolkata*
Gopal K. BASAK, *Indian Statistical Institute, Kolkata*
Samarjit DAS, *Indian Statistical Institute, Kolkata*

In this paper, we propose a new least absolute deviation (LAD) regression estimator based on cluster averaging. The estimator is shown to be consistent under three types of cross-sectional dependence, viz., strong, semi-strong and weak dependence. The asymptotic properties and the asymptotic distribu-

tion of the estimator is studied. A consistent estimate of the asymptotic covariance matrix is also proposed. A detailed simulation study shows the efficacy of the proposed estimator. As expected, the proposed estimator is robust to outliers and heavy-tailed distributions with cross-sectional dependence.

143. Bayesian Regularized Multivariate Regression: Unveiling Master Predictors in High-Dimensional Data

[03.M1.I47, (page 33)]

Tanujit DEY, *Brigham and Women's Hospital and Harvard Medical School*
Sounak CHAKRABORTY, *University of Missouri*
Priyam DAS, *Virginia Commonwealth University*

We focus on finding effective ways to uncover sparse relationships between predictors and responses, especially identifying "master predictors" in high-dimensional models. To tackle this, we introduce a new Bayesian penalized regression approach that combines L1 and squared L2 penalties to manage overall sparsity and predictor inclusion. We use Gibbs sampling to implement this method and validate its effectiveness with simulated data. We also apply our approach to identify master predictors in a microbiome cohort, predicting metabolomics in colorectal cancer patients. The results highlight the strength and usefulness of our Bayesian penalized regression model in revealing important relationships in complex datasets, with potential insights into the connection between microbiome composition and metabolomic profiles in colorectal cancer.

144. Optimal life testing plan in presence of hybrid censoring using compound optimal design strategy

[02.E1.C7, (page 28)]

Vaibhavbhai DHAMELIYA, *SVNIT Surat, Gujarat (395007)*
Vaibhavbhai DHAMELIYA, *SVNIT Surat, Gujarat (395007)*
Raj Kamal MAURYA, *SVNIT Surat, Gujarat (395007)*

Choosing the right life test plan is essential for improving product reliability and minimizing testing costs. In this talk, we will discuss about optimizing life tests by adjusting key factors, specifically within the Weibull model under hybrid censoring. The study strikes a balance between cost and variability, ensuring that the test designs are both economically viable and statistically robust. By using compound optimal designs, which take into account various criteria such as cost, trace, variance, and the determinant of the inverse Fisher information matrix, to develop comprehensive test plans. Using a graphical method, we find optimal designs that are not only straightforward to interpret but also deliver accurate results. Moreover, a sensitivity analysis is performed to evaluate the reliability of the proposed optimal designs, highlighting their advantages and potential drawbacks. Our findings suggest that compound optimal designs offer substantial benefits over conventional constrained designs, presenting a more holistic and effective strategy. To further support our approach, we apply the compound optimal design to a real-world data set, demonstrating its practical application in research. This study highlights the importance of considering multiple objectives in life test planning and shows how compound optimal designs can lead to more efficient and dependable outcomes.

Keywords: Weibull distribution, Hybrid censoring scheme, Cost function, Compound optimal design, Constraint optimal design.

145. Identifying arbitrary transformation between the slopes in functional regression

[03.M2.I50, (page 34)]

Subhra Sankar DHAR, *IIT Kanpur, India*
Pratim GUHA NIYOGI, *Johns Hopkins Bloomberg School of Public Health*

In this talk, we study whether the quantile curves of the slopes in the scalar on functional regression are the same or not up to a non-linear transformation. In order to carry out this test, we formulate a test statistic based on the estimated second derivative of the non-linear transformation, and the asymptotic distributions of the test statistic under null and a certain local alternatives have been derived. The proposed test performs reasonably well for various simulated and real data sets.

146. Deciding Optimal Number of Segments in DCA by Bivariate Clustering Approach

[04.E1.C26, (page 57)]

Shantaram DHUM, *Department of Statistics, Savitribai Phule Pune University, Pune*
Akanksha KASHIKAR, *Department of Statistics, Savitribai Phule Pune University, Pune*

In Ecology, Correspondence Analysis (CA) and Detrended Correspondence Analysis (DCA) are commonly used techniques in the ordination of species X sites (abundance data). This data is analogous to contingency table data; data in each table cell are frequencies. CA has two main drawbacks, both of which interdict the interpretation of results. I) Arch Effect or horseshoe effect II) CA does not preserve ecological distances. Out of these, the first is more serious. To eliminate the arch/ horseshoe effect, two methods of detrending, namely I) Detrending by segments, II) Detrending by polynomials, have been suggested in the literature. This gives birth to DCA. In detrending by segments, the first correspondence axis is divided into an arbitrary number of segments. The mean of scores along the second axis within each segment is made zero. This method arbitrarily chooses the number of segments, and different segmentation leads to different ordinations along the second correspondence axis. Twenty-six is the recommended number of segments; no theoretical justification is provided for this number. This paper presents a modified segmentation method to carry out the detrending process. This method also provides an approach to determine the optimal number of segments. The proposed method is compared with the existing one via simulations and a real data application.

147. Data-Driven Insights: Chemometric Analysis of Spectroscopic Data

[04.E1.C26, (page 57)]

Shreya DHUREKAR, *Department of Statistics Savitribai Phule Pune University*

The tablet under investigation is composed of two primary components, denoted as C and C . To thoroughly understand the composition, several mixtures with varying concentrations of these components were prepared. The dataset captures absorbance measurements of C and C across a wavelength range of 220-350 nm, with measurements taken at 4 nm intervals, ensuring detailed spectral resolution. To analyze this spectroscopic data, a novel chemometric analysis technique was developed, closely related to Principal Component Regression (PCR) but incorporating enhanced features to improve prediction accuracy. This technique utilizes calibrations derived from the absorbance data and concentration matrix datasets, allowing for the precise prediction of unknown concentrations of C and C within the tablet mixture. Originally 36 combinations of mixtures of components are available. The model with the first 2 principal components is fitted to the training data. The proposed method was rigorously tested, and it demonstrated highly accurate predictions, achieving a mean accuracy of approximately 100% for both C and C for the training data set. These results underscore the technique's robustness and reliability in quantifying the components, making it a valuable tool for complex spectroscopic analyses in pharmaceutical and chemical research.

148 . Predicting Response to Immunotherapy in Cancer Using Multi-Omic Data Integration and Machine Learning

[01.E1.C3, (page 12)]

MAHESH DIVAKARAN, *Amity University, Lucknow*
Gunjan SINGH, *Amity University, Lucknow*
Jayadevan SREEDHARAN, *Gulf Medical University*

Immunotherapy offers promising cancer treatments, but predicting patient response remains challenging. This study explores the integration of multi-omic data—genomics, transcriptomics, and proteomics—with machine learning to improve response prediction. Using publicly available datasets from The Cancer Genome Atlas (TCGA), the Gene Expression Omnibus (GEO), and the Human Protein Atlas, we combine these data sources to develop predictive models.

We preprocess genomic variants, RNA-seq expression profiles, and protein levels, followed by feature selection with methods like LASSO regression and ensemble feature importance. Various machine learning models, including Random Forests, Gradient Boosting, and deep learning techniques, are then applied to predict immunotherapy outcomes.

Model performance is evaluated using metrics such as ROC-AUC and precision-recall curves. Our findings demonstrate that integrating multi-omic data significantly enhances predictive accuracy, identifying key biomarkers and pathways linked to treatment response. This approach provides insights into personalized immunotherapy, offering potential for more effective patient stratification and treatment optimization in oncology.

149. Change Point Detection for Random Objects using Distance Profiles

[03.M1.I44, (page 31)]

Paromita DUBEY, *University of Southern California*
Minxing ZHENG, *University of Southern California*

We introduce a new powerful scan statistic and an associated test for detecting the presence and pinpointing the location of a change point within the distribution of a data sequence where the data elements take values in a general separable metric space (Ω,d). These change points mark abrupt shifts in the distribution of the data sequence. Our method hinges on distance profiles, where the distance profile of an element Ω is the distribution of distances from as dictated by the data. Our approach is fully nonparametric and universally applicable to diverse data types, including distributional and network data, as long as distances between the data objects are available. From a practicable point of view, it is nearly tuning parameter-free, except for the specification of cut-off intervals near the endpoints where change points are assumed not to occur. Our theoretical results include a precise characterization of the asymptotic distribution of the test statistic under the null hypothesis of no change points and rigorous guarantees on the consistency of the test in the presence of change points under contiguous alternatives, as well as for the consistency of the estimated change point location. Through comprehensive simulation studies encompassing multivariate data, bivariate distributional data and sequences of graph Laplacians, we demonstrate the effectiveness of our approach in both change point detection power and estimating the location of the change point. We apply our method to real datasets, including U.S. electricity generation compositions and Bluetooth proximity networks, underscoring its practical relevance.

150. Optimal Iterative Algorithms for Structured PCA with Invariant Noise

[04.M1.I65, (page 45)]

Rishabh DUDEJA, *UW Madison*
Songbin LIU, *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*
Junjie MA, *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*

We consider the problem of recovering a low-rank signal matrix from a noisy observed matrix corrupted with additive noise. When the noise matrix is i.i.d. Gaussian, a rich line of work has characterized the information-theoretic limits for this problem and determined the smallest possible estimation error achievable by computationally efficient estimators. The i.i.d. noise model constrains the eigenvalue spectrum of the observed matrix to follow the semi-circle law, which may not accurately represent all datasets. We study a flexible generalization of the i.i.d. Gaussian noise model known as the rotationally invariant noise model, which can capture noise spectrums beyond the semi-circle law. We develop a new class of approximate message-passing algorithms for this problem and characterize their dynamics. These algorithms leverage prior knowledge about the noise and signal structures by iteratively applying non-linear denoisers to the eigenvalues of the observed matrix and the previous iterates. We identify the optimal choices for these denoisers and provide evidence suggesting that the resulting algorithm is a natural candidate for the optimal computationally efficient algorithm by showing that it achieves the smallest possible estimation error among a broad class of iterative algorithms under a given iteration budget.

151. Bayesian Model Averaging for Implicit Generative Neural Model

[04.A1.I80, (page 51)]

Ritabrata DUTTA, *University of Warwick*
Sherman KHOO, *University of Bristol*
Shreya SINHA ROY, *University of Warwick*

This work presents a novel application of Bayesian Model Averaging to implicit, high- dimensional generative models parameterized with neural networks (GNN), which are char- acterized by intractable likelihoods. We leverage a likelihood-free, generalized Bayesian inference approach using Scoring Rules. However, model selection in such a setting remains a challenge. To tackle this problem, we adopt a principled approach using a continuous shrinkage prior, specifically, the horseshoe prior. For inference, we introduce an innovative blocked sampling scheme, offering compatibility with both the Boomerang Sampler for exact but slower inference and with Stochastic Gradient Langevin Dynamics (SGLD) for faster yet biased posterior inference. Further, we illustrate how the horseshoe prior can help us to ap- proximate the preconditioning matrix to facilitate the Markov chain Monte Carlo sampling for high-dimensional posterior. This approach serves as a versatile tool bridging the gap between intractable likelihoods and robust Bayesian model selection within the generative modelling framework.

152. Joint modeling of playing time & purchase propensity in massively multiplayer online role-playing games

[02.M2.I28, (page 18)]

Shantanu DUTTA, *University of Southern California*
Gourab MUKHERJEE, *University of Southern California*
Trambak BANERJEE,

Abstract: Massively Multiplayer Online Role-Playing Games (MMORPGs) offer a unique blend of a personalized gaming experience and a platform for forging social connections. Managers of these digital products usually rely on predictions of key player responses, such as playing time and purchase propensity, to design timely interventions for promoting, engaging, and monetizing their playing base. However, the longitudinal data associated with these MMORPGs not only exhibit a large set of potential predictors to choose from but also present several other distinctive characteristics that pose significant challenges in developing flexible statistical algorithms capable of generating efficient predictions of future player activities. For instance, the existence of virtual communities or guilds in these games complicates prediction since players who are part of the same guild exhibit correlated behaviors, and the guilds themselves evolve over time, thereby having a dynamic effect on the future playing behavior of their members. In this paper, we develop a Crossed Random Effects Joint Modeling (CREJM) framework for analyzing correlated player responses in MMORPGs. Contrary to existing methods that assume player independence, CREJM is flexible enough to incorporate both player dependence and time-varying guild effects on the future playing behavior of guild members. Using large-scale data from a popular MMORPG, CREJM simultaneously selects fixed and random effects in high-dimensional penalized multivariate mixed models. We study the asymptotic properties of the variable selection procedure in CREJM and establish its selection consistency. Beyond providing superior predictions of daily playing time and purchase propensity compared to competing methods, CREJM also predicts player correlations within each guild, which are valuable for optimizing future promotional and reward policies for these virtual communities.

153 . On Exact Feature Screening in Ultrahigh-dimensional Classification

[01.E1.I20, (page 10)]

Subhajit DUTTA, *ISI Kolkata and IIT Kanpur*

In this talk, we first motivate and analyze the well-known average distance classifier and its variants in the high-dimensional scenario. We will then discuss a new model-free feature screening method based on energy distances for ultrahigh-dimensional binary classification problems. Unlike existing methods, the cut-off involved in our procedure is data adaptive. With a high probability, our procedure retains only relevant features after discarding all the noise variables. The proposed screening method is also extended to identify pairs of variables that are marginally undetectable but have differences in their joint distributions. Finally, we build a classifier that maintains coherence between the proposed feature selection criteria and discrimination method and also establish its risk consistency. A numerical study shows clear and convincing advantages of our classifier over existing state-of-the-art methods.

154. Forecasting Lung Cancer Incidence in South Kerala, India using Bootstrap Methods

[03.E1.C15, (page 41)]

JINTO E G, *Regional Cancer Centre, Thiruvananthapuram, Kerala*
Aleyamma MATHEW, *Professor and Head Division of Cancer Epidemiology & Biostatistics Regional Cancer Centre, Thiruvananthapuram*
Preethi Sara GEORGE, *Additional Professor In Biostatistics Division of Cancer Epidemiology & Biostatistics Regional Cancer Centre, Thiruvananthapuram*

Lung cancer (LC) incidence has been increasing over the past decades in India. This study aims to analyse time-trends of LC and forecast through 2030 in South Kerala. Data were obtained from Cancer Registry, Kollam (1991-2020) and Thiruvananthapuram (2005-2020) in Kerala. Joinpoint regression was used to identify the time-trends of LC crude incidence (CR per 105) and estimated by average annual percent change (AAPC). Bootstrap methods were used for forecasting rates till 2030. Accuracy of the model was measured by normalized root mean squared error. This analysis included 4,243 LC cases, CR increased from 13.6 (in 1991) to 40.3 (in 2020) in Kollam and 14.2 (in 2005) to 26.2 (in 2020) in Thiruvananthapuram among men, and increased from 1.0 (1991) to 8.0 (2020) in Kollam and 2.4 (2005) to 8.4 (2020) in Thiruvananthapuram among women. CR was higher among 60yrs in both gender. AAPC among women was 6.6 in Kollam and 8.9 in Thiruvananthapuram with 12.4 in <60 years whereas the AAPC was 4.1 among men in both areas. The fore-

casted CR using the best bootstrap model for men is 52.7 in Kollam and 34.4 in Thiruvananthapuram, and for women 10.4 in Kollam and 12.3 in Thiruvananthapuram for 2030. In Kerala, LC incidence rates increased significantly in both gender and higher the rate of increase among younger women. More research works are needed to identify the reasons for the increased incidence particularly among women.

155. Bivariate Cumulative Residual Entropy of Equilibrium Distribution of Order n

[03.M1.C11, (page 33)]

Rajesh G, *Cochin University of Science and Technology*
Unnikrishnan NAIR N, *Cochin University of Science and Technology*
Sajily V.S., *Cochin University of Science and Technology*

Nair et al. (2023) [Communications in Statistics-Theory and Methods, 52(3):851-863.] have introduced the cumulative residual entropy (CRE) of order n and studied its importance in reliability theory. The present paper addresses extending this measure to higher dimensions and studies its properties. We use this measure to characterize some well-known bivariate lifetime models and study their relations with reliability measures such as product moment residual life and vector-valued failure rates. Several properties are obtained, including monotonicity and bounds based on well-known Fréchet- Hoeffding bounds. Moreover, we also find an implication between bivariate CRE and positively (negatively) quadrant-dependent PQD (NQD) distributions.

156. Rhetoric and Reality: Tracing the Evolution of Policy Discourse, Agenda Setting, and Populism in Indian Parliamentary Speeches

[02.M2.I27, (page 17)]

Kausik GANGOPADHYAY, *Indian Institute of Management Kozhikode*
Swarn RAJAN, *Indian Institute of Management Kozhikode*
Anirban GHATAK, *Indian Institute of Management Kozhikode*

In a parliamentary system, where prime minister is the head of government, delivers speeches on the house of parliament to convey their stance and their party position on various policy matters. This study innovatively analyzes the evolution of policy discourse and agenda-setting in India through Prime Ministers' parliamentary speeches from 1952 to 2024. Utilizing a novel methodological framework of text analysis, it combines Dynamic Topic Modeling (DTM), Structural Topic Modeling (STM), and hierarchical clustering to examine speeches across 17 Lok Sabhas and 15 Prime Ministers. The application of DTM, with a one-year window topic analysis, reveals year-on-year shifts in political focus, providing granular insights into the evolution of political agendas. STM, enhanced with metadata, highlights specific topics emphasized during different tenures, offering a nuanced understanding of the political agenda under various leaderships. Hierarchical clustering categorizes topics into broader themes, revealing structural underpinnings of political communication.

We have also used advanced text analysis technique of Word-embedding to understand the distance between various Prime ministers over different Lok Sabhas on various policy issues of Economy, National security, International Relation and Culture. The findings indicate a punctuated rise in populism and change in policy focus, marked by periods of intensity and retreat, reflecting complex dynamics in Indian politics. Our analysis of election period speeches suggests sudden change in policy focus in recent elections while more continuity in the earlier elections. Our results also suggests significant change in the policy discourse on matter of culture especially with changing party in power, while suggesting a gradual change and even continuity on the economic, security and international relation policy front. This research provides valuable insights into the analysis of political discourse, populism and policy discourse dynamics, benefiting policymakers, scholars, and the public by deepening understanding of India's democratic dialogue.

157. Statistical hydrology in the age of AI

[04.M1.I66, (page 45)]

Auroop GANGULY, *Northeastern University*

This presentation explores the evolution of the latest-generation machine learning and artificial intelligence in earth systems sciences and engineering from the perspective of the decades of development in statistical hydrology, including data-driven sciences in water, weather, and climate. The connections across the natural, human-engineered, and social sciences are discussed, leading to a notion of conver-

gence of complexities. Resilience to global change, despite the challenges of nonlinearity, extremes, interconnectedness, incentivization, and policy myopia are discussed. Potential solutions that integrate human knowledge and diverse suite of models with data-driven sciences and engineering are discussed with illustrative case studies.

158 . On improved estimation of the larger location parameter

[04.E1.C27, (page 58)]

Naresh GARG, *Indian Statistical Institute, Delhi Centre*
Lakshmi Kanta PATRA, *Indian Institute of Technology Bhilai, Bhilai, India*
Neeraj MISRA, *Indian Institute of Technology Kanpur, Kanpur, India*

This paper investigates the problem of estimating the larger location parameter of two general location families of distributions from a decision-theoretic perspective. The criteria of minimizing the risk function and the Pitman nearness, under a general bowl-shaped loss function, are considered. Inadmissibility of certain location and permutation equivariant estimators is proved and dominating estimators are obtained. It follows that a natural estimator (a plug-in estimator based on the best location equivariant estimators of (unordered) location parameters) is inadmissible, under certain conditions on underlying densities, and the loss function. A class of dominating estimators is provided. We also consider a class of linear and, location and permutation equivariant estimators and obtain a subclass of estimators that are admissible within the full class of estimators. We observe that the natural estimator is a boundary estimator in obtained subclass of admissible estimators. Further, using the IERD technique of Kubokawa (1994), we obtain an estimator dominating over another natural estimator. Additionally, under the generalized Pitman nearness criterion with a general bowl-shaped loss function, we show that two natural estimators are inadmissible and obtain improved estimators. The results are applied to specific loss functions, and explicit expressions for dominating estimators are obtained. We illustrate applications of these results to normal and exponential distributions for specified loss functions. A simulation study is also conducted to compare risk performances of different competing estimators.Finally, we present a real-life data analysis to illustrate a practical application of the findings of the paper.

159 . Evaluating Randomness Assumption: A Novel Graph Theoretic Approach for Linear and Circular Data

[04.E1.C27, (page 57)]

Shriya GEHLOT, *Indian Institute of Management Ahmedabad*
Arnab Kumar LAHA,

Randomness or mutual independence is an important underlying assumption for most widely used statistical methods in both linear and circular contexts. However, not many tests are available to check for randomness, particularly for circular data. In this paper, we introduce a new approach for developing non-parametric tests for linear and circular data. We introduce a new concept of Random Circular Arc Graphs (RCAG) for circular data analogous to that of Random Interval Graphs (RIG) for linear data. We examine various properties of the RCAGs, including edge probability, vertex degree distribution, maximum and minimum degrees, and the presence of Hamiltonian cycles. Then, we use them to create randomness tests for circular data. Similar ideas lead to new tests of randomness for linear data. For linear data, we demonstrate that our test outperforms most of the standard parametric and non-parametric tests available in the literature, including the Runs test. Similarly, we substantiate the effectiveness of our tests for circular data through extensive simulations. Several real-world applications of these tests are discussed.

160 . Dynamic Survival Prediction by Landmarking Using Parametric Proportional Hazards Models

[04.E1.C24, (page 55)]

JENET GEORGE, *Cochin University of Science and Technology*
SREEDEVI E. P., *Cochin University of Science and Technology*

Landmarking is a dynamic prediction technique used for analyzing time to event data with time-dependent covariates. The landmark approach enables us to update the survival probability and the hazard rate of patients, as new information on covariate values is available during the study period. These dynamic updations can be made by fitting survival models for individuals who are still at risk at each landmark time point. In this study, we present a parametric proportional hazards model for exploring the relation between lifetime and time-dependent co-

variates, which updates the survival probability and the hazard rate using landmarking approach. Inference procedures are carried out using maximum likelihood estimation method. We conduct simulation studies to validate the finite sample behavior of the proposed dynamic prediction model. We also illustrate the practical applicability of the model by applying it to a real data set.

161. Advances in the species distribution modelling for alien invasive weeds

[01.M2.I4, (page 5)]

Yogita GHARDE, *ICAR-Directorate of Weed Research*

The alien invasive species (AIW) are those, which are introduced into places outside their natural range, adversely affecting native biodiversity, ecosystem or human well-being. These species are introduced purposefully or accidentally outside their natural habitat, where they exhibit the ability to establish themselves, invade, out-compete native weeds and take over the new environment within short span of time. Further, their uncontrolled expansion in agriculture ecosystem may cause huge crop yield losses. Nowadays, ecological niche models (ENMs) are widely being used by researchers to examine the species' geographical distribution and to predict the probable shifts in species ranges under climate change. ENMs are important tools for predicting species ranges by correlating different types of variables including environmental, topographical, human-related, with the known distribution of a species. On the other hand, Species Distribution Modelling is a method of estimating fundamental ecological niches of alien invasive species based on association between point-occurrence data and raster data layers giving environmental information and the models used for this purpose are called Species Distribution Models (SDMs). These models are set of computer algorithms that are used to predict the species distribution in geographic space based on mathematical representation of their known distribution in ecological niche. As the climate is known to be the most significant factor affecting the growth and development, invasive weeds are heavily influenced by climate change and can extend their range, thereby causing increased damage to ecosystem and agricultural production. Studying the effects of future climate change on distribution of species is one of the fundamentals to manage informative activities for conservation of biodiversity. These models have become the extensively useful tool to know the relationships between species and their environments, and are used to predict extreme impacts of climate change, biogeographic studies, improve species management and answer conservation biology questions. Therefore, there is always need for precise prediction on dynamics of AIW under future climate change scenarios in order to manage such weeds.

162. Markov chains for spin systems from random initializations

[02.M2.I31, (page 19)]

Reza GHEISSARI, *Northwestern University*

Markov chains are ubiquitous as models for physical time-evolutions in statistical physics, as natural sampling schemes for high-dimensional distributions, and as soft optimization algorithms. A family of rich high-dimensional distributions that arise in various fields and for which Markov chains are especially pertinent are spin systems (e.g., the Ising and Potts models). Classical Markov chain analysis is focused on mixing times (i.e., time to approximate stationarity) from a worst-case initialization, but tools for understanding mixing times from natural (e.g., uniform-at-random or extremal) initializations are more limited. We describe some recent results on regimes where spin system dynamics are exponentially slow to mix from worst-case initializations, but where mixing from random or special initializations can be analyzed and shown to be significantly faster to equilibrate.

163 . Zero-One Modified Poisson INAR(1) Process

[04.E1.C25, (page 56)]

Aishwarya GHODAKE, *PhD Scholar*
Manik AWALE, *Associate Professor*

In this paper, we propose an INAR model of order one with a zero-one modified Poisson marginal distribution (ZMPoINAR(1)), based on a binomial thinning operator. The proposed model is capable of capturing both underdispersion and overdispersion, which may be caused by the deflation or inflation of zeros and/or ones. The authors have studied the probabilistic and inferential properties of the model. We used prediction root mean squared error (PRMSE), prediction mean absolute error (PMAE), and percentage of true prediction (PTP) measures to evaluate the forecast performance of the proposed model. The forecast performance of the model is also compared with that of competitive models. A real

data set has been analyzed using the proposed model.

Keywords: Integer-valued autoregressive (INAR); Zero-one modified Poisson INAR model; thinning operator; Yule-Walker; Conditional least squares; Conditional maximum likelihood.

164 . Functional proportional hazards mixture cure model

[04.M1.I67, (page 45)]

Rahul GHOSAL, *University of South Carolina*
Marcos MATABUENA, *Harvard University*
Jiajia ZHANG, *University of South Carolina*

We develop a functional proportional hazards mixture cure (FPHMC) model with scalar and functional covariates measured at the baseline. The mixture cure model, useful in studying populations with a cure fraction of a particular event of interest, is extended to functional data. We employ the EM algorithm and develop a semiparametric penalized spline-based approach to estimate the dynamic functional coefficients of the incidence and the latency part. The proposed method is computationally efficient and simultaneously incorporates smoothness in the estimated functional coefficients via roughness penalty. Simulation studies illustrate a satisfactory performance of the proposed method in accurately estimating the model parameters and the baseline survival function. Finally, the clinical potential of the model is demonstrated in a real data example that incorporates rich high-dimensional biomedical signals as functional covariate. In particular, we analyze minute-by-minute physical activity data from the National Health and Nutrition Examination Survey (NHANES) 2003-2006 to study the association between diurnal patterns of physical activity (PA) at baseline and all cancer mortality through 2019 while adjusting for other biological factors. Our findings provide novel epidemiological insights into the association between daily patterns of PA and cancer mortality.

165 . Heterogeneous Graphon JSQ(d) model

[03.A1.I57, (page 37)]

Arka GHOSH, *Iowa State University*
Ruoyu Wu - Iowa State University (United States) YANHAN CHEN - IOWA STATE UNIVERSITY (UNITED STATES),

A variation of the supermarket model is considered in which a task arriving at a dispatcher is routed to one of its neighborhood servers based on the JSQ(d) strategy. Both heterogeneous dispatchers and servers are considered whose neighborhood relationships are described by a deterministic graphon. The evolution of the queue length for each server is described in the form of stochastic differential equations in which the interaction between servers exists. The law of large number results, both locally and globally, are established as the size of the system grows and the underlying graphons converge. The interacting system is proven to converge to an independent but heterogeneous system.

166. Provable and Efficient Algorithms for Heterogeneous and Byzantine Robust Federated Learning

[04.A1.I81, (page 52)]

Avishek GHOSH, *IIT Bombay*
Dong YIN, *Research Scientist, Apple AI*
RajKumar MAITY , *Data Scientist, Microsoft*

In recent years, we have seen a huge interest in the field of large scale distributed optimization. Federated Learning is a distributed learning framework that leverages the compute power of users' personal devices (mobile, laptop) to solve an overall optimization problem. However, there are several caveats in Federated Learning. In this talk, we will address 3 such issues: (i) data heterogeneity across user devices; (ii) communication/uplink cost from users to server and (iii) Byzantine attacks. In particular, we will propose several iterative algorithms that are provable, computationally efficient and practical. We obtain sharp convergence rates of such algorithms and show that they yield optimal statistical (error) rates. Furthermore, we run experiments on real-world dataset (LEAF database, which is standard for Federated Learning) and compare with several baselines.

167. Regularized Additive Matrix Autoregressive Model

[01.E1.C4, (page 12)]

Debika GHOSH, *Indian Institute of Management, Udaipur*
Samrat ROY, *Indian Institute of Management, Ahmedabad*
Nilanjana CHAKRABORTY, *Indian Institute of Management, Udaipur*

High dimensional time series has diverse applications in macro-econometrics and finance. Recent

models for capturing the temporal dependence employs a bilinear representation in case of the matrix time series, or the Tucker-decomposition based representation in case of tensor time series. A bilinear or Tucker-decomposition based temporal effect is difficult to interpret on many occasions, along with its computational complexity due to the non-convex nature of the underlying optimization problem. Moreover, existing matrix case models do not explore the possibilities of imposing any lower dimensional pattern on the coefficient matrices. In this work, we propose a regularized additive matrix autoregressive model with additive interaction effect, that offers more interpretability, less computational burden due to its convex nature and estimation of the underlying low rank plus sparse pattern of its coefficient matrices. We address the issue of identifiability of various components in our model and subsequently develop a scalable Alternating Block Minimization algorithm for estimating the parameters. We provide finite sample error bounds under high dimensional scaling for the model parameters. Finally, the efficacy of the proposed model is demonstrated on synthetic and real data.

Key words: High-Dimensional, temporal dependence, low rank, sparse

168. A novel longitudinal rank-sum test for multiple primary endpoints in clinical trials

[03.M1.I44, (page 31)]

Dhrubajyoti GHOSH, *Duke University*
Sheng LUO, *Duke University*
Xiaoming XU,

Neurodegenerative disorders such as Alzheimer's disease (AD) present a significant global health challenge, characterized by cognitive decline, functional impairment, and other debilitating effects. Current AD clinical trials often assess multiple longitudinal primary endpoints to comprehensively evaluate treatment efficacy. Traditional methods, however, may fail to capture global treatment effects, require larger sample sizes due to multiplicity adjustments, and may not fully exploit multivariate longitudinal data. To address these limitations, we introduce the Longitudinal Rank Sum Test (LRST), a novel nonparametric rank-based omnibus test statistic. The LRST enables a comprehensive assessment of treatment efficacy across multiple endpoints and time points without multiplicity adjustments, effectively controlling Type I error while enhancing statistical power. It offers flexibility against various data distributions encountered in AD research and maximizes the utilization of longitudinal data. Extensive simulations and real-data applications demonstrate the LRST's performance, underscoring its potential as a valuable tool in AD clinical trials. We have also provided an extension of the LRST to multi-arm clinical trials, along with a Power Analysis and Sample Size (PASS) calculations.

169 . Global Local Priors for Spatial Small Area Estimation

[02.M2.I26, (page 17)]

Malay GHOSH, *University of Florida*
Xueying TANG, *University of Arizona*

The paper introduces global local priors in the context of spatial small area estimation. One big advantage of such priors is that they can capture variation between the different small areas very well in contrast to the classic Fay-Herriot model which includes only a global parameter. The results are illustrated with American Community Survey Poverty Data for states and counties.

170 . Professor P.K. Sen-A Friend , Guide and Philosopher

[Memorial Session 1, (page 26)]

Malay GHOSH, *University of Florida*

TBA

171 . Personalized Medicine: Adaptive Randomization in SMART Designs

[03.A1.I60, (page 38)]

Palash GHOSH, *IIT Guwahati*

Personalized medicine is a paradigm shift from the traditional way of "one size fits all" to individualized, patient-centric management that addresses the heterogeneity across patients and considers the variability within patients over time. It generally refers to tailoring using the genetic profile of an individual, but it is also common to personalize and allocate treatments to patients based on medical history, which could be time-varying. In this talk, we will focus on the Sequential Multiple Assignment Randomized Trials (SMARTs) that offer a rich source of data for developing, evaluating, and comparing adaptive intervention (treatment) sequences or dynamic treatment regimes (DTR) as a sequence of decision rules. First, we will discuss the need for adaptive randomization in SMART designs. Then, we will develop

the methodology that addresses the multi-stage randomization processes. We will discuss the optimality of the developed method. Using simulation and real data, we will show that adaptive randomization can be useful in treating more patients with a better treatment sequence (DTR), during the trial period.

172 . Optimal Adaptive Strategies in SMART

[03.E1.C13, (page 40)]

Rik GHOSH, *Indian Institute of Technology, Guwahati*
Bibhas CHAKRABORTY, *Centre for Quantitative Medicine, Duke-NUS Medical School, National University of Singapore, Singapore*
Inbal, Megan E. NAHUM-SHANI, PATRICK, *Institute for Social Research, University of Michigan*
Palash GHOSH, *Indian Institute of Technology Guwahati, Assam*

In a sequential multiple-assignment randomized trial (SMART), a sequence of treatments is given to a patient over multiple stages. In each of the stages, randomization may be done to allocate patients to the different treatment groups that have been considered for the SMART. In spite of SMART designs getting popular among clinicians/clinical researchers, the methodologies for the adaptive randomization at different stages of a SMART are still few and not sophisticated enough to deal with the complexity of optimal allocation of treatments at each and every stage of the trial. Lack of optimal allocation methodologies can raise serious concerns about SMART designs from an ethical point of view. In this work, we have developed an optimal adaptive allocation procedure to minimize the expected number of treatment failures for a SMART which has a binary primary outcome. Issues related to optimal adaptive allocations are explored theoretically along with detailed supporting simulations. The applicability of the proposed methodology is also demonstrated using a recently conducted SMART study named M-Bridge for developing an universal and resource-efficient dynamic treatment regimes (DTRs) for incoming first-year college students as a bridge to desirable treatments to address alcohol-related risks for those students.

173 . Signal-to-noise-ratio aware minimax analysis of sparse linear regression

[Student Poster Competition, (page 20)]

Shubhangi GHOSH, *Columbia University*
Yilin GUO, *Two Sigma*
Haolei WENG, *Michigan State University*
Arian MALEKI, *Columbia University*

The minimax framework has been one of the cornerstones of theoretical statistics, and has contributed to the popularity of many well-known estimators, such as the regularized M-estimators and regularized linear regression estimators for high-dimensional problems. In this paper, we demonstrate that numerous theoretical results within the classical minimax framework are inadequate in explaining empirical observations. In some instances, these minimax outcomes offer insights that contradict empirical findings. For example, although LASSO has been proven to be minimax optimal for the sparse linear regression problem, numerous empirical studies have shown its suboptimal performance across various signal-to-noise (SNR) levels. In this study, we aim to introduce an enhanced version of the minimax framework that not only elucidates these disparities but also offers more precise insights into the optimality of different estimators.

Our novel approach has two two distinctive components: (1) it integrates the signal-to-noise ratio into the construction of the parameter space. (2) It obtains accurate approximation of the minimax risk through asymptotic arguments. The theoretical findings derived from this refined framework provide new insights and practical guidance. For instance, in the context of estimating sparse signals under the linear regression model, our approach demonstrates that in the low SNR, ridge regression surpasses all other estimators, even when the regression coefficients are sparse.

174 . Bayesian Semi-supervised Multi-category Classification under Nonparanormality

[Student Paper Competition 1, (page 15)]

Shuvrarghya GHOSH, *North Carolina State University*
Rui ZHU, *Google Inc.*
Subhashis GHOSAL, *North Carolina State University*

Semi-supervised learning is a model training method that uses both labeled and unlabeled data. This paper proposes a fully Bayes semi-supervised learning algorithm that can be applied to any multi-category classification problem. We assume the labels are missing at random when using unlabeled data in a

semi-supervised setting. Suppose we have K classes in the data. We assume that the observations follow K multivariate normal distributions depending on their true class labels after some common unknown transformation is applied to each component of the observation vector. The function is expanded in a B-splines series, and a prior is added to the coefficients. We consider a normal prior on the coefficients and constrain the values to meet the normality and identifiability constraints requirement. The precision matrices of the Gaussian distributions are given a conjugate Wishart prior, while the means are given the improper uniform prior. The resulting posterior is still conditionally conjugate, and the Gibbs sampler aided by a data-augmentation technique can thus be adopted. An extensive simulation study compares the proposed method with several other available methods. The proposed method is also applied to real datasets on diagnosing breast cancer and classification of signals. We conclude that the proposed method has a better prediction accuracy in various cases.

175. Moment inequality for decreasing mean time to failure distributions with hypothesis testing application

[03.M1.I46, (page 32)]

Shyamal GHOSH, *School of Dada Science, IISER Thiruvananthapuram*

Based on a moment inequality, a family of test statistics for testing exponentiality against DMTTF alternatives is proposed. The asymptotic distribution of the test statistics is derived under the null and alternative hypothesis, and the consistency of the test is shown by exploiting the U-statistics theory. Comparisons with competing tests are made in terms of Pitman Asymptotic Relative Efficiency (PARE). Additionally, an adapted version of the test under random censorship is explored. The performance of the proposed test has been accessed by means of a simulation study and through application to some real-life data sets.

176. The Curious Problem of the Inverse Mean

[Student Poster Competition, (page 22)]

Soham GHOSH, *University of Wisconsin, Madison*
Uttaran CHATTERJEE, *School of Industrial Engineering, Purdue University*
Jyotishka DATTA, *Department of Statistics, Virginia Tech*

In astronomical observations, the estimation of distances from parallaxes is a challenging task due to the inherent measurement errors and the non-linear relationship between the parallax and the distance. This study leverages ideas from robust Bayesian inference to tackle these challenges, investigating a broad class of prior densities for estimating distances with a reduced bias and variance. Through theoretical analysis, simulation experiments, and the application to data from the Gaia Data Release 1 (GDR1), we demonstrate that heavy-tailed priors provide more reliable distance estimates, particularly in the presence of large fractional parallax errors. Theoretical results highlight the "curse of a single observation," where the likelihood dominates the posterior, limiting the impact of the prior. Nevertheless, heavy-tailed priors can delay the explosion of posterior risk, offering a more robust framework for distance estimation. The findings suggest that reciprocal invariant priors, such as the Half-Cauchy and Product Half-Cauchy, are particularly well-suited for this task, providing a balance between bias reduction and variance control.

177. Novel and Efficient Pipeline for Metagenomics Binning

[02.E1.C8, (page 28)]

SUBHAM GHOSH, *ICAR-Indian Agricultural Research Institute*
ULAVAPPA ANGADI, *Principal Scientist, Division of Agricultural Bioinformatics,ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*
DIPRO SINHA, *Research Associate,Division of Agricultural Bioinformatics,ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*
MIR ASIF IQUEBAL, *Principal Scientist,Division of Agricultural Bioinformatics,ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*

Metagenomics delves into the examination of microorganisms, and a pivotal aspect of this field involves piecing together the genetic makeup of distinct organisms. This task proves challenging due to the complexities of isolating and cloning certain organisms under in-vitro conditions. Metagenomics is alternatively termed environmental genomics, ecogenomics, or community genomics. To reconstruct the fragmented sequences obtained from shotgun sequencing, the process heavily relies on genome assembly. However, a significant hurdle arises when attempting to segregate and reassemble genomes from various organisms. The abundance of these genomes

and the intermingling of genomics reads present a formidable challenge. Shotgun sequencing produces genomic reads that contain fragments originating from diverse microorganisms' genomes. To facilitate reconstruction, it becomes imperative to classify these reads into separate bins corresponding to distinct microorganisms. For this purpose, various clustering techniques have emerged for the categorization of these intertwined genomes. These techniques encompass binning, boosting, bagging, and stacking. Among these, binning has gained prominence as the most extensively utilized algorithm in contemporary times. To put it differently, genomes are categorized into operational taxonomic units (OTUs) to facilitate subsequent taxonomic profiling and subsequent functional analysis. This process of OTU clustering is commonly referred to as binning. In this clustering process, binning employs a variety of clustering methods such as k-means, k-medoids, Hidden Markov Model (HMM), and hierarchical clustering. However, each of these clustering approaches comes with its own limitations and drawbacks. There is a no research on motif-based binning in the existing ones. Here an approach is given for metagenomic binning by constructing frequency table of motif or segments by using local alignment with gap by dynamic programming technique, the segments/motifs are non-overlapped. K-means clustering, PAM clustering and DBSCAN clustering are applied to cluster the contigs based on the segments/motifs. But K-means clustering has performed the best. The rand indexes for this approach are tend to 1. So, this approach is good for metagenomics binning. And it is also performing better than the existing binning tools, i.e., MaxBin and MetaBat. And this approach has a lot of scope. In the place of simple K-means clustering, many advanced clustering can be used for better performance. GC content, tetra-nucleotide frequency can be added for getting better performance. This approach also highlights the mutation concepts and conserved regions, which are too much necessary to get the idea of evolutionary biology.

