IIIA 2022 CONFERENCE

Dec 26th - 30th, 2022

Programme Book

National Science Seminar Complex
Indian Institute of Science
Bengaluru

International Indian Statistical Association
Annual Conference
National Science Seminar Complex
Indian Institute of Science, Bengaluru
26 - 30 December, 2022

Scientific Programme Committee

Chair: Subhashis GHOSHAL, North Carolina State University
Co-chair: Shailaja SURYAWANSHI, Bristol Myers Squibb
Co-chair: T. V. RAMANATHAN, Savitribai Phule University Pune

Bhaswati GANGULY, Calcutta University
Peihua QIU, University of Florida
Pierpaolo de BLASI, University of Torino
Bibhash CHAKRABORTY, National University of Singapore
Kingshuk ROY CHOUDHURY, Amazon
Abdus WAHED, University of Pittsburg
Kaushik JANA, Ahmedabad University
Moumanti PODDER, IISER, Pune
Arusharka SEN, Concordia University
Mahesh IYER,
Dootika VATS, IIT, Kanpur
Peter MÜLLER, University of Texas Austin
Aarti SINGH, Carnagie Melon University

Local Organising Committee
Convenor: Chiranjit MUKHOPADHYAY, IISc, Bengaluru
Sibasish DASGUPTA, Pfizer
Chiranjib BHATTACHARYYA, IISc, Bengaluru
Srikanth K. IYER, IISc, Bengaluru
K. Sampath KUMAR, Telengana University
R. GANGAI SELVI, Tamil Nadu Agricultural University
Soudeep DEB, IIM, Bengaluru
Firoz AHMAD, IISc, Bengaluru
Subhadip BANDYOPADHYAY, Ericsson, Bangalore
Angshuman SARKAR, GSK India Global Service Private Ltd

International Organizing Committee:
Snigdhansu CHATTERJEE, University of Minnesota
Subrata KUNDU, George Washington University
Sandeep MENNON, Pfizer
Abhyuday MANDAL, University of Georgia
Hiya BANERJEE, Eli Lilly
Sanjay CHAUDHURI, National University of Singapore
Arka Prasanna GHOSH, Iowa State University
Debashis MONDAL, Washington University of St. Louis

IISc Student and Staff Volunteers:
Anantha PADMANAVAN, Mohamed MEERAN, Thirunavukkarasu R, Akshay DIXIT, Anirudha DANDEKAR, Devadharsh VELUCHAMY, M. Pavan Kumar REDDY, Mehak AHUJA, Pranshu RAGHUVANSHI, Sam Sundar SINGH, Sayan ROY, Sumanjay DUTTA, Jharna BADULA, Pravin Gautam GAIKAWAD, Susmit DAS .

Cover Design: Jyotishka DATTA, Virginia Tech.

Sponsorships and Endorsements:

ASA

IASCT
Indian Association for Statistics in Clinical Trials

STATISTICAL
INDIAN INSTITUTE
भिन्नेष्वैक्यस्य दर्शनम्
UNITY IN DIVERSITY

MSD
INVENTING FOR LIFE

NSF
National
Science
Foundation

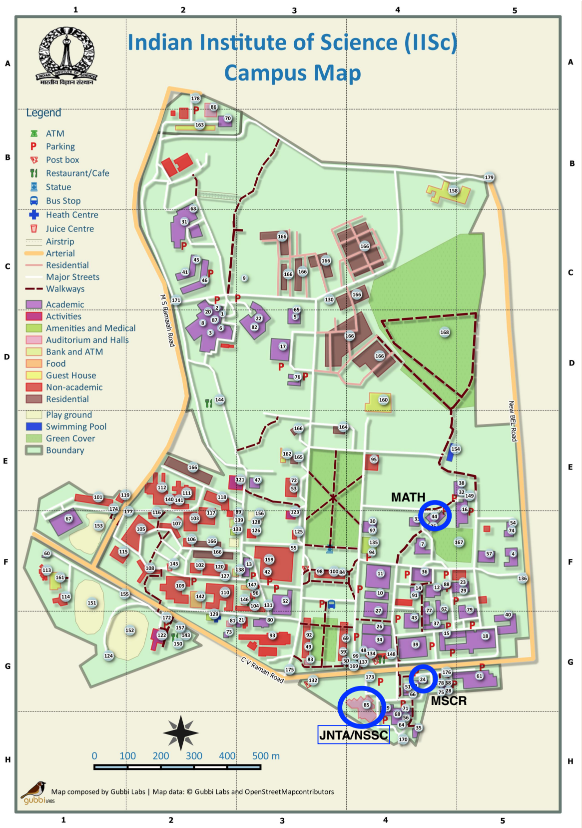
National Board for Higher Mathematics

Pfizer

R consortium

RStudio

Springer



Venues:

National Science Seminar Complex (NSSC)

1. JNTA: J. N. Tata Auditorium
2. NSSC-A
3. NSSC-B
4. NSSC-C

MSCR

1. MSSH
2. MSCR-1
3. MSCR-2

Mathematics Department (MATH)

1. MATH-LH1
2. MATH-LH2
3. MATH-LH3
4. MATH-LH4

Instructions:

- The talks with titles in magenta are virtual. The sessions with titles in magenta have at least one virtual session.
- The Programme book contains links that can be used to navigate its pages.

Glossary

M1	Morning Session 1	M2	Morning Session 2
A1	Afternoon Session 1	A2	Afternoon Session 2
I	Invited Session	C	Contributed Session

From The President, India Chapter

Dear all,

On behalf of the International Indian Statistical Association, it is my pleasure to welcome you to our Annual Conference, being held at the esteemed Indian Institute of Science, Bengaluru during December 26 - 30, 2022.

This year marks the return to an in-person annual event in India, and we are excited to bring you a full and diverse scientific agenda, with short courses, plenary lectures, invited and contributed sessions, panel discussions, poster competition and special sessions, over five days. The Scientific Program Committee chaired by Prof. Subhashis Ghoshal, and the Local Organizing Committee led by Prof. Chiranjit Mukhopadhyay have worked diligently over the past several months to organize this event, and we hope you will take advantage of this opportunity to gain insights on current state of research in various areas relating to theoretical considerations and applications of Statistics. Our conference also provides networking opportunities for researchers to connect with each other, and to meet up with friends you may not have seen in a while. Let me also take this opportunity to thank our sponsors for their support to the conference, and specifically for enabling many students to attend the event.

For those who are new to Bengaluru, it is one of Indias fastest growing megacities, famous as a technology hub and startup capital of the country. It is also home to several premier educational institutions. Bengalurus location at an elevation of about 900m on the Deccan Plateau blesses it with cool and moderate temperatures year-round. We hope you can take some time to explore this beautiful city of parks, gardens, and cosmopolitan lifestyle, though be aware that commute times can be long even for relatively short distances!

Welcome once again to this event, and to Bengaluru. I hope you have a great conference, and best wishes for 2023!

Debjit Biswas
President, India Chapter
International Indian Statistical Association

ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

November 22, 2022

Dear Conference Participants and Conference 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 in Bengaluru at the Indian Institute of Science for the 2022 IISA Conference. The IISA is a great partner of ASA, and we are pleased to support the outstanding conferences IISA hosts.

One of your important objectives is to foster the exchange of information and scholarly activities. After being shut down by the pandemic for two years, it is wonderful to be able to hold this meeting in person. Our time apart reminded us how valuable it is to be together. With the excellent program you have organized and the energy the participants will bring, we have no doubt this will be a productive and memorable conference. I congratulate you on creating a program that celebrates the importance of our science.

Sincerely,

Katherine B. Ensor

Katherine B. Ensor, 2022 President, ASA

Contents

Programme Overview

11:00 - 12:20

14:00 - 15:00

15:10 - 16:30

17:10 - 18:30

9:00 - 10:20

11:00 - 12:20

14:00 - 15:00

15:10 - 16:30

17:10 - 18:30

9:00 - 10:20

11:00 - 12:20

Programme

Monday December 26

Short Course 1 Artificial Intelligence (AI) in Precision and Digital Health *Venue:* NSSC-A

Time : **9:00 - 12:10**

- **Bibhas CHAKRABORTY**, National University of Singapore & Duke University

Lunch (Only for short course attendees) 12:10 - 13:15 Venue: NSSC

Conference Inauguration *Venue:* JNTA

Time : **13:45 - 14:00**

Plenary Lecture 1 Kalyan Das *Venue:* JNTA

Chair : Rituparna SEN, Indian Statistical Institute, Bangalore

14:00 **Milestones in Personalized Medicine An overview on subgroup identification and some potential directions** [**Abstract 90**]

Kalyan DAS, *kalyanstat@gmail.com*
Ujjwal DAS, *OM,QM & IS Area IIM Udaipur*

Switching time 15:00 - 15:10

Panel Discussion 1 Statistics of Cricket *Venue:* JNTA

Organizer : Arunabha SENGUPTA,

- **Srinivas BHOGLE**,
- **Rajeeva KARANDIKAR**,
- **Abhishek MUKHERJEE**,

26.A1.I1 Survival Analysis *Venue:* NSSC-A

Chair : P. G. SANKARAN, Cochin University of Science and Technology
Organizer : Sudheesh KATTUMANNIL, Indian Statistical Institute