178 . Bayesian model and uncertainty quantification for Time of Emergence

[02.M2.I29, (page 19)]

Sujit GHOSH , *NC State University*
Suresh IYYAPPAN, *Digital University of Kerala*
Radhendushka SRIVASTAVA , *Indian Institute of Technology Bombay*

In recent years, the study of climate change and its impacts has emphasized the importance of accurately determining the Time of Emergence (ToE) of significant climate signals from the background noise. This talk focuses on the application of Bayesian change point methods and Markov Chain Monte Carlo (MCMC) techniques to estimate the ToE in the context of sea surface temperature (SST) data from the Arabian Sea and Bay of Bengal. By utilizing a Bayesian framework, we can effectively quantify the uncertainty associated with ToE estimates, providing a more robust understanding of the timing of climate shifts.The Arabian Sea and Bay of Bengal are critical regions for understanding the broader impacts of climate change due to their significant influence on the South Asian monsoon system and regional weather patterns. This study involves analyzing long-term SST datasets from these regions, applying Bayesian change point models to detect shifts in climate signals, and employing MCMC methods to estimate the posterior distributions of ToE. Our approach enables the integration of prior knowledge with observed data, allowing for a probabilistic assessment of when climate signals emerge beyond natural variability. The results of this analysis not only provide insights into the timing of significant climatic changes in the studied regions but also highlight the importance of uncertainty quantification in climate change studies.

179. Resampling-free Inference for Infinite Dimensional Parameters in Time Series via Sample Splitting and Self-Normalization

[05.M1.C30, (page 61)]

Deep GHOSHAL, *University of Illinois urbana-Champaign*
Xiaofeng SHAO, *University of Illinois Urbana-Champaign*

We study nonparametric inference for infinite-dimensional parameters in stationary nonlinear time series, focusing on testing for goodness-of-fit, the presence of a change point in the marginal distribution, and the independence of two time series. Traditional methodologies address these problems often by relying on bandwidth-dependent bootstrap methods or stringent assumptions about the data-generating process. While the former approach is computationally expensive, the latter restricts the applicability of the methods. To overcome these limitations, we propose some novel kernel-based tests that are easy to implement and incur much lower computational

costs. By embedding the data into a reproducing kernel Hilbert space, we construct test statistics using sample splitting, projection, and self-normalization techniques. We use a new conditioning technique to show that these test statistics have pivotal limiting null distributions under absolute regularity and mild moment conditions. We also examine the limiting power of our tests under local alternatives. Finally, we showcase the comparable finite sample performance of our methods relative to some existing approaches.

180. Coverage of Credible Sets for Regression under Variable Selection

[01.E1.I18, (page 9)]

Subhashis GHOSHAL, *North Carolina State University*
Samhita PAL, *North Carolina State University*

We study the asymptotic frequentist coverage of credible sets based on a novel Bayesian approach for a multiple linear regression model under variable selection. We initially ignore the issue of variable selection, which allows us to put a conjugate normal prior on the coefficient vector. The variable selection step is incorporated directly in the posterior through a sparsity-inducing map and uses the induced prior for making an inference instead of the natural conjugate posterior. The sparsity-inducing map minimizes the sum of the squared ℓ_2-distance weighted by the data matrix and a suitably scaled ℓ_1-penalty term. We obtain the limiting coverage of various credible regions and demonstrate that a modified credible interval for a component has the exact asymptotic frequentist coverage if the corresponding predictor is asymptotically uncorrelated with other predictors. Through extensive simulation, we provide a guideline for choosing the penalty parameter as a function of the credibility level appropriate for the corresponding coverage. We also show finite-sample numerical results that support the conclusions from the asymptotic theory.

181. Modern Problems Require Precise Solutions: Beating Traditional RCTs by Leveraging External Clinical Data and Adaptive Enrichment

[Student Poster Competition, (page 21)]

Souradipto GHOSH DASTIDAR, *University of Minnesota Twin Cities*
Jialing LIU, *University of Minnesota Twin Cities*
Aidan NEHER, *University of Minnesota Twin Cities*
Steffen VENTZ, *University of Minnesota Twin Cities*

Most randomized controlled trials (RCTs) are designed to estimate average treatment effects for the entire patient population, which can lead to low statistical power when treatment effects are heterogeneous. Enrichment trial designs address this by identifying patient subgroups that benefit from the therapy during the RCT and then focusing future enrollments on these subgroups. This approach limits final testing and effect estimation to the identified subgroup. Enrichment based on limited interim data is challenging, leading to increased interest in using external data from completed trials and electronic health records to enhance clinical decision-making and drug development. We propose an optimal Bayesian enrichment design that leverages external data from completed clinical trials to estimate HTEs during ongoing trials and identifies an optimal patient subgroup for enrichment. Our design uses dynamic data-borrowing based on non-local priors and a decision-theoretic framework to define optimal enrichment decisions. Extensive simulations and patient-level data analysis from a surgery study are used to show that our enrichment design is more powerful and precise in detecting subgroup-specific treatment effects than other trial designs.

182. On a Unified Class of Bivariate Distributions Characterized by a Functional Equation via a Binary Operator

[01.A1.I17, (page 9)]

Asha GOPALAKRISHNAN, *Cochin University of Science & Technology*
Durga VASUDEVAN, *Cochin University of Science & Technology*

In this talk a generalization of a class of bivariate distributions using the addition equation is discussed. We use a binary operator which is reducible and associative corresponding to a continuous and strictly increasing function $g(\cdot)$ to generalise the addition equation. Conditions for the general solution of this equation to form classes of probability distributions is studied. This class includes a number of well-studied classes of bivariate distributions. Results based on properties of these classes are presented. Finally we illustrate how these results help to construct distributions belonging to the general class of distributions proposed.

183. Optimizing Drug Labelling using Gain based Graphical MCP

[04.A1.C22, (page 53)]

Budhaditya GOSWAMI, *Sr. Manager Biostatistics, Pfizer*
Pranab GHOSH, *Director Biostatistics, Pfizer*
Margaret GAMALO, *VP Biostatistics, Pfizer*
Abhishek BHATTACHARJEE, *Sr Manager Biostatistics, Pfizer*

In 2017, the U.S. Food and Drug Administration (FDA) issued guidelines for handling multiple endpoints in registrational trials. The graphical testing procedure by Bretz (2009) was highlighted for its innovative, yet simple method of managing complex testing scenarios. However, outcomes are sensitive to initial weights, often chosen based on practical rather than statistical considerations. Thus, finding a unique graph remains crucial. We propose optimizing the graphical procedure through 'gain', which attaches predefined values to each hypothesis. This gain can reflect the additional market value from approving new treatments. By weighting the testing outcome by gain and optimizing graph parameters, we aim to maximize this gain. Since the objective function lacks a closed form, we use artificial neural networks as function approximators, with the network's input being the graph's parameters and the output being the expected gain. We then apply a custom grid search-based optimization to determine the inputs that ensure maximum gain. Proper gain specification makes this approach effective, especially in studies with numerous endpoints and high costs, where current methods often result in suboptimal graphs.

184. On the Poisson phase-type process and its application in shock models

[05.M1.C30, (page 61)]

Dheeraj GOYAL, *Indian Institute of Technology Kanpur*
Dheeraj GOYAL, *Department of Mathematics and Statistics, Indian Institute of Technology Kanpur*
Nil Kamal HAZRA, *IIT Jodhpur*
Maxim FINKELSTIEN, *University of Free State, South Africa*

Although Poisson processes are widely used in various applications for modeling of recur-rent point events, there exist obvious limitations. Several specific mixed Poisson processes that were recently introduced in the literature overcome some of these limitations. In this talk, we define a general mixed Poisson process with the phase-type (PH) distribution as the mixing one. As the PH distribution is dense in the set of lifetime distributions, the new process can be used to approximate any mixed Poisson process. We define a general mixed Poisson process with the PH distribution as the mixing one. As the PH distribution is dense in the set of lifetime distributions, the new process can be used to approximate any mixed Poisson process. We study some basic stochastic properties of the new process and its compound version, and discuss relevant applications by considering the extreme shock model and cumulative shock model.

185. Varextropy of k-Record Values and its Applications

[04.E1.C27, (page 58)]

Annie GRACE, *University of Kerala*
Manoj CHACKO, *University of Kerala*

In this paper, the varextropy measure based on k-record valuess is considered. The expressions for the varextropy measure for the nth upper and lower k-record values have been established. Additionally, we establish the varextropy measure of residual and past lifetimes based on k-record values. We also consider the problem of estimation of the varextropy measure for a two-parameter Weibull distribution based on the upper k-record values. Maximum likelihood estimation and Bayes estimation for varextropy measure have been considered based on upper k-record values. Bayes estimators are obtained using Markov Chain Monte Carlo (MCMC) method. A simulation study is performed to determine the performance of the estimators developed in this paper. Inferential procedures developed in this paper have also been illustrated using real data.

186. Modeling periodic and nearly periodic signals

[02.M1.I25, (page 16)]

Rhythm GROVER, *Indian Institute of Technology Guwahati*
Debasis KUNDU, *IIT Kanpur*

Many natural and artificial phenomena exhibit some kind of periodic behavior. Knowing the period of a phenomenon facilitates a better understanding and also aids in making predictions. Often, however, the period is unknown and varies with time, and it must therefore be estimated from the available data.

In this talk, we will discuss some of the fundamental stochastic models that are used to understand periodic and nearly periodic data. We will mainly discuss a new model called a chirp-like model. Further, we will discuss some classical estimation methods based on probability theory and their advantages and limitations.

187. Bayesian approaches for Modeling Brain Network Dynamics

[02.M2.I30, (page 19)]

Sharmistha GUHA, *Texas A&M University*

TBA

188. Nonparametric Bayes Differential Analysis of Multigroup DNA Methylation Data

[04.M1.I63, (page 44)]

Subharup GUHA, *University of Florida*
Chiyu GU, *Bayer Crop Science*
Veerabhadran BALADANDAYUTHAPANI, *University of Michigan*
Subharup GUHA, *University of Florida*

DNA methylation datasets in cancer studies are comprised of measurements on a large number of genomic locations called cytosine-phosphate-guanine (CpG) sites with complex correlation structures. A fundamental goal of these studies is the development of statistical techniques that can identify disease genomic signatures across multiple patient groups defined by different experimental or biological conditions. We propose BayesDiff, a nonparametric Bayesian approach for differential analysis relying on a novel class of first order mixture models called the Sticky Pitman-Yor process or two-restaurant two-cuisine franchise (2R2CF). The BayesDiff methodology flexibly utilizes information from all CpG sites or biomarker probes, adaptively accommodates any serial dependence due to the widely varying inter-probe distances, and makes posterior inferences about the differential genomic signature of patient groups. Using simulation studies, we demonstrate the effectiveness of the BayesDiff procedure relative to existing statistical techniques for differential DNA methylation. The methodology is applied to analyze a gastrointestinal (GI) cancer dataset exhibiting serial correlation and complex interaction patterns. The results support and complement known aspects of DNA methylation and gene association in upper GI cancers.

189. TBA

[02.M2.I30, (page 19)]

Rajarshi GUHANIYOGI, *Texas A&M University*

TBA

190. Totally Concave Regression

[05.M1.I88, (page 59)]

Aditya GUNTUBOYINA, *University of California, Berkeley*
Dohyeong KI, *University of California Berkeley*

Shape constraints offer a compelling advantage in nonparametric regression by enabling the estimation of regression functions under realistic assumptions, devoid of tuning parameters and stringent parametric constraints. However, most existing shape-constrained nonparametric regression methods, with the exception of additive models, impose too few restrictions on the regression functions. This often leads to suboptimal performance, such as overfitting, in multivariate contexts due to the curse of dimensionality. On the other hand, additive shape-constrained models are too restrictive because they fail to capture interactions among the covariates.

In this talk, I will introduce a novel approach for multivariate nonparametric regression based on shape constraints which allows interactions without suffering from the usual curse of dimensionality. Our approach is based on the notion of total convexity originally due to T. Popoviciu and more recently described in a book by Gal. I discuss the characterization and computation of the least squares estimator over the class of totally convex functions, and derive rates of convergence under standard assumptions. The rate of convergence depends on the underlying number of covariates only logarithmically and the estimator, therefore, is guaranteed to avoid the usual curse of dimensionality to some extent. Total convexity can be justified for many real-world regression examples and I will validate the efficacy of our approach through empirical studies on various real-world datasets. Connections between our approach and nonparametric regression techniques based on mixed partial derivative constraints will also be discussed. This presentation is based on joint work with Dohyeong Ki (UC Berkeley).

191. Construction of Cyclic Minimal Balanced and Cyclic Minimal Strongly Balanced Crossover Designs

[Student Poster Competition, (page 23)]

Ashish GUPTA, *ICAR-IASRI*
Baidya Nath MANDAL, *ICAR-IARI, Jharkhand*
Rajender PARSAD, *ICAR-IASRI*
Cini VARGHESE, *ICAR-IASRI*

Crossover designs form an important class of designs that has been advantageously used in dairy feeding trials, long-term experiments, clinical trials, etc. In these designs, each experimental unit receives a sequence of treatments over a number of periods with observations recorded in each period. In this study, two new methods of construction for cyclic minimal balanced crossover designs (Cyclic MBCOD) and cyclic minimal strongly balanced crossover designs (Cyclic MSBCOD) have been proposed. The necessary condition for parameters (v,p,n) for existence of such designs has also been derived. These designs are catalogued within a restricted parameter range (n 30,p 7,v 20). The Efficiency Cyclic MBCOD and cyclic MSBCOD designs is assessed using relative efficiency of direct effect (RE_direct) or residual effect (RE_res) with respective to their Row-column design. This study provides the future potential of crossover designs, including the development of non-cyclical designs and designs with unequal period sizes. In a rapidly evolving landscape of experimental research, the findings of this study provide a valuable reference for making innovative experimental design choices.

192 . Spatial Modeling of Residential Property Launch Prices in Bengaluru, India

[03.E1.C17, (page 42)]

Kapil GUPTA, *IIM Bangalore*
Soudeep DEB,

Understanding the dynamic patterns of property prices across different locations has become a focal point in real estate markets. In this study, we introduce an innovative statistical model designed to capture complex spatial dependencies in real estate launch prices effectively. Our model employs a Gaussian spatial process, incorporating an additive mean structure alongside a random error component that follows a t-distribution rather than the conventional normal distribution. Implemented within a Bayesian framework, this approach offers both flexibility and computational efficiency over traditional methods. As a practical application, we apply this model to granular-level transaction data from Bengaluru, encompassing 1127 society projects and 1255 launches from January 2017 to June 2023. Our model demonstrates superior efficiency in predicting launch prices, highlighting its potential utility in real estate market analysis.

193. Time changed Counting processes

[02.E1.C6, (page 27)]

Neha GUPTA, *IIT Kanpur*
Neha GUPTA, *IIT Kanpur*

The Poisson process is an important Lévy process commonly used to model the count data. Recently, the subordinated counting processes, namely, the tempered space-fractional Poisson process, generalized space fractional Skellam process, etc, have drawn the significant attention of several researchers. The talk will present a few time-changed variant of the counting process and their applications, Two time-changed versions of the space-time tempered fractional Poisson process, fractional Skellam process of order k will be discussed where an independent Lévy subordinator and its inverse does time-changes. Later, a generalized version of the fractional counting process and its variants that perform jump of sizes 1; 2; 3; : : : ; k will be presented.

194. Enhanced Breast Cancer Diagnosis Using a Novel Hybrid Deep Learning Approach

[01.E1.C5, (page 13)]

Balakrishna GURMITKAL, *Yenepoya (Deemed to be University)*
Ismail B, *Department of Statistics, Yenepoya (Deemed to be University)*
N/A N/A, *N/A*
N/A N/A, *N/A*

Our research introduces a novel approach Swarm Optimized Convo-Neuro recurrence net [SO-CNRN], based on deep learning principles, to predict breast cancer. We collected a dataset from online source. Then, implement a step preprocessing methodology, including Z-score normalization to extract the mean and standard values in data. To extract meaningful features, we employ Principal Component Analysis [PCA] which aid in reducing dimensionality while retaining essential features. Next, the features are selected by using a Lasso Regression which selects the particular features. Finally, the classification task is accomplished using the proposed model, inherently considering temporal dependencies in the data. This is particularly crucial to detect Breast Cancer,

where historical patterns significantly influence outcomes. The proposal research is implemented is assessed through accuracy, precision, recall, F1 score, and AUC. In comparison to existing methods, our comprehensive approach aims to enhance breast cancer detection with accuracy 97.03%, precision with 97%, recall value 97%, F1-score with 97% and AUC with 0.99.

195. On Solving Uncertain Transportation Problem

[01.E1.C5, (page 13)]

Ummey HABIBA, *Indian Institute of Information Technology Guwahati*
Masihuddin MASIHUDDIN, *Indian Institute of Information Technology Guwahati*

In this talk, I will address a novel and uncertain extension of the classical transportation problem, termed the 'Neutrosophic Pentagonal Solid Interval Transportation Problem'(NPSITP). In this variant, the uncertainties associated with source, destination and conveyance parameters are characterized using pentagonal neutrosophic numbers. Additionally, the uncertain costs of transportation are represented by estimated ranges, or interval numbers. The primary goal, centered around the interval cost objective, has been subdivided into two comparable crisp objectives. This subdivision is achieved by incorporating the principles of both expected value and the quantification of uncertainty associated with interval numbers. The constraints involving pentagonal neutrosophic quantities have been transformed into crisp constraints through the application of a score function. This conversion process enhances the clarity and manageability of the constraints, providing a more interpretable framework for analysis. A fuzzy programming approach has been employed to drive the Pareto optimal solution for the modified bi-objective crisp transportation problem. To exemplify the computational steps involved in the proposed variant, a real life problem has been presented. The resolution of the problem has been facilitated through the utilization of LINGO software.

196. Asymptotic Tests for Tree Ordered Effects in One-Way ANOVA

[01.E1.C1, (page 11)]

Subha HALDER, *IIT Kharagpur*
Somesh KUMAR, *IIT Kharagpur*

Due to underlying conditions it is observed in many experiments that treatment effects in One-Way ANOVA follow some order restrictions. These restrictions may be of different nature such as monotone, tree, umbrella, nested etc. There are some studies where test procedures have been proposed and developed for monotone ordered effects. Here, we investigate the problem when the alternative hypothesis contains tree ordered effects. We develop asymptotic tests based on likelihood ratio and simultaneous comparisons. For implementation of the tests procedures, parametric bootstrap methods are used. Robustness of tests is also investigated under departure from normality. The procedures are illustrated with real data sets.

197. A Bayesian Joint Modelling of Current Status and Current Count Data

[04.E1.C25, (page 56)]

Pavithra HARIHARAN, *Department of Statistics, Cochin University of Science and Technology, Cochin*
P. G. SANKARAN, *Senior Professor and Vice-Chancellor, Department of Statistics, Cochin University of Science and Technology, Cochin*

Current status censoring or case I interval censoring takes place when subjects in a study are observed just once to check if a particular event has occurred. If the event is recurring, the data are classified as current count data; if non-recurring, they are classified as current status data. Several instances of dependence of these recurring and non-recurring events are observable in epidemiology and pathology. Estimation of the degree of this dependence and identification of major risk factors for the events are the major objectives of such studies. The current study proposes a Bayesian method for the joint modelling of such related events, employing a shared frailty-based semiparametric regression model. Computational implementation makes use of an adaptive Metropolis-Hastings algorithm. Simulation studies are put into use to show the effectiveness of the method proposed and fracture-osteoporosis data are worked through to highlight its application.

198. Product Optimization through unified DoE and ML techniques

[04.E1.I85, (page 54)]

Mohd HARUN, *ICAR-IASRI*

Designing of experiments (DoE) is being used as one of the classical statistical techniques for many decades to study and quantify the relationship among

factors influencing output quantity and quality based on underlying models. However, with a bloom in industrial sector leading to an increase in available data, machine learning (ML) techniques are being used significantly for handling and analyzing data. Both the techniques, DoE and ML, being irreplaceable for their own effectiveness raises the attention of researchers to investigate the joint application of both technique in screening factors, optimization, and prediction. The main idea behind this study is to automate the human-based parts of DoE using ML, optimize the training process of ML algorithms using DoE, an analysis of DoE data through application of ML. In this context some newly constructed and already well established statistical designs are evaluated and compared.

199. Flexible Modeling of Nonstationary Extremal Dependence Using Spatially-Fused LASSO and Ridge Penalties

[03.M2.I48, (page 34)]

Arnab HAZRA, *Indian Institute of Technology Kanpur*
Xuanjie SHAO, *King Abdullah University of Science and Technology*
Jordan RICHARDS, *University of Edinburgh*
Raphael HUSER, *King Abdullah University of Science and Technology*

Statistical modeling of a nonstationary spatial extremal dependence structure is challenging. Parametric max-stable processes(MSPs) are common choices for modeling spatially-indexed block maxima, where an assumption of stationarity is usual to make inference feasible. However, this assumption is unrealistic for data observed over a large or complex domain. We develop a computationally efficient method for estimating extremal dependence using a globally stationary but locally stationary MSP construction, with the spatial domain divided into a fine grid of subregions, each with its own dependence parameters. We use LASSO or ridge penalties to obtain spatially-smooth parameter estimates. We then develop a novel data-driven algorithm to merge homogeneous neighboring subregions. The algorithm facilitates model parsimony and interpretability. To make our model suitable for high-dimensional data, we exploit a pairwise likelihood to perform inference and discuss its computational and statistical efficiency. We apply our proposed method to model monthly maximum temperature data at over 1400 sites in Nepal and the surrounding Himalayan and sub-Himalayan regions; we show significant improvements in model fit compared to a stationary model. Furthermore, we demonstrate that the estimated merged partition is interpretable from a geographic perspective and leads to better model diagnostics by adequately reducing the number of parameters.

200. Statistical Analysis on In-Context Learning

[02.M1.I23, (page 15)]

Masaaki IMAIZUMI, *Dept. of Statistics, University of Tokyo*

Deep learning and artificial intelligence technologies have made great progress, and the usage of foundation models has attracted strong attension by its general ability. Motivated by this fact, mathematical understanding is required to efficiently control and develop these technologies. In this talk, I will present a statistics-based analysis of a scheme called in-context learning, which is an useful framework of meta-learning to describe foundation models. I argue that in-context learning can efficiently learn the latent structure of the data, using the property of transformers used in the learning scheme can efficiently handle the distribution of observations.

201. A New Generalization of Thinning-Based Integer-Valued Autoregressive Models for Count Data

[05.M1.I87, (page 59)]

M.R. IRSHAD, *Cochin University of Science and Technology*

The analysis of count-based time series is an emerging field in scientific research. Thinning operators are crucial in the analysis of integer-valued time series of counts, and among them, the binomial thinning operator is the most frequently employed. In this study, we introduce a reparametrized version of the generalized binomial thinning operator and develop a new INAR(1) model using the Poisson-Lindley distribution as the innovation distribution. The statistical properties of the resulting INAR(1) process, as well as the estimation methods, are examined. Monte Carlo simulation experiments are performed to assess the consistency of the estimates for the new INAR(1) model. Finally, the INAR(1) model is applied to a dataset, and a comparative study is conducted with other existing INAR(1) models.

202. Stochastic comparison of extreme order statistics in Archimax copula

[04.E1.C28, (page 58)]

SARIKUL ISLAM, *Indian Institute of Technology Kharagpur*
Nitin GUPTA, *Indian Institute of Technology Kharagpur*

Generalization and construction of Archimax copulas for higher dimensions random variables were introduced in the literature for describing dependency structures among more than two random variables and to get rid of the setbacks of Archimedean copulas. Archimax copula is the most generalized model for describing multivariate dependency structure among random variables. In this article, we take Archimax copula as a standard model and present some important findings concerning the stochastic comparison of extreme order statistics in Archimax copula. The stochastic comparison of sample extremes for the proportional hazard model is generalized and proved in the Archimax copula. Examples are presented for illustration purposes. Our theoretical findings will advance the knowledge and existing theory in the fields of system reliability and the lifetime of coherent systems.

203. TBA

[04.A1.I77, (page 50)]

Mahesh IYER, *BMS*

TBA

204 . Climatologist's perspective of emergence of climate change signals

[02.M2.I29, (page 18)]

Suresh IYYAPPAN, *Digital University Kerala*

The rate of warming of Earth's temperature has exhibited significant variability over time, with a notable increase in recent decades. Observations suggest a long-term warming trend of approximately 0.18°C per decade since 1973, with a marked acceleration in warming rates post-1990. Climate projections indicate that this warming will continue to intensify throughout the 21st century, driven by the unabated emissions of greenhouse gases. Furthermore, observations suggest that certain regions are experiencing significantly higher rates of warming, indicating spatial heterogeneity in the global warming pattern. In addition to rising temperatures, anthropogenic climate change has a wide array of effects, including shifts in regional climate and weather patterns, an increase in the frequency and severity of extreme events, and profound impacts on ecosystems and public health. To inform adaptation strategies and facilitate attribution studies, it is crucial to accurately extract and identify climate change signals and their emergence time. However, recent research indicates that observational uncertainties and natural climate variability can obscure climate change signals, with internal climate variability acting as noise that complicates the detection of these signals and their time of emergence. A variety of methods, each with varying levels of complexity, have been developed to estimate the time of emergence of climate change signals from the background noise. However, there is currently a lack of consistency among these methodologies. This talk will review the existing methods and their shortcomings, emphasizing the need to develop robust statistical techniques to reliably estimate climate change signals. Such advancements are essential to evaluate the realism of current climate models and to improve predictive capabilities regarding future climate scenarios.

205. Two-dimensional stochastic volatility models for image processing

[02.E1.C9, (page 29)]

FATHIMA JAFNA, *DEPARTMENT OF STATISTICS, UNIVERSITY OF CALICUT*
KRISHNARANI S. D., *DEPARTMENT OF STATISTICS, UNIVERSITY OF CALICUT*

The paper aims to explore the application of stochastic volatility models in the context of image processing, specifically for noise reduction in images using wavelet transforms. We develop a new two-dimensional stochastic volatility model and study its application in image denoising. The one-dimensional stochastic volatility models have wider applicability and are commonly used for modeling financial time series. Extending this one-dimensional stochastic volatility model into two dimensions can be used to statistically model the wavelet coefficients. This model can account for significant features of wavelet coefficients, including their non-stationarity, heavy-tailed marginal distribution, and the dependencies between the coefficients, and it provides a plausible model to estimate their variances. The volatilities are then estimated using the Kalman filtering technique and we employ the minimum mean square error estimator to estimate the coefficients of the clean wavelet image coefficients. Results are compared with other image-denoising techniques to demonstrate the efficacy of the proposed method.

206. Efficient and Elastic LLMs

[04.M2.I76, (page 48)]

Prateek JAIN, *Google Research, India*

TBA

207. A tensor-based algorithm for partitioning asymmetric measure of predictability: Delta index

[01.E1.C1, (page 11)]

Riya JAIN, *Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University Jalgaon, Maharashtra, India*
Kirtee KAMALJA, *Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India*

Many techniques exist in literature for studying the association among categorical variables. The Goodman-Kruskal -index is often a more appropriate measure than the classical symmetric measure of association, Pearson chi-squared statistic for analyzing the asymmetric (predictor-response type) association between two categorical variables. This index measures the predictability of one categorical variable given the other. Extensions of the -index for three variables such as Gray-Williams index, Marcotorchino index, and Delta index have been proposed depending on the number of response variables in the contingency table. Specifically, for dealing with two responses and one predictor, the Delta index is used to study their predictability. We generalize the Delta index for four-way data with three responses and one predictor variable. Further, we partition this index to study the influence of a predictor variable on each of the response variable individually. We propose an efficient algorithm based on tensor operations to partition this index. A software code is developed for obtaining the proposed partition and is demonstrated with an example to confirm the results. Keywords: Symmetric and Asymmetric measures, Delta index, Tensor, CATANOVA.

208. Spectrum of random Centrosymmetric matrices

[01.A1.I15, (page 9)]

Indrajit JANA, *IIT Bhubaneswar*
Sunita RANI, *IIT Bhubaneswar*

We analyze the asymptotic fluctuations of linear eigenvalue statistics of random centrosymmetric matrices with i.i.d. entries. We prove that for a complex analytic test function, the centered and normalized linear eigenvalue statistics of random centrosymmetric matrices converge to a normal distribution. We find the exact expression of the variance of the limiting normal distribution via combinatorial arguments. Moreover, we also argue that the limiting spectral distribution of properly scaled centrosymmetric matrices follows the circular law.

209. Nonparametric quantile regression for time series with replicated observations and its application to climate data

[04.M2.I75, (page 48)]

Kaushik JANA, *Ahmedabad University*
Soudeep DEB, *Indian Institute of Management, Bangalore*

We proposes a model-free nonparametric estimator of conditional quantile of a time series regression model where the covariate vector is repeated many times for different values of the response. This type of data abounds in climate studies. Although the use of quantile regression is standard in such studies, the opportunity to improve the results using the replicated nature of data is increasingly realized. The proposed method exploits this feature of the data and improves on the restrictive linear model structure of conventional quantile regression. Relevant asymptotic theories for the nonparametric estimators of the mean and variance function of the model are derived under a very general framework. We conduct a detailed simulation study that demonstrates the gain in efficiency of the proposed method over other benchmark models, especially when the actual data-generating process entails a nonlinear mean function and heteroskedastic pattern with time-dependent covariates. The predictive accuracy of the nonparametric method is remarkably high compared to other approaches when attention is on the higher quantiles of the variable of interest. The usefulness of the proposed method is then illustrated with two climatological applications, one with a well-known tropical cyclone wind-speed data and the other with an air pollution data.

210. Central Limit Theorem for Exponential GREM

[02.M2.I31, (page 19)]

Nabin Kumar JANA, *National Institute of Science Education and Research, Bhubaneswar*

Generalised random energy model was introduced

to study hierarchical structure of energy landscape shown by mean-field models in statistical physics. In this presentation, we want to show what we should expect when we study CLT for this model. The result predicted is reminiscent to that of Landau theory or renormalisation group flow in physics.

211. Generalized Multivariate Analysis of Variance (GMANOVA) models for volatile data

[02.M1.I21, (page 14)]

Sayantee JANA, *IIT Hyderabad*

TBA

212. A general theory for robust clustering: from initialization to mislabeling minimization

[03.M2.I50, (page 34)]

Soham JANA, *University of Notre Dame*
Jianqing FAN, *Princeton University*
Sanjeev KULKARNI, *Princeton University*

Clustering is a fundamental tool in statistical machine learning in the presence of heterogeneous data. Many recent results, such as the performances of the Lloyd algorithm and spectral clustering techniques, focus primarily on optimal mislabeling guarantees when data are distributed around centroids with sub-Gaussian errors. Yet, the restrictive sub-Gaussian model is often invalid in practice since various real-world applications exhibit heavy tail distributions around the centroids or suffer from possible adversarial attacks that call for robust clustering with a robust data-driven initialization. In this work, we introduce a novel hybrid clustering technique to produce optimal mislabeling guarantees under a weak initialization condition for general error distributions around the centroids and in the presence of adversarial outliers. To solve the problem thoroughly, we also present a novel data-driven robust initialization technique and show that, with probabilities approaching one, these initial centroid estimates are sufficiently good for the subsequent clustering algorithms to achieve optimal guarantees. Both simulated and real data examples support our robust algorithms.

213 . A Multiple Testing Approach for Estimating the Number of Blocks in Degree-Corrected Stochastic Block Models

[01.M2.I5, (page 5)]

Chetkar JHA, *Ahmedabad University*
Ian BARNETT, *University of Pennsylvania*
Debashis MONDAL, *Washington University in St Louis*

Network-based community detection approaches tend to be more robust compared to the model based clustering approaches making them more popular. However, community detection approaches invariably require estimating the number of communities. Moreover, as Jha et al. (2024) pointed out, the standard toolboxes for estimating the number of communities tend to be sensitive to the out-in ratio for stochastic block models. With the growing adoption of network based models in biomedical applications, it is important that methods for estimating the number of blocks/communities in flexible block models (such as degree-corrected stochastic block models) are suitable for more sparse and noisy graphs. Unfortunately, Jha et al. (2024)'s work is not applicable for degree-corrected stochastic block models. To this end, we propose a new approach for estimating the number of blocks/communities in the degree-corrected stochastic block models including stochastic block models. We validate our approach through numerical simulations and benchmark data analysis. Lastly, we apply our approach for the analysis of real single-cell data analysis.

214. Two-stage Circular-circular Regression with Zero-inflation: Application to Medical Sciences

[03.M1.I45, (page 32)]

Jayant JHA, *Indian Statistical Institute*
Prajamitra BHUYAN , *IIM Calcutta*

This work considers the modeling of zero-inflated circular measurements concerning real case studies from medical sciences. Circular-circular regression models have been discussed in the statistical literature and illustrated with various real-life applications. However, there are no models to deal with zero-inflated response as well as a covariate simultaneously. The Möbius transformation based two-stage circular-circular regression model is proposed, and the Bayesian estimation of the model parameters is suggested using the MCMC algorithm. Simulation results show the superiority of the performance of the proposed method over the existing competitors. The method is applied to analyse real datasets on astigmatism due to cataract surgery and abnormal gait

related to orthopaedic impairment. The methodology proposed can assist in efficient decision making during treatment or postoperative care.

215 . Estimation Based on Extended Sequential Order Statistics for Lindely Distribution

[01.E1.C1, (page 10)]

Jaaziah P JOHN, *University of Kerala*
Manoj CHACKO, *University of Kerala*
Jaaziah P JOHN, *University of Kerala*

In this paper, we considered the problem of estimation of parameters of Lindely distribution based on Extended Sequential Order Statistics. Under Conditional Proportional Hazard Rate model, a proportional link hazard rate function is used. Maximum Likelihood and Bayesian estimation techniques are applied for estimating the unknown parameters. Markov Chain Monte Carlo method using Metropolis-Hastings algorithm is implemented for obtaining Bayes estimators. An intensive simulation study is carried out to assess the performance of different estimators discussed in this paper.

216. Estimation of reliability in a multicomponent stress-strength model based on power transformed Perks distribution

[04.L1.C19, (page 49)]

SUDHEEP JOSE, *MADRAS CHRISTIAN COLLEGE*
Venugopal HARIDOSS, *Madras Christian College*
Thomas XAVIER, *Novartis*

The article explores a new distribution and examines its relationship with the Weibull and Perks distributions. It also analyzes the reliability of a multicomponent stress-strength model, assuming that the strength components are independently and identically distributed according to a power-transformed Perks distribution. The strength components which are subject to a common stress is assumed to be independent with power transformed Perks distribution. The maximum likelihood estimates of the multicomponent stress-strength reliability and its asymptotic confidence interval are obtained. To evaluate the performance of the procedure simulation study is considered. For illustration purpose of the proposed model, one real life example is given..

Key words: Power transformed Perks distribution; multicomponent stress-strength reliability; Monte-Carlo approximation; maximum likelihood estimation.

217. Nonparametric estimation of Matusita's measure for residual and past lifetimes

[01.E1.C2, (page 11)]

Chinu JOSEPH, *Mahatma Gandhi University,Kottayam,Kerala,Imdia*
Chinu JOSEPH, *Mahatma Gandhi University, Kottayam,Kerala, India*
Angel MATHEW, *Mahatma Gandhi University, Kottayam,Kerala,India*

In reliability and survival studies, assessing the similarity between probability distributions of the residual lifetimes or the past lifetimes of two populations is of great importance. Matusita's overlapping measure is a widely used similarity measure based on the concept of distance between two probability distributions. In this paper, we consider the Matusita's measure for residual and past lifetimes and study some of its properties. We propose nonparametric estimators for the residual and past Matusita's measures using the kernel method. The asymptotic properties of the proposed estimators are also studied. Finally, a simulation study is conducted to assess the performance of the proposed estimators.

218 . Adaptive sampling strategies for estimating the parameter of an Inverse Maxwell Distribution

[04.M2.I73, (page 47)]

Neeraj JOSHI, *Indian Institute of Technology Delhi*

This paper addresses the optimal parameter estimation problems related to the distributions that exhibit upside-down bathtub-shaped hazard rates. We consider an inverse Maxwell distribution and aim at estimating its scale parameter under the minimum risk point estimation constraint. Consideration is given to the squared-error loss and linear cost functions. As no fixed-sample procedure can solve this estimation problem, we propose optimal sequential sampling strategies to address the issue. The proposed sequential procedures are seen to be cost and time-efficient and satisfies several interesting first and second-order asymptotic properties. The extensive simulation analyses well support the theoretical results and two real data sets based on tax revenues and lifetimes of car brakes are analyzed to showcase

the practical importance of the suggested techniques.

219. Hazard Regression Change Point Model with Applications to Bone marrow transplant data

[02.A1.I38, (page 25)]

Savitri JOSHI, *Indian Institute of Information Technology Allahabad*

In this talk, the various hazard change point models namely Exponential, Lindley, Exponential Lindley, Weibull will be discussed. The applications of these models to real life data sets such as bone marrow transplant, heart transplant and others will be elaborated. The future research problems and scope in the field of hazard change point models will also be highlighted.

220. Comparison of Bayesian and Frequentist Logistic Regression: Predictors of Smoking Among Elderly in India

[04.E1.C25, (page 56)]

Kripa JOSTEN, *Manipal College of Health Professions, Manipal Academy of Higher Education, Manipal*

Background: Statistics is dominated philosophically and practically by frequentist and Bayesian ideologies. Notably, statistics was (definitionally) founded as a Bayesian endeavour and stayed that way for around 150 years. The discipline once was referred to as "the theory of inverse probability," not as Bayesian analysis .

A broad range of statistical studies are presented as model selection or parameter estimation issues. While Bayesians have techniques for computing the posterior probability of models and parameters, frequentists have procedures for selecting models and estimating parameters. When there exist real models and parameters, any kind of process could make sense. However, this is not always the case, and to suggest otherwise is to distract from the frequently important work of identifying many high-quality models, which can result in inferential mistakes .

Smoking is a major risk factor leading to morbidity and mortality in India which accounts for nearly 1.35 million deaths every year . The secondary data used for the study, records the self-reported users of tobacco users in India. The study aims to find various predictors involved in smoking among the elderly in India. By leveraging both Bayesian and Frequentist approaches, the study aims to provide comprehensive insights into the methodological strengths and limitations of each approach and the predictors of smoking.

Materials & Methodology: This study provides a comparative analysis of Bayesian and frequentist logistic regression methods, focusing on their application using WHO-Study on Global Ageing and Adult Health (SAGE) India dataset wave 2 conducted in 2015 for the outcome variable smoking status for elderly . The predicators included in the study are age, gender, Education status, Alcohol Consumption, and Chronic health Conditions such as Self-reported Angina, Diabetes Mellitus and Hypertension.

4180 samples were taken from the total data set using the inclusion and exclusion criteria. The outcome variable is binary indicating whether the participants are smokers or non-smokers. Using the various predictor variables such as gender, specific behaviors, and relationship status, regression models were constructed under both paradigms. The Bayesian logistic regression incorporated prior distributions and yielded posterior summaries, offering a probabilistic interpretation of the model coefficients and their uncertainties. The posterior summaries were obtained using Markov Chain Monte Carlo (MCMC) methods and model comparison was done using Bayes Factor (BF10). While for frequentist method the model fit was done using maximum likelihood estimation and model performance was evaluated using Area Under the Curve (AUC), precision, recall, and F1 score and odds ratio was calculated.

Results: The total smokers in the elderly (60 & above) in India as reported by WHO-SAGE wave – 2 were 35.89%. Majority of the smokers were Males (1095 out of 2152). Self-reported prevalence among the smokers for patients with Angina, Diabetes Mellitus and hypertension were 4.67 %, 9. 47 % and 22.33% respectively. With a posterior probability of 0.327 and a Bayes Factor (BF10) of 1.000, the model with the greatest posterior probability among the evaluated models is Gender, Alcohol Consumption, and Diabetes Mellitus. Conversely, the frequentist logistic regression suggests females have lower odds of smoking to the males ($p<0.001$). Alcoholics also have 5.58 times higher odds ($p<0.001$). The best frequentist model, evaluated using AIC (4765.299) and BIC (4835.017), identified Gender, Alcohol consumption and depression as significant predictors, highlighting their impact on the response variable.