15:10 **Semiparametric regression analysis of doubly censored recurrent event data** [**Abstract 276**]

P. G. SANKARAN, *Cochin University of Science and Technology*
Hari S., *Cochin University of Science and Technology*
Sreedevi E. P., *Maharaja's College, Ernakulam*

15:35 **Semiparametric transformation model for competing risks data with cure fraction** [**Abstract 185**]

Sudheesh KATTUMANNIL, *Indian Statistical Institute*

16:00 **Semiparametric Transformation Cure model for Current Status data under Competing risks.** [**Abstract 117**]

Sreedevi E P., *Maharaja's College , Ernakulam*
Sudheesh KATTUMANNIL, *ISI, Chennai*
Sankaran P. G., *Cochin University of Science and Technology, Cochin*

26.A1.I2 Discrete probability: random graphs and processes on random graphs *Venue:* `NSSC-B`

Chair and Organizer : Moumanti PODDER, Indian Institute of Science Education and Research (IISER) Pune

15:10 De-preferential attachment model through preferential attachment [Abstract 18]

Antar BANDYOPADHYAY, *Indian Statistical Institute, Delhi Centre*
Somak LAHA, *Indian Statistical Institute, Kolkata*

15:35 ON BAYESIAN BEST RESPONSE LEARNING IN GAMES [Abstract 301]

Souvik ROY, *ISI Kolkata*
Sayan MUKHERJEE, *ISI Kolkata*

16:00 Combinatorial games with "jump" moves on rooted Galton-Watson trees [Abstract 277]

Moumanti PODDER, *Indian Institute of Science Education and Research (IISER) Pune*
Dhruv BHASIN, *Indian Institute of Science Education and Research (IISER) Pune*

26.A1.I3 Recent theoretical and methodological advances in high-dimensional inference *Venue:* `NSSC-C`

Chair : Bikram KARMAKAR, University of Florida
Organizer : Jyotishka DATTA, Virginia Tech

15:10 An Empirical Bayes Approach to Controlling the False Discovery Exceedance [Abstract 29]

Pallavi BASU, *Indian School of Business*

15:35 Asymptotics of Large Autocovariance Matrices [Abstract 38]

Monika BHATTACHARJEE, *Department of Mathematics, IIT Bombay*

16:00 An approximation algorithm for blocking of an experimental design [Abstract 180]

Bikram KARMAKAR, *University of Florida*

26.A1.I4 Machine Learning Models for Business Applications *Venue:* `MSSH`

Chair and Organizer : Rituparna SEN, Indian Statistical Institute, Bangalore

15:10 A Predictive Analytics Approach for the Impact Assessment of Patent Characteristics on the Renewal Life of a Patent [Abstract 286]

Pritam RANJAN, *pritamr@iimidr.ac.in*
Ashit KUMAR, *IIM Indore*
Shadab DANISH, *BASE University*

15:35 Care2Vec: A Deep learning approach for the classification of self-care problems in physically disabled children [Abstract 280]

Sayan PUTATUNDA, *sayanp@iima.ac.in*

16:00 Preventive Maintenance of a Machine: A statistical approach to model and analysis of recurrent events from machine logs to predict failures [Abstract 263]

Sandip PAL, *S&P Global*
Pritam RANJAN, *IIM Indore*
Arnab KOLEY, *IIM Indore*
Debasis KUNDU, *IIT Kanpur*

26.A1.I5 Probabilistic Modeling and Theoretical Statistics: Recent Advances *Venue:* MSCR-1

Chair : Srinivasan BALAJI, Department of Statistics, George Washington University
Organizer : IISA,

15:10 The containment profile of hyper-recursive trees [Abstract 15]

Srinivasan BALAJI, *Department of Statistics, George Washington University*
Hosam MAHMOUD, *Department of Statistics, George Washington University*
Joshua SPARKS, *Department of Statistics, George Washington University*

15:35 Transfer learning under posterior drift [Abstract 26]

Moulinath BANERJEE, *University of Michigan*
Subha MAITY, *University of Michigan*
Yuekai SUN, *University of Michigan*

16:00 Tampered Random Variable Modeling for Multiple Step-stress Life Test [Abstract 356]

Farha SULTANA, *Assistant Professor*
Anup DEWANJI, *Professor*

26.A1.I6 Bayesian modeling of structured data *Venue:* MSCR-2

Chair and Organizer : Rajarshi GUHANIYOGI, Texas A & M University

15:10 Bayesian regression analysis of skewed tensor responses [Abstract 350]

DEBAJYOTI SINHA, *FLORIDA STATE UNIVERSITY*
Inkoo LEE, *Rice University, USA*
Dipankar BANDYOPADHYAY, *Virginia Commonwealth University*
Qing MAI, *Florida State University*

15:35 Bayesian Multivariate Regression Models for Tropical Cyclone Activity [Abstract 134]

Joyee GHOSH, *The University of Iowa*
Xun LI,
Gabriele VILLARINI,

16:00 A Graphical Lasso model for Hermitian matrices to detect global time-lagged Teleconnections [Abstract 312]

Indranil SAHOO, *Virginia Commonwealth University*
Joe GUINNESS, *Cornell University*
Brian REICH, *North Carolina State University*

Coffee 16:30 - 17:10 Venue: NSSC

26.A2.I7 Random Graphs and Processes *Venue:* JNTA

Chair : Laura ESLAVA, IIMAS-UNAM
Organizers : *Moumanti PODDER, Indian Institute of Science Education and Research (IISER) Pune* and *Parthanil ROY, Indian Statistical Institute*

17:10 Inhomogeneous Erds-Rnyi random graphs: bulk and edge of the spectrum [Abstract 58]

Arijit CHAKRABARTY, *Indian Statistical Institute, Kolkata*

17:35 **Poisson process approximation under stabilization** [**Abstract 384**]

D YOGESHWARAN, *Indian Statistical Institute*

18:00 **High-degree vertices in weighted recursive trees** [**Abstract 118**]

Laura ESLAVA, *IIMAS-UNAM*
Bas LODEWIJKS, *Institut Camille Jordan*
Marcel ORTGIESE, *University of Bath*

26.A2.I8 Recent Developments In Mathematical Statistics

Venue: `NSSC-A`

Chair : Snigdhansu CHATTERJEE, University of Minnesota
Organizer : IISA,

17:10 **Fay-Herriot Area Level Model with Arc-Sin transformation for Sample Proportions** [**Abstract 161**]

Masayo HIROSE, *Institute of Mathematics for Industry, Kyushu University*
Malay GHOSH, *Department of Statistics, University of Florida*
Tamal GHOSH, *Citibank, N.A*

17:35 **Regridding uncertainty for statistical downscaling of solar radiation** [**Abstract 19**]

Soutir BANDYOPADHYAY, *Colorado School of Mines*
Maggie BAILEY, *Colorado School of Mines*
Douglas NYCHKA, *Colorado School of Mines*
Manajit SENGUPTA, *National Renewable Energy Laboratory (NREL)*

18:00 **A Modern CLT for the classical AIPW: variance inflation, cross-fit covariance and beyond** [**Abstract 357**]

Pragya SUR, *Harvard University*
Kuanhao JIANG, *Harvard University*
Subhabrata SEN, *Harvard University*
Rajarshi MUKHERJEE , *Harvard University*

26.A2.I9 Advances in SMART Design - interim monitoring, non-inferiority, and missing data

Venue: `NSSC-B`

Chair : Samiran GHOSH, University of Texas School of Public Health
Organizer : Abdus WAHED, University of Pittsburgh

17:10 **Some aspects of SMART design: Methodological Developments with an Application in Non-inferiority Trial** [**Abstract 138**]

Samiran GHOSH, *University of Texas School of Public Health*

17:35 **Imputation-Based Q-Learning for Optimizing Dynamic Treatment Regimes with Right-Censored Survival Outcome** [**Abstract 79**]

Yu CHENG, *University of Pittsburgh*
Lingyun LYU, *University of Pittsburgh*
Abdus WAHED, *University of Pittsburgh*

18:00 **Interim Monitoring in Sequential Multiple Assignment Randomized Trials** [**Abstract 376**]

Abdus WAHED, *University of Pittsburgh*
Liwen WU, *University of Pittsburgh*
Junyao WANG, *University of Pittsburgh*

26.A2.I10 Quantitative Decision making *Venue:* NSSC-C

Chair : Debjit BISWAS,

Organizer : Angshuman SARKAR, GSK India Global Service Private Ltd

17:10 **Quantitative Decision Making in Early Stages of Drug Development** [**Abstract 262**]

RUPAM PAL, *GSK Pharmaceutical*

17:35 **Fractile Graphical Analysis in comparing conditional distributions: Applications in Financial Economics** [**Abstract 132**]

Aurobindo GHOSH, *Singapore Management University*
ANIL BERA,

18:00 **Application of Statistical Machine Learning in Biomarker Selection** [**Abstract 97**]

Shibasish DASGUPTA, *Pfizer and Chennai Mathematical Institute (CMI)*

26.A2.I11 Advancements in Spatio-Temporal Modeling and Monitoring *Venue:* MSSH

Chair : Anjishnu BANERJEE, Medical College of Wisconsin

Organizer : Kai YANG, Medical College of Wisconsin

17:10 **Greater than the sum of its parts learning from piecewise data in spatial contexts** [**Abstract 21**]