Conclusion: The two approaches find comparable significant predictors (e.g., gender, alcohol con-

sumption etc..), but the entire posterior distribution for each parameter and the incorporation of prior information in Bayesian logistic regression allow for a more comprehensive interpretation. When compared to conventional Frequentist techniques, this makes it a very effective tool for decision-making in complex data environments. The comparative study shows that by using Bayes factors, Bayesian logistic regression provides a deeper, more intuitive knowledge of parameter uncertainty and model comparisons. On the other hand, the frequentist method requires less computing effort and is quicker, making it appropriate for applications involving huge datasets and easy significance testing. In contrast to the Bayesian approach, the frequentist method relies solely on the observed data and provides confidence intervals and p-values for hypothesis testing.

221. Sea Root Clustering / RouteFinder : A Novel Approach to Estimate Vessels Route by Trajectory Clustering based on AIS Data

[03.E1.C13, (page 40)]

Patra JYOTIRMOY, *S&P Global*
Shubham S, *S&P Global*
Sandip PAL,

In recent years, there has been a rise in maritime traffic, leading to ships deviating either completely or partially from their usual courses. Understanding the ships' anomalous behaviour during the journey is crucial for ensuring their safety, security, and accurate estimates of arrival time. In this paper, we propose a novel approach termed "Sea Root Clustering" utilizing graph partitioning with the Girvan-Newman method to analyse Automatic Identification System (AIS) data. AIS data provides information about vessel movements, and by converting this data into polygons, we obtain a spatial representation of maritime traffic. We calculate grid-wise distances between these polygons to construct a similarity matrix, capturing the spatial relationships between different areas of maritime activity. The Girvan-Newman approach is subsequently utilized to split the graph obtained from this similarity matrix, effectively identifying clusters of marine transportation routes. This approach enables the identification of significant maritime traffic routes and their clustering patterns, providing valuable insights for maritime management, route optimization, and risk assessment. We demonstrate the efficacy of our method through experiments on real-world AIS data, showcasing its potential for enhancing maritime domain awareness and decision-making processes.

222 . Estimation of weighted extropy with a focus on its use in reliability modeling

[04.E1.C27, (page 57)]

Archana K, *Cochin University of Science and Technology, Kochi, Kerala*
Irshad M. R., *Cochin University of Science and Technology*

In the literature, the estimation of weighted extropy has received minimal attention. In this paper, we present several non-parametric estimators of weighted extropy. These estimators are validated and compared through simulation studies, and their effectiveness is demonstrated using real data sets. We also demonstrated the application of proposed estimators in reliability modelling.

223 . A Study on some new Reliability Measures in the Univariate Case

[03.E1.C18, (page 43)]

FATHIMA K, *University of Kerala*
E I Abdul SATHAR, *University of Kerala*

We introduce a new measure to describe the failure pattern of components/devices, studied its properties including uniqueness and applications in empirical and non-parametric kernel estimation, both in the case of complete and censored samples, and carried out simulation and a real data analysis to study the performance of the estimators.

224 . A BIVARIATE INTEGER-VALUED BILINEAR AUTOREGRESSIVE MODEL WITH RANDOM COEFFICIENTS

[04.M1.I68, (page 46)]

Jayakumar K, *Senior Professor, Department of Statistics, University of Calicut*

Modelling time series of counts has been an interesting topic for many researchers since this type of time series can be found in various scientific fields. Many models of integer-valued time series rely on the thinning operator. Time series can be represented by linear or non-linear models. Bilinear time series models are the simple type of non-linear models and are useful for the series that display shocks such as earthquakes and disease outbreaks. Also, there are some events which assume non-linearity, whose oc-

currences are characterized by human factors, such as criminal records. Bilinear time series have some similarities with linear autoregressive moving average models which can help in getting many properties as in autoregressive models. A popular class of models that we focus on are autoregressive models, where the auto regression is achieved through the thinning operator. Note that, the bivariate non-negative bilinear time series models do not represent non-Gaussian bivariate processes because the normality assumption of the innovation process of the model is unavailable in non-negative models. Therefore, such models are interesting in their own right and of major importance in modeling paired correlated count data. Motivated by this, in this paper we develop and study a bivariate integer valued bilinear autoregressive model with random coefficients.

225. Autoregressive Time Series Models with Generalized Laplacian Bilateral Gamma marginal distribution and Their Applications

[05.M1.I87, (page 59)]

Jose K.K., *Schoolof Mathematics and Statistics, Mahatma Gandhi University, Kottayam, India*

In this talk we will discuss the emerging developments in non-Gaussian autoregressive time series modeling. The importance of self-decomposability, geometric infinite divisibility and random self-decomposability in autoregressive time series moeling is established. As an illustration we consider generalized Laplacian Bilateral Gamma distribution and develop an autoregressive time series model with this stationary marginall distribution.The geometric generalized Laplacian distribution is also introduced and studied. The problem of parameter estimation and related issues are also addressed.The model is applied to real data sets from financial contexts.

226 . DEFECTIVE REGRESSION MODELS FOR CURE RATE MODELLING WITH INTERAL CENSORED COMPETING RISKS DATA

[04.E1.C24, (page 56)]

Silpa K, *Cochin University of Science and Technology*
Sreedevi E. P., *Cochin University of Science and Technology*
P. G. SANKARAN, *Cochin University of Science and Technology*

In this work, we present two defective regression models for the analysis of interval-censored competing risk data in the presence of cured individuals, viz. defective Gompertz distribution and defective inverse Gaussian distribution. The proposed models enable us to estimate the cure fraction directly from the model. Simultaneously, we also estimate the regression parameters corresponding to each cause of failure using the method of maximum likelihood. The finite sample behaviour of the proposed models is evaluated through Monte Carlo simulation studies. We illustrate the practical applicability of the models using a real-life data set (HIV data).

227 . Trade Durations and Liquidity Analysis of Government Bonds: An Econometric Study

[01.M2.I6, (page 5)]

IRFANA K A, *federal bank*

This paper investigates the relationship between trade durations and liquidity of Government bonds in India using Autoregressive Conditional Duration (ACD) models. We employ the ACD models to decompose trade durations into two components: the expected and the unexpected durations. Then we analyze whether trade durations affect liquidity with regressions. We find that there exists a strong dependence between consecutive durations especially for liquid bonds. The findings indicate that the ACD models are alternative to traditional liquidity measures employed in Government bond markets.

228. Enhancing Pediatric Drug Development through Bayesian Dynamic Borrowing: A Case Study in Rare and Serious Diseases

[01.A1.I11, (page 7)]

Lijina KAIPRATH, *GSK, India*

Pediatric drug development poses unique challenges, particularly in rare and serious diseases where data availability is limited. In this project, we address this challenge by employing Bayesian Dynamic Borrowing, a statistical methodology that leverages adult and adolescent data to inform pediatric extrapolation. Focusing on a rare and serious pediatric disease, our approach strengthens available data, enhancing decision-making despite limited pediatric data availability. Our study showcases the significance of Bayesian Dynamic Borrowing in im-

proving the robustness of pediatric drug development strategies. By incorporating adult and adolescent PD data judiciously, we demonstrate how this methodology enables informed decision-making and supports regulatory requirements. Through this presentation, we aim to share insights into leveraging statistical methodologies to optimize pediatric drug development processes, particularly in rare and serious diseases.

229 . Modelling the Predictability of categorical response variable through Non-Symmetric Correspondence Analysis

[04.M2.I74, (page 48)]

Kirtee KAMALJA, *Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India*
Riya JAIN, *Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India*

Non-symmetric Correspondence Analysis (NSCA) is a powerful statistical technique designed for modeling, interpreting and visualizing the asymmetric association among response and predictor variables which are categorical in nature. NSCA is often considered as the categorical equivalent of regression analysis which is used for modelling the continuous response variable. NSCA for three-way contingency table (CT) visualizes the dependence among the categories of variables via modeling the centered column-tube profile array using tucker3 model approximation. The predictability in three-way CT is traditionally modelled using Kronecker product. We generalize the theory of three-way NSCA to multi-way case using higher order singular value decomposition. In order to model the relationships among the categories of more than three CVs, we propose to use tensor-based algorithm. The components in the model for centered predictor profile array and the coordinates for Interactively-coded predictor isometric Biplot are obtained using tensor operations. We present NSCA to analyze four-way and five-way CTs. Further, an efficient tensor-based algorithm is developed for modelling the predictability in multi-way CTs. The R codes are developed for the implementation of the proposed tensor-based algorithm and demonstrated through numerical examples.

Keywords: Tucker3 model, Higher order SVD, Non-Symmetric Correspondence analysis, Tensor operations.

230 . Inferring the Effect of a Confounded Treatment by Calibrating Resistant Population's Variance

[03.E1.I61, (page 39)]

Bikram KARMAKAR, *University of Florida*
Zikun QIN, *University of Florida*
Bikram KARMAKAR, *University of Florida*

In a general set-up that allows unmeasured confounding, we show that the conditional treatment effect can be identified as one of two possible values. Unlike existing causal inference methods, we do not require an exogenous source of variability in the treatment, e.g., an instrument or another outcome unaffected by the treatment. Instead, we require (a) a nondeterministic treatment assignment, (b) that all effect modifiers are measured, and (c) a resistant population that was not exposed to the treatment or, if exposed, is unaffected by the treatment. Assumptions (a) and (b) are mild and (b) can be relaxed. For (c), which is a new assumption, we show that a resistant population is often available in practice. We develop a large sample inference methodology and demonstrate our proposed method in a study of the effect of surface mining in central Appalachia on birth weight that finds a harmful effect.

231. Modeling Marriage Rate Fluctuations Using Survival Analysis Approach

[04.M2.I74, (page 48)]

Akanksha KASHIKAR, *Department of Statistics, Savitribai Phule Pune University*
Neela GULANIKAR, *Dept of Statistics, Savitribai Phule Pune University*

We use NFHS-IV data to study variations in marriage rates based on year of birth, education level, and gender. The log-rank test provides a way of comparing these rates when the Kaplan-Meier curves are parallel. However, in the case of intersecting curves, the test performance is not up to the mark. Hence, this paper discusses alternatives and methods to test the assumption of parallelism in the Kaplan-Meier curves.

232 . Two-stage drop-the-losers design for the selection of effective treatments and estimating their worth

[02.E1.C10, (page 30)]

Yogesh KATARIYA, *Indian Institute of Technology Kanpur*

In drug development clinical studies, multiple treatments or different dose levels of a drug are often of interest for evaluation and selection. Due to limited resources such as time, patient availability, budget, and health risk factors, it is important to consider only treatments that provide the best-observed response. Low-efficiency treatments should be dropped out early in the study to eliminate ineffective or unsafe treatments. In clinical studies, drop-the-losers design (DLD) is one way to screen out ineffective treatments, and it is also helpful in estimating the effectiveness of the selected treatments. In this talk, we will consider a two-stage DLD on different treatments whose effects are described by independent Gaussian responses with different unknown means and common variance. In the first stage of this design, data are collected to determine a subset of effective treatment(s) so that the inferior treatment groups will be dropped at the end of the first stage based on some pre-specified criterion, e.g. one can select the subset by using the subset selection rule of Gupta (1956,1965). The selected treatment(s) in the subset will then proceed to the second stage to estimate the effectiveness of the selected treatment(s) through the point estimation of the worth of the selected effective treatments, where worth is defined as the arithmetic average of the mean effect of the selected treatments. Since the bias of an estimator is an important criterion in clinical studies, we will discuss the Uniformly Minimum Variance Conditionally Unbiased Estimator (UMVCUE) of the worth of the selected effective treatments. The mean squared error (MSE) and the bias performances of the derived UMVCUE are compared via a simulation study with the naive estimator (Maximum Likelihood Estimator).

233. Modified semi-distance correlation and its use in testing indepndence

[03.M2.I52, (page 35)]

Sudheesh KATTUMANNIL, *Indian Statistical Institute, Chennai*

TBA

234. Statistical Challenges and Opportunities in Drug Development

[04.E1.I84, (page 54)]

Amarjot KAUR, *Merck Research Labs*

TBA

235 . Unveiling Cancer Complexity Through Integrative Multi-View Machine Learning

[03.M1.I43, (page 31)]

Aparajita KHAN, *Indian Institute of Technology Roorkee*

Aparajita KHAN, *Indian Institute of Technology Roorkee*

In today's era of precision medicine, the abundance of heterogeneous data in cancer research offers great promise for enhancing patient outcomes. Integrating and analyzing these diverse multi-view oncology data, encompassing multimodal genomics (copy number variation, gene expression, DNA methylation), radiology imaging and reports (from CT/PET scans), and electronic health records (EHRs), among others, demand cutting-edge machine learning algorithms. While multiple views are expected to provide more information for an improved learning performance, they pose their own set of unique challenges, like high-dimensional and heterogeneous nature of views, leveraging cross-view complementarity, and mitigating noise propagation from individual views during data integration. In this talk, first we will discuss our graph Laplacian approximation and fusion method to address these challenges and perform multi-view eigenspace clustering, effectively identifying distinct cancer subtypes. Next, we will present a hybrid modeling approach that combines structured EHR data with natural language processing of radiology reports, enabling automated abstraction of CT imaging indications. Here, our investigation explores the impact of surveillance imaging on long-term outcomes for lung cancer survivors. Overall, the talk will underscore the significance of harnessing multi-view learning methodologies to drive advancements in cancer research and ultimately improve patient care and outcomes.

236 . A unified Bayesian approach to transcriptome-wide association study

[Student Poster Competition, (page 21)]

Arnab Kumar KHAN, *Indian Statistical Institute, Kolkata*

Tanushree HALDAR, *Institute for Human Genetics, University of California San Francisco, San Francisco, California 94143, USA*

Arunabha MAJUMDAR, *Department of Mathematics, Indian Institute of Technology Hyderabad, Kandi, Telangana 502285, India*

Transcriptome-wide association study (TWAS)

has shed light on molecular mechanisms by examining the roles of genes in complex disease etiology. TWAS facilitates gene expression mapping studies based on a reference panel of transcriptomic data to build a prediction model to identify expression quantitative loci (eQTLs) affecting gene expressions. These eQTLs leverage the construction of genetically regulated gene expression (GReX) in the GWAS data and a test between imputed GReX and the trait indicates gene-trait association. Such a two-step approach ignores the uncertainty of the predicted expression and can lead to reduced inference accuracy, e.g., inflated type-I error in TWAS. To circumvent a two-step approach, we develop a unified Bayesian method for TWAS, combining the two datasets simultaneously. We consider the horseshoe prior in the transcriptome data while modeling the relationship between the gene expression and local SNPs and the spike and slab prior while testing for an association between the GReX and the trait. We extend our approach to conducting a multi-ancestry TWAS, focusing on discovering genes that affect the trait in all ancestries. We have shown through simulation that our method gives better estimation accuracy for GReX effect size than other methods. In real data, applying our method to the GEUVADIS expression study and the GWAS data from the UK Biobank revealed several novel genes associated with the trait body mass index (BMI).

237 . New characterization driven test for Pareto family

[03.E1.C18, (page 43)]

Sakshi KHANDELWAL, *Central University of Rajasthan*
Deepesh BHATI, *Central University of Rajasthan*

The Pareto distribution, a prominent member of distribution classes utilized across numerous disciplines, is the focus here. Despite numerous tests for assessing its goodness-of-fit, this article introduces a novel test grounded in a fresh characterization. We delve into the properties of this new test and compare its empirical size and power against existing tests. In many cases, the novel tests either outperform or match the performance of existing ones. Real data applications are considered for illustrative purposes.

238 . Bayesian inference in high-dimensional mixed frequency regression

[02.M2.I26, (page 17)]

Kshitij KHARE, *University of Florida*

Technological advancements in recent years have enabled organizations to collect, organize, store and analyze very large amounts of data from variables that are available at different temporal frequencies - eg. monthly, weekly, daily. Such data is commonly referred to as mixed frequency time series data. We will focus on mixed frequency regression, where the response variable and the covariates are available at different frequencies (for example, quarterly vs. monthly). We will present novel Bayesian methodology for (sparse) estimation of the regression coefficients and of the (autoregressive) lag length using a Bayesian adaptation of the nested lasso framework. The talk is based on joint work with Satyajit Ghosh and George Michailidis.

239 . Bayesian Inference for Stress-Strength Reliability Models Using Phase-Type Distributions

[02.A1.I33, (page 23)]

Joby K JOSE, *Kannur University*

Stress-strength reliability models has achieved considerable attention in recent years due to its applicability in various areas like engineering, quality control, psychology, biology, genetics, medicine etc. Phase-type distribution is a generalized class of distributions that is closed under several mathematical operations like maxima, minima, convolution, finite mixture etc and any discrete or continuous probability distributions on the positive real line can be represented as phase-type. Hence stress-strength reliability models based on phase-type distributions give a generalized structure for the stress-strength models. Moreover, matrix representation of the parameters helps in their flexible evaluation and easy manipulation. In Bayesian inference, we combine the prior knowledge with the information provided by the set of current observations to make more reliable inferences. Bayesian approach has the advantage of providing more meaningful inferences by making use of all available information. In this paper, we assume that both the strength of the system and the stress imposed on it are phase-type random variables and explore Bayesian inference for stress-strength reliability. We also consider Bayesian inference for phase-type stress-strength models under a progressive type-II right censoring scheme. The paper includes a detailed explanation of computing Bayes estimates for stress-

strength reliability using the Markov Chain Monte Carlo method, addressing both continuous and discrete phase-type distributions.

240. Analysis of life testing data in presence of competing risks

[04.A1.I78, (page 50)]

Arnab KOLEY, *IIM Indore*

In this study, an analysis of life testing data in presence of two competing risks is considered. The random variables associated with the different causes are assumed to follow Weibull distributions with different scale parameters. It is also assumed that the experiment starts under a normal environment which further goes into a stressed environment at a pre-fixed time point. The experiment stops when all the units fail. The whole analysis has been conducted under the classical framework.

241 . Spatial Regression: The Curious Case of Negative Spatial Dependence

[03.M2.I48, (page 34)]

Malabika KOLEY, *IIT, Kanpur*
Yu-Hsien KAO,
Anil BERA,

Positive spatial dependence is predominant in the spatial data analyses. Therefore, it is not surprising that most of the methodological papers are concerned with the positive spatial dependence when evaluating estimation and testing. However, prevalence of negative spatial dependence is not uncommon as evidenced in many applications; just to mention a few: Saavedra (2000), Boarnet and Glazer (2002), Filiztekin (2009), Pavlyuk (2011), Garrett and Marsh (2002), Garretsen and Peeters (2009) and Basdas (2009). Previous studies on testing spatial models usually concentrated only on positive spatial dependence. Anselin and Rey (1991), however, considered both negative and positive values in their Monte Carlo studies to compare the properties of Moran's I and Rao's score (RS) test separately for spatial error and spatial lag dependence. Anselin, Bera, Florax and Yoon (1996) considered the joint presence of lag and error dependence, but not negative parameter values. This paper is concerned with the case of negative spatial dependence and its consequence on estimation, specification tests and calculation of impact effects. We will extend the theory and Monte Carlo results considered in Anselin et al. (1996) by including negative coefficients. We then suggest how to alter the standard methodologies for model estimation and evaluation in the presence of negative spatial dependence.

242 . Current Status Data with two Competing risks and Missing failure types: A Non-Parametric Approach

[03.M1.I46, (page 32)]

Tamalika KOLEY, *Indian Institute of Management Lucknow*

This study focuses on non-parametric maximum likelihood estimation of sub-distribution functions based on current status data with two competing risks and a general pattern of missing failure types without assuming that the failure types are missing at random (MAR). Various estimation techniques are proposed based on whether the monitoring time variable is fixed or random. Estimation under fixed monitoring time can be extended to include discrete random monitoring time variable with a finite range. This paper also examines the most general situation, in which the monitoring time for each individual is random. Non-parametric method of estimation is employed and the standard errors of the maximum likelihood estimators are estimated using the subsampling bootstrap. Several simulation studies are carried out in order to investigate the finite sample properties of the maximum likelihood estimators. Finally, the methods are illustrated through the analysis of a real-life data set on hearing loss.

243. Learning from heterogeneous preferences

[01.A1.I16, (page 9)]

Ramya KORLAKAI VINAYAK, *University of Wisconsin-Madison*

Large pre-trained models trained on internet-scale data are often not ready for safe deployment out-of-the-box. They are heavily fine-tuned and aligned using large quantities of human preference data, usually elicited using pairwise comparisons. While aligning an AI/ML model to human preferences or values, it is worthwhile to ask whose preference and values we are aligning it to? The current approaches of alignment are severely limited due to their inherent uniformity assumption and the need for plurality , i.e., capturing the diversity in human preferences and values – is getting recognized as an important challenge to address in this arena. There is also rich literature on learning preferences from human judgements us-

ing comparison queries. It plays a crucial role in several applications ranging from cognitive and behavioral psychology, crowdsourcing democracy, surveys in social science applications, and recommendation systems. However, the models in this literature often focus on learning average preference over the population due to the limitations on the amount of data available per individual or on learning an individual's preference using a lot of queries. Furthermore, the knowledge of the metric, i.e., the way humans judge similarity and dissimilarity, is assumed to be known which does not hold in practice. We aim to overcome these limitations by building mathematical foundations for learning from diverse human preferences.

In this talk, I will discuss some recent results that focus on how we can reliably capture diversity in preferences while pooling together data from individuals to learn a common metric. In particular, I will talk about fundamental questions regarding simultaneous metric and preference learning where the goal is to learn an unknown but shared unknown metric from preference queries while the preferences are diverse and also unknown.

244. On estimation of multicomponent stress-strength reliability for Bilal distribution

[01.E1.C2, (page 11)]

Ashly Elizabeth KOSHY, *University of Kerala*
Manoj CHACKO, *University of Kerala*

In this paper, we consider multicomponent system reliability where the system has k components with independent and identical strengths and each component is experiencing a random stress. The reliability of such a system is derived when both stress and strength variables follow Bilal Distribution. The maximum likelihood and Bayes estimators are obtained. The Bayes estimators are derived using MCMC method. Simulation studies are also done to find the performance of the estimators developed in this paper.

245. Nonparametric estimation of doubly truncated stress-strength reliability

[03.A1.C12, (page 38)]

DEEPA K R, *MG UNIVERSITY KOTTAYAM*
ANGEL MATHEW, *MAHATMA GANDHI UNIVERSITY, KOTTAYAM*

The study of stress-strength models has a significant role in the field of reliability. Numerous researchers have conducted several studies to expand the well-known concept of stress-strength reliability. In reliability and survival analysis, there are scenarios where lifetime data is only available between two specific time points. In such cases, it is necessary to use a doubly truncated lifetime variable for studying various reliability measures. In this paper, we consider the stress-strength reliability for doubly truncated random variables and examine its various properties. The significance of the proposed measure is also illustrated using examples. Further, we propose a non-parametric estimator for the doubly truncated stress-strength reliability using the kernel method and establish its asymptotic properties under certain regularity conditions. Finally, a simulation study is conducted to assess the performance of the suggested estimator. Keywords: Stress-strength reliability, doubly truncated random variables, kernel estimator.

246. Choice of baseline hazards in joint modelling of longitudinal markers and time-to-event data

[04.E1.C28, (page 58)]

K M Jagathnath KRISHNA, *Regional Cancer Centre, Thiruvananthapuram*
Anand HARI, *Regional Cancer Centre, Thiruvananthapuram*
Divya DENNIS, *Regional Cancer Centre, Thiruvananthapuram*

Longitudinal time-to-event analysis is a statistical method to analyze data where covariates are measured repeatedly. In survival studies, the risk for an event is estimated using Cox-proportional hazard model or extended Cox-model for exogenous time-dependent covariates. However, these models are inappropriate for endogenous time-dependent covariates like longitudinally measured biomarkers, Carcinoembryonic Antigen (CEA). Joint models that can simultaneously model the longitudinal covariates and time-to-event data have been proposed as an alternative. The present study highlights the importance of choosing the baseline hazards to get more accurate risk estimation. The study used colon cancer patient data to illustrate and compare four different joint models which differs based on the choice of baseline hazards [piecewise-constant Gauss–Hermite (GH), piecewise-constant pseudo-adaptive GH, Weibull Accelerated Failure time model with GH & B-spline GH]. We conducted simulation study to

assess the model consistency with varying sample size (N = 100, 250, 500) and censoring (20 %, 50 %, 70 %) proportions. In colon cancer patient data, based on Akaike information criteria (AIC) and Bayesian information criteria (BIC), piecewise-constant pseudo-adaptive GH was found to be the best fitted model. Despite differences in model fit, the hazards obtained from the four models were similar. The study identified composite stage as a prognostic factor for time-to-event and the longitudinal outcome, CEA as a dynamic predictor for overall survival in colon cancer patients. Based on the simulation study Piecewise-PH-aGH was found to be the best model with least AIC and BIC values, and highest coverage probability (CP). While the Bias, and RMSE for all the models showed a competitive performance. However, Piecewise-PH-aGH has shown least bias and RMSE in most of the combinations and has taken the shortest computation time, which shows its computational efficiency. This study is the first of its kind to discuss on the choice of baseline hazards.

247. Some non-standard inference problems

[05.M1.I88, (page 59)]

Arun KUCHIBHOTLA, *Carnegie Mellon University*

Valid inference is of paramount importance in any scientific study. In this talk, I will present some methods of inference for a class of non-standard problems.

248. Evaluating an intervention targeting early detection of a disease in RCTs –evidence synthesis under the framework of multistate models

[04.E1.I84, (page 54)]

Sangita KULATHINAL, *University of Helsinki*
Aapeli NEVALA,

The conduct and analyses of randomized control trials studying disease histories are often complicated due to different factors, such as early withdrawal, loss to follow-up, intercurrent events, recurrent and terminal events. The ICH-E9(R1) addendum (2019) states that the description of an estimand should reflect the clinical question of interest in respect to these intercurrent events … and that the statistical analysis of clinical trial data should be aligned to the estimand. The addendum also clarifies the role of sensitivity analysis to explore the robustness of conclusions from the main statistical analysis. [Committee IES. ICH E9 (R1): addendum to statistical principles for clinical trials on choosing appropriate estimands and defining sensitivity analyses in clinical trials. Paper presented at: International Conference on Harmonization; 2019.] Motivated by the Finnish population-based randomized colorectal cancer (CRC) screening study with primary outcomes all-cause mortality or CRC mortality, we aim to evaluate the effectiveness of CRC screening under the multistate models as the underlying framework. We will define various estimands to answer the clinical questions of interest and discuss their estimation, in the Bayesian setting, by properly synthesizing the existing evidence and different data sources.

249. Parameter Estimation for Multi-state Coherent Series and Parallel Systems having Random Degradation Rates

[02.A1.I33, (page 23)]

Leena KULKARNI, *NMIMS (Deemed-to-be) University*
Sanjeev SABNIS, *IIT Bombay*

Any multi-state system having multi-state components can be viewed as a degradation system. The degradation model with random degradation rates for the multistate system proposed by Eryilmaz (2016) did not take into account the multi-state system's structure function into consideration. The main focus of this paper is to remove this lacuna and compare the theoretical results obtained by taking into account multi-state system's structure function with the corresponding results given in Eryilmaz (2016) paper, numerically and graphically. In this work, the degradation rates of the components are modelled using the FGM distribution with power series marginals. Additionally, the parameter estimation problem is attempted in the Bayesian framework. The numerical results show a significant difference between the respective relevant probabilities.

Key Words: Degradation System, Multi-state Series System, Multi-state Parallel System, Farlie-Gumbel-Morgenstern distribution.

250. Navigating a Dynamic Career in Statistics: Insights, Adaptation, and Building the Future

[Plenary Lecture 1, (page 6)]

Pandurang KULKARNI, *Eli Lilly*

In this plenary session, Dr. Pandu Kulkarni will share his fascinating career journey, highlighting his experiences across both academia and industry. Reflecting on pivotal moments, he will offer valuable lessons learned and advice for early-career statisticians, emphasizing the importance of adaptability and strategic decision-making. Pandu will also explore the essential skillsets for statisticians in the pharmaceutical industry, particularly in the evolving landscape of data science. He will address the role of academic institutions in preparing the next generation of statisticians, discuss strategies for bridging the skills gap in India's growing pharma sector, and shed light on the influence of professional organizations like IISA in shaping future leaders. With insights into the changing dynamics of the industry and the evolving roles of statisticians, this session will serve as a blueprint for building a successful and fulfilling career in biostatistics.

251. Truncation Effect and it's Adjustment via Parametric and Non Parametric approaches: Illustration through First Birth Interval Data

[02.A1.I38, (page 25)]

Anup KUMAR, *Department of Biostatistics and Health Informatics Sanjay Gandhi Postgraduate Institute of Medical Sciences Lucknow, Uttar Pradesh*

TBA

252. On Alternative Hyper-Poisson Distribution and its Related Versions

[01.M2.I7, (page 5)]

C. Satheesh KUMAR, *Department of Statistics, University of Kerala*

The Poisson distribution is based on the assumption that events occur under the principle of complete randomness. But in practice this principle does not hold. Consequently the Poisson distributions has been generalised in several ways. Among them the alternative hyper-Poisson distribution is of particular interest. The present talk focusses on some important aspects of alternative hyper-Poisson distribution and its related versions developed recently in the literature.

253. Redundancy Allocation for Series and Parallel Systems: A Copula-based Approach

[03.E1.C18, (page 43)]

Ravi KUMAR, *IIT Hyderabad*

The allocation of redundant components to a system is a common method for enhancing the system's lifetime. This study explores the optimal allocation of redundancies in series and parallel systems with two components by assuming components and redundancies are dependent. That is, we perform the stochastic comparisons of the series (parallel) systems in the case of two redundancies at the component level. Specifically, we examine the stochastic comparisons across three scenarios: (i) components (and redundancies) have dependent lifetimes but are independent of each other, and components (redundancies) have identical marginal distributions in the two generated systems; (ii) components (and redundancies) have dependent lifetimes and are independent of each other, but the marginal distributions of components (redundancies) are different in the two generated system; and (iii) components and redundancies are interdependent and the marginals of the components (redundancies) in the two generated systems are same. In this study, we model the dependency using the concept of copula and perform the desired stochastic comparisons using generalized distorted distribution functions. Furthermore, we demonstrate our findings through various examples and counterexamples. Finally, we provide a simulation-based study to illustrate our findings.

254. Bivariate Empirical Mode Decomposition Based Models for Agricultural Price Forecasting

[04.E1.C23, (page 55)]

Rounak KUMAR, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*
Girish Kumar JHA, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*
Rajeev Ranjan KUMAR, *ICAR - Indian Agricultural Statistics Research Institute, New Delhi – 110012*

The present research focus on concept of decomposition-based models for interval valued time series forecasting. Bivariate Empirical Mode Decomposition (BEMD), introduced by Rilling et al. 2007, particularly in the context of interval forecasting. BEMD extends Empirical Mode Decomposition (EMD) to two-dimensional data, enabling the capture of more intricate patterns and interactions within the data. This results in more reliable and accurate interval forecasts, making BEMD a superior choice for forecasting agricultural prices. In

this study, two BEMD based neural network models BEMD-TDNN and BEMD-LSTM are proposed for interval-valued agricultural price forecasting and are compared with EMD based models like EMD-TDNN and EMD-LSTM, respectively, to show the effectiveness of BEMD in case bivariate data in terms of theil statistic, IMSE and IMAE.

255 . Estimation Of FGMBEW and FGMBGE distributions under Progressive censored data

[02.E1.C8, (page 29)]

Saneesh KUMAR, *M G University*
Ansa Alphonsa ANTONY, *Mg University*

The Bivariate Exponentaited Weibull distribution and Bivariate generalized exponential distribution are important lifetime distributions used in survival analysis.Farlie-Gumbel-Morgenstern (FGM) copula and Exponentaited Weibull marginal distribution are used for creating a Bivariate distribution which is called FGM Bivariate Exponentaited Weibull (FGMBEW) distribution while FGM copula and generalized exponential marginal distribution are used for creating a Bivariate distribution which is called FGM Bivariate Generalized Exponential distribution (FGMBGE). These distributions are used for describing Bivariate data that we have weak correlation between variables in lifetime data.There are many scenarios in life-testing and reliability experiments in which progressive censored data is available. Works are available for estimation of FGMBEW and FGMBGE distributions under uncensored data. But there is no literature available for estimation of FGMBEW and FGMBGE distributions under Progressive censored data.So a study on this area will be of great significance to contribute to the estimation under Progressive censored data. In this paper theoretical discussion of FGMBEW and FGMBGE distributions with maximum likelihood method estimation under Progressive censored sampling is performed.The estimation is demonstrated using some real data sets.

256. Bivariate Distribution with Singular Component and its real-life Application

[04.A1.C21, (page 52)]

Sanjay KUMAR, *PhD candidate, Department of Mathematics & Statistics, IIT Kanpur*
Debasis KUNDU, *Professor, Department of Mathematics & Statistics, IIT Kanpur*
Sharmishtha MITRA, *Professor, Department of Mathematics & Statistics, IIT Kanpur*

Real-life data sets with ties arise quite commonly in reliability and survival analysis. We attempt to model such types of data sets using bivariate distributions with singular components. For this purpose, we consider mainly two types of approaches, namely the "Minimization approach" and the "Maximization approach". Using the minimization approach the bivariate modified Weibull (BMW) distribution is derived. The BMW is more general distribution and it reduces to the Marshall-Olkin bivariate exponential (MOBE) and Marshall-Olkin bivariate Weibull (MOBW) distributions under certain parameter restrictions. Some distributional, modal and aging properties of BMW will be discussed. Finally we will discuss about the maximum likelihood estimation of parameters of BMW distribution via EM algorithm. We will give some numerical results and comparisons.

257. On the two sample test for persistent homology.

[04.A1.C22, (page 53)]

Satish KUMAR, *Indian Institute of Technology Kanpur*
Subhra Sankar DHAR, *Indian Institute of Technology Kanpur*

In this talk, I shall talk about the problem of testing the hypothesis that the shapes underlying two data sets are the same up to their persistent homology. Persistent homology is a key tool in topological data analysis. In this talk, I shall briefly describe persistent homology, the formulation of the problem, test statistics, and the hypothesis testing procedure.

258 . Estimation and Inference for Change Points in Functional Regression Time Series

[Student Paper Competition 2, (page 18)]

Shivam KUMAR, *PhD Candidate*
Haotian XU, *University of Warwick*
Haeran CHO, *University of Bristol*
Daren WANG, *University of Notre Dame*

In this paper, we study the estimation and inference of change points under a functional linear regression model with changes in the slope function. We present a novel Functional Regression Binary Segmentation (FRBS) algorithm which is computationally efficient as well as achieving consistency in multiple change point detection. This algorithm utilizes

the predictive power of piece-wise constant functional linear regression models in the reproducing kernel Hilbert space framework. We further propose a refinement step that improves the localization rate of the initial estimator output by FRBS, and derive asymptotic distributions of the refined estimators for two different regimes determined by the magnitude of a change. To facilitate the construction of confidence intervals for underlying change points based on the limiting distribution, we propose a consistent block-type long-run variance estimator. Our theoretical justifications for the proposed approach accommodate temporal dependence and heavy-tailedness in both the functional covariates and the measurement errors. Empirical effectiveness of our methodology is demonstrated through extensive simulation studies and an application to the Standard and Poor's 500 index dataset.

259 . STATISTICAL ANALYSIS OF AREA, PRODUCTION AND PRODUCTIVITY OF MAJOR PULSES IN TELANGANA STATE

[04.L1.C19, (page 49)]

VANKUDOTH KUMAR, *ICAR-IASRI, NEW DELHI*
A S KAMBLE,

Pulses play a vital role in global agriculture due to their high protein content and their benefits for sustainable farming. They require less water and fewer inputs than other crops, and they contribute to soil health by fixing atmospheric nitrogen. To study the trends in pulse crop cultivation, a 30-year dataset (1993-2022) was used, focusing on three districts in Telangana state: Adilabad, Mahbubnagar, and Nizamabad. Various statistical models, such as linear, quadratic, Generalized Additive Model (GAM), exponential, and monomolecular, were employed. The analysis, based on Minimum Mean Absolute Percentage Error (MAPE) values, shows that the Generalized Additive Model (GAM) is the most effective for analyzing pulse crop production, productivity, and cultivation area across all three districts. GAM, in particular, proves to be highly accurate for both production and cultivation areas in Adilabad, Mahbubnagar, and Nizamabad, highlighting the significance of precise models for predicting and understanding pulse crop cultivation dynamics in these regions.

260. General weighted extropy of minimum and maximum ranked set sampling with unequal samples

[04.E1.C23, (page 55)]

Santosh KUMAR CHAUDHARY, *INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR*
Nitin GUPTA, *Indian Institute of Technology Kharagpur*

In industrial, environmental, and ecological investigations, ranked set sampling is a sample method that enables the experimenter to use the whole range of population values. The ranked set sampling process can be modified in two extremely helpful ways: maximum ranked set sampling with unequal samples and minimum ranked set sampling with unequal samples. They permit an increase in set size without too many ranking errors being introduced. In this paper, we are defining general weighted extropy (GWJ) of minimum and maximum ranked set samples when samples are of unequal size (minRSSU and maxRSSU, respectively). Stochastic comparison and monotone properties have been studied under different situations. Additionally, we compare the extropy of these two sampling data with that of ranked set sampling data and simple random sampling data. Bounds of GWJ of minRSSU and maxRSSU have been obtained. Finally, we investigate the weighted discrimination information between simple random sampling, ranked set sampling and minimum and maximum ranked set sampling with unequal sample sizes. Some results for equality of GWJ of minRSSU and maxRSSU under symmetric assumption are also obtained.

261. Matrix-Variate K-Means: An Extended K-Means Approach using Euclidean and Mahalanobis Distances

[03.E1.C13, (page 40)]

Shiva Kumar KURVA, *Pondicherry University*
Kiruthika C, *Pondicherry University*

Matrix-variate data, where each observation is a matrix rather than a vector, is increasingly common in various fields. Effective clustering of such data is crucial for uncovering underlying patterns and structures. In this study, we extend the classical K-means clustering to effectively cluster matrix-variate data by incorporating both Euclidean and Mahalanobis distance metrics. The traditional K-means clustering, typically designed for vector-based data, struggles with the complexities of matrix-variate structures. Our proposed extension of the K-means captures the

inherent correlations and variances within matrix-variate datasets, thus enhancing its applicability and performance. This extension not only provides a clustering technique for matrix-variate data but also a good initialization strategy for finite mixture models. To evaluate the effectiveness of our approach, we conducted simulation and real data studies. The results demonstrate that the modified K-means clustering performs robustly across various scenarios, achieving high-quality clustering.

262. A neural network approach to modelling count time series data

[03.M1.C11, (page 33)]

Divya KUTTENCHALIL ANDREWS, *Cochin University of Science and Technology, Kochi, India.*
Narayana BALAKRISHNA, *Indian Institute of Technology, Tirupati, India; Cochin University of Science and Technology, Kochi, India.*

Time series of counts are often analyzed using generalized integer-valued autoregressive models with conditional heteroskedasticity (INGARCH). These models utilize response functions to map a vector of past observations and past conditional expectations to the current observation's conditional expectation. This paper demonstrates how INGARCH models can be integrated with artificial neural network (ANN) response functions to create a class of nonlinear INGARCH models. The ANN framework enables the interpretation of many existing INGARCH models as a simplified version of a corresponding neural model. An empirical analysis of count time series is conducted.

263. Interval-Valued Data Analysis with Applications

[03.A1.I57, (page 37)]

Arnab LAHA, *Indian Institute of Management Ahmedabad*
MAHESH, K. C.,

Interval-valued data refers to observations which are not single numbers but are finite closed intervals of real numbers. Such data arises in many fields including meteorology, engineering, finance, economics etc. In the context of big data, such as those coming from high frequency trading in stock markets or from sensors, it is often convenient to record only the highest and lowest value of the observations within a time period. Since the finite intervals of real numbers can be uniquely specified by their centers (midpoints) and ranges, summary intervals based on these had been proposed in the literature. In this paper, we propose a new criteria for evaluating the performance of a summary interval and study its properties. We then propose two new methods of creating summary measures for interval valued data and compare their performance using the above criteria through extensive simulation. Some advantages and shortcomings of these new methods are discussed. Some real-life applications are discussed.

264. Bayesian prior for achieving multiple goals and its relationship with the adjusted maximum likelihood method

[02.A1.I39, (page 25)]

Parthasarathi LAHIRI, *University of Maryland College Park*
Masayo HIROSE, *Kyushu University, Japan*

The two-level normal hierarchical model has played an important role in statistical theory and applications. In this paper, we first introduce a general adjusted maximum likelihood method for estimating the unknown variance component of the model and the associated empirical best linear unbiased predictor of the random effects. We then discuss a new idea for selecting prior for the hyperparameters. The prior, called a multi-goal prior, produces Bayesian solutions for hyperparmeters and random effects that match (in the higher order asymptotic sense) the corresponding classical solution in linear mixed model with respect to several properties. Moreover, we establish for the first time an analytical equivalence of the posterior variances under the proposed multi-goal prior and the corresponding parametric bootstrap second-order mean squared error estimates in the context of a random effects model. This is a joint work with Dr. Masayo Hirose.

265 . Asymptotic methods in high dimensional regression under dependence

[02.M1.I25, (page 16)]

Samriddha LAHIRY, *National University of Singapore*
Pragya SUR, *Harvard University*
Jean BARBIER, *International Centre for Theoretical Physics*
Daria TIEPLOVA, *International Centre for Theoretical Physics*

In recent years, the field of supervised learning has witnessed the emergence of a novel paradigm known

as the proportional asymptotics regime, where the number of features and samples both diverge, at a rate proportional to each other. This approach has gained traction for its ability to deliver precise asymptotic results without imposing sparsity assumptions on the signal. However, extending these techniques to address dependent setups remains a significant challenge.

In this presentation, we aim to bridge this gap by exploring two distinct models of dependence. First, we examine a scenario where the columns of the design matrix exhibit a dependence structure. We establish a universality result in this context and show how it can be applied to accurately characterize the risk of estimators in nonlinear regression involving high-dimensional covariates.

In the second part of the talk, we investigate a time series model where the rows of the design matrix are dependent. Here, we adopt a Bayesian perspective and characterize the asymptotic behavior of normalized mutual information and minimum mean square error within this framework.

266 . Next generation statistical workflows for multimodal data integration

[01.E1.I19, (page 10)]

Leo LAHTI, *University of Turku, Finland*

Open data science infrastructures have become critical components of reproducible research, and a vast body of contemporary data science applications rely critically on open algorithmic methods and workflows. In order to enjoy the benefits of automation while maintaining flexibility, data science experts routinely remix and assemble algorithmic software libraries into custom workflows in modern computational environments. Seeking the optimal balance between automation and domain expertise remains a key challenge in applied statistics, however. We will show how the potential of open data science developer networks can be harnessed as a powerful innovation ecosystem to address these challenges. Heterogeneous and hierarchically structured multi-assay data arises in several fields from natural to social sciences and humanities. Next-generation of data containers have emerged to store relations between interlinked data tables in a highly optimized way, providing functions to access, manipulate and integrate data at varying resolutions for seamless downstream analysis within statistical and probabilistic programming workflows. Whereas this approach has been shown to facilitate collaborative development of improved statistical methods and workflows, its adoption has been limited to specific problem domains. We will show how statistical and probabilistic programming workflows can be support by this technology, leading to better explainable and verifiable analyses. The work is motivated by timely challenges in two distinct but complementary fields, computational humanities and microbial ecology.

267 . Asymptotic non-identifiability of peer effects in linear-in-means models

[01.M2.I5, (page 5)]

Keith LEVIN, *University of Wisconsin–Madison*
Alex HAYES,

Many network autoregressive models seek to account for peer effects such as contagion and interference, in which node-level responses or behaviors may influence one another. These models are frequently deployed by practitioners in sociology and econometrics, typically in the form of linear-in-means models, in which node-level covariates and local averages of responses are used as predictors. In highly structured networks, previous work has shown that peer effects in linear-in-means models are collinear with other regression terms, and thus cannot be estimated, but this collinearity is widely believed to be ignorable, as peer effects are typically identified in empirical networks. In this work, we show a concerning negative result: under linear-in-means models, when node-level covariates are independent of network structure, peer effects become increasingly collinear with other regression terms as the network size (i.e., number of nodes) grows, and are inestimable asymptotically. We also show a narrow positive result: under certain latent space network models, some peer effects remain identified as the network size grows, albeit under rather stringent conditions. Our results suggest that linear models for peer effects are appropriate in far fewer settings than was previously believed.

268. Conformal Prediction for Network-Assisted Regression

[01.M2.I5, (page 5)]

Robert LUNDE, *Washington University in St Louis*

TBA

269 . Exploring the Estimation Techniques and Stress-Strength analysis on

PGDUS-powered inverse Rayleigh Distribution

[01.E1.C4, (page 12)]

AMRUTHA M, *Research Scholar, Department of Statistics, St. Thomas College (Autonomous), Thrissur, Kerala, India*
V M CHACKO, *Professor, Department of Statistics, St. Thomas College (Autonomous), Thrissur, Kerala, India.*

We choose the Power Generalized DUS (PGDUS) transformation because it can improve the flexibility and fit of baseline distributions, especially when dealing with lifetime and reliability data. Its ability to control tail behavior, provide parsimonious models, and offer a range of estimation methods makes it a powerful tool for statistical modeling and analysis of a parallel system. The empirical success and adaptability of the PGDUS transformation further highlight its value in diverse applications. with parameters alpha, beta, and theta, we propose a new distribution called the PGDUS-powered inverse Rayleigh distribution (PGDUS-PIR) by using the powered inverse Rayleigh (PIR) distribution as the baseline distribution. Its statistical features are derived in part. The unknown parameters are estimated using maximum likelihood estimation (MLE), maximum product spacing estimation (MPSE), and least square methods (LSE). By conducting a simulation study, we can gain deeper insights into the behavior and applicability of the PGDUS-PIR distribution, leading to more accurate and reliable statistical modeling and analysis. A real data set is then used to compare the performance of the proposed distribution over some existing distribution. Stress-Strength reliability is an essential aspect of lifetime distribution analysis, providing a systematic approach to evaluating the reliability and durability of components and systems under varying stresses. Here we derive the stress-strength reliability of the PGDUS-PIR distribution and estimating the reliability by using different estimation techniques and conducting simulation studies based on these techniques. Keyword:- powered inverse Rayleigh distribution, PGDUS transformation, MLE, MPSE, LSE, Stress-Strength Reliability.

270 . An autoencoder based semi-supervised approach for accurate binning of metagenomics data

[04.L1.C19, (page 49)]

Deeksha P M, *ICAR-Indian Agriculture Research Institute, New Delhi*
Shashi Bhushan LAL, *1. ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
Anu SHARMA, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
Dwijesh Chandra MISHRA, *3. ICAR-National Academy of Agricultural Research Management, Hyderabad*

Microorganisms thrive in diverse environments, from soil to extreme locations like hot springs and glaciers, showcasing remarkable metabolic and physiological diversity. Microbes, too small to see with the naked eye, have been traditionally studied through culture-based methods. However, traditional culture-based methods have limited our understanding of these organisms, as most remain uncultured. Metagenomics, a culture-independent approach, has revolutionized microbial research by directly analysing genetic material from environmental samples, revealing genetic diversity, functional potential, and ecological interactions of entire microbial communities. Despite advancements, metagenomic data analysis faces challenges, particularly in clustering and assembling genomes from diverse organisms. This study introduces a novel approach using deep learning to enhance clustering in metagenomic data through the K-means algorithm. The study explores the use of single-copy phylogenetic marker genes for accurate species delineation. A semi-supervised clustering approach was applied to a metagenome dataset, initially reducing the high-dimensional feature matrix to 512 dimensions using an autoencoder. The produced clusters were analysed using marker genes data identified in the contigs by running Prodigal, FetchMG and USEARCH sequentially. K-means clustering analysis revealed 8 clusters with a Rand index of 0.85 and an F1 score of 0.61. Post-correction, 10 clusters were identified, significantly improving clustering quality with overall accuracy of 0.62. These findings highlight the importance of rigorous cluster correction in metagenomic analysis, enhancing the precision and significance of clustering outcomes.

271. Change point detection and inference

[04.E1.I83, (page 54)]

Carlos Misael MADRID PADILLA, *Washington University in St Louis*

This talk explores two complementary approaches to change point analysis. The first focuses on detecting and localizing change points in functional

data observed on d-dimensional domains. We present the Functional Seeded Binary Segmentation (FSBS) algorithm, which is efficient, handles sparsely or densely sampled data, and achieves sharp localization error rates, with applications demonstrated in detecting sea surface temperature changes linked to El Niño. The second part extends the discussion to multivariate time series, introducing methods for inferring change points with short-range dependence. These include consistent estimators and a novel variance estimator, providing theoretical guarantees and advancing the challenging task of inference beyond localization.

272. A rank-based test of independence between nonparametric covariate and error in semiparametric regression with missing at random response

[01.M2.I9, (page 6)]

Saran Ishika MAITI, *Visva-Bharati*
Sthitadhi DAS,

Traditional semiparametric regression model combines the flexibility of nonparametric regression and parametric linear regression. The basic model is additive in form of nonparametric and parametric covariates and the random error. The model assumes non-association (or independence) between error and these pair of covariates. However, in applied work, this assumption of no-association might fail. Again in regression theory, issue on missing observations is a genuine concern quite often. An intrinsic quest would be the validity of assumptions in presence of missingness. In addressing it, this article unravels a rank based test procedure under missing at random (MAR) situation occurred in response Y under semiparametric regression. At the initial step, parametric coefficient is estimated through residual regression technique. Nadaraya-Watson kernel density and imputed local linear smoothing method are used for imputing the unknown regression function. Then using complete set of observations, hence obtained, three types of test statistics are proposed and studied on contiguous alternative set-up. Further, the power performance of newly introduced test is investigated under the finite sample simulation study accompanied by a neat real data analysis.

273. A Bayesian method for estimating gene-level polygenicity under the framework of transcriptome-wide association study

[03.M1.I45, (page 32)]

Arunabha MAJUMDAR, *IIT Hyderabad*

TBA

274. A stochastic comparison study on replacement policies

[02.A1.I33, (page 23)]

Priyanka MAJUMDER, *Indian Institute of Science Education and Research Thiruvananthapuram*
Akhila Anna VARGHESE, *IISER Thiruvananthapuram*

To ensure uninterrupted service, a proper replacement policy should be followed based on the failure mechanism of the system. In this talk, a brief review of various stochastic comparison results under various replacement policies like age replacement policy, block replacement policy, and replacement model with random replacement time will be presented for various nonparametric aging classes together with some new results in the context of increasing convex average order.

275 . InterSpatial: Bi-directional Integration of Single Cell RNA-seq and Spatial Transcriptomic Data

[01.M2.I1, (page 3)]

Tuhin MAJUMDER, *Duke University*
Cliburn CHAN, *Duke University*
Jichun XIE, *Duke University*

To understand the interactions and dynamics between cells, it is essential to integrate spatial transcriptomic with single cell RNA-seq data. While existing spatial transcriptomic methods mainly focus on gene expression patterns in different locations of a tissue, our novel approach provides a spatial probability distribution for each cell in the single cell data across the tissue space. By explicitly modeling probability distributions, we can compare our results with existing methods of cell type distribution and unlock a broad scope for computing cell-cell interactions. We use the spatial coordinates of the spots to compute the Wasserstein distance between the spatial distributions of two different cells, eventually forming a cell-cell network that helps identify co-localized cell types. We demonstrate our method's effectiveness on several datasets. First, we integrate spatial transcriptome and scRNA-seq data from the human lung to to locate the senescent cells in human lungs, responsible for aging. We also identify the spatial distributions of different cell types, the co-localization patterns between immune cells and the senescent cells,

explaining the ligand-receptor relationship. We also use our method identifying the neighboring cell types of senescent cells. which is crucial for preventive measures as these cells induce their neighbors to become senescent over time. Second, we analyze the cell-cell co-localization between microglia's and excitatory neurons in the middle temporal gyrus of patients with Alzheimer's disease. We focus on the microglia's role on Alzheimer's propagation and neurodegeneration. In summary, our novel method provides a more comprehensive understanding of cell-cell interactions and dynamics in a pan-tissue setup, unlocking new insights into the functioning of biological tissues.

276. Impact of Psychological Contract Breach on Innovative Behavior and Well-being amongst Academicians

[01.A1.I14, (page 8)]

Sunita MALL, *MICA,Ahmedabad*
Anushree KARANI , *Shri Jairambhai Patel Institute of Business Management and Computer Applications, Gandhinagar, India*
Sunita MALL, *MICA,Ahmedabad*
Revati DESHPANDE, *Independent Researchers, Gujarat, India*

The study aimed to understand the relationship between psychological contract breach well-being indicators, i.e., subjective well-being and, mental well-being, and innovative behavior. Data were collected via a structured questionnaire through Google Docs from 238 academicians working in different capacities in Indian academic industries. The hypotheses were tested using structural equation modelling. Psychological contract breach positively impacted occupational stress, and occupational stress negatively impacted work engagement. Work engagement positively impacted innovative behaviour and well-being indicators. Occupational stress and work engagement mediated the relationship between psychological contract breach, well-being, and innovative behaviour.

277. Re-purposing without Reinventing the Wheel - Ensemble Models for Differential Analysis

[02.A1.I36, (page 24)]

Himel MALLICK, *Cornell University*

Inspired by ensemble models in machine learning, we propose a general framework for aggregating multiple diverse base models to boost the power of published differential association analysis (DAA) methods. We demonstrate this approach by augmenting popular DAA models with one or more biologically motivated alternatives. This creates an ensemble that bypasses the challenge of selecting an optimal model but instead combines the strengths of complementary statistical models to achieve superior performance. Our proposed ensemble learning approach is platform-agnostic and can augment any existing DAA method, providing a general and flexible framework for various downstream modeling tasks across domains and data types. We performed extensive benchmarking across both simulated and experimental datasets from single-cell to bulk RNA-Seq to microbiome profiles, where the ensemble strategy vastly outperformed non-ensemble methods, identified more differential patterns than the competitors, and displayed good control of false positive and false discovery rates across diversified scenarios. In addition to highlighting a substantial performance boost for state-of-the-art DAA methods, this work has practical implications for mitigating the so-called reproducibility crisis in omics data science. Lightweight R/Python implementations of the ensemble strategy are publicly available.

278. Credible Distributions of Overall Ranking of Entities

[04.M1.I64, (page 44)]

Abhyuday MANDAL, *University of Georgia*

Inference on overall ranking of a set of entities, such as athletes or players, schools and universities, hospitals, cities, restaurants, movies or books, companies, states, countries or subpopulations, based on appropriate characteristics or performances, is an important problem. Estimation of ranks based on point estimates of means does not account for the uncertainty in those estimates. Treating estimated ranks without any regard for uncertainty is problematic. We propose a novel solution using the Bayesian approach. It is easily implementable, competitive with a popular frequentist method, more effective and informative. Using suitable joint credible sets for entity means, we appropriately create credible distributions (CDs, a phrase we coin), which are probability distributions, for the rank vector of entities. As a byproduct, the supports of the CDs are credible sets for overall ranking. We evaluate our proposed procedure in terms of accuracy and stability using a number of applications and a simulation study. While the fre-

quentist approach cannot utilize covariates, the proposed method handles them routinely to its benefit.

279. Network sampling based inference for subgraph counts and clustering coefficient in a Stochastic Block Model framework with some extensions to a sparse case

[Student Paper Competition 2, (page 17)]

Anirban MANDAL, *TSMU, Indian Statistical Institute, Delhi*
Anirban MANDAL, *Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Delhi*
Arindam CHATTERJEE, *Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Delhi*

Sampling is frequently used to collect data from large networks. In this article, we provide valid asymptotic prediction intervals for subgraph counts and clustering coefficient of a population network when a network sampling scheme is used to observe the population. The theory is developed under a model-based framework, where it is assumed that the population network is generated by a Stochastic Block Model (SBM). We study the effects of induced and ego-centric network formation, following the initial selection of nodes by Bernoulli sampling, and establish the asymptotic normality of sample-based subgraph count and clustering coefficient statistic under both network formation methods. The asymptotic results are developed under a joint design and model-based approach, where the effect of sampling design is not ignored. In case of the sample-based clustering coefficient statistic, we find that a bias correction is required in the ego-centric case, but there is no such bias in the induced case. We also extend the asymptotic normality results for estimated subgraph counts to a mildly sparse SBM framework, where edge probabilities decay to zero at a slow rate. In this sparse setting, we find that the scaling and the maximum allowable decay rate for edge probabilities depend on the choice of the target subgraph. We obtain an expression for this maximum allowable decay rate and our results suggest that the rate becomes slower if the target subgraph has more edges in a certain sense. The simulation results suggest that the proposed prediction intervals have excellent coverage, even when the node selection probability is small and unknown SBM parameters are replaced by their estimates. Finally, the proposed methodology is applied to a real data set.

280. Generalized propensity scores to obtain causal estimates in observational studies of environment, microbiome and health

[02.A1.I36, (page 24)]

Siddhartha MANDAL, *Center for Chronic Disease Control, New Delhi, India; Center for Health Analytics and Trends, Ashoka University, India*

TBA

281. Estimating Shannon Entropy of the Selected Population

[03.E1.C14, (page 40)]

Masihuddin MASIHUDDIN, *Indian Institute of Information Technology Guwahati*
Neeraj MISRA, *IIT Kanpur*

In this talk, I will discuss the problem of estimating the Shannon entropy of the selected gamma population (referred as the selected entropy) under the mean squared error criterion. In order to improve upon various naive estimators of the selected entropy, we derive a class of shrinkage estimators that shrink various naive estimators towards the central entropy. For this purpose, we first consider a class of naive estimators comprising linear, scale and permutation equivariant estimators and identify optimum estimators within this class. The class of naive estimators considered by us contains three natural plug-in estimators. To further improve upon the optimum naive estimators, we consider a general class of equivariant estimators and obtain dominating shrinkage estimators.

282. Predicting Immune Checkpoint Blockade by Stereotactic Radiation and In Situ Virus Gene Therapy in Metastatic Triple-Negative Breast Cancer

[01.E1.I19, (page 10)]

Sunil MATHUR, *Houston Methodist Neal Cancer Center*
Ethan BURNS, *Houston Methodist NEal Cancer Center*
Shreya MATHUR, *Boston Medical Center*
Jenny CHANG, *Houston Methodist Neal Cancer Center*

Background/Introduction Triple-negative breast cancer (TNBC) presents a significant challenge due to its aggressive and resistant nature. For individuals with locally advanced unresectable or metastatic TNBC (mTNBC), the median overall survival (OS) ranges between 8 to 13 months, highlighting the ur-

gent need for effective treatments. Evaluating drug efficacy for TNBC is crucial in cancer research, providing a foundation for informed medical decisions, healthcare planning, and patient empowerment in their healthcare choices. The resulting improvements in care quality, error reduction, fewer adverse events, increased efficiency, cost-effectiveness, and higher satisfaction among both providers and patients emphasize the importance of effective treatments. Given the complexity of the disease, the patient population often does not follow a normal distribution. Methods To address this challenge, we propose a test based on ranked-set empirical distribution functions that incorporates external information. This approach constructs a distribution-free alternative to traditional hypothesis testing procedures under non-normal conditions, overcoming several limitations of earlier tests. Specifically, we define the proposed test statistic using the power divergence between two empirical distribution functions. Results Our test statistic, which is based on a power divergence between empirical distribution functions derived from two independent samples, proves particularly potent in comparison to alternatives under both heavy-tailed and light-tailed distributions. Employing the permutation principle fortifies the test's implementation. Utilizing the Monte Carlo method, empirical power computations demonstrate the superiority of our test across heavy-tailed, light-tailed, and even elliptically asymmetric population distributions. We show the implementation of the proposed test using real data. Conclusion In contrast to Hotelling's T^2 and Chatterjee and Sen's bivariate Wilcoxon rank sum test, the innovative approach we propose eliminates the need for estimating a dispersion matrix. This presents a significant advantage over existing methods, as the sample dispersion matrix may become singular, rendering its inverse nonexistent and thereby hindering the implementation of the test. Unlike the Kolmogorov-Smirnov test, which utilizes the empirical distribution function in a univariate setting for both one and two-sample problems, its multivariate extension is limited to one sample. To the best of our knowledge, there is currently no test available for two samples, apart from our proposed test, that employs the empirical distribution function, addressing a notable gap in the existing literature. In summary, our proposed test consistently exhibits superior power, regardless of the underlying population characteristics, offering a robust solution to drug efficacy assessment, and the proposed test can be applied to other situations.

283. Analyzing Progressive Censoring in k out of n Load Sharing Systems under Exponential Distribution

[03.A1.C12, (page 39)]

Anil MAURYA, *Sardar Vallabhbhai National Institute Technology*

Anil MAURYA, *Sardar Vallabhbhai National Institute of Technology, Surat, 395007, Gujarat, India*

Raj Kamal MAURYA, *Sardar Vallabhbhai National Institute of Technology, Surat, 395007, Gujarat, India*

In this talk, the estimation of model parameters of a k-out-of-n load-sharing system under a progressive censoring scheme will be discussed. This load-sharing system consists of n identical components whose lifetimes follow an exponential distribution. The system works as long as at least k out of n components are operational. When a component fails, its load is redistributed among the remaining surviving components, and the failure rate of the surviving component has been increased due to increasing load. Such systems are known as load-sharing systems. Despite its complexity, this load-sharing mechanism is crucial for understanding the reliability of systems in which components do not fail independently. The estimation of model parameters has been discussed under classical and Bayesian frameworks. In the classical approach, we utilize maximum likelihood estimation and maximum product spacing estimation. Further, we have constructed the approximate confidence intervals based on these estimates to enhance our analysis. The Bayesian framework estimates model parameters via the likelihood function and the product of the spacing function under the squared error loss function. The Metropolis-Hastings algorithm is employed to obtain these parameter estimates. Additionally, the derived Bayes estimates have been used to construct the Highest Posterior Density intervals. A comprehensive Monte Carlo simulation study is conducted to evaluate the performance of these estimation methods. Furthermore, the proposed methodologies are validated using real data, demonstrating their practical applicability and effectiveness in estimating model parameters.

Keywords: load-sharing system, progressive censoring scheme, maximum likelihood estimation, maximum product spacing estimation, Bayesian estimation.

284. Structural Aspects of Steady-state Characters in Queuing Systems

[01.A1.I17, (page 9)]

Manoharan MAVILA, *University of Calicut*

This paper brings together relevant materials on identifiability and characterizations of queueing systems based on certain steady-state characters of the system. These questions are motivated by enumerable instances and manifestation of queuing models in diverse fields such as population dynamics, inventory systems, telecommunications, stochastic networks etc. A brief review of the characterizations of queueing systems based on departures initiates the discussion followed by that based on infinite divisibility and unimodality property. We investigate the infinite divisibility property of the steady state distribution of certain queue characteristics like queue length, waiting time, number of customers served during the busy period etc. of certain discrete and continuous time single/bulk arrival queueing models with general service time distribution. Our investigation is further extended to multiple vacation queueing models, revealing the structural aspects of their steady state distributions.

285. Detecting and Managing Malnutrition with AI

[Special Invited Session 1, (page 14)]

Madhu MAZUMDAR, *Icahn School of Medicine at Mount Sinai*
Melanie BESCULIDES, *Institute for Healthcare Delivery Science, Icahn School of Medicine at Mount Sinai*
Sara WILSON, *Institute for Healthcare Delivery Science, Icahn School of Medicine at Mount Sinai*
Pramathamesh PARCHURI , *Institute for Healthcare Delivery Science, Icahn School of Medicine at Mount Sinai*

Malnutrition is a prevalent but often missed condition in hospitalized patients with potentially severe consequences. Accurate identification and coding are crucial for understanding malnutrition's impact on patient outcomes, including mortality, length of stay, and readmissions. This study will develop and evaluate a machine learning tool, MUST-Plus, to detect malnutrition. Phase I will build the tool using retrospective data. Phase II will assess its performance in various hospital settings. Phase III will examine how dietitians integrate the tool into their workflow. Ultimately, this presentation will demonstrate the value of machine learning and mixed methods research in developing and implementing AI tools.

286. Identifying key drivers of extinction for Chitala populations: Data-driven insights from an intraguild predation model using a Bayesian framework

[03.E1.C16, (page 42)]

Dipali V. MESTRY, *Department of Mathematics, Institute of Chemical Technology, Mumbai*

The fish species Notopterus chitala is a freshwater fish that is widely distributed in African and Asian countries, including India, Pakistan, Bangladesh, Sri Lanka, Nepal, Thailand, and Indonesia. This species has been categorized as endangered (EN) in the Conservation Assessment and Management Plan. The study aims to investigate the cause of the species' decline in their natural habitat. Using mathematical models supported by empirical data analysis, we explore the interaction of the species with other tropic levels and discover important parameters that may be attributed to the rapid decline. Based on the literature, we considered an intraguild predation (IGP) system consisting of three species, namely Chitala (IG predator), Mugil (IG prey), and shrimp (resource). Two variants of IGP models governed by three coupled differential equations are considered for data modeling purposes. Chitala depends only on Mugil and shrimp in one model. An alternative food source is available to Chitala in the second model. The models are estimated using the Bayesian modeling framework. Posterior estimates of the parameters for each model were obtained using the Gibbs algorithm, and the reversible-jump Markov chain Monte Carlo method has been utilized for posterior model inference. Our findings suggest that the primary reason for the decline in Chitala is due to the reduced nutritional gain from the Mugil and reduced predation efficiency in acquiring shrimp as a food source in the unavailability of Mugil. This study may be useful to develop management strategies for Chitala conservation by emphasizing the regeneration of Mugil populations.

287 . Investigating the price determinants of the European Carbon Trading System: a non-parametric approach

[04.M2.I70, (page 46)]

Antonietta MIRA, *Università della Svizzera italiana and Insubria University*
Cristiano SALVAGNIN, *University of Brescia*
Aldo GLIELMO, *Team, Directorate General for IT, Banca d'Italia*
Maria Elena DE GIULI, *University of Pavia*

The European carbon market (EUA) plays a pivotal role in the European Union's ambitious target of achieving carbon neutrality by 2050. Understanding the intricacies of factors influencing EUA market prices is paramount for effective policy making and strategy implementation. Employing the Information Imbalance, a nonparametric measure, the study delves into the relationships among macroeconomic, economic, uncertainty, and energy variables concerning EUA prices. Beyond variable selection, we extend the Information Imbalance methodology to time scale selection, identifying the weekly time scale as the most informative. This insight enhances our understanding of the temporal dynamics shaping EUA prices. Additionally, we employ Gaussian Processes for nowcasting and forecasting predictions, emphasising the precision achieved by focusing on the three most informative variables identified by the Information Imbalance model. In essence, the study provides nuanced insights into the evolving dynamics of the EUA market, shedding light on the changing significance of key variables across different phases of the market.

288 . A multivariate approach to joint testing of main genetic and gene-environment interaction effects

[02.E1.C8, (page 29)]

Saurabh MISHRA, *Indian Institute of Technology, Hyderabad*
Arunabha MAJUMDAR, *Indian Institute of Technology Hyderabad*

Gene-environment (GxE) interactions crucially contribute to complex phenotypes. The statistical power of a GxE interaction study is limited mainly due to weak GxE interaction effect sizes. To utilize the individually weak GxE effects to improve the discovery of associated genetic loci, Kraft et al. proposed a joint test of the main genetic and GxE effects for a univariate phenotype. We develop a testing procedure to evaluate combined genetic and GxE effects on a multivariate phenotype to enhance the power by merging pleiotropy in the main genetic and GxE effects. We base the approach on a general linear hypothesis testing framework for a multivariate regression for continuous phenotypes. We implement the generalized estimating equations (GEE) technique under the seemingly unrelated regressions (SUR) setup for binary or mixed phenotypes. We use extensive simulations to show that the test for joint multivariate genetic and GxE effects outperforms the univariate joint test of genetic and GxE effects and the test for multivariate GxE effect concerning power when there is pleiotropy. The test produces a higher power than the test for multivariate main genetic effect for a weak genetic and substantial GxE effect. For more prominent genetic effects, the latter performs better with a limited increase in power. Overall, the multivariate joint approach offers high power across diverse simulation scenarios. We apply the methods to lipid phenotypes with sleep duration as an environmental factor in the UK Biobank. The proposed approach identified six independent associated genetic loci missed by other competing methods.

289 . Is deep learning-based climate downscaling faithful to extreme events?

[02.E1.I41, (page 26)]

Adway MITRA, *Indian Institute of Technology Kharagpur*
Sumanta Chandra MISHRA SHARMA, *IIT Kharagpur*
Deepayan CHAKRABORTY, *IIT Kharagpur*

Many climate scientists are increasingly using Machine Learning over climate simulations by Earth System Models such as GCMs, or Weather Forecasting Models. The aims can be to obtain estimates at higher resolutions than the original (process-based) models, or to develop low-cost "surrogate" models that can emulate the climate simulators for a small subset of variables. However, questions remain about the statistical properties of such estimates. It has been observed that while ML models, especially those based on Deep Neural Networks can capture the mean climatology well, they often fall short in terms of the variabilities, and particularly the extreme values. In this talk, we will demonstrate the failures of Convolution-based Neural Networks to represent the extreme values of forecasted/simulated climate variables, and strategies to improve the same.

290. Causal Inference 3 in Drug Development

[01.M2.I3, (page 4)]

Jit MITRA, *Eli Lilly*

TBA

291. Trends of Deep learning at single-cell level for biomarker based drug dis-

covery and development

[01.E1.C3, (page 12)]

Sharda MOHAK, *Glaxo Smith Kline*

Many disease treatments including cancer have faced failures owing to various reasons including drug resistance. This can be majorly attributed to a focus on tissue-level drug development studies. With the availability of high throughput experimental setups, it is now possible to carry out cell-drug outcome predictions at a single cell level allowing one to incorporate cell heterogeneity. Because of the high dimensional and multi-modal nature of the resulting datasets, conventional statistical methods (including conventional machine learning models like linear regression, SVMs, tree-based models) fall short of analyzing such mammoth datasets. Various deep learning methods come to rescue. In this talk, I would talk about the importance, trends, potentials and challenges of deep learning methods in single cell and multi-modal data analyses and how that could accelerate drug development pipelines, resulting in better informed and successful biomarkers.

292. Bayesian Beta Regression for Analyzing Well-Being Patterns in ZIP Codes

[02.A1.I34, (page 24)]

Shariq MOHAMMED, *Boston University*

Individual-level assessment of well-being can be used to develop community-level indices that measure wellness and health risks for different geographical regions. These indices typically have bounded support and are updated over time to reflect changes in responses by individuals or the characteristics of the region. In this paper, we present a Bayesian Beta regression framework that uses individual-level survey data to model the overall well-being for ZIP codes and well-being from five subscales (Physical, Financial, Social, Community, Purpose). Our hierarchical prior formulation incorporates spatial and temporal information. ZIP Code neighborhood information is included through a graph Laplacian matrix constructed using driving time between ZIP Code population centroids. Additionally, we incorporate a temporal component, in which posterior estimates of ZIP Code effects from one year are used to inform prior distributions for subsequent years. This allows estimation of ZIP Code effects on well-being using individual data as well as borrowing data from neighboring ZIP Codes and past surveys. We deploy our model on well-being data from 2021 to 2024 for US States and Census Bureau–designated regions. We find that our model captures the effects of demographic features and identifies ZIP Codes well-being patterns over time.

293. Extreme Value Analysis for Attribution of Floods and Heat Waves to Climate Change

[02.E1.I41, (page 26)]

Arpita MONDAL, *Indian Institute of Technology Bombay*
Roshan JHA, *Indian Institute of Technology Bombay*
Manish Kumar DHASMANA, *Indian Institute of Technology Bombay*

Extreme events such as heat waves, floods and droughts cause significant societal, environmental and economic impacts. In recent years, such extremes have been reportedly more frequent, more intense and have larger extents. Based on physical understanding of the earth system, global warming and climate change are further likely to exacerbate such extremes. Identifying the role of anthropogenic climate change in extreme events has been the subject of attribution science. It is known with high confidence that regional heat waves are made more intense by global warming. Warming-induced thermodynamic changes also lead to an increase in heavy rainfall, though dynamic changes are hard to quantify and are often uncertain at regional scales. This study will present the use of Statistical Extreme Value Analysis in probabilistic event attribution analysis to delineate the role of human-induced climate change in extreme events in India, focusing on recent heat waves and heavy flooding in India.

294. Analyzing soccer shots with spatial latent trait models

[03.A1.I59, (page 38)]

Debashis MONDAL, *Washington University in St Louis*
Sayan DAS, *Washington University in St Louis*

Soccer, the world's most popular sport, captivates millions with its dynamic gameplay and strategic depth. In sports analytics, analyzing soccer shot data provides valuable insights into team performance and player efficiency. In this talk, we focus on spatial latent trait models, which expand upon the classical Rasch models, to explore shot data from the English Premier League. By incorporating different types of

symmetry—such as axial, and reflectional—we investigate how shot success probabilities vary across different areas of the field, quantify teams' attacking and defensive abilities, and analyze the effects of formations, home-field advantage, seasonal trends, and time effects. Additionally, we compare open-play and set-piece situations and analyze how success probability shifts when shots are taken with the right foot versus the left.

295. Computationally efficient Bayesian joint modeling of mixed-type multivariate spatial data

[03.E1.C16, (page 42)]

Arghya MUKHERJEE, *Indian Institute of Technology Kanpur*
Arnab HAZRA, *Indian Institute of Technology Kanpur*
Dootika VATS, *Indian Institute of Technology Kanpur*

Multivariate spatial data, where multiple responses are recorded across spatially indexed observational units, are routinely collected in various disciplines. Recent research emphasizes the advantages of jointly modeling mixed-type responses, leading to more robust inference and improved prediction. Joint spatial modeling approaches offer two key benefits: they enable the estimation of associations among variables precluded in univariate models and enhance predictive accuracy by leveraging dependencies across variables. As a result, the residuals from joint models better capture random noise by accounting for multiple sources of dependence. Despite extensive research in spatial statistics, most spatial generalized linear mixed-effect models have been confined to marginal modeling of univariate responses within the exponential family of distributions. While joint modeling offers significant benefits, developing methods for large-scale multivariate spatial data has been limited due to the computational demands of fitting latent multivariate GP models, which require substantial floating point operations and memory. We propose a novel and computationally efficient methodology for jointly modeling mixed-type responses using a latent multivariate GP framework within a Bayesian paradigm. Our approach employs a Markov chain Monte Carlo (MCMC) inference method that utilizes a Metropolis-within-Gibbs sampling strategy with blocking, which enhances mixing and overall model performance. The proposed method offers substantial computational and inferential advantages over fitting separate models for each response type. We illustrate the effectiveness of the proposed method through numerical examples and application on real datasets.

296. Unveiling Bias: A Statistician's Quest for Data Equity in Health Research

[Plenary Lecture 3, (page 62)]

Bhramar MUKHERJEE, *Yale University*

Despite numerous proposed strategies to enhance diversity in scientific research, a significant portion of the world's health research data continues to be derived from populations with historical privileges. In this presentation, I will delve into the crucial concept of data equity and highlight the obvious: algorithms developed on exclusionary datasets can yield erroneous conclusions and exacerbate health disparities. However, while we strive for the ideal scenario of globally representative and extensive datasets, statisticians play a pivotal role in mitigating selection bias and handling missing data—an expertise that few other quantitative disciplines possess.

Drawing from my own journey as a statistician, I will showcase instances where timely statistical analysis with imperfect data resulted in enhanced inference and influenced policy outcomes. Utilizing examples from analyses of COVID-19 and biobanks linked with electronic health records, I will demonstrate how statistical methodologies can navigate the complexities of real-world data to inform decision-making and drive positive change.

In conclusion, I urge statisticians to proactively lead efforts in curating, and collecting new data, launching their own studies while also spearheading innovative scientific inquiries. It is time for statisticians to step out from the fringes of design and analysis and assert their independent leadership in shaping the trajectory of research and policymaking driven by data.

297. statistical theory of deep neural network: estimating composite functions to solving inverse problem via physics informed neural networks

[03.E1.I62, (page 39)]

Debarghya MUKHERJEE, *Boston University*

TBA

298. Empirical Bayes Estimation with Side Information

[03.A1.I60, (page 38)]

Gourab MUKHERJEE, *University of Southern California*
Jiajun LUO,
Trambak BANERJEE,
Wenguang SUN,

We investigate the problem of compound estimation of normal means while accounting for the presence of side information. Leveraging the empirical Bayes framework, we develop a nonparametric integrative Tweedie (NIT) approach that incorporates structural knowledge encoded in multivariate auxiliary data to enhance the precision of compound estimation. Our approach employs convex optimization tools to estimate the gradient of the log-density directly, enabling the incorporation of structural constraints. We conduct theoretical analyses of the asymptotic risk of NIT and establish the rate at which NIT converges to the oracle estimator. As the dimension of the auxiliary data increases, we accurately quantify the improvements in estimation risk and the associated deterioration in convergence rate. The numerical performance of NIT is illustrated through the analysis of both simulated and real data, demonstrating its superiority over existing methods.

299. A feature-based image comparison method

[03.M1.I45, (page 32)]

Partha Sarathi MUKHERJEE, *Indian Statistical Institute, Kolkata*
Anik ROY, *Indian Statistical Institute, Kolkata*
Partha Sarathi MUKHERJEE, *Indian Statistical Institute, Kolkata*

Image comparison is an important step for monitoring images which has wide applications in satellite imaging, medical research, defence applications, and many others. Since the image intensity functions are discontinuous, and the observed images often contain noise, the problem of image comparison is challenging. Most state-of-the-art methods in the literature are intensity-based which is often inappropriate in real life situations where minor changes in the background may not indicate a meaningful change in the images as long as the boundaries of the image objects remain unaltered. In this talk, I will discuss a feature-based image comparison method based on local pixel clustering. Numerical examples and statistical properties show that this method works well in many real life applications.

300. Nuisance Function Tuning for Optimal Doubly Robust Estimation

[03.M2.I50, (page 34)]

Rajarshi MUKHERJEE, *Harvard T.H. Chan School of Public Health*

TBA

301 . Stabilizing Non normal matrices and Davies Conjecture

[03.A1.I58, (page 37)]

Satyaki MUKHERJEE, *National University of Singapore*
Jess BANKS, *University of California, Berkeley*
Archit KULKARNI,
Nikhil SRIVASTAVA, *University of California, Berkeley*

Normal matrices have the property that they are orthogonally diagonalizable. This makes calculating functions on normal matrices remarkably stable. We prove that every matrix is quantitatively close to a normal matrix, a fact conjectured by E.B.Davies. In particular we show that for every matrix there is a close enough matrix whose eigenvector matrix have bounded condition number, where this bound and the aforementioned closeness are polynomially related to each other. Our proof uses tools from random matrix theory to show that the pseudospectrum of A can be regularized with the addition of a complex Gaussian perturbation. Along the way, we explain how a variant of a theorem of Sniady implies a conjecture of Sankar, Spielman and Teng on the optimal constant for smoothed analysis of condition numbers.

302. Relative Efficiency of Estimators in tensor Ising models

[01.M2.I8, (page 6)]

Somabha MUKHERJEE, *National University of Singapore*
Jaesung SON, *Columbia University*
Swarnadip GHOSH, *Radix Trading LLC*
Sourav MUKHERJEE, *Bristol Myers Squibb*

In this talk, I will discuss about two popular parameter estimates in tensor Curie-Weiss models, namely the maximum likelihood and pseudolikelihood (MLE and MPLE) estimates. After briefly talking about their properties and phase transitions, I will discuss about their relative efficiency in the Bahadur sense. We will see that for the classical two spin models, they are equally efficient at all temperatures, but for higher order tensor models, a surprising discrep-

ancy appears when the null inverse temperature lies between the estimation threshold of the tensor model and log(2). Our results are slightly more universal in the sense that they can be extended to dense block Ising models, such as the Erdos-Renyi Ising models.

303. A test for the temporal changepoint detection in spatiotemporal data

[02.A1.I39, (page 25)]

Minerva MUKHOPADHYAY, *IIT Kanpur, ISI Kolkata*
Arnab HAZRA,

Change point detection is a challenging problem, especially when considered in a non-Euclidean setup. In this work, we focus on a spatiotemporal setup, where data are available across sites within a spatial domain at each time point, where the number of sites can be large, and the sites can change across time as well. Our goal is to test if the underlying data-generating process remains the same throughout, or it changes after a certain time point. Traditional tests in the literature for changepoint estimation can either apply to multivariate data, or to univariate functional data. Considering a graph-based test for changepoint detection, we show its applicability to spatial stochastic processes where realizations are available over time. The proposed method is nonparametric and can apply to generic spatial processes without distributional assumptions like Gaussian processes. We validate the efficacy of the proposed method via various simulated datasets. We implement our methodology to the Global Precipitation Climatology Centre (GPCC) gridded annual precipitation dataset available at 2749 pixels of resolution 0.5X0.5 degrees across mainland Australia for 1891–2019.