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

17:35 **Effective spatio-temporal disease surveillance by using covariate information** [**Abstract 383**]

Kai YANG, *Medical College of Wisconsin*
Peihua QIU, *University of Florida*
Kai YANG, *Medical College of Wisconsin*

18:00 **Online Nonparametric Monitoring of Heterogeneous Data Streams with Partial Observations based on Thompson Sampling** [**Abstract 377**]

Xiaochen XIAN, *University of Florida*

26.A2.I12 Statistical Methods for Financial Risk *Venue:* MSCR-1

Chair : Sourish DAS, CMI

Organizer : Rituparna SEN, Indian Statistical Institute, Bangalore

17:10 **Network Modeling of Contagion in High Dimensional Financial Time Series** [**Abstract 328**]

Rituparna SEN, *Indian Statistical Institute, Bangalore*
Sumanta BASU, *Cornell University*
Georgia SMITS, *Cornell University*

17:35 **On Distance to Default for Unlisted Firm using Hierarchical Bayesian Models** [**Abstract 93**]

Sourish DAS, *CMI*
Susan THOMAS, *XKDR Research Forum*

18:00 **Estimation of Spectral Risk Measure for Left Truncated and Right Censored Data** [**Abstract 51**]

Suparna BISWAS, *Indian Statistical Institute, Bangalore*
Rituparna SEN,

26.A2.C1 Contributed Session 1

Venue: Math-LH1

Chair : Richa PANCHGAUR, Research Scholar

17:10 **A bootstrap control chart for some well known finite range distributions.** [**Abstract 236**]

BIDHAN MODOK, *RAIGANJ UNIVERSITY*
Shovan CHOWDHURY, *Indian institute of Management Kozhikode, Kozhikode, India.*
Amarjit KUNDU, *Raiganj University, Raiganj, India*

17:25 **A Fiducial Confidence Interval for Agreement Measurement Among Raters in a Generalized Linear Mixed Effects Model** [**Abstract 315**]

Soumya SAHU, *University of Illinois Chicago*
Dulal BHAUMIK, *University of Illinois Chicago*
Thomas MATHEW, *University of Maryland Baltimore County*

17:40 **Random slopes** [**Abstract 126**]

Disha GHANDWANI, *Stanford University*
Trevor HASTIE, *Stanford University*
Swarnadip GHOSH, *Radix Trading*
Art OWEN, *Stanford University*

17:55 **Estimation of the Selected Treatment Mean in Two Stage Drop-the-Losers Design** [**Abstract 226**]

MASIHUDDIN, *INDIAN INSTITUTE OF TECHNOLOGY KANPUR*
Neeraj MISRA,

18:10 **Inflated binomial and Beta Binomial Model to study the Distribution of Rural Out migrants** [**Abstract 265**]

Richa PANCHGAUR, *Research Scholar*
Alok KUMAR,

26.A2.C2 Contributed Session 2

Venue: Math-LH3

Chair : Ummey HABIBA, Integral University, Lucknow

17:10 **On computationally efficient algorithm to estimate the two-dimensional chirp model parameters with the interaction term in phase** [**Abstract 340**]

Abhinek SHUKLA, *IIT Kanpur*
Rhythm GROVER, *Mehta Family School of Data Science and Artificial Intelligence, Indian Institute of Technology Guwahati, Assam-781039, India*
Debasis KUNDU, *Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur - 208016, India*
Amit MITRA, *Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur - 208016, India*

17:25 **Structurally incomplete 2-part designs involving treatment combinations** [**Abstract 181**]

Sayantani KARMAKAR, *ICAR-Indian Agricultural Research Institute*
Cini VARGHESE, *ICAR-Indian Agricultural Statistics Research Institute*
Mohd HARUN, *ICAR-Indian Agricultural Statistics Research Institute*
Vinaykumar L.N., *ICAR-Indian Agricultural Research Institute*

17:40 Optimal PBIB designs in two-replicates [Abstract 212]

Vinaykumar L.N., *ICAR - Indian Agricultural Research Institute*
Cini VARGHESE, *ICAR - Indian Agricultural Statistics Research Institute*
Harun MOHD, *ICAR - Indian Agricultural Statistics Research Institute*
Sayantani KARMAKAR , *ICAR - Indian Agricultural Research Institute*

17:55 Construction Methods for Doubly Nested Block Designs [Abstract 5]

Nehatai AGASHE, *PGS IARI, New Delhi*
Cini VARGHESE, *IASRI, New Delhi*
Harun MOHD, *IASRI, New Delhi*
Vinay KUMAR, *PGS IARI, New Delhi*