304. Inferences on Lifetime Performance Index of Weibull Products under Step-Stress Setup

[01.E1.C5, (page 13)]

Sarat Sindhu MUKHOPADHYAY, *Indian Statistical Institute, Bangalore*
Viswakala K.V., *Research Associate, SQC & OR Unit, Indian Statistical Institute, Bangalore*
Gijo E.V., *Faculty member, SQC & OR Unit Indian Statistical Institute, Bangalore*

Tests based on the exponential distribution provide inaccurate assessments of the lifetime performance index (C_L) for products with non-constant hazard rate. To address this issue, a hypothesis testing procedure is proposed for C_L under a hazard rate-based step-stress model with Type II censoring, assuming a two-parameter Weibull distribution that accounts for both increasing and decreasing hazard rates. This procedure can be easily applied to assess whether the lifetime of products meets the requirements. The maximum likelihood (ML) estimator and the uniformly minimum variance unbiased estimator of C_L are derived and compared, highlighting the advantages of the ML estimator. For known shape parameter, hypothesis tests using the ML estimator are presented, whereas for unknown shape parameter, two bootstrap-based techniques are implemented. Additionally, lower bounds for C_L are established. The robustness of the suggested process is shown by validating it on two real-life datasets and a simulated dataset. This highlights the risk of overestimating product lifetime when applying exponential distribution assumptions to data with non-constant hazards.

305 . On Mis-specified Newsvendor Models: A Precision and Complexity Comparison

[02.A1.I37, (page 24)]

Sujay MUKHOTI, *Indian Institute of Management Indore*
Soham GHOSH, *Indian Institute of Technology Indore*

The assumption of completely specified demand distribution is a popular practice in newsvendor literature. Multi-parameter families of distributions provide more realistic demand model in terms of shape and tail fatness, although the computation becomes challenging, especially when the parameters are unknown and have to be estimated from the data. Computational complexity plays a critical role in the big data newsvendor problems, which is becoming increasingly important. In order to elude this computational complexity, a simpler demand distribution like single parameter ones may be used, despite they being unrealistic and possibly mis-specified. In this work we study the impact of distribution mis-specification on the optimal order quantity estimation. The precision-complexity trade-off between the correct and mis-specified newsvendor models is presented using simulation.

306. The Max Chart for Joint Monitoring of the Two Parameters of a Zero-

inflated Poisson(ZIP) Process Using the Folded Normal Distribution

[04.E1.C26, (page 57)]

Subhradeep MUNIYAN, *Indian Statistical Institute*

In the high yield processes, the defects data are usually modelled by the zero inflated Poisson(ZIP) distribution. A ZIP model has two parameters - a zero inflation parameter and a Poisson parameter . It is necessary to monitor both the parameters for the stability of the process. There are some works on the progressive joint monitoring of both the parameters(,), estimated using accumulated samples(i.e., cumulatively combined individual samples over time). But it has led to some critical issues like auto-correlation in the plots over time. Also no closed forms of UCL, -error were derived. In this article, group sampling with large sample size is considered for monitoring of the ZIP process. Here the parameters are estimated only based on the individual samples. So the plots are independent over time and thus the control chart is easier to interpret. Here the folded normal distribution is applied to get the distribution of the charting statistic M. Also the UCL, -error are derived in the closed forms. Thus the chart is more applicable in practical situations. Also in case of a shift in the joint chart, the testing procedure for shift detection in the individual parameters is given, thus making it easier to get to the cause of the shift. Implementation of the chart is demonstrated through an application in one real life example.

307. Length Biased Weighted Ishita Distribution and Its Applications on Real Life Data Sets

[05.M1.C29, (page 60)]

Mustafa MUSTAFA, *Department of Statistics, University of Delhi*
Mustafa MUSTAFA, *Department of Statistics, University of Delhi*

In this investigation, we unveil a fresh extension within the realm of statistical distributions, introducing the captivating "length-biased Ishita distribution." This addition holds a special allure as it falls under the esteemed category of weighted distributions, particularly the length-biased iteration. With a meticulous scrutiny, we delve into the mathematical and statistical complexities of this novel distribution reveal its distinct traits. Employing the robust methodology of maximum likelihood estimation, we adeptly ascertain the model parameters, thereby enriching our comprehension of its behavioural patterns. To showcase the practical utility and merits of the length-biased Ishita distribution, we present its effectiveness through a real-world temporal dataset. Through this empirical exploration, we unravel its superior performance and adaptability, furnishing invaluable insights into its multiple potential applications across diverse domains. Keywords: Ishita distribution, length-biased distribution, new weighted length-biased Ishita distribution, parameter estimation, percentage competency

308. How does the training loss function affect in-distribution and out-of-distribution learning in high dimensional linear models?

[04.M2.I76, (page 48)]

Vidya MUTHUKUMAR, *Georgia Institute of Technology*

TBA

309. Penalized regression with multiple loss functions and variable selection by voting

[03.M2.I53, (page 35)]

Ursula U. MÜLLER, *Texas A&M University*

We consider a sparse linear model with a fixed design matrix in a high dimensional scenario. We introduce a new variable selection procedure called "voting", which combines the results from multiple regression models with different penalized loss functions to select the relevant predictors. A predictor is included in the final model if it receives enough votes, i.e. is selected by most of the individual models. By employing multiple different loss functions our method takes various properties of the error distribution into account. This is in contrast to the standard penalized regression approach, which typically relies on just one criterion. When that single criterion is not met the standard approach is likely to fail, whereas our method is still able to identify the underlying sparse model. Working with the voting procedure reduces the number of predictors that are incorrectly selected, which simplifies the structure and improves the interpretability of the fitted model. We prove model selection consistency and illustrate the advantages of our method numerically using simulated and real data sets.

This talk is based on joint work with Guorong Dai (Fudan University) and Raymond Carroll (Texas

A&M University).

310 . Estimation for parameter-driven count time series

[04.M1.I68, (page 46)]

N. BALAKRISHNA, *Indian Institute of Technology Tirupati*

One way of characterizing a count time series is through its intensity function. The intensity function at any time may be specified as observation-driven or parameter-driven. In this talk, we describe some of the parameter-driven models and discuss their estimation. In particular, we focus on zero-inflated Poisson and zero-inflated negative binomial time series with intensities generated by non-negative Markov sequences. Parameters of the model are estimated by the method of estimating equations which is facilitated by expressing the model in a generalized state space form. The latent intensities required for estimation are extracted using generalized Kalman filter. The applications of the proposed model and its estimation methods are illustrated using simulated and real data sets.

311 . Biomarker Selection in Randomized Clinical Trials with Survival Outcomes: Addressing High-Dimensional and Low Sample Size Challenges

[02.E1.C8, (page 28)]

Vipin N, *Novartis Healthcare Private Limited*
Vipin N,

Identifying biomarkers in randomized clinical trials (RCTs) with survival outcomes is crucial, yet challenging due to the high-dimensional nature of omics data and limited sample sizes. This study employs simulation and real data analysis to evaluate the performance of advanced statistical techniques in variable selection, specifically tailored for high-dimensional data. The methodology incorporates innovative data reduction strategies, such as feature selection and dimensionality reduction, enabling the inclusion of numerous potential biomarkers without compromising statistical power. Additionally, we address bias resulting from the limited sample size. Applying our approach to a real-life RCT dataset, we identify a subset of robust biomarkers associated with survival outcomes, emphasizing their potential in personalized treatment strategies. Comprehensive validation analyses, including resampling methods, affirm the reproducibility and stability of the selected biomarkers, reinforcing their clinical utility. By effectively addressing challenges associated with high-dimensionality and low sample size, our study contributes to precision medicine by facilitating the identification of reliable biomarkers for personalized treatment approaches.

312 . A Multi-objective Economic Statistical Design of the CUSUM chart: NSGA II Approach

[05.M1.C29, (page 60)]

Sandeep , *Indian Statistical Institute, Kolkata*
Arup RANJAN MUKHOPADHYAY,

This paper presents an approach for the economic statistical design of the Cumulative Sum (CUSUM) control chart in a multi-objective optimization framework. The proposed methodology integrates economic considerations with statistical aspects to optimize the design parameters like the sample size (n), sampling interval (h), and decision interval (H) of the CUSUM chart. The Non-dominated Sorting Genetic Algorithm II (NSGA II) is employed to solve the multi-objective optimization problem, aiming to minimize both the average cost per cycle (C_E) and the out-of-control Average Run Length (ARL_δ) simultaneously. The effectiveness of the proposed approach is demonstrated through a numerical example by determining the optimized CUSUM chart parameters using NSGA II. Additionally, sensitivity analysis is conducted to assess the impact of variations in input parameters. The corresponding results indicate that the proposed methodology significantly reduces the expected cost per cycle by about 43% when compared to the findings of the article by M. Lee in the year 2011. No comparison of ARL_δ could be made due to the non-availability of references. This highlights the practical relevance and potential of this study for the right application of the technique of the CUSUM chart for process control purposes in industries.

313. TBA

[Special Session 2, (page 49)]

P. Unnikrishnan NAIR, *Cochin University of Science and Technology*

TBA

314. Community Detection With Contextual Multilayer Networks

[01.A1.I12, (page 7)]

Sagnik NANDY, *University of Chicago*
Zongming MA, *Yale University*

In this talk, we shall study community detection when we observe m sparse networks and a high dimensional covariate matrix, all encoding the same community structure among n subjects. In the asymptotic regime where the number of features p and the number of subjects n grow proportionally, we derive an exact formula of asymptotic minimum mean square error (MMSE) for estimating the common community structure in the balanced two block case using an orchestrated approximate message passing algorithm. The formula implies the necessity of integrating information from multiple data sources. Consequently, it induces a sharp threshold of phase transition between the regime where detection (i.e., weak recovery) is possible and the regime where no procedure performs better than random guess. The asymptotic MMSE depends on the covariate signal-to-noise ratio in a more subtle way than the phase transition threshold. In the special case of m=1 , our asymptotic MMSE formula complements the pioneering work Deshpande et al., (2018) which found the sharp threshold when m=1 . A practical variant of the theoretically justified algorithm with spectral initialization leads to an estimator whose empirical MSEs closely approximate theoretical predictions over simulated examples. This talk based on a joint work with Prof. Zongming Ma.

315. Jackknife empirical likelihood inference of risks measures

[04.M2.I72, (page 47)]

Rupel NARGUNAM, *Madras School of Economics*

TBA

316. "Bayesian clustering with feature selection."

[02.A1.I34, (page 23)]

Leelavati NARLIKAR, *IISER Pune*

TBA

317. A note on Quasi Periodic Gaussian Process

[Student Poster Competition, (page 21)]

Unnati NIGAM, *IIT Bombay, Mumbai, India*
Radhendushka SRIVASTAVA, *IIT Bombay, India*
Michael BURKE, *Monash University, Australia*
Faezeh MARZBANRAD, *Monash University, Australia*

Many biomedical signals exhibit periodic patterns. A natural regression model for this type of signal consists of a periodic mean functional under additive noise. In such scenarios, researchers often model the additive noise as a periodic Gaussian process (PGP) with special covariance structures like Mackay's kernel, Cosine Kernel and Matern's kernel. Typically, the residuals of regression model do not exhibit periodic pattern whereas sample paths of PGP's are periodic. To address this issue, one often models the noise as quasi-periodic Gaussian process (QPGP). There are several ways to define the QPGP in literature. In this work, we describe a structural equation model to describe the QPGP that reduces to a popularly used QPGP defined via covariance structure in the field of biomedical signal processing. The new QPGP model is advantageous to describe the likelihood function leading to an efficient estimation strategy of the parameters. We also present a model based bootstrap procedure to describe the uncertainty of the model parameters.

318. Efficient Designs for Multivariate Crossover Trials

[02.E1.C10, (page 30)]

Shubham Sanjay NIPHADKAR, *Indian Institute of Technology Bombay*
Siuli MUKHOPADHYAY, *Indian Institute of Technology Bombay*

This article aims to study efficient/trace optimal designs for crossover trials with multiple responses recorded from each subject in the time periods. A multivariate fixed effects model is proposed with direct and carryover effects corresponding to the multiple responses. The corresponding error dispersion matrix is chosen to be either of the proportional or the generalized Markov covariance type, permitting the existence of direct and cross-correlations within and between the multiple responses. The corresponding information matrices for direct effects under the two types of dispersions are used to determine efficient designs. The efficiency of orthogonal array designs of Type I and strength 2 is investigated for a wide choice of covariance functions, namely, Mat(0.5), Mat(1.5) and Mat(∞). To motivate these multivariate crossover designs, a gene expression dataset in a 3×3 framework is utilized.

319 . STOCHASTIC MODELING ON RAINFALL VARIABILITY IN

NORTHERN NIGERIA

[03.E1.C14, (page 41)]

John Olutunji OLAOMI, *University of South Africa, Department of Statistics*
James Babatunde EHIMONY, *Department of Statistics, University of South Africa, and Department of Statistics, Kogi State Polytechnic, Lokoja, Nigeria*

This study is aim at exploring the stochastic process to model the distribution of rainfall from some selected stations in Northern region of Nigeria, using continuous-time Markov chain to determine the level of its persistence based on its relative frequencies. The rainfall data were collected from the stations spread across the Sahel, Sudan, and Guinea savannah. An Exponential probability distribution was used to model the distribution of rainfall intensity after clustering the average rainfall experienced in all the stations. The extreme rainfall and the intensity dryness, over the recorded period, across all the stations in the Northern region were observed. We observed that the change in climatic conditions of each station depend on the amount of rainfall experience annually. This study will help in simulating the likely rainfall to be experienced in each Station of the Northern Nigeria. It will assist the Meteorological and Agricultural practitioners to make short-term probabilities prediction for aviation sector and agricultural production. Key words: Continuous – time Markov Chain, Clustering, Rainfall, Exponential probability distribution, Persistence

320 . On some properties and applications of the Modi-Frechet distribution

[04.L1.C19, (page 49)]

AKHILA P, *GOVT. ARTS AND SCIENCE COLLEGE CALICUT, KOZHIKODE, KERALA, INDIA*
GIRISH BABU M, *Govt. Arts and Science College Calicut, Kozhikode, Kerala, India*

This paper discuss a special case of the Modi family, namely, Modi Frechet distribution with four parameters. Some statistical properties like moments, stochastic ordering, and entropy are studied in this paper. The estimation of unknown parameters of the proposed distribution is derived. To study the application and flexibility of the proposed distribution, two real-life data one from the reliability sector and other from the biomedical industry were analyzed. A comprehensive simulation study is also conducted and showed the estimation techniques' accuracy and consistency.

Keywords— Entropy, Frechet distribution, Hazard rate, Maximum likelihood estimation, Modi Family

321 . Inference Procedures on Power Generalized DUS Transformation of Inverse Weibull Distribution

[04.E1.C24, (page 56)]

GAUTHAMI P, *St. Thomas College (Autonomous), Thrissur, University of Calicut, Kerala*
CHACKO V M, *St. Thomas College (Autonomous), Thrissur, University of Calicut, Kerala*

The primary objective of this research work is to introduce and extensively analyze the Power Generalized DUS-Transformed Inverse Weibull (PGDUS-IW) Model, a novel lifetime distribution with significant applicability in reliability theory, particularly for parallel systems. We derived and explored the fundamental statistical properties of the PGDUS-IW Model, including moments, moment generating function (MGF), entropy, extropy and more along with plotting the density and failure-rate function. To estimate the model parameters, we employed Maximum Likelihood (ML) Estimation and Maximum Product Spacing (MPS) Estimation methods, providing detailed procedures and simulation study to validate the estimation techniques. Furthermore, we derived the stress-strength parameters, both single and multi-component under this model and estimated these parameters. We conducted an extensive data analysis and compared the PGDUS-IW model's performance against other competing distributions, demonstrating it's efficiency in fitting various datasets. This comprehensive study, together with introducing the PGDUS-IW model, establishes it's robustness and applicability, providing a valuable tool for researchers in reliability theory and related fields.

322 . Inaccuracy and divergence measures based on survival extropy, their properties and applications in testing and image analysis

[04.E1.C24, (page 55)]

SARANYA P, *COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY*
SUNOJ S M, *COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY*

The article introduces novel measures of inaccuracy and divergence based on survival extropy and their dynamic forms and explores their properties

and applications. To address the drawbacks of asymmetry and range limitations, we introduce two measures: the survival extropy inaccuracy ratio and symmetric divergence measures. The inaccuracy ratio is utilized for the analysis and classification of images. A goodness-of-fit test for the uniform distribution is developed using the survival extropy divergence. Characterizations of the exponential distribution are derived using the dynamic survival extropy inaccuracy and divergence measures. The article also proposes non-parametric estimators for the divergence measures and conducts simulation studies to validate their performance. Finally, it demonstrates the application of symmetric survival extropy divergence in failure time data analysis.

323. Optimal cohort Stepped Wedge design for unequal cluster size

[03.E1.C17, (page 42)]

Soumadeb PAIN, *Indian Institute of Technology Kanpur*
Satya Prakash SINGH,

Recently, there has been a growing interest in designing cluster trials using stepped wedge design (SWD). Stepped wedge randomised trial designs involve sequential roll-out of an intervention to participants (individuals or clusters) over a number of time periods. By the end of the study, all participants will have received the intervention, although the order in which participants receive the intervention is determined at random. However, the Hussey and Hughes model (Hussey and Hughes, 2007) with a single random cluster effect is quite simple and does not consider cohort designs with additional correlation parameters. Recent reviews of stepped wedge cluster randomized trials indicate that cohort designs are common in practice. Girling and Hemming (2016) considered optimal cohort SW designs, but they discussed a larger class of designs that includes hybrid designs having both parallel and stepped wedge components. Li et al. (2018) found out the optimal design for cohort SW design assuming equal cluster size. The clusters will usually differ in size but this is overlooked in much of the existing literature. This poster considers the case when clusters have different sizes and determines how efficient cohort SW designs can be found.

324 . Approximate Bayesian inference for high-resolution spatial disaggregation using alternative data sources

[Student Poster Competition, (page 21)]

Anis PAKRASHI, *PhD Student in Statistics, Department of Statistics, the Pennsylvania State University, USA*
Arnab HAZRA, *Assistant Professor of Statistics, Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, India*
Sooraj M RAVEENDRAN, *Senior Consultant - Urban Informatics Lab, Indian Institute of Human Settlements, Bengaluru, India*
Krishnachandran BALAKRISHNAN, *Lead - Research, Indian Institute of Human Settlements, Bengaluru, India*

This paper addresses the challenge of obtaining precise demographic information at a fine-grained spatial level, a necessity for planning localized public services such as water distribution networks, or understanding local human impacts on the ecosystem. While population sizes are commonly available for large administrative areas, such as wards in India, practical applications often demand knowledge of population density at smaller spatial scales. We explore the integration of alternative data sources, specifically satellite-derived products, including land cover, land use, street density, building heights, vegetation coverage, and drainage density. Using a case study focused on Bangalore City, India, with a ward-level population dataset for 198 wards and satellite-derived sources covering 786,702 pixels at a resolution of 30m×30m, we propose a semiparametric Bayesian spatial regression model for obtaining pixel-level population estimates. Given the high dimensionality of the problem, exact Bayesian inference is deemed impractical; we discuss an approximate Bayesian inference scheme based on the recently proposed max-and-smooth approach, a combination of Laplace approximation and Markov chain Monte Carlo. A simulation study validates the reasonable performance of our inferential approach. Mapping pixel-level estimates to the ward level demonstrates the effectiveness of our method in capturing the spatial distribution of population sizes. While our case study focuses on a demographic application, the methodology developed here readily applies to count-type spatial datasets from various scientific disciplines, where high-resolution alternative data sources are available.

325. Spectra of adjacency and Laplacian matrices of Erdős-Rényi hypergraphs

[Student Paper Competition 2, (page 18)]

Dipranjan PAL, *Indian Statistical Institute, Kolkata*
Soumendu Sundar MUKHERJEE, *Indian Statistical In-*

stitute, Kolkata
Himasish TALUKDAR, *Indian Statistical Institute, Kolkata*

We study adjacency and Laplacian matrices of Erdős-Rényi r-uniform hypergraphs on n vertices with hyperedge inclusion probability p, in the setting where r can vary with n such that $r/n \to c \in [0, 1)$. Adjacency matrices of hypergraphs are contractions of adjacency tensors and their entries exhibit long range correlations. We show that under the Erdős-Rényi model, the expected empirical spectral distribution of an appropriately normalised hypergraph adjacency matrix converges weakly to the semi-circle law with variance $(1 - c)^2$ as long as $\frac{d\_avg}{r^7} \to \infty$, where $d\_avg = \binom{n-1}{r-1}p$. In contrast with the Erdős-Rényi random graph ($r = 2$), two eigenvalues stick out of the bulk of the spectrum. In a Gaussian version of the model we show that an appropriately scaled largest (resp. smallest) eigenvalue converges in probability to $\frac{c}{2} + \sqrt{c - \frac{3c^2}{4}}$ (resp. $\frac{c}{2} - \sqrt{c - \frac{3c^2}{4}}$) and for $c = 0$, a rescaled version converges in probability to 1 (resp. -1). We also establish analogous results for the bulk and edge eigenvalues of the associated Laplacian matrices.

326. Realtime Mortality Prediction of Left Truncated Right Censored (LTRC) MIMIC-IV data Via Piecewise Linear Approximation (PLA) of Hazard Function in Presence of Competing Risks

[02.E1.C7, (page 28)]

Sandip PAL, *S&P Global Inc*
Arnab KOLEY, *Indian Institute of Management, Indore*
Debasis KUNDU, *Indian Institute of Technology, Kanpur*
Gayatri PAL, *Tesco India*

In recent years, real-time health monitoring and collecting health records have become important phenomenon in medical research. Electronic Health Record (EHR) systems provide rich, valuable and high-frequency critical data on real-time which can be used to analyze and predict the possible health hazards for patients. In this article, we leverage the MIMIC-IV (Medical Information Mart for Intensive Care) database to develop a robust model for mortality prediction in critically ill patients. The unique challenges of left-truncated and right-censored data are addressed, which is vital given that patients often enter in the study period after the onset of their critical condition and have different length of observation periods. Additionally, the presence of competing risks is incorporated into the predictive model to enhance its accuracy and applicability. Our approach ensures that the time-dependent nature of patient data and the multi-state outcomes are accurately captured and analysed. Key variables from the MIMIC-IV database, including demographic information, vital signs, laboratory results, and clinical interventions, are utilized to demonstrate that the model significantly improves mortality prediction accuracy compared to traditional methods that do not account for left truncation and competing risks. This improvement is crucial for clinical decision-making, allowing healthcare providers to identify high-risk patients promptly and tailor interventions accordingly. The proposed approach into mortality prediction models using MIMIC-IV data presents a more comprehensive approach to understanding patient outcomes in critical care settings. The findings underscore the importance of considering the full complexity of patient data to enhance not only predictive accuracy but also support better clinical management strategies.

Keywords: MIMIC-IV, Left-truncated Right-censored data, Competing risks, Likelihood function, Cox-Proportional Hazard Model

327. Novel Machine Learning-Based Cure Models

[02.M1.I24, (page 16)]

Suvra PAL, *University of Texas at Arlington*

I will start the talk by giving a brief overview of cure models and discuss few applications. I will present a real data that motivates the need for cure models, as opposed to the standard survival models. Next, I will present a novel cure model that uses the support vector machine (SVM) to model the incidence. The proposed model provides flexibility in capturing non-linearity in the data. For the estimation of model parameters, I will make use of a computationally efficient expectation maximization (EM) algorithm. Next, I will present results of a detailed simulation study. Finally, I will analyze a real cancer survival data and show that the proposed SVM-based model results in improved predictive accuracy for cure. I will end my talk by discussing some interesting ongoing and future research works.

328. Le Cam meets Turing: Computationally efficient reductions between some statistical models

[04.M1.I65, (page 45)]
Ashwin PANANJADY, *Georgia Institute of Technology*

TBA

329 . Joint Modeliing of Longitudinal PSA Levels and Survival Data on Prostate Cancer with Missing Grades

[02.A1.I38, (page 25)]
Mahaveer Singh PANWAR, *Banaras Hindu University, Varanasi*
Vikas BARNWAL, *Department of Biostatistics, Boston University, USA*

Joint modelling of longitudinal measurements and time-to-event data has received considerable attention, especially in the field of public health studies. In this work, we jointly analyze the longitudinal process and competing risks event process with a linear mixed effect model and cause-specific hazard model, respectively. The two processes are associated with the shared random effect approach. In the competing risks data, the cause of the event is not known always, leading to incomplete data with respect to the cause. The cause of events for such individuals is said to be masked. We assume that the masking is not independent of the causes and hence, our proposed joint model also deals with cause-dependent masking situations in competing risks data. The estimation of model parameters is carried out under the Bayesian paradigm as it is computationally more flexible against the model complexity. An extensive numerical study is performed to evaluate the efficacy of estimators obtained under the joint model. The established methodology subsuming the dependent masking in competing risks data is illustrated with the Prostate, Lung, Colorectal, Ovarian (PLCO) cancer prevention trial dataset.

330 . Current Account Deficit Sustainability: The Indian Case

[01.A1.I14, (page 8)]
Savita PAREEK, *Auburn University*

TBA

331 . A Feasibility Study to assess the adequacy of historical clinical trial data using an external control to inform the Development of an Innovative treatment for a rare disease

[01.M2.I3, (page 4)]
Rashidkhan PATHAN, *Novartis Healthcare Pvt. Ltd.*
Gerhild ANGYALOSI , *Novartis AG Basel*
Kirsten CARTER , *Novartis AG Basel*
Rima IZEM, *Novartis AG Basel*

To address unmet medical need in a rare disease, a clinical study comparing the new treatment to the current standard of care is under discussion. Due to reduced number of cases in recent years there may be a challenge to recruit the adequate number of patients in the study. To address this, we aspire to incorporate external control data to accelerate evidence generation. By integrating individual patient data from this external control with the clinical trial data, the non-inferiority of a new treatment can be determined more precisely. The research objectives include: - Identification of subject-level data with patient characteristics to inform heterogeneity - Evaluation of feasibility of including external data to supplement internal control - Implementation framework including developing appropriate statistical methods

Methods: The methodology will proceed in two stages, an exploratory stage to evaluate the feasibility of using the external data, and an analytical stage to leverage external data with clinical trial data. A target trial emulation framework will be applied to select patients from the external control to limit the selection bias. Summaries of baseline characteristics and comparison to the same characteristic in ongoing study for various parameters will be presented. Balance in each characteristic will also be implemented using standardized mean differences. To explore balance in covariates across all baseline characteristics, a propensity score model using a logistic regression including all the potential confounders will be fitted. Bayesian borrowing methods such as propensity-based meta-analytic predictive prior will be considered to augment an external control arm data.

332. Robust Probabilistic Inference via a Constrained Transport Metric

[04.M2.I70, (page 47)]
Debdeep PATI, *University of Wisconsin-Madison*
Abhisek CHAKRABORTY, *Eli Lilly and Company*
Anirban BHATTACHARYA, *Texas A&M University*

Flexible Bayesian models are typically constructed using limits of large parametric models with a multitude of parameters that are often uninterpretable. In this talk, we offer a get-around by proposing an exponentially tilted empirical likelihood

carefully designed to concentrate near a simpler parametric family of distributions of choice with respect to a novel variant of the Wasserstein metric. It finds applications in a wide variety of robust inference problems, where we intend to make inference on the parameters associated with the centering distribution in presence of outliers. Interestingly, we demonstrate equivalence of our approach with a nonparametric Bayesian formulation under a suitable asymptotic framework, testifying to its flexibility. The constrained entropy maximization setup that sits at the heart of our likelihood formulation finds its utility beyond robust Bayesian inference; illustrations are provided in the field of demographic parity in machine learning.

333. Spectra of sequential sample covariance matrices with application to detecting structural changes

[Special Invited Session 3, (page 44)]

Debashis PAUL, *Indian Statistical Institute, Kolkata*
Nina DÖRNEMANN, *Aarhus University*

Suppose we have i.i.d. observations, indexed by time, from a multivariate distribution. The corresponding sequential sample covariance matrices, defined as the sample covariance matrix computed from observations up to time t, form a matrix-valued stochastic process indexed by time. We study the spectral statistics of such matrices as stochastic processes, when the dimension of the data vectors is comparable to the sample size. First results of this kind have recently been established by Dörnemann and Dette (2023), who derived the asymptotic distribution of linear spectral statistics associated with the sequential covariance matrices. We study a related problem, that of characterizing the behavior of normalized spiked eigenvalues of the sequential covariances when the population covariance matrix is a finite rank perturbation of the identity matrix, and the spiked eigenvalues are sufficiently big. We establish the existence of a limiting process associated with the normalized eigenvalues. We also consider an application of the results to constructing rotation-invariant tests for detecting change points in the structure of the population covariance matrix when such changes only affect the sizes of the spiked eigenvalues.

334. Inference for Median and a Generalization of HulC

[02.E1.C9, (page 29)]

Manit PAUL, *Wharton School, University of Pennsylvania*
Arun Kumar KUCHIBHOTLA, *Carnegie Mellon University*

Constructing distribution-free confidence intervals for the median, a classic problem in statistics, has seen numerous solutions in the literature. While coverage validity has received ample attention, less has been explored about interval width. Our study breaks new ground by investigating the width of these intervals under non-standard assumptions. Surprisingly, we find that properly scaled, the interval width converges to a non-degenerate random variable, unlike traditional intervals. We also adapt our findings for constructing improved confidence intervals for general parameters, enhancing the existing HulC procedure. These advances provide practitioners with more robust tools for data analysis, reducing the need for strict distributional assumptions.

335. Study on fractional order entropy from a decision-making perspective

[03.E1.C14, (page 41)]

Poulami PAUL, *Rajiv Gandhi Institute of Petroleum Technology, Jais*

The fractional order entropy proposed by Ubriaco for discrete random variable was designed with a specific objective and motivation. Our primary goal is to investigate the utility of this entropy in studying uncertainty due to randomness in the level of risk tolerance of individuals engaged in real world decision-making process. To this aim, we carry out an exhaustive study of its properties and examine the influence of variations in human risk tolerance towards uncertainty through adjustment of the fractional parameter alpha. By choosing the values of alpha close to one, high risk tolerance of individuals is represented whereas the risk averse nature is portrayed through the values of alpha closer to zero. In order to validate our theoretical investigations, we have used this entropy as a portfolio selection criterion for large-cap, mid-cap and diversified portfolios.

336. Nonparametric estimation of residual extropy function under length-biased sampling.

[04.E1.C27, (page 57)]

Vaishnavi PAVITHRADAS, *Department of Statistics, Cochin University of Science and Technology, Kochi*
Rajesh G, *Department of Statistics, Cochin University of*

Science and Technology, Kochi
Richu RAJESH, *Department of Statistics, Government Victoria College, Palakkad*

This article presents nonparametric estimators for residual extropy under length-biased sampling. It validates the consistency and asymptotic normality of the proposed estimators under suitable regularity conditions for large sample sizes. The performance of the estimators is evaluated using simulated observations and compared based on root mean squared errors across different sample sizes. Furthermore, a confidence interval for the nonparametric kernel estimator is suggested and its performance is evaluated through simulation studies. Finally, a real data application is conducted to demonstrate the usefulness of the estimators.

337. Regularized empirical likelihood for Bayesian inference: theory and applications

[01.A1.I10, (page 7)]

Mario PERUGGIA, *The Ohio State University*
Eunseop KIM, *The Ohio State University*
Steven MACEACHERN, *The Ohio State University*

Bayesian inference with empirical likelihood faces a challenge as the posterior domain is a proper subset of the original parameter space due to the convex hull constraint. We propose a regularized exponentially tilted empirical likelihood to address this issue. Our method removes the convex hull constraint using a novel regularization technique, incorporating a continuous exponential family distribution to satisfy a Kullback-Leibler divergence criterion. The regularization arises as a limiting procedure where pseudo-data are added to the formulation of an exponentially tilted empirical likelihood in a disciplined way. We show that this regularized exponentially tilted empirical likelihood retains certain desirable asymptotic properties of (exponentially tilted) empirical likelihood with improved finite sample performance. Simulations and data analysis demonstrate that the proposed method provides a suitable pseudo-likelihood for Bayesian inference.

338. On On higher order approximation of Bayesian procedures through empirical Bayes

[04.M2.I70, (page 46)]

Sonia PETRONE, *Bocconi University, Milan. Italy*
Stefano RIZZELLI, *University of Padova*
Judith ROUSSEAU, *CEREMADE, Université Paris Dauphine and Oxford University*

Bayesian procedures are often optimal but analytically complex. Moreover, even when the prior law is carefully specified, it may be delicate to fix the prior hyperparameters, so that, to bypass further computational issues, it is tempting to fix them from the data, obtaining a so-called empirical Bayes posterior distribution. Although questionable, this is a common practice. We show that, in regular cases, the EB posterior distribution gives a fast approximation of an oracle Bayesian posterior (within the given class of priors). This is faster than Bernstein-von Mises Gaussian approximations. We however also disentangle cases where asymptotic agreement fails.

339. Generalized inaccuracy measure of order α in order statistics

[02.E1.C6, (page 27)]

TINCY PHILIP, *COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY*
RAJESH G, *COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY*

In this paper, we introduce a measure based on generalized inaccuracy measure of order α between distributions of the ith order statistic and the parent random variable. This measure uniquely characterizes the distribution function of the parent random variable . We show that the generalized inaccuracy measure of order α is invariant under scale transformations but not under location transformations. Additionally, we derive the exponential mean value and establish a lower bound for the proposed measure. Nonparametric estimators for this measure are developed, and their performance is evaluated through a Monte Carlo simulation study. The simulation results show that the estimator using the reflection boundary technique for probability density function estimation and the empirical method for cumulative distribution function estimation, yields the best performance. Furthermore, we apply these estimators to a real dataset to demonstrate their practical utility.

340 . An autoencoder based semi-supervised approach for accurate binning of metagenomics data

[04.E1.C26, (page 57)]

Deeksha P M, *ICAR-Indian Agricultural Statistics Re-*

search Institute, New Delhi

Microorganisms thrive in diverse environments, from soil to extreme locations like hot springs and glaciers, showcasing remarkable metabolic and physiological diversity. Microbes, too small to see with the naked eye, have been traditionally studied through culture-based methods. However, traditional culture-based methods have limited our understanding of these organisms, as most remain uncultured. Metagenomics, a culture-independent approach, has revolutionized microbial research by directly analyzing genetic material from environmental samples, revealing genetic diversity, functional potential, and ecological interactions of entire microbial communities. Despite advancements, metagenomic data analysis faces challenges, particularly in clustering and assembling genomes from diverse organisms. This study introduces a novel approach using deep learning to enhance clustering in metagenomic data through the K-means algorithm. The study explores the use of single-copy phylogenetic marker genes for accurate species delineation. A semi-supervised clustering approach was applied to a metagenome dataset, initially reducing the high-dimensional feature matrix using an autoencoder. The produced clusters were analyzed using marker genes data identified in the contigs by running Prodigal, FetchMG and USEARCH sequentially. K-means clustering analysis revealed 8 clusters with a Rand index of 0.85 and an F1 score of 0.61. Post-correction, 10 clusters were identified, significantly improving clustering quality with overall accuracy of 0.62. These findings highlight the importance of rigorous cluster correction in metagenomic analysis, enhancing the precision and significance of clustering outcomes.

341. GCP Erlang Queues

[04.E1.C28, (page 58)]

Rohini Bhagwanrao POTE, *Indian Institute of Technology Bhilai*
Kuldeep Kumar KATARIA, *Indian Institute of Technology*

We introduce and study a queue with Erlang service system and whose arrivals are governed by a counting process in which there is a possibility of more than one arrival at any instant. We call it the generalized counting process (GCP) Erlang queue. We derive a system of differential equations for its transient probabilities. Its probability generating function is obtained from which the explicit expressions of its transient probabilities are derived. Also, the probability of zero customer at any instant is derived. Further, we define queue length for GCP Erlang queue and obtain a system of differential equations for its mean queue length. We derive the explicit expression for the mean queue length and obtain its Laplace transform.

342 . The Reversed Aging Intensity Function- A Quantile Approach

[03.E1.C13, (page 40)]

ANJALI P R, *Research Scholar, Department of Statistics, University of Calicut*
Dileepkumar M, *Assistant Professor, Department of Statistics, University of Calicut*

Reversed aging intensity functions, introduced and studied by Majid Rezaei, and Vahideh Alhrari Khalaf (2014) has several applications in the context of reliability theory. For models defined through quantile functions having no closed form distribution function, we cannot apply this measure in its present form. This calls for an analogues definition of reversed aging intensity function in quantile set up. The present study introduces Quantile based Reversed Aging Intensity (QRAI) function and establish various properties of it. Aging and ordering properties based on the proposed measure are also discussed.

343. A quantile-based bivariate distribution

[02.E1.C6, (page 27)]

Shifna P R, *Cochin University of Science and Technology*
N. Unnikrishnan NAIR, *Cochin University of Science and Technology*
S. M. SUNOJ, *Cochin University of Science and Technology*

In this paper we present a flexible bivariate distribution specified by a quantile function. The distribution contains as special cases new bivariate exponential, Pareto I, Pareto II, beta, power, log logistic and uniform distributions and also can approximate many other continuous models. Various L-moment based properties of the distribution such as covariance, coskewness, cokurtosis, L-correlation, etc are discussed. The distribution is used to model two real data sets.

344 . Towards Strategic Management for Controlling Invasive Alien Plants in India: Integrating Ensemble Based Species Distribution Models

[03.E1.C16, (page 42)]

Jyoti J PRAJAPATI, *Department of Mathematics, Institute of Chemical Technology, Mumbai*

The study provides a thorough national-scale assessment, rather than concentrating on particular region, by encompassing all 218 plant species that have been classified as invasive aliens throughout India. The extensive coverage enables a generalized comprehension of the dynamics of invasive species dynamics in diverse ecological zones. A more nuanced knowledge of invasive species dynamics is provided by the thorough examination of invasive alien plant species in India through the integration of ensemble species distribution models, their characteristics, and growth patterns. Strategic mitigation activities guided by the forecasts under climatic scenarios (RCP 4.5 and RCP 6.0 for 2050 and 2070), which make easier to identify possible invasive hotspots. The study provides a comprehensive understanding of how different characteristics affect species expansion, which is frequently not addressed in isolation, by combining projections under each category. It allows for the early identification of possible invasive hotspots and guide strategic mitigation actions. The emphasis on comparing initial and ongoing area changes deepens our understanding of the expansion patterns and trajectories of invasive species holistically. Identifying high-risk locations and species rapid expansion facilitates the more efficient allocation of resources and the prioritization of management actions.

345. Inference on progressive-stress accelerated life testing under generalized adaptive progressive hybrid censoring

[02.E1.C7, (page 28)]

Aman PRAKASH, *NIT Surat, Gujarat (395007)*
Aman PRAKASH, *NIT Surat*
Raj Kamal MAURYA, *NIT Surat*

The accelerated life testing (ALT) approach has been used in several fields to obtain failure time information faster than testing under normal operating conditions. In this talk, we have considered that the lifetime of test units follows an inverted exponential Rayleigh distribution (IERD). The cumulative exposure model has been considered when test units have an IERD lifetime and the scale parameter of the distribution obeys the inverse power law. In the classical framework, maximum likelihood estimation (MLE) and maximum product spacing estimation (MPSE) are used for point estimates. Asymptotic confidence intervals have been constructed using MLE estimates. Further, in the Bayesian framework, point estimates have been obtained using the Metropolis-Hasting algorithm under informative, non-informative, and kernel prior. Additionally, we constructed the credible intervals for the model parameters using Bayes estimates under informative and non-informative prior. The proposed estimation methods are compared using Monte Carlo simulations. Moreover, simulation studies are also used to compare two different designs of progressive-stress ALTs. The first design is a simple ramp-stress test with two stress levels. Secondly, a multiple ramp-stress test with three and four stress levels. Finally, to validate our work in real-life scenarios, one actual data set is analyzed to determine the applicability of the proposed model.

Keywords: Generalized adaptive progressive hybrid censoring (GAPHC); Accelerated life testing (ALT); Maximum likelihood estimation (MLE); Maximum product spacing estimation (MPSE); Kernal prior (KP).

346 . Improving the Adaptive Gauss-Hermite Quadrature approach for Approximate Bayesian Inference

[Student Poster Competition, (page 20)]

Priyanka PRIYANKA, *Indian Institute of Technology Hyderabad*
Patrick BROWN, *University of Toronto*
Sayantee JANA, *Indian Institute of Technology Hyderabad*

Laplace approximation is more commonly used approximated inference method when the exact inference is infeasible due to the large datasets or model complexity. In this study we propose two methods relatively more reliable, based on the Laplace approximation used for marginals by Integrated Nested Laplace ap- proximation (INLA) and quadrature technique to obtain the joint posterior mean. The accuracy of the proposed methods is demonstrated through two real exam- ples based on the Latent Gaussian Models (LGMs) and Extended Latent Gaussian Models (ELGMs) including both the fixed and the unknown hyper-parameters. We provide computational benefit of the proposed methods over Markov Chain Monte Carlo (MCMC) and compare the results with INLA and Adaptive Gauss-Hermite

Quadrature (AGHQ).