18:10 A Novel Approach to Solve Fixed Charge Interval Transportation Problems [Abstract 156]

Ummey HABIBA, *Integral University, Lucknow*
Ummey HABIBA, *Integral University, Lucknow*
Abdul QUDDOOS,

26.A2.C3 Contributed Session 3

Venue: Math-LH4

Chair : Asmita SHUKLA, Banaras Hindu University

17:10 Statistical power and sample size requirements to detect an intervention by time interaction in four-level longitudinal cluster randomized trials [Abstract 217]

Priyanka MAJUMDER, *Postdoctoral Fellow*
Samiran GHOSH, *Professor, Department of Family Medicine and Public Health Sciences, Wayne State University, Detroit, MI, USA.*
Siuli MUKHOPADHYAY, *Professor, Department of Mathematics, Indian Institute of Technology Bombay, Mumbai, India*
Bo WANG, *Professor, Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts, USA*

17:25 Ordering and ageing properties of sequential order statistics governed by the Archimedean copula [Abstract 314]

Tanmay SAHOO, *Indian Institute of Technology Jodhpur*
Nil Kamal HAZRA, *Indian Institute of Technology Jodhpur*

17:40 A general class of shock models with dependent inter-arrival times [Abstract 144]

Dheeraj GOYAL, *Department of Mathematics, Indian Institute of Technology Jodhpur, 342037, Karwar, Rajasthan, India*
Nil Kamal HAZRA, *Department of Mathematics, Indian Institute of Technology Jodhpur, 342037, Karwar, Rajasthan, India*
Maxim FINKELSTEIN, *Department of Mathematical Statistics and Actuarial Science, University of the Free State, 339 Bloemfontein 9300, South Africa*

17:55 Bayesian Weibull Quantile Regression and Variable Selection with Application Uterine Serous Carcinoma patient Survival [Abstract 364]

Tripti TRIPATHI, *Banaras Hindu University*
Rakesh RANJAN, *DST-CIMS, BHU, Varanasi*

18:10 Bayes Analysis of Generalized Gamma and its component models using gradient based Markov Chain Monte Carlo Methods [Abstract 341]

Asmita SHUKLA, *Banaras Hindu University*
Rakesh RANJAN, *Banaras Hindu University*
Richa SRIVASTAVA, *Jaipuria Institute of Management, Lucknow*

Tuesday December 27

27.M1.I13 New biostatistical methods of analyzing complex data *Venue:* `JNTA`

Chair : Subharup GUHA, University of Florida
Organizer : Peihua QIU, University of Florida

9:00 **Regression Analysis of a Future State Entry Time Distribution Conditional on a Past State Occupation in a Progressive Multistate Model** [**Abstract 102**]

Somnath DATTA, *University of Florida*

9:25 **A zero-inflated Bayesian nonparametric approach for identifying differentially abundant taxa in multigroup microbiome data with covariates** [**Abstract 149**]

Subharup GUHA, *University of Florida*
Archie SACHDEVA, *University of Florida*
Somnath DATTA, *University of Florida*
Subharup GUHA, *University of Florida*

9:50 **Soft-thresholded scalar on vector-valued image regression with application to DTI data** [**Abstract 294**]

Arkaprava ROY, *University of Florida*
Zhou LAN,

27.M1.I14 Recent Advances in Data Integration and Distributed Computing *Venue:* `NSSC-A`

Chair : Priyam DAS, Department of Biostatistics, Virginia Commonwealth University
Organizer : Dipankar BANDYOPADHYAY, Virginia Commonwealth University

9:00 **Estimating the optimal linear combination of predictors using spherically constrained optimization** [**Abstract 92**]

Priyam DAS, *Department of Biostatistics, Virginia Commonwealth University*
Debsurya DE, *Indian Statistical Institute*
Raju MAITI, *Indian Statistical Institute*
Christine PETERSON, *University of Texas MD Anderson Cancer Center*

9:25 **Recent Advances in Data Integration and Distributed Computing** [**Abstract 64**]

Sounak CHAKRABORTY, *University of Missouri*
Tanujit DEY, *Harvard University*
Anjishnu Banerjee BANERJEE, *Medical College of Wisconsin*
Yilun HUANG, *University of Missouri*

9:50 **Bayesian Learning for disparities from Integrated Data Sources Related to the Rare Events** [**Abstract 110**]

Tanujit DEY, *Harvard Medical School*
Anjishnu BANERJEE, *Medical College of Wisconsin*
Sounak CHAKRABORTY, *University of Missouri*

27.M1.I15 Applications of Random Matrix Theory *Venue:* `NSSC-B`

Chair : Indrajit JANA, IIT Bhubaneswar
Organizer : Madhuchhanda BHATTACHARJEE, University of Hyderabad, Hyderabad, India