347. Inference on the Polychoric Correlation using the Density Power Divergence

[04.A1.C21, (page 53)]

Arijit PYNE, *Novartis*
Abhik GHOSH, *Indian Statistical Institute, Kolkata*
Ayanendranath BASU, *Indian Statistical Institute, Kolkata*

In the context of polychoric correlation, estimators and test statistics based on the likelihood function are known to be vulnerable to outliers that defy the model assumptions. This poses a serious challenge, because, in practice, we do not have any prior information about the underlying model that generates data. To circumvent this problem, we adopt the density power divergence (DPD) in our setup and derive a class of estimators that become a robust alternative to the maximum likelihood estimators (MLE). We also develop a robust version of the Wald test statistic. Asymptotic properties of these estimators and tests are rigorously developed. Also, their stability properties are studied through the influence function and breakdown point analysis. Extensive simulation studies substantiate the theoretical results. Using the DPD instead of the usual likelihood function has myriad benefits– the performance of the proposed estimators and the test statistics closely match those based on the MLE for pure data; however, they become far more stable under different types of data contamination. Finally, we wrap up this article with an application on a real-life data set providing further assertion of the validity of the theoretical results established in this paper.

348. Reliability properties and applications of proportional reversed hazards in reversed relevation transform

[01.E1.C2, (page 11)]

Anand R, *University of Calicut*
Dileepkumar M, *University of Calicut*
Sankaran P.G, *Cochin Univeristy of Science and Technology*

The concept of reversed relevation transform was introduced by Di Crescenzo and Toomaj (2015). In this article, we study important reliability properties of reversed relevation transform under the proportional reversed hazards assumption. Certain results regarding information measures are presented. Various ageing concepts and stochastic orders are discussed. A new flexible generalization of the Fréchet distribution is introduced using the proposed transformation, and reliability properties and applications of the same are discussed.

349. Clustering high-dimensional data

[03.M2.I53, (page 35)]

Shahina RAHMAN, *Amazon LLC*

TBA

350. Application of personalization in Enhancing customer engagement

[05.M1.I89, (page 59)]

Saurabh RAJ, *Google*

In the current era of rapid innovations & stiff competition; where users are overwhelmed by options, businesses need to look beyond the generic offerings and personalize every step of the customer journey, from advertising to in-app experience. This becomes far more critical for businesses or apps that provide multiple products and services. This talk will explore the applications of personalization on in-app contents in order to enhance customer engagement. Additionally, it also covers techniques like Thompson Sampling to achieve personalization. It concludes with some challenges in implementing this technique as well as recommendations for overcoming these challenges.

351. Discriminating between Weibull and log-normal distributions in the presence of hybrid censoring

[05.M1.C29, (page 61)]

Ojasvi RAJPUT, *IIT kanpur*
Ojasvi RAJPUT, *IIT kanpur*
Debasis KUNDU, *IIT Kanpur*
Sharmishtha MITRA, *IIT Kanpur*

Weibull and log-normal distributions are the two most commonly used distributions for analyzing lifetime data. Both the distributions share number of interesting properties, and for a certain range of parameters their cumulative distributions functions can also be similar in nature. However, we will see why selecting a more appropriate distribution is an important issue. Further, when the data are observed in the presence of some censoring scheme the problem becomes more challenging. It is assumed that the data are coming either from log-normal or Weibull distributions and they are hybrid censored(both Type-I

and Type-II hybrid censoring has been considered). We use the difference of the maximized log-likelihood functions, in discriminating between the two distribution functions. We have also used Bayesian decision criterion for the above problem. We obtain the asymptotic distribution of the ratio of maximized likelihood(RML) statistic. It is used to determine the probability of correct selection and minimum sample size determination in this discrimination process. We perform some simulation studies to observe how the asymptotic results work for different sample sizes, for different degree of censoring and different choices of censoring time. It is observed that the asymptotic results work quite well even for moderate sample sizes. Along with that we have also studied the effect of model misspecification on p^{th} quantile, mean residual life and prediction for future failures. A real data set is analyzed for illustrative purpose

352 . Applications of Dynamic Linear Models in Marketing

[02.M2.I27, (page 17)]

Balaji RAMAN, *Cogitaas*

TBA

353 . The robust desparsified lasso and the focused information criterion for high-dimensional generalized linear models

[01.M2.I9, (page 6)]

T. V. RAMANATHAN, *Savitribai Phule Pune University & Plaksha University*

The classical lasso estimation for sparse, high-dimensional regression models is typically biased and lacks the oracle properties. The desparsified versions of the lasso have been proposed in the literature that attempt to overcome these drawbacks. In this paper, we propose the outliers-robust version of the desparsified lasso for high dimensional generalized linear models. The robustness, consistency and high dimensional asymptotics are investigated rigorously in a general framework of M-estimation under potential model misspecification. The desparsification mechanism is subsequently utilized to construct the focused information criterion (FIC) thereby facilitating robust, focused model selection in high dimensions. The applications are demonstrated with the Poisson regression under robust quasilikelihood loss function. The empirical performance of the proposed methods is examined via simulations and a real data example

354. Bayesian Variable Selection in Survival Regression using Gaussian Process Prior

[04.E1.C25, (page 56)]

Rakesh RANJAN, *Banaras Hindu University*
Sourabh BHATTACHARYA, *Indian Statistical Institute*

This paper considers the Bayesian variable selection in the nonparametric survival regression model when the lifetime follows the Weibull distribution. The variable selection in the model is derived using a Gaussian process prior. The model comparison uses the posterior inclusion probability obtained from the Trans-dimensional Transformation-based Markov Chain Monte Carlo (TTMCMC) output. Theoretical results for the proposed methodology have also been derived. Numerical illustrations are provided for simulated and real datasets. Real data illustrates the selection of genes that can predict the survival of uterine-serous carcinoma patients.

355 . A FUN PROBLEM FROM MATHEMATICAL STATISTICS

[02.A1.I35, (page 24)]

Marepalli RAO, *University of Cincinnati*
Neelakshi CHATTERJEE, *University of Cincinnati*
Zhaochong YU, *University of Cincinnati*
Shesh RAI, *University of Cincinnati*

The focus is on dissecting probabilistically and statistically a British TV Game. Three contestants are on the stage. Each one chooses an integer from 0 to 10000 independently. Some one will win if the chosen numbers are distinct. A partcipant wins if his chosen number is in between the other two. The winning amount is the chosen number in British pounds. What is the best strategy in selecting a number for the partcipants? A continuus analogue provides intriguing and challenging problems

356. Exact MCMC for Bayesian Inference for Privatized Data

[04.M1.I63, (page 44)]

Vinayak RAO, *Purdue University*
Nianqiao JU, *Purdue University*
Jordan AWAN, *Purdue University*
Ruobin GONG, *Rutgers University*

Differentially private mechanisms protect privacy by introducing additional randomness into the data. Restricting access to only the privatized data makes

it challenging to perform valid statistical inference on parameters underlying the confidential data. Specifically, the likelihood function of the privatized data requires integrating over the large space of confidential databases and is typically intractable. For Bayesian analysis, this results in a posterior distribution that is doubly intractable, rendering traditional MCMC techniques inapplicable. We propose an MCMC framework to perform Bayesian inference from the privatized data, which is applicable to a wide range of statistical models and privacy mechanisms. Our MCMC algorithm augments the model parameters with the unobserved confidential data, and alternately updates each one conditional on the other. For the potentially challenging step of updating the confidential data, we propose a generic approach that exploits the privacy guarantee of the mechanism to ensure efficiency. In particular, we give results on the computational complexity, acceptance rate, and mixing properties of our MCMC. Time permitting, I will also discuss some extensions to Dirichlet Process mixture models. This is joint work with Nianqiao Ju and Jordan Awan from Purdue, Ruobin Gong from Rutgers and Mario Beraha and Stefano Favaro from the University of Turin.

357. Regularization Schemes for Statistical Inverse Problems

[04.M2.I69, (page 46)]

Abhishake RASTOGI, *LUT university, Finland*

TBA

358. LLM based approach for binary classification in peptide bioactivity

[Student Poster Competition, (page 20)]

Ravi , *Bioinformatics Division, ICAR- Indian Agricultural Statistics Research Institute*

Predicting potent peptides through computational models can significantly cut down on the need for experimental lab work. Yet, traditional model development often involves complex and time-consuming processes due to issues like peptide representation, feature selection, model selection, and tuning hyper parameters. In this study, an ESM2 pre-trained embedding model was used for feature extraction, and later. Various machine leaning models like Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost), on ten different BPs datasets available on GitHub dzjxzyd/UniDL4BioPep: webserver, were used for prediction of bioactivity. It is observed from this study that machine learning, deep learning and large language models are emerging as potential future technologies in the area of computational biology.

359. Block Design for Two-level factorial experiments in block size four

[02.E1.C10, (page 30)]

Anurag RAWAT, *ICAR-IASRI*
Sukanta DASH, *Senior Scientist*

In experimental scenarios characterized by one source of heterogeneity within the experimental material, block designs offer significant value. Exploring the optimal number of replicates required for factorial experiments, conducted in blocks of size four has garnered significant attention among researchers. While experiments in blocks of size two have been extensively studied, there is growing recognition that experiments in blocks of size four might offer greater utility in practical applications. Particularly, when estimating main effects and specific two-factor interactions from two-level factorial experiments conducted within blocks, a considerable number of replicates may be necessary. This article delves into the exploration of designs that minimize the required number of replicates for factorial experiments conducted in blocks of size four. The article presents methodologies aimed at obtaining such designs, which hold promise for enhancing the efficiency and effectiveness of experimental investigations.

360. PACE : Privacy Aware Collaborative Estimation for Heterogeneous GLMs

[Student Poster Competition, (page 20)]

Bhaskar RAY, *North Carolina State University*
Srijan SENGUPTA, *North Carolina State University*
Aritra MITRA, *North Carolina State University*

Statistical research with sensitive data collected across various sites is often hindered by data sharing restrictions. Federated Averaging (FedAvg) was one of the first introduced methods to perform collaborative estimation in such settings. However, collaborative algorithms fitting a global model can lead to erroneous parameter estimation or fail to converge under model heterogeneity across sites. We propose a novel method of parameter estimation for a broad class of Generalized Linear Models with sites obtain-

ing data based on the same distribution, the true parameter values varying across clusters. Our method accounts for the uncertainty in the local empirical risk minimizer and that in the optimization algorithm iterates and leverages established concentration inequalities to provide non-asymptotic risk bounds. We conduct a hypothesis test-type clustering based on one-shot estimation and utilize the inference to conduct a decentralized collaborative estimation, provably improving upon local estimation with high probability. Besides, we prove asymptotic accuracy of the clustering algorithm and the consistency of the estimates, validating our results with simulation studies.

361. TBA

[Special Invited Session 2, (page 31)]

Debashree RAY, *Johns Hopkins University*

362. Ising inverse critical temperature for preferential attachment models

[02.A1.I40, (page 26)]

Rounak RAY, *TU Eindhoven*

TBA

363. Insufficient Gibbs sampling

[04.A1.I80, (page 51)]

Christian ROBERT, *Université Paris Dauphine PSL & University of Warwick*
Antoine LUCIANO, *Université Paris Dauphine*
Robin RYDER, *Université Paris Dauphine*

In some applied scenarios, the availability of complete data is restricted, often due to privacy concerns; only aggregated, robust and inefficient statistics derived from the data are made accessible. These robust statistics are not sufficient, but they demonstrate reduced sensitivity to outliers and offer enhanced data protection due to their higher breakdown point. We consider a parametric framework and propose a method to sample from the posterior distribution of parameters conditioned on various robust and inefficient statistics: specifically, the pairs (median, MAD) or (median, IQR), or a collection of quantiles. Our approach leverages a Gibbs sampler and simulates latent augmented data, which facilitates simulation from the posterior distribution of parameters belonging to specific families of distributions. A by-product of these samples from the joint posterior distribution of parameters and data given the observed statistics is that we can estimate Bayes factors based on observed statistics via bridge sampling. We validate and outline the limitations of the proposed methods through toy examples and an application to real-world income data.

364 . Stochastic Optimization Algorithms for Instrumental Variable Regression with Streaming Data

[02.M2.I28, (page 18)]

Abhishek ROY, *Texas A&M University*
Xuxing CHEN, *University of California Davis*
Yifan HU, *College of Management of Technology, EPFL, Department of Computer Science, ETH Zurich, Switzerland*
Krishnakumar BALASUBRAMANIAN, *University of California Davis*

We develop and analyze algorithms for instrumental variable regression by viewing the problem as a conditional stochastic optimization problem. In the context of least-squares instrumental variable regression, our algorithms neither require matrix inversions nor mini-batches and provides a fully online approach for performing instrumental variable regression with streaming data. When the true model is linear, we derive rates of convergence in expectation, that are of order O(log T/T) and O($1/T^{1a}$) for any a $>$ 0, respectively under the availability of two-sample and one-sample oracles, respectively, where T is the number of iterations. Importantly, under the availability of the two-sample oracle, our procedure avoids explicitly modeling and estimating the relationship between confounder and the instrumental variables, demonstrating the benefit of the proposed approach over recent works based on reformulating the problem as minimax optimization problems. Numerical experiments are provided to corroborate the theoretical results.

365. Detecting localized dependence in bivariate datasets

[04.M1.I66, (page 45)]

Angshuman ROY, *Indian Institute of Technology Tirupati*
Angshuman ROY, *Indian Institute of Technology*
Sagnik DAS, *Indian Institute of Science Education and Research Kolkata*

In this talk, the problem of testing independence between two random variables is considered. We approach the problem by examining neighborhoods of different sizes within the dataset and then combining

all the results. We show that this method is especially effective for identifying functional dependence and more generally for detecting localized dependence patterns. We propose a general testing framework where the power of an existing test of independence can be improved for detecting functional/ localized dependence. A special test method is proposed which is shown powerful for detecting functional dependence between two continuous random variables. We discuss the consistency of these method under appropriate assumptions. We compare our method with existing method on several simulated and real data examples.

366. Rotation and Translation Invariant Monitoring of Shape and Size in Image Data: An Application in Satellite Imaging

[03.E1.C17, (page 42)]

Anik ROY, *Indian Statistical Institute, Kolkata*
Partha Sarathi MUKHERJEE, *Indian Statistical Institute, Kolkata*

Sequential monitoring of the shape and size of an image object is an emerging research area in statistics and machine learning which has applications in satellite imaging, medical diagnostics, manufacturing industries, and so forth. The shape and size monitoring of an image object is challenging mainly because the images are often not geometrically aligned, and hence, the problem of rigid-body image registration is also inherently associated with it. Most methods in the literature require image registration as a preprocessing step. In this paper, we propose a shape and size monitoring method which is invariant to rotation and translation of the image object, and thus capable of disregarding the changes due to rigid-body image transformation. For comparing the boundaries of an image object in two images, we construct a test statistic based on the distribution of the distances from the centroid of the image object to its boundary. For online monitoring, we propose a CUSUM control chart based on that statistic. The primary advantage of the proposed method is that the complications associated with the performance of image registration do not arise. Theoretical justifications and numerical studies show that the proposed method works well in many practical applications.

367. Adaptive Rate-Optimal Lack-Of-Fit Testing For Systems Of Ordinary Differential Equations

[Student Paper Competition 1, (page 15)]

Archi ROY, *Doctoral student*
Itai DATTNER, *Proffessor*
Moumanti PODDER, *Assistant Proffessor*

Ordinary differential equations (ODEs) serve as essential tools for modeling complex dynamics across a wide range of scientific disciplines. Despite their importance, evaluating how well an observed dataset aligns with a given ODE in the presence of statistical noise remains a challenging and underexplored area in the literature. Common model selection tools like the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are frequently employed, but their application to ODE models is not always straightforward. For instance, in modeling predator-prey interactions, various Lotka-Volterra type systems attempt to enhance the basic model by addressing its unrealistic aspects based on empirical phenomena. However, assessing the relative significance of these modifications through mere parameter counting lacks objectivity. In this paper, we introduce a novel nonparametric test for detecting model misspecification in the context of ODEs. This test, based on local smoothing techniques, is theoretically demonstrated to adapt to the smoothness level of alternative models. We establish the test's rate-optimality and its consistency in detecting model misspecification against fixed alternatives, sequences of local alternatives, and sequences of smooth alternatives with unknown differentiability order. To our knowledge, this is the first work addressing such a broad range of alternatives in ODE models. Our simulations validate the test's bandwidth adaptivity when applied to fixed alternatives. We believe that this test will be a valuable tool for model selection in ODE contexts, with potential applications spanning biology, population sciences, environmental modeling, and machine learning.

368. A Regularized Low Tubal-Rank Model for High-dimensional Time Series Data

[01.M2.I8, (page 6)]

Samrat ROY, *Indian Institute of Management Ahmedabad*
George MICHAILIDIS, *University of California Los Angeles*

High dimensional time series analysis has diverse applications in macroeconometrics and finance. Recent factor-type models employing tensor-based de-

compositions prove to be computationally involved due to the non-convex nature of the underlying optimization problem and also they do not capture the underlying temporal dependence of the latent factor structure. This work leverages the concept of tubal rank and develops a matrix-valued time series model, which first captures the temporal dependence in the data, and then the remainder signals across the time points are decomposed into two components: a low tubal rank tensor representing the baseline signals, and a sparse tensor capturing the additional idiosyncrasies in the signal. We address the issue of identifiability of various components in our model and subsequently develop a scalable Alternating Block Minimization algorithm to solve the convex regularized optimization problem for estimating the parameters. We provide finite sample error bounds under high dimensional scaling for the model parameters.

369. On the Computational Complexity of Private High-dimensional Model Selection

[03.E1.C14, (page 40)]

Saptarshi ROY, *University of Texas, Austin*
Zehua WANG, *University of Michigan, Ann Arbor*
Ambuj TEWARI, *University of Michigan, Ann Arbor*

We consider the problem of model selection in a high-dimensional sparse linear regression model under privacy constraints. We propose a differentially private best subset selection method with strong utility properties by adopting the well-known exponential mechanism for selecting the best model. We propose an efficient Metropolis-Hastings algorithm and establish that it enjoys polynomial mixing time to its stationary distribution. Furthermore, we also establish approximate differential privacy for the estimates of the mixed Metropolis-Hastings chain. Finally, we perform some illustrative experiments that show the strong utility of our algorithm.

370. Interpretable classification of high-dimensional time series in spectral domain

[03.A1.I54, (page 36)]

Sarbojit ROY, *King Abdullah University of Science and Technology*
Malik SULTAN, *King Abdullah University of Science and Technology*
Hernando OMBAO, *King Abdullah University of Science and Technology*

Interpretable classification of time series presents significant challenges in high dimensions. Traditional feature selection methods in the frequency domain often assume sparsity in spectral density matrices (SDMs) or their inverses, which can be restrictive for real-world applications. We propose a model-based approach for classifying high-dimensional stationary time series by assuming sparsity in the difference between inverse SDMs. The proposed approach emphasizes the interpretability of model parameters, making it especially suitable for fields like neuroscience, where understanding differences in brain network connectivity across various states is crucial. The estimators for model parameters demonstrate consistency under appropriate conditions. Additionally, we introduce a method to screen the most discriminatory frequencies for classification, which exhibits the sure screening property under general conditions. The flexibility of the proposed model allows the significance of covariates to vary across frequencies, enabling nuanced inferences and deeper insights into the underlying problem. The novelty of our method lies in the interpretability of the model parameters, addressing critical needs in neuroscience. The proposed approaches have been evaluated on simulated examples and the 'Alert-vs-Drowsy' EEG dataset.

371. TAVIE: Tangent Approximation for Variational Inference in different Exponential Families

[02.E1.C9, (page 29)]

Somjit ROY, *Department of Statistics, Texas A&M University*
Pritam DEY, *Department of Statistics, Texas A&M University*
Debdeep PATI, *Department of Statistics, University of Wisconsin-Madison*
Bani K. MALLICK, *Department of Statistics, Texas A&M University*

Jaakkola and Jordan [1997] suggested an exact variational inference (VI) algorithm based on the tangent transformation technique for the logistic regression problem with Gaussian priors over the parameters, leveraging on a quadratic variational lower bound of the log-likelihood. We propose TAVIE, an extension of the existing framework to encompass a general class of exponential families. Specifically, we explore (i) families parametrized by a positive parameter viz., Poisson, Gamma and Inverse-Gamma distributions; and (ii) super-Gaussian families of distribu-

tion. Further, we demonstrate that TAVIE is readily applicable to Neural Networks having response distributed as Negative-Binomial distribution or one of (i) and (ii). We introduce an efficient Expectation-Maximization (EM) algorithm—with provable theoretical guarantees—specifically designed to facilitate Bayesian inference in our setup. Finally, we establish the superior performance of our proposed framework with respect to state-of-the-art competing models through extensive simulation studies and real data applications.

372. Optimal Reliability Sampling Plans under Progressive Type-I Interval Censoring Schemes

[02.A1.I37, (page 25)]

Soumya ROY, *Indian Institute of Management Kozhikode*

In reliability studies, experiments are often administered under various censoring schemes, resulting in incomplete data sets. Among these censoring schemes, Type-I and Type-II schemes are perhaps the most popular among reliability engineers. However, these traditional schemes assume continuous inspection during the experiment, which may not be feasible due to resource constraints. Thus, interval censoring schemes are proposed in the literature. A major limitation of these traditional interval censoring schemes is their inability to accommodate intermediate withdrawals of test units during the experiment. Because of this, Progressive Type-I Interval (PIC-I) censoring schemes are introduced in the literature. This work deals with such PIC-I schemes and presents optimal reliability sampling plans for systems with multiple competing failure modes.

373. Robust Principal Component Analysis using Density Power Divergence

[Student Paper Competition 1, (page 15)]

Subhrajyoty ROY, *Indian Statistical Institute, Kolkata*
Ayanendranath BASU, *Indian Statistical Institute, Kolkata*
Abhik GHOSH, *Indian Statistical Institute, Kolkata*

Principal component analysis (PCA) is a widely employed statistical tool used primarily for dimensionality reduction. However, it is known to be adversely affected by the presence of outlying observations in the sample, which is quite common. Robust PCA methods using M-estimators have theoretical benefits, but their robustness drop substantially for high dimensional data. On the other end of the spectrum, robust PCA algorithms solving principal component pursuit or similar optimization problems have high breakdown, but lack theoretical richness and demand high computational power compared to the M-estimators. We introduce a novel robust PCA estimator based on the minimum density power divergence estimator. This combines the theoretical strength of the M-estimators and the minimum divergence estimators with a high breakdown guarantee regardless of data dimension. We present a computationally efficient algorithm for this estimate. Our theoretical findings are supported by extensive simulations and comparisons with existing robust PCA methods. We also showcase the proposed algorithm's applicability on two benchmark datasets and a credit card transactions dataset for fraud detection.

374. Dynamical Survival Analysis of SIR model

[Student Poster Competition, (page 20)]

Suchismita ROY, *Duke University*
Jason XU, *Duke University*
Alexander FISHER, *Duke University*

The SIR model is a widely used framework that models the spread of infectious diseases within a closed population, dividing it into three compartments: susceptible, infected, and recovered. The dynamical survival analysis of the SIR model introduced a new interpretation of this model that enabled us to apply survival analysis tools to epidemic data. This interpretation identifies the distribution of infection and recovery time of a randomly chosen individual under certain assumptions of a large population. We developed a method to generate independent samples from the distribution of infection times. Additionally, dynamic survival analysis-based models require continuous infection and recovery times that are rarely available because epidemic data are often reported as daily or weekly infection counts. Using the new interpretation, we modified the DSA-based likelihood and inferred the infection and recovery rates using incidence data without any approximation. Our approach was evaluated on 300 simulated datasets across three different parameter sets, using Maximum Likelihood Estimation and posterior sampling through Robust Adaptive Metropolis and Hamiltonian Monte Carlo, and their performance in estimating infection and recovery rates was compared.

375. Space filling designs based on optimal covariate designs

[04.L1.C20, (page 49)]

NEETHU RS, *ICAR IASRI*
Cini VARGHESE, *Principal Scientist*
Mohd. HARUN, *Scientist*
Anindita DATTA, *Scientist*

Optimal covariate designs (OCDs) are widely accepted because they can control certain sources of variation by incorporating additional information into the model and provide most efficient estimators of the parameters of interest by enabling orthogonal estimation of treatment/ block effects and covariate effects. Here, optimal covariate PBIB designs using Hadamard matrices and orthogonal polynomials. Space filling property ensures more effectiveness to the design by allowing efficient use of resources in which the experimental units are not overly concentrated in specific areas. The superiority of the optimal covariate designs over traditional designs was evaluated based on the space-filling ability of the design points, using maxpro criterion and appropriate plotting. Such space-filling OCDs can be advantageously used to utilize the regions in the design space that are more favourable to the response variable.

376. CUSUM Control Chart for Monitoring Mean of the INAR(1)PCJ Process.

[03.M1.C11, (page 33)]

Aswathy S, *Cochin University of Science and Technology*
Irshad M. R. ,

A novel lifetime distribution is presented, derived from compounding the Poisson and two-parameter Chris-Jerry distributions. Various of its statistical characteristics are determined, encompassing aspects such as the moments, probability generating function, and hazard rate function. Statistical inference regarding the model parameters is explored through maximum likelihood estimation. A simulation study is carried out to evaluate the bias and mean square error of the estimated values. The importance of the proposed distribution is illustrated within the framework of an integer-valued first-order autoregressive process, which we shall call the INAR(1) process. In the realm of statistical quality control, we devise a cumulative sum control chart for monitoring the process mean. To illustrate its applicability, we conduct both simulation and real-data analysis.

377. On Spatio-temporal Autoregressive Models and their Applications

[05.M1.I87, (page 59)]

Krishnarani S. D., *Department of Statistics, University of Calicut*

Spatio-temporal models are being used in an increasing number of research studies to examine location-specific and temporal impacts in real-world data sets. This paper will provide an overview of spatio-temporal autoregressions and the methods used to investigate them. Some of the conventional autoregressive time series models are modified to include a spatial component in order to produce new models. Estimating methodologies are developed, and simulation studies are carried out to assess the estimation procedures. The model applications are demonstrated using actual data sets.

378 . Bivariate Generalized Geometric Distribution

[04.E1.C28, (page 58)]

Dr Harisankar S, *Asian School of Business, Technocity, Trivandrum*
Dr. C. Satheesh KUMAR,

Here we introduce a bivariate version of the generalized geometric distribution of Kumar and Harisankar (Jour. Statistist. Comput. Simul., 2020) through the name, the bivariate generalized geometric distribution (BGGD), which we obtained as the distribution of the random sum of certain types of independent and identically distributed bivariate Bernoulli random variables. A genesis of the distribution and explicit closed form expressions for its probability mass function, factorial moments and p.g.f's of conditional distributions and marginal distributions are derived. Certain recurrrence relations for probabilities, raw moments and factorial moments of the BGGD are also obtained. The method of maximum likelihood estimation is employed for estimating the parameters of the distribution. A Monte Carlo simulation study empirically evaluates and compares the performance of the proposed estimators in terms of bias and standard error. Two data sets have been analyzed to show how the proposed model and the method work in practice. We will see that the performances are quite satisfactory.

379 . On Parameter Estimation of Gompertz Distribution under Constant Stress Accelerated Life Testing using Adaptive type-II progressive censoring

[01.E1.C2, (page 11)]

Sreelakshmi S, *University of Kerala, Trivandrum, Kerala*
Manoj CHACKO, *University of Kerala, Trivandrum, Kerala, India*

In this paper, the problem of estimation of parameters of Gompertz distribution under constant stress accelerated life testing model using adaptive type-II progressive hybrid censored sample is considered. The maximum likelihood and Bayesian methods are used for estimating the unknown parameters. Markov Chain Monte Carlo techniques are employed to carry out Bayesian estimation. A simulation study is carried out to assess and compare the different estimators discussed in this paper. Finally, a real data is analyzed to illustrate the results.

380 . Multiscale Geographically Weighted Regression Modeling Approach for Assessing the Localized Spatial Effects of Lung Cancer Incidence

[03.E1.C17, (page 42)]

SRUTHI S, *Regional Cancer Centre, Thiruvananthapuram, Kerala*
Preethi Sara GEORGE, *Additional proferssor, Division of Cancer Epidemiology & Biostatistics, Regional Cancer Centre, Thiruvananthapuram*
Aleyamma MATHEW, *Professor and Head, Division of Cancer Epidemiology & Biostatistics, Regional Cancer Centre, Thiruvananthapuram*

The local Geographically weighted regression (GWR) model explores spatial non-stationarity by estimating relationships that vary across different locations within the study area. As GWR is considered as a single-scale model, it assumes that processes all operate at the same scale. However, this posits a limitation in modeling potentially multi-scale processes which are more often seen in cancer research. Therefore, we introduced Multiscale Geographically Weighted Regression (MGWR) as an extension to the GWR model which provides information about the different scales of predictor-to-response relationships. The global regression models capture the effect of spatial autocorrelation by estimating the relationship between variables across the entire study area. This study explores the spatial variation of lung cancer incidence based on several environmental variables using global and local spatial regression models. The results suggest that the MGWR model not only achieves higher goodness of fit but also performs better at alleviating residual autocorrelation than GWR and different global models. This study also identifies the high-risk areas of lung cancer incidence at different scales, corresponding to each environmental and socio-demographic factor.

381. Notes on an Existing Power Function Distribution Process and Introducing a Novel Approach for Modelling Proportional Reverse Hazard Processes.

[04.A1.C22, (page 53)]

Sachin SACHDEVA, *PhD Research Scholar at School of Mathematics and Statistics , University of Hyderabad*
B. G. MANJUNATH, *Associate Professor, Deptt. of Statistics, Rajiv Gandhi University of Arunachal Pradesh.*
Barry C. ARNOLD,
DISTINGUISHED PROFESSOR, DEPTT. OF STATISTICS, UNIVERSITY OF CALIFORNIA, RIVERSIDE, USA.,

It is known that all the proportional reversed hazard (PRH) processes can be derived by a marginal transformation applied to a power function distribution (PFD) process. Kundu [1] investigated PRH processes that can be viewed as being obtained by marginal transformations applied to a particular PFD process that will be described and investigated and will be called a Kundu process. In the present note, in addition to studying the Kundu process, we introduce a new PFD process having Markovian and stationarity properties. We discuss the distributional features of such processes, explore inferential aspects and include an example of applications of the PFD processes to real-life data. Keywords: Power Function Distribution Process, Pareto Distribution, Moment Methods, Auto-correlation, Stationarity, Markovian Property.

Key References : [1] Kundu, Debasis. (2022), Stationary GE-Process in its applications in analyzing gold Price data, Sankhya, 84-B, 575-595. [2] Gupta, R.C. and Gupta, R.D. (2007), Proportional reversed hazard rate model and Its applications, J. Stat. Plann. & Inf., 137, 3525–3536.

382. Asymptotic Properties of Generalized Elephant Random Walks

[Student Poster Competition, (page 22)]
Tamojit SADHUKHAN, *Indian Statistical Institute, Kolkata*
Krishanu MAULIK, *Indian Statistical Institute, Kolkata*
Parthanil ROY, *Indian Statistical Institute, Bangalore*

Elephant random walk is a special type of random walk that, unlike the simple random walk with steps independent of each other, incorporates the memory of the past steps to determine its future steps. It has drawn attention of the Statistical Physics literature as it exhibits anomalous diffusion. Several variations of the classical elephant random walk have also been studied. The walker in the elephant random walk is referred to as an elephant as it retains and uses the memory of the entire history of the walk. The classical elephant random walk is a one-dimensional discrete-time random walk in which the elephant moves along the integer line one step at a time. At each epoch, the elephant chooses one of its previous steps uniformly at random and then either repeats it with probability p, or moves one step in the direction opposite to the chosen step with probability $1-p$. The asymptotic behavior of the walk depends on the value of the parameter $p \in (0,1)$, which indicates the strength of the elephant's memory. The walk undergoes a phase transition at $p = 3/4$, where its rate of growth changes from diffusive ($p < 3/4$) to superdiffusive ($p \geq 3/4$). However, despite seemingly using the entire history of the walk for determining the future steps, the walk has an inherent Markovian structure, even though it is time-inhomogeneous. Namely, the probability of taking a particular step ($+1$ or -1) at a time point, conditioned on the entire history, depends on a linear function of the proportion of steps of that type till that time point. Further, at any time point, the total number of steps of either type is the elapsed time, while the difference of the number of steps of the two types gives the present location. Thus, the probability of taking a step of a particular type at a time point is a function of the present location of the walk (Markovian) as well as the time point (time-inhomogeneous).

The above structure of the elephant random walk is useful for modeling situations where a sequence of actions is performed dynamically, choosing between two (or more) options, with each subsequent action depending on the relative frequencies of the actions taken previously. However, a limitation of using the elephant random walk to model these types of situations is that the relative frequencies of past actions might not be directly observable. Instead, what can be observed is a potentially nonlinear function of these relative frequencies. For example, consider a market with two competing brands of the same product where a customer can choose between the brands before buying the product. A customer buys the product of a brand over the competing brand based on how often the products of that brand were purchased in the past. We call the (relative) frequency of past purchases of the products of a brand as the present *market share* of that brand. However, the market shares of the brands are usually not known to the customers, but the current price of the products of a brand, which can be viewed as a nonlinear function of the present market share of the corresponding brand, is always known.

Motivated by this example, in this work, we consider a generalization of the elephant random walk where we investigate how its dynamics will change if we replace the underlying linear function with a generic map satisfying some analytic conditions. We propose a new model, called the multidimensional generalized elephant random walk, that incorporates the above-mentioned generalizations of not only the classical elephant random walk but also many of its one and multidimensional variants. Using tools from the theory of stochastic approximation, we derive the limiting speed of the random walk and study the phase transition (from the diffusive to the superdiffusive regime) of the fluctuations of the speed around its (almost sure) limit. In the one-dimensional case, we also derive the law of iterated logarithm which enables us to investigate the recurrence and transience of the walk. Additionally, in this case, we present an expansion of the speed of the walk around its (almost sure) limit in the superdiffusive regime. The order of the expansion depends on the smoothness of the underlying function used to obtain the probability of the next step. We present several intriguing and illustrative examples that may be of independent interest. We also mention a few open problems in this context.

383. On adaptivity of nearest neighbors in non-smooth factor models

[02.E1.C9, (page 29)]
Tathagata SADHUKHAN, *Cornell University*
Manit PAUL, *University of Pennsylvania*
Dwivedi RAAZ, *Cornell University*

Nearest neighbor (NN) algorithms have been extensively used for missing data problems in recommender systems and sequential decision-making sys-

tems. Prior theoretical analysis has established favorable guarantees for NN when the underlying data is sufficiently smooth and the missingness probabilities are lower bounded. Here we analyze NN with non-smooth non-linear functions with vast amounts of missingness. In particular, we consider matrix completion settings where the entries of the underlying matrix follow an unknown non-linear latent factor model, with the non-linearity belonging to a Holder function class that is less smooth than Lipschitz. Our results establish following favorable properties for a suitable two-sided NN: (1) The mean squared error (MSE) of NN adapts to the smoothness of the non-linearity, (2) under certain regularity conditions, the NN error rate matches the rate obtained by an oracle equipped with the knowledge of both the row and column latent factors, and finally (3) NN's MSE is non-trivial for a wide range of settings even when several matrix entries might be missing deterministically. We support our theoretical findings via extensive numerical simulations and a case study with data from a mobile health study, HeartSteps.

384. On Interval Entropy Loss Functions

[01.E1.C1, (page 11)]

P M SAFWANA, *University of Calicut*

In many parameter estimation problems, the parameter's support is either inherently constrained (as in cases of probability or variance) or deliberately restricted by the researcher based on prior knowledge. One way to account for this restricted space is by adopting a Bayesian approach, where a uniform prior distribution is defined over the restricted space. After obtaining the posterior distribution, the squared error loss function is typically employed to make inferences about the parameter . However, squared error loss often fails to adequately penalize values near the boundaries, rendering it less suitable for many applications. To address this issue, an interval entropy loss function is proposed for cases where the parameter space is restricted to an interval (a, b). This loss function effectively penalizes boundary values, similar to how squared error loss operates on the entire real line. The benefits of this approach are demonstrated using two well-known examples.

385. A Class of tests for trend change in hazard rate function with random right censored data

[03.A1.C12, (page 39)]

Aritra SAHA, *Indian Statistical Institute, Kolkata*
Md. Zafar ANIS, *Indian Statistical Institute, Kolkata, India*

In this study, we expand upon the test family proposed by Majumder and Mitra (2019) for testing exponentiality against bathtub (BT) and upside-down bathtub (UBT) hazard rates by considering randomly right-censored cases. We derive the test statistic's asymptotic distribution and utilize Monte Carlo simulations to determine the empirical powers for certain alternative distributions. The asymptotic distribution of the test statistic proposed by Park (1988) is derived as a special case. Finally, real-life data examples are presented for illustrative purposes.

386 . Random Forests for Geospatial Data

[02.A1.I32, (page 23)]

Arkajyoti SAHA, *Department of Statistics, University of California, Irvine*
Sumanta BASU, *Department of Statistics and Data Science, Cornell University*
Abhirup DATTA, *Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

Due to recent advancements in geographical information systems, remote sensing technology, and affordable sensors, we are now faced with datasets that require us to consider spatial dependencies. These geospatial data are often analyzed using the linear mixed model framework, which includes a linear fixed covariate effect and a Gaussian Process (GP)-distributed spatial random effect. However, the assumption that covariate effects are linear is quite limiting. Non-linear modeling of spatial data is gaining popularity, and contemporary extensions of Random Forests (RF) for spatial data diverge from the mixed model setup, giving up inference on the fixed effects and other benefits of utilizing GP. By explicitly modeling the spatial random effects with a GP, we offer a unique and well-principled extension of RF for estimating nonlinear covariate effects in spatial mixed models. Our method extends RF in the same way generalized least square extends ordinary least squares to accommodate for dependence in linear models. For both estimation and prediction with spatial data, our method significantly outperforms classical RF in an extensive simulation study. We also discuss expanding our method to include binary geospatial data.

387. Probabilistic Guarantees on Sensitivities of Bayesian Neural Network

[04.E1.C25, (page 56)]

Diptarka SAHA, *University of Illinois, Urbana-Champaign*
Zihe LIU, *University of Illinois, Urbana-Champaign*
Feng LIANG, *University of Illinois, Urbana-Champaign*

The study of theoretical properties of wide and deep neural networks is a growing body of research that complements their empirical success. In this paper, we study the partial derivatives of a random, wide, fully connected neural network w.r.t. each individual feature, which we refer to as the feature sensitivities of that neural network. Under a set of general conditions, as the network widens, we show that these sensitivities are consistent around their mean. Moreover, we show that these sensitivities, as a random function of the features, converge in distribution to Gaussian processes under proper scaling. We discuss the ramifications of such behavior and how this can be leveraged to obtain robust estimates of feature importance and pruning strategies. We employ these strategies to prune larger models to obtain more concise models with equivalent predictive power. Our approach shows competitive efficacy compared to existing neural network feature relevance detection methods across various simulated and real-world datasets.

388 . SMAUG: Sample-specific Multiomic Association networks Using Gaussian graphical models

[01.A1.I12, (page 7)]

Enakshi SAHA, *University of South Carolina*
Katherine H. SHUTTA, *Channing Division of Network Medicine, Brigham and Women's Hospital*
John QUACKENBUSH, *Harvard T. H. Chan School of Public Health*

Gaussian Graphical Models (GGM) provide an invaluable tool for studying the interaction patterns between multiple omics modalities, as recently demonstrated by DRAGON (Determining Regulatory Associations using Graphical models on multi-Omic Networks), a method that enables joint analysis of transcriptomic and methylomic data. Existing methods employing GGMs for omics data, including DRAGON, estimate single aggregate network that approximate the average conditional molecular dependence structure across the entire population and thus fail to recognize the individual-specific variability in these interaction networks. To overcome this limitation, we propose an empirical Bayesian model SMAUG (Sample-specific Multiomic Association networks Using Gaussian graphical models) that recognizes individual-specific heterogeneity in molecular dependence by estimating sample-specific GGMs that capture the conditional dependence structure between different omics data types for each individual. By employing data-driven Individual-specific conjugate priors, SMAUG provides a scalable tool for deciphering variability in disease mechanisms across sex, age and other clinical variables, thereby providing a more nuanced understanding of diseases. In addition, SMAUG being a partial-correlation-based method, is better suited to distinguish between direct molecular dependence and spurious correlations, compared to existing methods for sample-specific network inference that employ Pearson's correlation as their foundation.

389. Bayesian compositional regression with flexible microbiome feature aggregation and selection

[02.A1.I34, (page 24)]

Satabdi SAHA, *The University of Texas MD Anderson Cancer Center*

TBA

390. Quantiles for Control: Robust inference in Deep Networks

[03.A1.I55, (page 36)]

Snehanshu SAHA, *APPCAIR and Dept. of CSIS, BITS Pilani Goa Campus*
Aditya CHALLA,
Soma DHAVALA,

While machine learning models have made significant strides in improving healthcare by detecting diseases, there is still hesitancy among both patients and doctors in adopting ML diagnostic tools and trusting their predictions.This is primarily due to the fact that the confidence scores reported by a model are not true representations of the likelihood of a diagnosis being correct. Our method solves this issue by providing an absolute measure of the chance that a sample was classified incorrectly by computing the uncertainty in the function learnt by the ML model. In addition, it provides additional resilience to mislabeled training data, as well as a visualization to both doctors and patients as to why the decision was made. This method serves as an add-on to any ex-

isting classifier, and thus augments any deep learning model that aims to answer a yes-or-no question. I will present a Quantile based approach and show that it achieves some of the stated goals.

391 . Statistical Analysis of Bivariate Left Truncated Right Censored Data

[05.M1.C29, (page 61)]

Sourav SAHA, *IIT Guwahati*
Ayon GANGULY, *IIT Guwahati*
Debanjan MITRA, *IIM Udaipur*

Truncation indicates the circumstance where certain subjects in the population are rejected due to their very short or very long survival times whereas, censoring indicates the circumstance where survival times of subjects under consideration are not completely observed. Here, we entertain ourselves with the analysis of bivariate data where both the random variables are left truncated and right censored. Here, we model left-truncated and right-censored bivariate data with the help of the family of Archimedean copulas, which comprises several useful copulas; thus, providing flexibility in modelling. We use Expectation-Maximization algorithm to perform likelihood inference. Louis's missing information principle is utilized to construct confidence intervals of model parameters. An extensive simulation study is performed to check adequacy of the proposed methods. Finally, a real dataset is analyzed for illustrative purposes.

392. Learning representation for mixed data types with a nonlinear deep encoder-decoder framework

[05.M1.I89, (page 60)]

Saswata SAHOO, *Google*

Representation of data on mixed variables, numerical and categorical types to get suitable feature map is a challenging task as important information lies in a complex non-linear manifold. The feature transformation should be able to incorporate marginal information of the individual variables and complex cross-dependence structure among the mixed type of variables simultaneously. In this work, we propose a novel nonlinear Deep Encoder-Decoder framework to capture the cross-domain information for mixed data types. The hidden layers of the network connect the two types of variables through various non-linear transformations to give latent feature maps. We encode the information on the numerical variables in a number of hidden nonlinear units. We use these units to recreate categorical variables through further nonlinear transformations. A separate and similar network is developed switching the roles of the numerical and categorical variables. The hidden representational units are stacked one next to the others and transformed into a common space using a locality preserving projection. The derived feature maps are used to explore the clusters in the data. Various standard datasets are investigated to show nearly the state of the art performance in clustering using the feature maps with simple K-means clustering.

393 . Estimating scale-disparity in the proportional hazards model using martingale residuals

[03.M1.I46, (page 32)]

Shyamsundar SAHOO, *Department of Statistics, Haldia Government College*

If the proportional hazards (PH) assumption fails, it is common to seek an alternative to the PH model when analyzing censored survival data. Sometimes, the effect of a particular covariate may be non-proportional, even if the PH assumption holds for other covariates. In this article, we look at the non-proportional effect of a single covariate in the form of a scale change on the failure time, while the remaining covariates have a proportional effect on hazard. If this alternative model holds, we estimate the corrective scale factor using a suitable transformation of the martingale residuals obtained by fitting Cox's proportional hazards regression model. After rescaling the data by the estimated scale factor, the PH model may be applied to the remaining covariates. The performance of the proposed method is studied via Monte Carlo simulations. The procedure is then illustrated through the analysis of a survival dataset.

394 . Cohesive response clustering via metric space embedding

[04.A1.C21, (page 53)]

Suvadip SANA, *Cornell University*

Clustering of responses is one of the important problems in unsupervised machine learning. We try to answer the following question: Given multiple responses and fixed covariates, how do the responses cluster based solely on the information contained in the covariates? Naturally, if we change the covariate

space, the clustering might change. Our approach includes computing the regression function for each response and considering a metric space consisting of regression functions with flexible metric choices. We then cluster the regression functions based on a "cohesion measure," which gives us a notion of binding forces between two points in a metric space. We apply a hierarchical agglomerative algorithm for our setup. We demonstrate the clustering method by applying it to various multivariate data sets. Interestingly, there is a duality between a cohesion measure and a distance measure. We also find that the usual definition of a cohesion measure can be extended to a notion we define as the random cohesion measure, which generalizes the theory of cohesion with respect to arbitrary discrete measures.

395. PGDUS Kumaraswamy Distribution: Properties and Reliability Analysis

[01.E1.C2, (page 11)]

ANN SANIA, *St.Thomas College(Autonomous), Thrissur, University of Calicut, Kerala*
V M CHACKO, *Professor, St Thomas College (Autonomous), Thrissur*

In this paper, we employ various estimation procedures to estimate the parameters of a new lifetime distribution derived by applying the Power Generalized DUS transformation to the Kumaraswamy distribution (PGDUS-K). Its capability to handle both monotone and non-monotone failure rate functions makes it a crucial tool for lifetime data analysis. We examined the statistical properties of this distribution. A notable characterization of the proposed distribution is its closeness under exponentiation: If $W = Z^p$ and $p > 0$, then $Z \sim$ PGDUS-K(λ, β, δ) if and only if $W \sim$ PGDUS-K$(\frac{\lambda}{p}, \beta, \delta)$. Furthermore, distribution of the maximum remains within the PGDUS-K family, thereby preserving the characteristics of the individual component distributions. The model's robustness is validated through using simulated datasets. Our proposed distribution outperforms existing distributions using real-world data. We also derive stress-strength reliability for single-component and multi-component systems, showing the distribution's versatility and efficacy in reliability analysis.

396. TBA

[01.A1.I16, (page 9)]

Kris SANKARAN, *University of Wisconsin-Madison*

TBA

397. Semiparametric regression analysis of mixed recurrent event and panel count data with multiple causes of failure

[04.A1.I82, (page 52)]

Sankaran PADUTHOL GODAN, *Cochin University of Science and Technology*

The regression analysis of mixed recurrent event and panel count data with multiple modes of failure. A semiparametric regression model that specifies the multiplicative covariate effect on the marginal mean function under different failure modes is introduced. The asymptotic properties of the estimators are studied. The performance of the estimators is evaluated using simulation studies, and a real data is used to illustrate the proposed techniques in real life applications.

398. Asymptotic Inference in Genetic Association Studies Leveraging Genetic Correlations

[Student Poster Competition, (page 22)]

Madhav SANKARANARAYANAN, *Harvard T.H. Chan School of Public Health*
Rajarshi MUKHERJEE, *Harvard T.H. Chan School of Public Health*
Tamar SOFER, *Beth Israel Deaconess Medical Center*
Yana HRYTSENKO, *Beth Israel Deaconess Medical Center*

Genetics is being increasingly used to study disease biology, for prediction, and for elucidating causal relationships between risk factors to health and disease outcomes. Proteins mediate the effect of genetic variation on disease, and recent advances reveal the potential of proteins to elucidate mechanics of diseases, including diabetes, by acting as mediators of genetic effects. Thus, a particular problem of interest is to estimate the association of genetic determinants of proteins (protein quantitative trait loci, pQTLs) with metabolic traits, which characterize diabetes. If accurate, these associations can be later used to predict metabolic traits and diabetes, and to design protein-targeting interventions to alleviate the health burden of these phenotypes. Our primary goal is to improve the estimation of pQTL associations with metabolic traits by using (1) the estimated genetic correlation between proteins and traits, (2) pQTL-protein effect sizes. The incorporation of these exter-

nal data in our analysis posits a "constrained" association problem, which requires tailored methodology, particularly in high dimensional cases (many proteins and/or many pQTLs). We work on the construction and show the optimality of estimators and algorithms for a range of dimensional cases. Finally, we extend the methodology to account for estimation errors and apply it to real genetic datasets.

399. Real-time claim risk assessment using ML models and technology

[01.M2.I6, (page 5)]

Omesh SARAF, *Bajaj Allianz Life Insurance Co Ltd*

Accurate assessment of early claim risk is critical to the sustainability of any life insurance company, which, understandably is a difficult event to predict - more so as the decision needs to be taken real-time within a few seconds during the proposal process. There are various factors that contribute to early claims, such as, age, quality of customer profile, fraud and health disclosures. For some of these, reliable risk markers or attributes may be available directly but for others, it may not be (e.g. fraud, affluence profile, health status). The solution, therefore, lies in 1) building a comprehensive profile of customer and associated eco-system using mostly alternate data and 2) leveraging advanced machine learning models to estimate probability of claim. The claim probability is subsequently used to categorize the customers into risk buckets for different treatments.

400 . Data Science and its future in emerging sports and markets

[03.A1.I59, (page 38)]

Subrat SARANGI, *MICA*
Subrat SARANGI,

Data analytics has significantly impacted sports marketing, the media industry, sports accessories, gifting, and society over the last decade. Sports viewership and consumption of sports artifacts, accessories, and memorabilia have defied market boundaries. Markets globally not known for sports consumption as much have become areas of focus for sporting events like Cricket, Kabaddi, Hockey, and martial arts, to name a few. Emerging sports like Cricket in its different formats (i.e., T10, Gulley Cricket, 90 Balls a side), Kabaddi, Kho-Kho, among others, are drawing huge fan bases and viewership among the emerging markets of India, Sri Lanka, Bangladesh, and South Korea. The talk shall present an exhibition of the application of data science using different models to cricket in different formats, explaining the behaviour of sports consumption and its impact on sports organizers and legislation bodies.

401 . A Machine Learning approach to predict Selective Sweep in Genomic Region

[04.E1.C23, (page 55)]

Abhik SARKAR, *ICAR- Indian Agricultural Statistics Research Institute*
Dwijesh Chandra MISHRA, *ICAR- Indian Agricultural Statistics Research Institute, New Delhi-110012*
Dipro SINHA, *ICAR- Indian Agricultural Statistics Research Institute, New Delhi-110012*
Girish Kumar JHA, *ICAR- Indian Agricultural Statistics Research Institute, New Delhi-110012*

Selective Sweep is a natural phenomenon crucial to evolution and survival, significantly affecting the adaptability of different species in various environmental conditions. Selective Sweep can be classified into two main types: Hard Selective Sweep and Soft Selective Sweep. Traditional approaches to selective sweep analysis often lack the advanced methodologies, such as machine learning, leading to poor prediction accuracy.

To enhance accuracy, this study implements various advanced learning models for predicting selective sweeps. The study begins with data simulation to generate population genetic information, followed by feature extraction. Eight different artificial intelligence-based machine learning classifiers such as Support Vector Machines (SVMs), Regression Tree, Random Forest, Naive Bayes, Multiple Logistic Regression, K-Nearest Neighbour (KNN), Gradient Boosting and Artificial Neural Networks (ANNs) are then employed for comparative analysis. Evaluation metrics such as Accuracy, Precision, Sensitivity, and Specificity are calculated. Bootstrap analysis and boxplot visualization are used to assess the stability of the models' accuracies. Based on performance evaluation measures, TOPSIS (The Technique for Order of Preference by Similarity to Ideal Solution) analysis is conducted to identify the best-performing model.

In this study, the Random Forest model outperformed other models, achieving a TOPSIS score of 0.938 and securing the first rank. Consequently, a robust Random Forest-based model for selective sweep classification has been developed, capable of classify-

ing selective sweeps into three categories: Hard Selective Sweep, Soft Selective Sweep, and No Selective Sweep. This model has demonstrated superior performance compared to existing approaches. Additionally, an R package named "SweepDiscovery: Selective Sweep Discovery Tool" has been developed based on the model. This advanced machine learning approach for selective sweep analysis excels in both accuracy and reliability.

Key Words: Selective Sweep, Simulation, Machine Learning, Random Forest, R package

402 . Predictive Inference in Linear Mixed Models

[04.E1.C23, (page 55)]

Abir SARKAR, *Senior Associate, Data Scientist, Capital One, India (DATALABS)*
Gourab MUKHERJEE, *Associate Professor, USC Marshall School of Business*
Ishan SENGUPTA, *PhD student, Rutgers University*
Keisuke YANO, *The Institute of Statistical Mathematics*

We consider predictive inference in Linear Mixed Models (LMMs) and study the problem of estimating the predictive density under Kullback-Leibler (KL) loss in LMMs with large number of units. We consider flexible classes of empirical Bayes (EB) predictive densities and develop a novel risk estimation based methodology for selecting hyper-parameters of EB predictive density estimates. Our risk estimation based hyper-parameter selection methodology uses the decision-theoretic identity that connects predictive KL risk for density estimation to Stein's unbiased estimate of the quadratic risk in point estimation. Direct construction of unbiased KL risk estimates is not possible in LMMs. We leverage information in the covariates and exchangeability of the individual effects to construct asymptotically effcient estimates of the KL risks for wide classes of predictive density estimators. We derive the rates of convergences of the proposed KL risk estimates (KLRE) and show that EB predictors calibrated by minimizing KLREs are asymptotically optimal in LMMs.

403. First Collision Time of simple symmetric random walks

[Special Invited Session 3, (page 44)]

Anish SARKAR, *Indian Statistical Institute*

We consider 3 independent simple symmetric random walks starting from $-2, 0$ and 2 respectively. We show that the expected time of their first collision is finite and actually takes the value 4. If time permits, we will talk about the use of the result in computation of density of coalescing simple symmetric random walks starting from all even integers.

404 . High-dimensional Bernstein Von-Mises theorems for covariance and precision matrices

[02.M1.I21, (page 14)]

Partha SARKAR, *Florida State University*
Kshitij KHARE, *University of Florida*
Malay GHOSH, *University of Florida*
Matt P. WAND, *School of Mathematical and Physical Sciences, University of Technology Sydney*

This paper aims to examine the characteristics of the posterior distribution of covariance/precision matrices in a "large p, large n" scenario, where p represents the number of variables and n is the sample size. Our analysis focuses on establishing asymptotic normality of the posterior distribution of entire covariance/precision matrices under specific growth restrictions on p_n and other mild assumptions. In particular, the limiting distribution turns out to be a symmetric matrix variate normal distribution whose parameters depend on the maximum likelihood estimate. Our results hold for a wide class of prior distributions which includes standard choices used by practitioners. Next, we consider Gaussian graphical models which induce sparsity in the precision matrix. The posterior contraction rates and asymptotic normality of the corresponding posterior distribution are established under mild assumptions on the prior and true data-generating mechanism.

405. On differentially private U statistics

[03.E1.I62, (page 40)]

Purnamrita SARKAR, *UT Austin*
Shourya PANDEY, *UT Austin*
Po-Ling LOH, *Cambridge*
Kamalika CHAUDHURI, *UCSD/Meta*

We consider the problem of privately estimating a parameter E[h(X1...Xk)] are i.i.d. data from some distribution and h is a permutation-invariant function. Without privacy constraints, the standard estimators for this task are U-statistics, which commonly arise in a wide range of problems, including nonparametric signed rank tests, symmetry testing, uniformity testing, and subgraph counts in random networks, and are the unique minimum variance un-

biased estimators under mild conditions. Despite the recent outpouring of interest in private mean estimation, privatizing U-statistics has received little attention. While existing private mean estimation algorithms can be applied in a black-box manner to obtain confidence intervals, we show that they can lead to suboptimal private error, e.g., constant-factor inflation in the leading term, or even rather than in degenerate settings. To remedy this, we propose a new thresholding-based approach that reweights different subsets of the data using local Hájek projections. This leads to nearly optimal private error for non-degenerate U-statistics and a strong indication of near-optimality for degenerate U-statistics.

406. High-dimensional Bayesian Compressed Mixed-Effects Models

[Student Poster Competition, (page 21)]

Sreya SARKAR, *University of Iowa*
Sanvesh SRIVASTAVA, *University of Iowa*
Kshitij KHARE, *University of Florida*

Linear mixed-effects models are fundamental in statistical methodology. Their high-dimensional extensions estimate the model parameters using regularized likelihood or maximum a posteriori estimation. There is extensive literature on sampling-based Bayesian inference in mixed models, but their high-dimensional extensions are relatively unexplored. The main reason is the computational bottlenecks of Monte Carlo methods in high-dimensional settings. We propose compressed mixed models for efficient prediction and fixed effects selection in high-dimensional settings. These models project a subset of the parameters into a low-dimensional space using random projection matrices, which yields a generalized likelihood. Data augmentation algorithms use this generalized likelihood to facilitate efficient posterior predictive inference and fixed effects selection using Gaussian and shrinkage priors, respectively. The linear compressed mixed model has better predictive accuracy, coverage, and selection guarantees than its competitors in diverse simulation setups and real data analyses.

407. Properties of Extropy and Its Weighted Version for Doubly Truncated Random Variables

[01.A1.I17, (page 9)]

Abdul SATHAR E I, *University of Kerala*

This work proposes interval extropy and its length-biased version to measure the uncertainty of a doubly truncated random variable. Properties and characterizations of some important life distributions in terms of the new measures are obtained. Non-parametric estimators for the proposed measures are also suggested, and their performance is verified using simulated and real data sets.

408. Bayesian Pseudo Posterior Mechanism under Asymptotic Differential Privacy

[02.M2.I26, (page 17)]

Terrance SAVITSKY, *U.S. Bureau of Labor Statistics*
Matthew WILLIAMS, *RTI International*
Monika HU,

We propose a Bayesian pseudo posterior mechanism to generate record-level synthetic datasets equipped with a differential privacy (DP) guarantee from any proposed synthesis model. The pseudo posterior mechanism employs a data record-indexed, risk-based weight vector with weights $\in [0, 1]$ to surgically downweight high-risk records for the generation and release of record-level synthetic data. The differentially private pseudo posterior synthesizer constructs weights using Lipschitz bounds for a log-pseudo likelihood utility for each data record, which provides a practical, general formulation for using weights based on record-level sensitivities that we show achieves dramatic improvements in the DP expenditure as compared to the unweighted posterior mechanism. By selecting weights to remove likelihood contributions with non-finite log-likelihood values, we achieve a local privacy guarantee at every sample size. We compute a local sensitivity specific to our Consumer Expenditure Surveys dataset for family income, published by the U.S. Bureau of Labor Statistics, and reveal mild conditions that guarantee its contraction to a global sensitivity result over the space of databases. We further employ a censoring mechanism to lock-in a local result with desirable risk and utility performances to achieve a global privacy result as an alternative to relying on asymptotics. We show that utility is better preserved for our pseudo posterior mechanism as compared to the exponential mechanism (EM) estimated on the same non-private synthesizer due to the use of targeted downweighting. Our results may be applied to any synthesizing model envisioned by the data disseminator in a computationally tractable way that only involves estimation of a pseudo posterior distribution for parameter(s) θ, unlike recent approaches that use naturally-bounded

utility functions under application of the EM.

409. Joint Modeling of binary longitudinal Response and Treatment Compliance in the presence of missing data in a Clinical Trial setup

[01.E1.C3, (page 12)]

Sarfaraz SAYYED, *Novartis Healthcare Pvt Ltd*
Ashwini MATHUR, *Onesto Consulting Limited*
Asha KAMATH, *Manipal University*

Background: Typically, during follow-up, clinical studies monitor the response variable and the occurrence of certain important outcomes multiple times in the trial. Repeated observations data of an outcome and a set of covariates for each of many subjects comprises of Longitudinal data. For follow-up assessments, patient need to visit the clinic and may skip certain follow-up visits resulting in missing responses. This missing response may be due to efficacy or safety or administrative related issues. Missing values in clinical trials is a prevalent problem and is common in statistical literature. Parameter estimation would be biased in the presence of missing data while interpreting the analysis results and this bias may increase with increase in percentage of missingness. Additionally due to lost to follow-up and treatment non-compliant patients, the trial losses information on the treatment effect. Treatment non-compliance would likely affect the response of the drug in patients. As the treatment compliance is highly correlated with the response it would be beneficial to study simultaneously the response and the treatment compliance in order to have better estimates about the treatment effect. Methods: The attempt in this article is to address the missing response issue in binary longitudinal response setting by joint modeling the response and longitudinal treatment compliance. The dependence between response and treatment compliance, where information from treatment compliance can be employed in suitable imputation of the missing response, may be instructive to consider given the variability of the longitudinal outcome. EM algorithm technique is used and the comparison with/without joint modeling is carried out to study the benefit of joint modeling. The effectiveness of proposed estimator is evaluated by applying proposed method on simulation study creating artificial missingness and compliance. Results: The proposed idea of joint modeling along with the implementation of EM algorithm technique was used to analyze the simulated data and compare with the results to that of without joint modeling estimates. Joint modeling analysis produced smaller absolute bias and lower mean squared error, compared to analyzing data without joint modelling. Conclusions: Results seem to be improved with the proposed method of joint modelling providing better estimates compared to analysis where joint modeling was not done. Since missing response and non-compliance are often observed in clinical trials, the research and proposed methodology developed would help in better decision making.

410. Statistical inference and optimal censoring scheme for generalized exponential distribution under improved adaptive type-II progressive censoring

[02.E1.C10, (page 30)]

Shilpa S DEV, *University of Kerala*
Manoj CHACKO, *University of Kerala*

This paper explores an approach to analyse the estimations of generalized exponential distribution using an improved adaptive type-II progressive censoring scheme. This scheme guarantees that the experimental time will not exceed a pre-fixed time. The estimation of unknown parameters, reliability, and hazard rate functions using classical methods such as maximum likelihood and maximum product of spacing are considered. Bayesian estimation is also considered under both symmetric and asymmetric loss functions. Also, we propose different optimality criteria to find the optimal sampling scheme. A Monte Carlo simulation study is conducted to compare the performance of the proposed estimates, and the effectiveness of the estimation approach is illustrated using a real dataset.

411. Fractional cumulative entropy function in quantile framework

[04.A1.C21, (page 52)]

Iona Ann SEBASTIAN, *Department of Statistics, CUSAT*
Iona Ann SEBASTIAN, *Department of Statistics, CUSAT*
Sunoj S M, *Department of Statistics, CUSAT*

Fractional cumulative residual entropy is a powerful tool for the analysis of complex systems. Quantile functions are efficient and equivalent alternatives to distribution functions in the modeling and analysis of statistical data. We introduce a quantile-based fractional cumulative entropy measure for the residual

and past lifetime random variables. We study various properties of these measures and examine their usefulness in different applied fields.

412 . Gene identification through a sparse correlation approach

[01.E1.I19, (page 10)]

Ananda SEN, *University of Michigan*

TBA

413. Long-term impact of short-term experiments

[05.M1.I89, (page 60)]

Deborshee SEN, *Google*

A/B experimentation is a workhorse of innovation in the tech industry. Experiments can typically only be run for short durations as it is costly to run long-term experiments. However, it is often of interest to know what the long-term impact of an experiment will be. An example of this is when a tech company wants to launch a new feature and runs an A/B experiment to study its impact on user behaviour. However, when deciding to launch the feature, it is not enough to rely on short-term impacts (i.e., the impact observed during the experiment), and instead it is important to know the long-term impact of the new feature. The surrogate index (Athey et al., 2019) is a statistical method that uses short-term outcomes to predict long-term outcomes. This talk shall discuss the problem mentioned and techniques to mitigate it, including the use of surrogate indices.

414 . Random Matrices with independent entries: Beyond non-crossing partitions

[01.A1.I15, (page 8)]

Priyanka SEN, *IIT Bombay*
Arup BOSE, *ISI Kolkata*
Koushik SAHA, *IIT Bombay*
Arusharka SEN, *Concordia University*

The scaled standard Wigner matrix is known to have the semi-circular distribution as its LSD. The moments of the LSD are described with the help of non-crossing pair partitions. There have been several extensions of this result. We shall discuss the LSD of symmetric matrices with independent entries under certain moment assumptions and find a suitable set of partitions that help describe the moments of the LSD. The set of partitions found is generally larger than the non-crossing partitions and poses some interesting combinatorial questions. The LSD result will also help us bring several existing LSD results under one umbrella. For example, results on the standard Wigner matrix, the adjacency matrix of a sparse homogeneous Erd˝os-R´enyi graph, heavy-tailed Wigner matrix, some banded Wigner matrices, Wigner matrices with variance profiles follow from our result.

415 . Community detection on multi-view networks

[02.M2.I31, (page 19)]

Subhabrata SEN, *Harvard University*
Xiaodong YANG, *Harvard University*
Buyu LIN,

The community detection problem seeks to recover a latent clustering of vertices from an observed random graph. This problem has attracted significant attention across probability, statistics and computer science, and the fundamental thresholds for community recovery have been characterized in the last decade. Modern applications typically collect more fine-grained information on the units under study. For example, one might measure relations of multiple types among the units, or observe an evolving network over time. In this talk, we will discuss the community detection problem on such 'multi-view' networks. We will present some new results on the fundamental thresholds for community detection in these models. Finally, we will introduce algorithms for community detection based on Approximate Message Passing.

416. Adaptive estimation using records data under asymmetric loss, with applications

[04.A1.I79, (page 51)]

Raghu SENGUPTA, *IIT Kanpur*
Saibal CHATTOPADHYAY, *IIM Calcutta*
Neeraj JOSHI, *IIT Delhi*

We consider a scenario where data are accessible in terms of record values which are quite frequent in problems of common occurrence. For example, one may consider the hottest day ever, the lowest stock market figure, auction prices of an item in bidding, etc. One can analyze such data as record values from a sequence of observations, where the upper value is larger and the lower value is smaller, respectively, than all previous observations. The literature on classical theory of records and its several variants is quite

rich. A significant literature also exists in reliability theory and associated areas. However not much work has, has been done so far using records data when over and under estimation of the parameter of interest attract unequal penalties, even though there is a compelling need for considering such an asymmetric loss function whenever the consequences of over and under estimation are not identical. This can happen in such diverse fields of application as real estate management, accounting, reliability analysis, and so on. We consider the estimation problem based on records data for the scale parameter of distribution coming from an exponential family under an asymmetric linear-exponential loss function. With a view to controlling the associated risk, we aim at ensuring a pre-assigned upper bound on it. Since there is no known solution to this problem with a fixed sample size, the present research addresses some adaptive sampling methodologies to estimate the unknown sample size by a random sample size, and estimate the scale parameter based on a random number of random observations. Performances of the methodologies have been compared using Monte-Carlo simulations.

417. A two-stage design for A/B testing with unknown network interference

[03.A1.I55, (page 36)]

Srijan SENGUPTA, *North Carolina State University*

A/B testing or online controlled experiment (OCE) is an ubiquitous technique for comparing the effect of two or more treatments in online settings. The classical randomized design approach to implementing controlled experiments relies on the stable unit treatment value assumption (SUTVA), which states that the outcome of an individual is independent of the treatment assigned to any other individual. It is well-known that SUTVA is violated in online platforms due to a phenomenon known as network interference, where individuals are connected via a social network, and the treatment assigned to one individual can influence the outcomes of its neighbors. Over the last decade or so, the classical randomized design has been largely supplanted by network clustering based designs in A/B tests to account for this phenomenon. We show that there are two problems with network clustering based designs. First, the network itself is often unobserved or challenging/expensive to measure, which means network clustering cannot be implemented to begin with. Second, there are almost always lurking variables, which are unobserved user features that influence both user response and network formation. We demonstrate that the presence of lurking variables makes network clustering based estimators biased. We propose a two-stage design and estimation technique called HODOR (Hold-Out Design for Online Randomized experiments) to address both problems. Remarkably, HODOR is based on the classical random design, albeit with a correction to deal with network interference. We carry out a theoretical analysis of HODOR to prove its unbiasedness and to establish its optimal configuration. We also develop a statistical inference framework based on HODOR to carry out hypothesis tests and construct confidence intervals. Through simulation studies and real world data examples, we compare the empirical performance of HODOR to the classical randomized as well as the network clustering based design.

418 . On the Statistical Complexity of Sample Amplification: Increasing Dataset Size even when Learning is Impossible

[04.M2.I76, (page 48)]

Vatsal SHARAN, *University of Southern California*
Brian AXELROD,
Yanjun HAN,
Shivam GARG,

Is learning a distribution always necessary for generating new samples from the distribution? To study this, we introduce the problem of "sample amplification": given n independent draws from an unknown distribution, D, to what extent is it possible to output a set of m $>$ n datapoints that are indistinguishable from m i.i.d. draws from D? Curiously, we show that nontrivial amplification is often possible in the regime where the number of datapoints n is too small to learn D to any nontrivial accuracy. We prove upper bounds and matching lower bounds on sample amplification for a number of distribution families including discrete distributions, Gaussians, and any continuous exponential family. This is based on joint work with Brian Axelrod, Yanjun Han, Shivam Garg and Greg Valiant.

419 . The Exponentiated Teissier Distribution and Its Bivariate Extension: Properties, Estimation Techniques, and Applications

[02.E1.C6, (page 27)]

Vikas Kumar SHARMA, *Banaras Hindu University,*

Varanasi
Sudhanshu V. SINGH, *IITRAM, Ahemedabad*
Komal SHEKHAWAT, *IITRAM, Ahmedabad*
Ashok Kumar PATHAK, *Central University of Panjab, Bathinda*

In this talk, we would discuss two-parameter exponentiated Teissier distribution (ETD) and its bivariate extension. It is the main advantage of the ETD to have increasing, decreasing and bathtub shapes for its hazard rate function. The expressions of some their important properties are derived. Estimators based on likelihood, least squares, weighted least squares and product spacings are constructed for estimating the unknown parameters of the ETD. The MLE and Bayes estimation is discussed for the bivariate extension. Simulation experiments are conducted to compare the performances of the considered estimators. Real data are fitted by using the proposed univariate and bivariate distributions over the competing distributions.

420. Bayesian biclustering and its application in education data analysis

[01.E1.I18, (page 10)]

Weining SHEN, *University of California, Irvine*

We propose a novel nonparametric Bayesian item response theory model to estimate clusters at the question level while simultaneously allowing for heterogeneity at the examinee level under each question cluster, characterized by the mixture of Binomial distributions. We present some theoretical results that guarantee the identifiability of the proposed model and show that the model can correctly identify question-level clusters asymptotically. We also provide a tractable sampling algorithm to obtain valid posterior samples from the proposed model. Compared to the existing methods, the model manages to reveal the multi-dimensionality of the examinee's proficiency level in handling different types of questions parsimoniously by imposing a nested clustering structure. The proposed model is evaluated via a series of simulations as well as applied to an English proficiency assessment data set. This data analysis example nicely illustrates how the model can be used by test makers to distinguish different types of students and aid in the design of future tests.

421. LSD of Large Kendall's Correlation Matrix and it's Application

[02.E1.C10, (page 30)]

Raunak SHEVADE, *IIT Bombay*
Monika BHATTACHARJEE, *IIT Bombay*
Raunak SHEVADE, *IIT Bombay*

The limiting spectral distribution (LSD) of high-dimensional Kendall's correlation matrix and tests of independence based on this matrix have been studied in the literature when the observations are absolutely continuous with respect to Lebesgue measure, and are independent and identically distributed. We investigate the LSD of this matrix under much weaker assumptions which accommodate discrete and/or non-identical distributions, and also identify the limit distribution by using free probability. We also propose a graphical test of independence under these assumptions.

422 . A generalized dynamic spatio-temporal model for panel count data

[03.E1.C17, (page 42)]

Shrinivas SHIRKE, *Department of Statistics , Savitribai Phule Pune University*

Spatial and spatio-temporal models for count data have gained traction and most of the models are developed adapting the approach based on space–time dynamic generalized linear models. In departure from existing approach, we develop a generalized dynamic spatio-temporal model for panel count data wherein we allow for dependence between the mean process and neighbouring observations in the previous time period as well as spatial locations using exponential feedback model. This model is also a generalization of the existing spatio-temporal models as it accounts for latent heterogeneity as well as serial error correlation in space-time domain. A Bayesian approach has been developed for model estimation. The models' usefulness is demonstrated simulated as well as real life application.

423. MCMC Importance Sampling via Moreau-Yosida Envelopes

[03.E1.C16, (page 42)]

Apratim SHUKLA, *Indian Institute of Technology Kanpur*
Dootika VATS, *Indian Institute of Technology Kanpur*
Eric CHI, *Rice University*

Gradient-based Markov chain Monte Carlo (MCMC) schemes are popular for Bayesian inference owing to faster mixing, however they require the underlying target distribution to be differentiable.

This has given rise to proximal MCMC algorithms which are exact/inexact and use tools from optimisation literature to construct smooth approximation to the non-differentiable targets. These approximations, known as the Moreau-Yosida (MY) envelopes, alleviate non-differentiability and promote good mixing. However, estimation still remains a challenge due to algorithms being exact/inexact. We propose an importance sampling based estimator that uses samples from a proxy (MY envelope) distribution to estimate functions with respect to original target. Further, we show that the estimator so constructed is guaranteed to have finite variance. We also highlight the improved mixing achieved by targeting the MY envelope of the target distribution. Moreover, the asymptotic normality of our estimator is also established here and we demonstrate its efficiency over other alternatives from proximal MCMC across a wide variety of high and low dimensional problems.

424. Statistical modeling and mitigation techniques for sensor data imputation

[04.M2.I75, (page 48)]

Sharvari SHUKLA, *Symbiosis Statistical Institute, Symbiosis International (Deemed University)*

TBA

425. L-Estimation in Instrumental Variables Regression for Censored Data in Presence of Endogeneity and Dependent Errors

[Student Poster Competition, (page 21)]

SWATI SHUKLA, *Indian Institute Of Technology Kanpur*
Subhra Sankar DHAR, *Indian Institute Of Technology Kanpur*

In this talk, I shall discuss L-estimators for unknown parameters in instrumental variables regression when dealing with censored data under endogeneity. The model permits dependencies among the random errors. The proposed estimation method involves a two-stage procedure, and the asymptotic properties of the estimators are derived.

426. The Spillover Effect of a Minimum Wage Increase on Commission based Sales Agents at a Multi-level Marketing Firm

[02.M2.I28, (page 18)]

Sivaramakrishnan SIDDARTH, *Marshall School of Business, University of Southern California*
Sung Joo KIM, *Purdue University*
Wreetabrata KAR, *SUNY Buffalo*
Dinesh PURANAM, *Marshall School of Business, USC*

Multi-level marketing (MLM) firms such as Avon or Amway rely on a large salesforce that is not subject to minimum wage regulations, and therefore termed "uncovered" workers. However, an increase in the minimum wage could have spillover effects on these firms due to changes in factors such as labor supply (impacting employee recruitment and retention) and/or consumer purchasing power (affecting sales). Should MLM managers and policy makers be concerned about such spillover effects? To address this question, we leverage a quasi-experiment in which the city of Chicago increased the minimum wage by 21%, while the rest of Illinois did not. Our results show that sales agent retention and sales volumes at a leading MLM firm declined by 14% and 13%, respectively. Further analysis of sales agents who stayed with the firm after the wage hike reveals that they retained fewer frontlines but maintained sales volumes and commission rates. An examination of heterogeneous effects across different segments of sales agents shows that frontline retention was higher among the more economically vulnerable segments of the firm's sales agents, including women and minorities. Some agent sub-groups such as low-income workers registered an increase in sales. We discuss implications for managers and policy makers.

427. Functional Autoregressive Model for Forecasting Prices of Major Pulses in India

[05.M1.C30, (page 61)]

AKARSH SINGH, *ICAR-indian Agricultural Research Institute*
Ranjit Kumar PAUL, *ICAR-Indian Agricultural Statistics Research Institute*
Md YEASIN, *ICAR-Indian Agricultural Statistics Research Institute*

Linear models often fall short in capturing the complexities of real-world phenomena, prompting the exploration of non-linear time-series analysis methods. One such versatile non-parametric model is the Functional Autoregressive (FAR) model, where coefficient functions evolve gradually rather than abruptly. This paper discusses the properties of the FAR model and outlines its coefficient estima-

tion procedure using local regression techniques. An empirical analysis is conducted on aggregate price data for three crops (Arhar, Gram, and Mung) to compare the FAR model with traditional autoregressive (AR) approaches. Performance evaluation using Root mean square error (RMSE), Mean absolute error (MAE), and Mean absolute percentage error (MAPE) metrics demonstrates that the FAR model consistently outperforms the AR approach in predicting crop prices accurately. These findings highlight the FAR model's potential to enhance multistep forecasts and support decision-making in agricultural markets.

428. Power optimal designs for comparing a set of controls to a set of treatments

[04.L1.C20, (page 49)]

Arpan SINGH, *IIT Hyderabad*
Satya Prakash SINGH, *IIT Kanpur*

In various experiments the researchers are interested in the pairwise comparison of a set of already existing treatments (controls) with a set of newly developed treatments. Usually, experimenters pair every control with several new treatment and then simultaneously compare these pairs. In such experiments, the following two hypotheses are generally of interest: (1) Is there a significant difference in *at least one* of the selected pairs? (2) Is there a significant difference in *all* selected pairs? These type of simultaneous comparisons can be seen as a *bi-partite* graph with one set of vertices as the controls and the other as the treatments. In this article, we propose max–min designs for the hypothesis testing problems of the type (1) and (2) for these bi-partite graphs based on an elegant game theoretic approach. Some of the max-min designs obtained are well known with respect to different optimality criteria, and other max–min designs are novel. Theoretical findings are supplemented with numerical studies based on a real data example.

429. Mappable Nearly Orthogonal Arrays using Projective Geometry

[01.M2.I7, (page 5)]

Poonam SINGH, *Department of Statistics, University of Delhi, Delhi*

Mukerjee et al. (2014) introduced the concept of Mappable Nearly Orthogonal Arrays (MNOAs) and constructed these arrays by using resolvable orthogonal arrays. These arrays are used in computer experiments and other fields as space filling designs due to their inherent space filling properties. In this work, a method using projective geometry is developed for the construction of MNOAs. The proposed method is illustrated through suitable examples and the constructed arrays are tabulated along with corresponding values of degree of orthogonality.

430. Batting strike rate inTwenty20 and One Day International Cricket: A comparison

[03.E1.C13, (page 40)]

RK Renin SINGH, *Chitkara University*
Subrat SARANGI, *MICA*
Dhawal JADHAV, *MICA*

Sporting events are an increasing source of revenue to the media and entertainment industry globally. Cricket as a sport has been growing significantly over the last decade in terms of sponsorship by corporates and adoption globally. Viewership in cricket has always been motivated by the excitement the game creates for fans through boundaries and sixes scored by batters. In this study we analyze the batting strike rate by national side, inning of play, country of play and across the One Day International (ODI) and Twenty20 (T20) formats for men's international cricket. We find that batting is more aggressive in T20 format and the batting strategy between T20 and ODI cricket is clearly differentiated in South Africa and Zimbabwe where the difference in the strike rates is high. Studying the interaction effect between the format and inning of play, we establish that while the overall model is statistically significant and independently as well batting strike rate varies with format and inning, the interaction effect of both is not significant. The study has implications for the organizers, sponsors, and coaching staff in planning and scheduling matches based on the excitement that the specific teams can create across the formats.

431. On max–min designs for multiple comparisons for Binary outcomes

[04.A1.I78, (page 51)]

Satya SINGH, *Indian Institute of Technology Kanpur*
Ori DAVIDOV, *University of Haifa Israel*
Dharm SINGH, *Indian Institute of Technology Kanpur*

Within various experimental frameworks, the focus often lies in comparing several pairs of treatments. In this paper the problem of optimal alloca-

tion of experimental units to the experimental groups is addressed for the binary responses. It utilises a power-based design criterion to determine optimal designs under various settings of multiple comparisons. A max–min approach is adopted to tackle the issue of parameter dependence on the design criterion. The optimal designs obtained are well known, but not recognized as optimal within the framework of binary outcomes.