9:00 **Estimation of spectra of linear processes in high dimensions** [**Abstract 272**]

Debashis PAUL, *University of California, Davis*
Jamshid NAMDARI, *University of California, Davis*
Alexander AUE, *University of California, Davis*

9:25 **A nonparametric test of co-spectrality of networks** [**Abstract 167**]

Indrajit JANA, *IIT Bhubaneswar*
Srijan SENGUPTA, *North Carolina State University*
Chetkar JHA, *University of Pennsylvania*

9:50 **Spectrum of High Dimensional Sample Covariance and Related Matrices: A Selective Review** [**Abstract 53**]

Arup BOSE, *Indian Statistical Institute*

27.M1.I16 Computational Bayes *Venue:* `NSSC-C`

Chair : Raffaele ARGIENTO, Università degli studi di Bergamo
Organizer : Pierpaolo DE BLASI, University of Torino and Collegio Carlo Alberto, Italy

9:00 **Bayesian Analysis of Weibull Quantile Regression model using Hamiltonian Monte Carlo Algorithm when the Observations are Interval Censored** [**Abstract 287**]

Rakesh RANJAN, *Benares Hindu University*

9:25 **Sampling using Adaptive Regenerative Processes** [**Abstract 292**]

Christian ROBERT, *Université Paris Dauphine PSL and University of Warwick*
Hector MCKIMM, *University of Warwick*
Andi WANG, *University of Bristol*
Murray POLLOCK, *University of Newcastle*

9:50 **Clustering grouped data via hierarchical normalized finite point process** [**Abstract 7**]

Raffaele ARGIENTO, *Università degli studi di Bergamo*
Alessandro COLOMBI, *Universit Milano Bicocca*
Federico CAMERLENGHI, *Universit Milano Bicocca*
Lucia PACI, *Universit Cattolica del Sacro Cuore*

27.M1.I17 Novel methods for shape-constrained distributions, nonparametric hypothesis testing and survival analysis with applications to biomedical data *Venue:* `MSSH`

Chair and Organizer : Sharmistha GUHA, Texas A&M University

9:00 **Nonparametric Hypothesis Testing in High Dimensions and GWAS Applications** [**Abstract 74**]

Snigdhansu CHATTERJEE, *University of Minnesota*

9:25 **Efficient Estimation of the Additive Risks Model for Interval-Censored Data** [**Abstract 351**]

Samiran SINHA, *Texas A&M University*
Tong WANG, *School of Statistics and Data Science, Nankai University, Tianjin, China*
Dipankar BANDYOPADHYAY, *Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA*

9:50 **An overview of application of Bayesian statistics in early drug development** [**Abstract 129**]

Nairita GHOSAL, *Merck & Co., Inc., Rahway, NJ, USA*

27.M1.I18 Statistical methods for modern healthcare applications *Venue:* `MSCR-1`

Chair : Abhirup DATTA, Johns Hopkins University

Organizer : Srijan SENGUPTA, North Carolina State University

9:00 **Statistical methods for finding cytometric markers associated with Systemic Lupus Erythematosus [Abstract 302]**

Pratyaydipta RUDRA, *Oklahoma State University*

9:25 **Bayesian validation framework for dynamic epidemic models [Abstract 96]**

Sayan DASGUPTA, *Fred Hutchinson Cancer Center*
Mia MOORE, *Fred Hutchinson Cancer Center*
Dobromir DIMITROV, *Fred Hutchinson Cancer Center*
James HUGHES, *University of Washington, Seattle*

9:50 **Assembling multiple data sources and machine learners to assess ambient PM2.5 in India and linking it with cardiovascular health [Abstract 221]**

Siddhartha MANDAL, *Center for Chronic Disease Control, New Delhi, India*
Poornima PRABHAKARAN, *Public Health Foundation of India*
Petter LJUNGMAN, *Karolinska Institute*
Joel SCHWARTZ, *Harvard TH Chan School of Public Health*

Student Paper Competition 1 *Venue:* `MSCR-2`

Chair : Adityanand GUNTUBOYINA, University of California Berkeley

9:00 **Minimax Nonparametric Estimation for Markov Decision Processes [Abstract 24]**

Imon BANERJEE, *Purdue University*
Vinayak Rao, Harsha Honnappa ,

9:15 **Stochastic Correlation modelling with von Mises Process [Abstract 216]**

Sourav MAJUMDAR, *Indian Institute of Management Ahmedabad*
Arnab Kumar Laha ,

9:30 **A Conditional Bayesian Approach with Valid Inference for High Dimensional Logistic Regression [Abstract 260]**