432. ESG Performance and Sustainable Investment - Company's Comparative Analysis using the TOPSIS Method

[01.A1.I14, (page 8)]

Rushina SINGHI, *NMIMS (Deemed-to-be) University, Mumbai*

TBA

433. Analysis of spatially clustered survival data with unobserved covariates using SBART

[02.E1.I42, (page 27)]

Debajyoti SINHA, *Florida State University*

TBA

434. A multi-region discrete time chain binomial model for infectious disease transmission

[02.E1.C8, (page 29)]

Pallab Kumar SINHA, *Indian Institute of Technology Bombay*
Siuli MUKHOPADHYAY, *Department of Mathematics, Indian Institute of Technology Bombay*

Conventional mathematical models of infectious diseases frequently overlook the spatial spread of the disease concentrating only on local transmission. However, spatial propagation of various diseases have been noted between geographical regions mainly due to the movement of infectious individuals from one region to another. In this work, we propose a multi-region discrete time chain binomial framework to model dependencies between the multiple infection time series from neighbouring regions. It is assumed that the infection counts in each region at various time points is not only governed by local transmissions but also by interactions of individuals between spatial units. Effect of intervention strategies like vaccination campaigns used in disease control and various other socio-demographic factors like, live births, population density, vaccination coverage, disruption in disease surveillance, sudden surge in birth rates have been taken into account while modelling the multiple infection time series. For estimating this multi-region chain-binomial model with spatial autocorrelation among the units, an appropriate likelihood function maximization approach is proposed. Simulation results considering effects of seasonal pattern in disease outbreaks, out of sync outbreaks in connected geographical regions, variations in reporting rates in connecting regions have been considered to depict realistic disease scenarios. Various methods of forecasting of infections and effect of spatial heterogeneity on future infections have also been studied. A real world application based on measles counts from adjoining spatial regions is presented to motivate the proposed modelling approach.

435 . CMPLE: Correlation Modeling to Decode Photosynthesis Using the Minorize–Maximize Algorithm

[02.E1.I42, (page 27)]

Samiran SINHA, *Texas A&M University*
Abhijnan CHATTOPADHYAY,
David KRAMER,
Tapabrata MAITI,

In a multivariate setup, modeling the correlation between phenotypes with genetic and environmental predictors presents significant challenges. These challenges stem from the need to satisfy the positive definiteness constraint and the complexity of estimating a large number of parameters. In this talk, I will present a novel approach to address these difficulties in this regression modeling. The proposed methodology is supported by a simulation study and an analysis of real data from a photosynthesis experiment.

436. Bayesian deep generative reinforcement learning

[Student Poster Competition, (page 20)]

Shreya SINHA ROY, *University of Warwick*
Ritabrata DUTTA, *University of Warwick*
Richard EVERITT, *University of Warwick*
Christian ROBERT, *Université Paris-Dauphine (France) , University of Warwick*

We present a Bayesian model-based reinforcement learning approach that integrates Bayesian statistics with reinforcement learning to enable decision-making in uncertain environments. Our method con-

sists of two main components: (1) inferring the true Data Generating Process (DGP) and (2) learning an optimal policy. We introduce a likelihood-free framework for Bayesian inference of the DGP, modelling the dynamics of the unknown environment using deep generative models. This is achieved by minimizing a class of objective functions defined via predictive-sequential (or prequential) scoring rules, under the assumption that the DGP follows a Markov decision process. To infer the model parameters, we employ Sequential Monte Carlo (SMC) samplers, combined with gradient-based MCMC kernels, to effectively handle the high-dimensional parameter space. We also establish a Bernstein-von Mises type theorem, demonstrating the contraction of the generalized posterior distribution towards the true parameter values. For policy learning, we explore Thompson sampling and its extensions, which are known for their strong performance and favourable regret bounds. These methods typically rely on a single sample drawn from the posterior distribution to guide policy exploration. However, there has been limited exploration in the literature regarding the potential benefits of using multiple samples from the inferred posterior. Building on this insight, we propose a novel policy learning approach called Expected Thompson Sampling. This method learns the optimal policy by maximizing the expected value function with respect to the inferred posterior distribution. Our results show significant performance improvements, and we provide a regret analysis to support the effectiveness of the proposed method.

437. Mathematical Modeling of Malaria Transmission in Mumbai

[03.E1.C15, (page 41)]

Adithya SOMARAJ, *National Disease Modeling Consortium, IIT Bombay*
Usha ANANTHAKUMAR, *Shailesh J. Mehta School of Management, IIT Bombay*

Malaria remains a significant public health challenge in Mumbai, Maharashtra. This study employs a susceptible-infected-recovered-susceptible (SIRS) compartmental mathematical model described in the literature to analyze malaria transmission dynamics in Mumbai. The model uses a saturated treatment function and logistic growth for the mosquito population, which is appropriate for resource-constrained regions like Mumbai. In this study, the estimation of the key parameter of the model pertaining to the transmission of Malaria resulting from the interaction of susceptible humans and infected vectors (1) is performed. This is carried out by fitting the model to the observed cumulative annual malaria cases in Mumbai from 2012 to 2019 using the least-squares curve fitting technique. The results of the model provide a good fit when compared with the actual malaria cases during the time period considered. Furthermore, using the estimated parameters, the basic reproduction number (R0) is calculated, and sensitivity analysis is performed on the model to identify the significant parameters influencing malaria transmission in Mumbai. This study emphasizes the critical role of effective treatment and vector control strategies in mitigating malaria spread. The findings provide a framework for understanding malaria dynamics in urban settings and inform public health interventions.

438. Semiparametric regression analysis of doubly censored recurrent event data

[04.A1.I82, (page 52)]

Sreedevi E. P., *CUSAT*
Sankaran P. G., *CUSAT*
Hari S., *CUSAT*

Recurrent event data are common in survival and reliability studies, where a subject experiences the same type of event repeatedly. There are situations, in which the event of interest can be observed only if they belong to a window of observational range, leading to double censoring of recurrent event times. In this paper, we study recurrent event data subject to double censoring. We propose a proportional mean model for the analysis of doubly censored recurrent event data based on the mean function of the underlying recurrent event process. The estimators of the regression parameters and the baseline mean function are derived and their asymptotic properties are studied. A Monte Carlo simulation study is conducted to assess the finite sample behaviour of the proposed estimators. Finally, the procedures are illustrated using two real life data sets, one from a bladder cancer study and the other from a study on chronic granulomatous disease.

439. Data-driven crime prediction with human inputss

[02.A1.I37, (page 25)]

Karthik SRIRAM, *IIM Ahmedabad*
Ankur SINHA, *IIM Ahmedabad*
Suvashis CHOUDHARY, *Police Department (retired)*

Predictive hotspot mapping is an important problem in crime prediction and control. An accurate hotspot mapping helps in appropriately targeting the available resources to manage crime in cities. With an aim to make data-driven decisions and automate policing and patrolling operations, police departments across the world are moving towards predictive approaches relying on historical data. Most approaches studied in the literature suffer from the drawback that they do not allow human inputs to be integrated in predictive hotspot mapping. In this paper, we make methodological contributions in the context of spatio-temporal kernel density estimation to make it amenable for the purpose of crime prediction based on historical data as well as human inputs. The proposed approach has been evaluated in a real-world setting by collaborating with the Delhi police department to make crime predictions that would help in effective assignment of patrol vehicles to control street crime. The results obtained in the paper are promising and can be applied in other settings.

440 . Robust Recovery of the Central Subspace for Regression Using the Influence Function of the Renyi Divergence

[04.E1.I83, (page 54)]

T. N. SRIRAM, *University of Georgia*
Ross IACI, *William and Mary*

A considerable amount of research in the literature has focused on quantifying the effect of extreme observations on classical methods for estimating the Central Subspace (CS) for regression through the study of influence functions and their sample estimates. Alternatively, a method that is inherently robust to data contamination is also important and desirable for the increased reliability in the estimation of the CS without relying on the identification and removal of influential values. To this end, we develop a new method that is innately resistant to outlying observations in recovering a dimension reduction subspace for regression based on the Renyi divergence. In addition to deriving the theoretical Influence Function (IF), the Sample Influence Function (SIF) values are directly utilized to provide new powerful and efficient methods for both estimating the dimension of the CS and selecting an optimal level of the tuning parameter to decrease the impact of extreme observations. The model-free approach is detailed theoretically, its performance investigated through simulation, and the application in practice is demonstrated through a real data analysis.

441. Robustness and Overparameterization

[03.A1.I58, (page 38)]

Piyush SRIVASTAVA, *Tata Institute of Fundamental Research*
Santanu DAS, *Tata Institute of Fundamental Research*
Jatin BATRA, *Tata Institute of Fundamental Research*

In contemporary deep learning practice, models are often trained to near zero loss i.e. to nearly interpolate the training data. However, the number of parameters in the model is usually far more than the number of data points, the theoretical minimum needed for interpolation: a phenomenon referred to as overparameterization. In an interesting piece of work that contributes to the considerable research that has been devoted to understand overparameterization, Bubeck and Sellke showed that for a broad class of covariate distributions (specifically those satisfying a natural notion of concentration of measure), overparameterization is necessary for robust interpolation i.e. if the interpolating function is required to be Lipschitz. However, their robustness results were proved only in the setting of regression with square loss. In practice, however many other kinds of losses are used, e.g. cross entropy loss for classification. We identify a bias-variance type decomposition that lies at the heart of the proof and Bubeck and Sellke, and use this observation to generalize Bubeck and Selke's result to Bregman divergence losses, which form a common generalization of square loss and cross-entropy loss.

442 . Statistical modelling and uncertainty quantification of temperature inversion in the Bay of Bengal

[02.M2.I29, (page 19)]

Radhendushka SRIVASTAVA, *Indian Institute of Technology Bombay*
Suresh IYYAPPAN, *Digital University Kerala*
Jovi D'SILVA, *CSIR-National Institute of Oceanography, Goa*

The Bay of Bengal (BoB) is periodically infused with large amount of fresh water from rivers and rainfall during the Indian summer monsoon season. This affects the near surface salinity stratification, leading to increasing temperatures with depth due to the formation of warm subsurface layer sandwiched be-

tween surface and subsurface colder waters. This special phenomenon in BoB is referred to as inversion in the vertical temperature profile that generally occurs during the winter season (November–March). In this work, we consider the vertical temperature profile from Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction (RAMA) buoys in the BoB since 2007. We estimate the long-term trend and periodic characteristics of the multivariate temperature inversion time series within the inversion layer for forecasting. We model the residual series as a stationary time series. We will describe the covariance structure of this residual series to estimate the standard errors of the inversion forecast. We also illustrate a model-based bootstrap method to quantify the uncertainty in the inversion forecast.

443. Precise generalization error of min-norm interpolants under transfer learning

[02.M1.I23, (page 15)]

Pragya SUR, *Harvard University*

TBA

444. Modelling and Analysis of Bivariate Lifetime Data using Additive Hazards Survival Model

[04.E1.C24, (page 56)]

Namitha SURESH, *Cochin University of Science and Technology*
S. M. SUNOJ, *Cochin University of Science and Technology*
N Unnikrishnan NAIR, *Cochin University of Science and Technology*

When each participant may encounter several events during a survival study, multivariate lifetime data, also known as correlated or clustered lifetime data, frequently emerge. When analyzing such data statistically, intra-cluster dependence must be taken into account. In this article, a bivariate additive hazards model that uses the vector hazard rate is examined. Ageing properties, dependence measures, and stochastic orderings related to the proposed model are examined. The model is validated using a real-life dataset.

445. Jackknife empirical likelihood ratio test for testing the equality of semivariance

[04.M2.I72, (page 47)]

Saparya SURESH, *Indian Institute of Management, Kozhikode*

TBA

446 . The Corrected Likelihood Approach for Adjusting Measurement Error in Cox's Model

[04.L1.C20, (page 50)]

Anu SUSAN GEORGE, *Cochin University of Science and Technology*
Asha GOPALAKRISHNAN, *Cochin University of Science and Technology*

While dealing with survival data, we may come across situations where some of the covariate(s) affecting the failure time are prone to measurement error. Further, the effect of the model covariates on the event time may vary for different groups in the underlying population. The paper discusses about modeling and estimating the effect of mis-measured model covariates on the event times, where the effects are dependent on an auxiliary covariate that categorizes the population into groups. A corrected likelihood approach that can eliminate the bias due to the measurement error, is used to estimate the parameters.

447 . Reliability Modeling of Unequal Load Sharing Systems Using the Accelerated Failure Time Model

[04.M2.I74, (page 48)]

Santosh SUTAR, *Shivaji University, Kolhapur, India*
Sukumar RAJGURU, *Shivaji University, Kolhapur*

In systems where components share unequal loads, the failure of one component can impact the load distribution among the remaining components, either increasing or decreasing their likelihood of failure. This article explores reliability modeling for such unequal k-out-of-m load-sharing systems, specifically focusing on a 2-out-of-4 system where component lifetimes are assumed to follow either exponential or Weibull distributions. The load-sharing phenomenon is modeled using the Accelerated Failure Time model, applying an unequal load-sharing rule. The article also covers inferential procedures and presents a simulation study to evaluate model performance, along with an analysis of a real-world dataset for practical illustration.

448. Bivariate Exponential Distribution through Entropy Optimization

[03.A1.I60, (page 38)]

Princy T, *Cochin University of Science and Technology*
Sneha BABU, *Cochin University of Science and Technology*

Bivariate exponential model is one of the most popular continuous bivariate distributions. In this work, we have derived a generalized version of bivariate exponential models through entropy optimization and we call these models the q-bivariate exponential models. One of the major properties of the q-bivariate exponential model is that its marginal densities are q-exponential distributions. The joint probability density function, the joint cumulative distribution function, and the joint survival distribution function can be expressed in compact forms. Several properties of this distribution have been discussed. One data set has been analysed for illustrative purposes.

449. Bulk Spectra of Truncated Sample Covariance Matrices

[02.E1.C6, (page 27)]

Himasish TALUKDAR, *Indian Statistical Institute, Kolkata*
Subhroshekhar GHOSH, *National University of Singapore*
Soumendu Sundar MUKHERJEE, *Indian Statistical Institute, Kolkata*

Determinantal Point Processes (DPPs), which originate from quantum and statistical physics, are known for modeling diversity. Recent research [Ghosh and Rigollet, 2020] has demonstrated that certain matrix-valued U-statistics (that are truncated versions of the usual sample covariance matrix) can effectively estimate parameters in the context of Gaussian DPPs and enhance dimension reduction techniques, outperforming standard methods like PCA in clustering applications. This paper explores the spectral properties of these matrix-valued U-statistics in the *null* setting of an isotropic design. These matrices may be represented as $XLX^\top$, where X is a data matrix and L is the Laplacian matrix of a random geometric graph associated to X. The main mathematically interesting twist here is that the matrix L is dependent on X. We give complete descriptions of the bulk spectra of these matrix-valued U-statistics in terms of the Stieltjes transforms of their empirical spectral measures. The results and the techniques are in fact able to address a broader class of kernelised random matrices, connecting their limiting spectra to generalised Marčenko-Pastur laws and free probability.

450 . Threshold Exceedance Modeling for Spatial Extreme Data Using Neural Network

[02.M1.I22, (page 15)]

Debjoy THAKUR, *Washington University in St. Louis*
Soumendra N. LAHIRI, *Washington University in St. Louis*

Most of the applications of neural networks in classical extreme value theory generally ignore the local spatial dependence in parameter estimation and sparse extremal regression. In this study, we develop a novel spatially regularized local neural network-based framework for threshold exceedance modeling of spatial extreme data. This article focuses on parameter estimation of the probability distribution and proposes a local extremal regression in threshold exceedance setup. Here we utilize the Peaks-Over-Threshold (POT) algorithm to model exceedances over the spatial surface incorporating the Generalized Pareto Distribution (GPD) to capture the spatial variability of extreme events. For parameter estimation we estimate the shape and scale parameter of GPD for every location exceeding the threshold incorporating the local spatial dependence in neural network framework. We have employed the spatially regularized neural network for the data exceeding the threshold. The dominance in accuracy, runtime efficiency, and spatial continuity of the proposed method in comparison with the traditional methods is validated by thorough simulations and real-world data from the Red Sea temperature. These findings account for local spatial dependence in threshold exceedance modeling, paving the way for more robust and accurate predictions of extreme spatial modeling.

451 . Statistical Methodologies for Pediatric Drug Development: Enhancing Safety and Efficacy

[01.A1.I11, (page 7)]

Anil Kumar THUKKAMATTATHIL, *Stat Analytics*

Pediatric drug development presents unique challenges due to the physiological differences and ethical considerations involved in testing medications in children. This presentation delves into the sta-

tistical methodologies crucial for ensuring the safety and efficacy of pediatric drugs. It explores strategies for extrapolating adult data to pediatric populations, adaptive trial designs, and innovative statistical approaches for handling small sample sizes. Additionally, the presentation addresses ethical considerations, regulatory requirements, and the role of real-world evidence in pediatric drug development. Through a comprehensive examination of statistical methodologies, this presentation aims to provide insights into advancing pediatric drug development while safeguarding the well-being of young patients.

452. A Study On Some New Bivariate Reliability Measures.

[04.E1.C24, (page 55)]

ATHIRA T S, *UNIVERSITY OF KERALA*
E I ABDUL SATHAR, *UNIVERSITY OF KERALA*

We introduce a new bivariate reliability measure which plays a crucial role in analyzing the reliability and survival characteristics of multi-component systems, where dependencies between lifetimes of components are considered. We define and study its properties, including uniqueness and applications in empirical and non- parametric kernel estimation. Simulation studies and real data analysis demonstrate the performance.

453. Exact Inference For The Exponential Distribution Using Type-I Censoring on Stage Life testing Experiments

[05.M1.C29, (page 61)]

Asad UDDIN, *Indian Institute of Information Technology Guwahati*
Erhard CRAMER, *RWTH Aachen University*
Farha SULTANA, *Indian Institute of Information technology Guwahati*

The distribution of maximum likelihood estimators (MLEs) are in general obtained using moment generating function. In this paper, we have used an alternative method known as the expected value approach, which was introduced by [Gorny2017b], to obtain the exact distribution of the MLEs. We have derived the exact inference for the exponential distribution when Type-I censoring is implemented on stage life testing (SLT) introduced by [Laumen2019]. We assume that the lifetime distribution of experimental units follows the exponential distribution in both stages. The distributions under two stress levels are connected through the cumulative exposure (CE) model. We derive the MLEs for the unknown model parameters and subsequently obtain the conditional moment generating function (CMGF) and the exact distribution of the MLEs. Additionally, we derive asymptotic confidence intervals (ACI) based on the asymptotic normality of MLEs, and we also employ bootstrap methods to obtain the Boot-p and Boot-t confidence intervals. Furthermore, we obtain Bayes estimators (BEs) under the squared error loss (SEL) function using informative prior for the unknown model parameters. A comparisons of the MLEs and BEs are made in terms of their mean squared errors (MSEs). An extensive simulation study and a data analysis are carried out to demonstrate the effectiveness of our approach.

454 . Competing for space: Self-organization among a group of crowd-avoiding agents

[04.A1.I81, (page 52)]

Sasidevan V., *Assistant Professor, Dept. of Physics, Cochin University of Science and Technology, Cochin*

TBA

455. Cross impact analysis based models for technology forecasting in agriculture

[01.M2.I4, (page 4)]

Ramasubramanian VAIDHYANATHAN, *ICAR-National Academy of Agricultural Research Management, Hyderabad*
Umesh HUDEDAMANI, *ICAR-National Academy of Agricultural Research Management, Hyderabad*
Abin GEORGE, *ICAR-National Academy of Agricultural Research Management, Hyderabad*
Mrinmoy RAY, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

One of the tools of technology forecasting for the purpose of systematically exploring, creating and testing both possible and desirable futures, that aids better decision making, is the Cross-Impact Analysis (CIA). Classical CIA techniques are of many types depending on their data requirements, methodological steps and the results obtained but can be broadly binned into three approaches viz., Structural (e.g., MICMAC-Matrice d'Impacts Croisés Multiplication Appliqués à un Classement i.e. Cross Impact Matrix Multiplication Applied to Classification, KSIM-Kane's Simulation Model etc.), Morphologi-

cal (e.g., CIB-Cross Impact Balances approach) and Probability (e.g., SMIC-Cross Impact Systems and Matrices, AXIOM- Advanced Cross-Impact Option Method, Markov based CIA etc.) based on the orientations of outputs they offer. In this study, an extensive review on CIA on the various techniques that exists till date and also case studies of CIA viz., KSIM, MICMAC, CIA with Time (CIAT) in agricultural and allied domains that were done, specifically in envisioning Bacillus Thuringiensis-Bt Technology based crops, Export/ Import of Cotton and separately Fisheries in Indian context have been discussed. The steps of SMIC have also been demonstrated which determines simple/ conditional probabilities of 'n' events and their specific combinations that, in turn, enables to choose from among the 2n possible scenarios. As the results obtained through conventional CIA remain constant over time, alternatively Markov chains have been also integrated into CIA to make it dynamic allowing the probabilities of events to change over time and have also been attempted in an agricultural context and compared with its conventional counterparts.

456. A statistical framework for analyzing shape in a time series of random geometric objects

[03.A1.I54, (page 36)]

Anne VAN DELFT, *Columbia University*
Anne VAN DELFT, *Columbia University*
Andrew J. BLUMBERG, *Columbia University*

We introduce a new framework to analyze shape descriptors that capture the geometric features of an ensemble of point clouds. At the core of our approach is the point of view that the data arises as sampled recordings from a metric space-valued stochastic process, possibly of nonstationary nature, thereby integrating geometric data analysis into the realm of functional time series analysis. Our framework allows for natural incorporation of spatial-temporal dynamics, heterogeneous sampling, and the study of convergence rates. Further, we derive complete invariants for classes of metric space-valued stochastic processes in the spirit of Gromov, and relate these invariants to so-called ball volume processes. Under mild dependence conditions, a weak invariance principle in $D([0,1] \times [0,R])$ is established for sequential empirical versions of the latter, assuming the probabilistic structure possibly changes over time. Finally, we use this result to introduce novel test statistics for topological change, which are distribution-free in the limit under the hypothesis of stationarity. We explore these test statistics on time series of single-cell mRNA expression data, using shape descriptors coming from topological data analysis.

457 . ANALYSIS OF NON-CRISP DATA DERIVED FROM DESIGNED EXPERIMENTS

[04.E1.I85, (page 54)]

Cini VARGHESE, *ICAR-IASRI, PUSA, Library Avenue, New Delhi - 110 012*

Many a times, data collected from scientific experimentation have some degree of indeterminacy. It is not advisable to ignore the vague values and taking the mid value. Fuzzy logic concept was introduced to deal with such situations, which in turn, further generalized to neutrosophic logic. Neutrosophic Statistics is a more generalized form of classical Statistics, where one deals with range values instead of crisp values. Rules pertaining to various operations on netrosophic data are different and exclusively defined. Neutrosophic Analysis of Variance (NANOVA) procedures for few basic experimental designs are developed in literature. Analytical procedures for neutrosophic data derived from different types of row column design set ups have been derived.

458. Quantile based Analysis of Residual Incomes

[01.E1.C5, (page 13)]

Ashlin VARKEY, *Farook College (Autonomous), Kozhikode*
Haritha N HARIDAS, *Farook College (Autonomous), Kozhikode*
Ashlin VARKEY, *Farook College (Autonomous), Kozhikode*

The concept of residual incomes helps to study poverty and affluence. In the present work, we define median residual income and, in general, quantile residual income for the poor and the rich. We have also derived these measures for some income distributions having explicit quantile functions.

459 . A New Formulation of Minimum Risk Fixed-Size Confidence Region (MRFSCR) Estimation Problems for a Multivariate Normal Mean with Illustrations Incorporating Covariance Structure via Positive Definite Matrix

[04.M2.I73, (page 47)]
Swathi VENKATESAN, *Fairfield University*
Nitis MUKHOPADHYAY, *University of Connecticut*

In traditional fixed-size confidence region (FSCR) estimation, ellipsoidal confidence regions of fixed diameter are often constructed arbitrarily for the mean vector of a multivariate normal distribution. However, it is more appropriate to determine the size of the confidence region based on the quality of available data. This paper introduces a novel formulation of FSCR for the multinormal mean, specifically considering the dispersion matrix in the form of $\sigma^2 H$, where the size of the confidence region is expressed as a function of σ. Unlike minimum risk point estimation (MRPE) problems, which incorporate explicit loss functions to balance estimation error and sampling cost, FSCR lacks such considerations. Motivated by this, a new framework is proposed: minimum risk fixed-size confidence region (MRFSCR) for the mean vector in multivariate normal distributions. Additionally, a unified structure for multistage sampling strategies to construct MRFSCRs is presented, with demonstrated asymptotic first-order and second-order properties. Illustrations and analyses using simulated data complement the theoretical and methodological discussions.

460. Exploring Quantile-Based Cumulative Residual Extropy of Record Values: A Comprehensive Analysis.

[01.E1.C1, (page 10)]
VEENA L VIJAYAN, *Department of Statistics, University of Kerala*
E I Abdul SATHAR, *University of Kerala*

Recent research has shown an increasing interest in employing quantile-based methods to evaluate uncertainty linked with random variables. In contrast to traditional distribution function techniques, quantile-based measurements offer distinct viewpoints. This paper examines the extropy of recorded values through the introduction of a novel quantile-based approach and explores its characteristics. Furthermore, a nonparametric estimator is introduced and applied to this new metric using distributions commonly utilized in lifetime data analysis.

461. Wavelet-Based Multiscale Analysis for Enhanced Detection of Heart Murmurs in Phonocardiographic Signals

[02.E1.I42, (page 27)]
Horahenage VIMALAJEEWA, *University of Nebraska-Lincoln*

TBA

462. On a class of AR(1) model for Z valued time series

[04.L1.C20, (page 49)]
Arathi VINAYAN, *Govt. Arts and science college Calicut, Calicut 673018*
Jilesh V, *Govt . Arts and Science and College Calicut, Calicut 673018*

This paper proposes a new stationary integer-valued autoregressive (1) model for Z valued time series. The proposed model is suitable for both positively and negatively correlated time series data. Various statistical properties of the model are examined. We considered different methods for estimating the parameters of the model and explored the properties of the estimators. The performance of the estimators are compared via simulation study. To illustrate the significance of the given model, a real data analysis is presented.

463. Generalized Birth-Death Process on Finite Lattice

[04.E1.C28, (page 58)]
Pradeep VISHWAKARMA, *Indian Institute of Technology Bhilai*
Kuldeep Kumar KATARIA, *Indian Institute of Technology Bhilai*

We consider a birth-death process whose state space is a finite subset of finite q-dimensional lattice. Here, it is assumed that at any instance there can be multiple but finitely many transition in every possible direction. Also, we assume that at any point of time transition will take place in exactly one direction with some positive probability, and the probability of simultaneous transition in more than one directions is zero. Such processes are of interest as their transition probability matrix is diagonalizable under suitable conditions. Thus, their k-step transition probabilities can be efficiently obtained. For a generalized birth-death process on two dimensional finite grid, we obtain a sufficient and necessary condition for the vertical and horizontal transition probability matrices to commute. Later, we extend these results to the case of q-dimensional finite grid. We obtain the minimal number of constraints required on transition probabilities that ensure the commutation of transi-

tion probability matrices for each possible direction.

464. Testing Exponentiality for Progressively Type II Censored Data Using Equilibrium Distributions

[03.A1.C12, (page 39)]

Sajily V S, *Cochin University of Science and Technology*
Rajesh G, *Cochin University of Science and Technology*

The relative entropy (Kullback-Leibler information) and relative extropy are tools for quantifying the dissimilarity between two probability density functions. We express both measures in terms of distribution functions by using equilibrium distributions instead of original density functions. And find the estimates of these two measures for progressively Type II censored data, utilizing them as a statistic for testing exponentiality, motivated by the fact that the exponential distribution is the only one for which the equilibrium distribution coincides with the original distribution. We conducted Monte Carlo simulations to evaluate the performance of the proposed test statistics. Also, we compared the proposed statistics with previously defined test statistics based on divergence measures for progressively Type II censored data. Finally, we illustrated the application of the proposed statistics with a real-life example.

465. Testing Informative Intra-Cluster Group Sizes in Clustered Data

[03.A1.I58, (page 37)]

Hasika WICKRAMA SENEVIRATHNE, *National University of Singapore*
Sandipan DUTTA, *Old Dominion University*

Clustered data is a special type of correlated data where units within a cluster are correlated while units between different clusters are independent. The number of units in a cluster (cluster size) can be associated with the cluster's outcome. This is known as informative cluster sizes (ICS) which impacts clustered data inference. Recently, a hypothesis testing procedure has been developed to test the presence of ICS. However, when comparing the outcomes from multiple groups of units in a clustered data, considering ICS may be insufficient. Here, the number of units in a group within a cluster (intra-cluster group size) can be correlated with the group's outcomes in that cluster causing informative intra-cluster group sizes (IICGS). IICGS can exist even in the absence of ICS in clustered data. Motivated by this, we developed a statistical hypothesis-testing method that can test a claim of IICGS in clustered data. Through simulated data and real data applications, we demonstrate that our proposed testing method can accurately identify IICGS, with substantial power, in clustered data.

466. Selection of optimal time window in restricted mean survival time under delayed treatment effect using statistical distributions

[01.E1.C3, (page 12)]

Thomas XAVIER, *Novartis, India*

An appropriate statistical method of a randomized controlled trial with a time-to-event endpoint under non-proportional hazards owing to a delayed effect has been the subject of major debate. Restricted mean survival time (RMST) is considered as one of the measures to handle such situations, but selection of time window has been an important concern for its implementation. The selection of time window can be subjective and hence motivating us to look at methods to support the selection of time window. We propose a method where we use statistical distributions and theory of Monte-Carlo simulations which can help to come up with an appropriate optimal time window.

467. Sparse Spatiotemporal Dynamic Generalized Linear Models for Inference and Prediction of Bike Counts

[02.A1.I32, (page 23)]

Rishikesh YADAV, *Indian Institute of Technology Mandi*

In this work, we propose a novel sparse spatiotemporal dynamic generalized linear model for efficient inference and prediction of bike count data across Montreal Island. Assuming bike counts follow a Poisson distribution conditional on spacetime-varying rate parameters, we incorporate spatiotemporal intercepts, dynamic temporal covariates, and site-specific covariates additively to model various desired components in the Poisson rate. Spatiotemporal dependence is modeled using a spacetime-varying intercept that evolves dynamically over time with spatially correlated errors, and coefficients of some temporal covariates including seasonal harmonics also evolve dynamically over time. Inference is performed following the Bayesian paradigm and uncertainty

quantification is naturally accounted for when predicting unobserved locations and future times of interest. To address the challenges of high-dimensional inference of spatiotemporal data in a Bayesian setting, we develop a customized hybrid Markov Chain Monte Carlo (MCMC) algorithm that utilizes the gradient and Hessian of the target distribution for model components without closed-form full conditionals. Furthermore, motivated by the high computational cost associated with dense covariance matrices, we extend our modeling framework to higher spatial dimensions using the sparse SPDE approach of Lindgren et al. (2011) and demonstrate its accuracy and scalability with synthetic data and real applications to bike datasets of Montreal Island. The proposed approach naturally provides missing value imputations, kriging, future forecasting, spatiotemporal predictions, and inference of model components. Moreover, it provides ways for the predictions of annual average daily bikes (AADB), a key metric often sought when designing bike networks. The proposed model is compared with several competing models including the recently developed R package BKTR, and showcases the superior performance.

468 . Two sample test of high-dimensional means test under missing observations

[03.E1.C14, (page 40)]

Shiv Kumar Yadav YADAV, *Indian Institute if Bombay*

Statistical test for high-dimensional means under missing observations seems to be very rare in the literature. I propose a new two sample test for high dimensional means based on independent observations with missing values. The critical region of the proposed test is based on a bootstrap estimate of the sample quantile of the proposed test statistic. Unlike the existing tests, this test does not require any distributional assumptions or any particular correlation structure of the covariance matrices. I establish a Gaussian approximation result for proposed test statistic which is a non-trivial extension of two-sample Gaussian approximation with no missing values. The rate of accuracy of the bootstrap approximation the sample quantile of the proposed test statistics also derived. This Gaussian approximations result and the accuracy of bootstrap estimators together provide the theoretical guarantees on the size and power the proposed test.

469 . A New Approach to Mean Estimation under Ranked Set Sampling for Skewed Populations

[04.E1.C23, (page 55)]

Tanushree YADAV, *University of Allahabad*
Priyanka SINGH, *University of Allahabad*
Girish CHANDRA, *Delhi University*

Bhoj and Chandra (2021) proposed the Ranked Set Sampling (RSS) procedure with the Lowest order statistics (RSSL) to estimate the population mean of Pareto distribution. This paper assesses the performance of the mean estimator based on the RSSL procedure for Weibull distribution. Further, a new RSS procedure using the Highest order statistics (RSSH) has also been proposed to estimate the population mean of Weibull distribution. The expressions for Bias and MSE of estimators based on RSSL and RSSH procedures have been derived. Different forms (based on positive and negative skewness) of Weibull distributions are considered for the purpose. The analysis indicates that the gains in the relative precisions of the estimators, based on both RSSL and the proposed RSSH method, consistently exceed the other RSS procedures in the case of skewed populations. A simulation study validated the theoretical results carried over a randomly generated population.

Keywords: Order Statistics, Ranked Set Sample, Weibull Distribution, Bias, Mean Square Error

470 . Emotion-Aware Banking: An Artificial Intelligence Framework to Elevate Customer Interaction Experience Using Contextual Insights and Emotion Ontologies

[01.M2.I6, (page 5)]

Dr Naveen YERI, *Wells Fargo-Rennes School of Business*

Customers' financial decisions are backed by reasons that analytical approaches may not comprehend. We sense it through their experiences and needs. It is in the emotions and trust we cultivate that customers find security in our services, not merely in rational explanations. This, then, is the essence of banking: understanding customers through their emotions and trust, creating a foundation of reliability beyond mere transactional logic. In my talk, I expose the need to understand the emotions behind banking relationships, discover features that can provide tremendous explanatory power in analytics, and recommend a

framework to leverage AI toward customers' success.

471. WEIGHTED QUASI SUJA DISTRIBUTION AND ITS PROPERTIES

[04.A1.C22, (page 53)]

Vidya YERNENI, *Symbiosis Statistical Institute, Symbiosis international (Deemed University)*
Afaq A. RATHER,

There is a close relation between probability and data science, as probability theory forms the foundation of many data science techniques be it machine learning, A/B testing, Risk Analysis. A solid understanding of probability is necessary to build predictive models. In this study we explored and obtained one of the probability models namely a three parameter weighted quasi suja distribution. We derived its statistical properties like moments, generating functions, harmonic mean, reliability, hazard rate, reverse hazard rate, order statistics, Entropy measures like Renyi entropy, Tsalli's entropy, bonferroni and lorenz curves. Discussed the estimation of parameters using maximum likelihood method. Tested the efficiency of parameters of the proposed model using simulation study. We also studied the fitness of the proposed model over other probability models. Finally applied the proposed probability model on real data which proved that weighted quasi suja probability model can be used effectively in analysing real life data and for prediction purpose.

472. The Nested Interpretable Neural Network Model and Its Application In Evaluating Digital Competitiveness of Manufacturing Enterprises in China

[01.E1.I18, (page 9)]

William Weimin YOO, *Heriot-Watt University Malaysia*
Siwei LIU, *Heriot-Watt University Malaysia*
Sarat Chandra DASS, *Heriot-Watt University Malaysia*

Motivated by the study of digital competitiveness of manufacturing firms in China, we develop a nested interpretable neural network (NINN) model to systematically analyze the impact of digital transformation on these firms. We capture this impact by computing total factor productivity values using NINN. Different neural network architectures are used to model multimodal data inputs consisting of financial reports and control variables, and they are stacked together into a nested structure such that intermediate layers encode crucial information relating to indices on digital transformation, thus make NINN interpretable to corporate stakeholders. We study and quantify the influence of data inputs on these indices and the total factor productivity values, by extending the LIME and SHAP algorithms to take into account of the nested architecture. We apply NINN to manufacturing firms in Shanghai and Shenzhen, and use the nested versions of LIME and SHAP algorithms to evaluate the effectiveness of various digitalization initiatives.

473. Modeling Spatio-temporal Extremes with Conditional Variational Autoencoders

[02.A1.I32, (page 23)]

Likun ZHANG, *University of Missouri - Columbia*
Xiaoyu MA, *University of Missouri - Columbia*
Christopher K. WIKLE, *University of Missouri - Columbia*

Extreme weather events are widely studied in the fields of agriculture, climatology, ecology, and hydrology, to name a few. Enhanced scientific understanding of the spatio-temporal dynamics of extreme events could significantly improve policy formulation and decision-making within these domains. We formulate a novel approach to model spatio-temporal extremes by conditioning on a time series (e.g., the El Niño-Southern Oscillation (ENSO) index) via a conditional variational autoencoder (extreme-CVAE). The prominent alignment of extremal dependences showcase the model's ability to be a spatio-temporal extreme emulator. Along with a decoding path, a convolutional neural network was built to investigate the relationship between the time series dynamics and parameters within the latent space, thereby inheriting the intrinsic temporal dependence structures. An extensive simulation validated the effectiveness and time efficiency of the model. We conducted an analysis of the monthly precipitation data which adequately demonstrates both the time efficiency and model performance of our approach in real-world scenarios.

Directory

International Indian Statistical Association (IISA) 2024 Code of Conduct

The International Indian Statistical Association (IISA) is committed to providing a safe, productive, and welcoming environment for all participants in any conference, workshop, or event organized or co-organized by IISA. The organization is dedicated to fostering an inclusive atmosphere that promotes the free exchange of ideas while ensuring respect and professionalism among all attendees, staff, and volunteers.

To achieve this goal, IISA adheres to the following Code of Conduct:

Be Professional
All IISA conference participants are expected to maintain the highest standards of professionalism and respect for others. The attendees are expected to demonstrate **respectful Interaction** by treating all participants with dignity and courtesy, regardless of their background, role, or level of experience. All participants are expected to conduct themselves positively reflecting positively on the individual and the association. IISA strongly promotes a welcoming and inclusive environment for individuals of all races, ethnicities, national origins, genders, sexual orientations, ages, abilities, religions, and other protected characteristics.

Language Considerations
We come from diverse backgrounds and have diverse expectations about professional conduct in our in-person and online spaces. Effective and respectful communication is key to a successful event, as it provides a welcoming atmosphere. IISA strongly prohibits the use of offensive or discriminatory language, including slurs, jokes, or comments that demean individuals or groups. Instead, use inclusive language that fosters mutual understanding and respect. Be mindful of cultural differences and language barriers that may influence communication.

Drinking and Drugs
Participants are expected to either abstain from alcohol or consume it responsibly while behaving appropriately at all times. Excessive intoxication from alcohol or legal drugs, visible drunkenness, impairment from behavior-altering substances, and the use of illegal drugs are strictly prohibited. Individuals deemed significantly under the influence at an IISA event will be escorted out and may not return until the following day. As substance use is often a precursor to Code of Conduct (CoC) violations, any participant found violating the CoC while under the influence may face suspension from all future IISA events. Intoxication or substance use will never be accepted as an excuse for CoC violations.

Harassment-Free Environment
The International Indian Statistical Association (IISA) is committed to maintaining a harassment-free environment and enforces a zero-tolerance policy for all forms of harassment. This includes verbal harassment, such as derogatory comments, slurs, or unwelcome advances; physical harassment, including unwelcome physical contact or threatening behavior; visual harassment, such as displaying offensive images or text in shared spaces; and sexual harassment, including unwanted sexual attention, inappropriate jokes, or suggestive remarks. IISA is dedicated to fostering a safe, respectful, and inclusive community for everyone.

Commercial Activities
IISA events and workspaces are meant to focus on the dissemination of knowledge. While we recognize that academic and commercial spaces commonly intersect in the field of statistics and data science, IISA events and workspaces should not be used to promote or sell unauthorized commercial services.

Reporting Code of Conduct Violations
Harassment and other code of conduct violations reduce the value of our programming for everyone. If someone makes you or anyone else feel unsafe, unwelcome, or demeaned, you are urged to report as soon as possible. However, if it takes months or years to feel comfortable, there is no statute of limitations on reporting. If you see others being harassed, please report it immediately and intervene if you feel comfortable ("see something, say something").

Procedure for Code of Conduct Violations

IISA has appointed **Professor Sayantee Jana** as the ombudsman for all reported conference Code of Conduct violations. Please contact Professor Jana (sayantee.jana@math.iith.ac.in; mobile number: +91 9607548883) or, Dr. Irshad (Local Organization Committee, loc@intindstat.org; mobile number: +91 8943745636), or any IISA volunteer if you witness any violations or wish to report an incident. If a matter is time-sensitive and the IISA conference staff cannot be reached, please contact the IISA President (saonli@umn.edu) or the security staff at the meeting venue.

We will take all good-faith reports of harassment by participants seriously. IISA reserves the right to exclude people from the conference and workshops based on the findings. IISA also reserves the right to reject any report we believe to have been made in bad faith. This includes reports intended to silence legitimate criticism. **If an offense is deemed serious, we will report it to the security officials at CUSAT, for proper investigation of any potential crime.** IISA will always cooperate with CUSAT authorities and local administration in all conference conduct-related matters.

ASA Code of Conduct: https://www.amstat.org/meetings/code-of-conduct
NSF Code of Conduct: It is National Science Foundation policy to foster harassment-free environments wherever science is conducted, including at NSF-sponsored conferences, workshops, and symposia.