Abhishek OJHA, *University of Illinois, Urbana-Champaign*
Naveen N. Narisetty ,

9:45 **A General Framework for Powerful Confounder Adjustment in Omics Association Studies [Abstract 295]**

Asmita ROY, *Texas A&M University*
Jun Chen, Xianyang Zhang ,

10:00 **A unified approach to a class of distributions [Abstract 368]**

Durga VASUDEVAN, *Department of Statistics, Cochin University of Science and Technology*
Dr. Asha Gopalakrishnan ,

27.M1.C4 Contributed Session 4 *Venue:* `Math-LH1`

Chair : Partha SARKAR, University of Florida

9:00 **Generating Optimal Designs for Mixture Experiments using Genetic Algorithm [Abstract 266]**

Mahesh Kumar PANDA, *Associate Professor of Statistics, Department of Statistics, Ravenshaw University, Cuttack, Odisha*

9:15 **Modeling of Load-sharing Systems using Piecewise Linear Approximation to Cumulative Hazard** [Abstract 50]

Shilpi BISWAS, *Research Scholar, IIT Guwahati*
Ayon GANGULY, *Assistant Professor, IIT Guwahati*
Debanjan MITRA, *Assistant Professor, IIM Udaipur*

9:30 **Shrinkage Estimators for order restricted Parameters under the Generalized Pitman Nearness Criterion** [Abstract 125]

Naresh GARG, *IIT Kanpur*
Neeraj MISRA, *IIT Kanpur*

9:45 **DUG-MLES: An efficient hybrid machine learning algorithm to classify Dual Prong Goods measure the compliance risk** [Abstract 197]

Rajat KUMAR, *S&P Global*
Suman SENAPATI, *S&P Global*
Byron MCKINNEY, *S&P Global*

10:00 **Posterior Consistency of Covariance matrices Under Some Classes of Noninformative Priors** [Abstract 319]

Partha SARKAR, *University of Florida*
Kshitij KHARE, *University of Florida*
Malay GHOSH, *University of Florida*

27.M1.C5 Contributed Session 5

Venue: Math-LH3

Chair : Santosh SUTAR, Shivaji University, Kolahpur

9:00 **Statistical Planning for Non-Proportional Hazards in Design and Analysis of Cancer Clinical Trials** [Abstract 80]

Suresh CHENJI, *Eli Lilly*

9:15 **Optimal Multivariate Crossover Designs** [Abstract 259]

Shubham NIPHADKAR, *niphadkarshubham@gmail.com*
Siuli MUKHOPADHYAY, *Indian Institute of Technology Bombay*

9:30 **Application of Statistics in the Cell and Gene Therapy modality** [Abstract 137]

Rochishnu GHOSH, *GSK India Global Services Pvt. Ltd.*

9:45 **Multiple inflated negative binomial regression for correlated multivariate count data** [Abstract 44]

SUMANGAL BHATTACHARYA, *Department of Mathematics and Statistics, IIT Tirupati*
Joseph MATHEWS,
Sumen SEN, *Associate Professor, Austin Peay State University*
Ishapathik DAS, *Asst. Professor, IIT Tirupati*

10:00 **Reliability modeling of a two component parallel load sharing system with non-identical components** [Abstract 358]

Santosh SUTAR, *Shivaji University, Kolahpur*

Coffee 10:20 - 11:00 Venue: NSSC

Special Invited Session 1 Prateek Jain, Pritam Gupta

Venue: JNTA

Chair : Abhinanda SARKAR, Great Learning

11:00 Reverse Experience Replay: An Efficient Way to Learn with Dependent Data. [Abstract 164]

Prateek JAIN, *Google Research India*
Dheeraj NAGARAJ, *Google Research India*
Praneeth NETRAPALLI, *Google Research India*
Naman AGARWAL, *Google Research Princeton*

11:40 Brief description of the key methodological aspects of a vaccine trials with an emphasis on the Covid-19 vaccine study in Pfizer [Abstract 155]

Pritam GUPTA, *Pfizer Healthcare India Pvt Ltd*

27.M2.I19 Sharp statistical guarantees for statistical machine learning methods

Venue: NSSC-A

Chair and Organizer : Nilanjana LAHA, Texas A&M University

11:00 Near Optimal inference in Adaptive Linear Regression [Abstract 187]

Koulik KHAMARU, *Rutgers University*
Yash DESHPANDE, *Voleon*
Tor LATTIMORE , *Deep Mind*
Lester MACKEY, *Microsoft Research*

11:25 Adaptive Estimation of Nonparametric Functionals [Abstract 246]

Rajarshi MUKHERJEE, *Harvard University*
Lin LIU, *Shanghai Jiao Tong University*
James ROBINS, *Harvard University*
Eric TCHETGEN TCHETGEN, *University of Pennsylvania*

11:50 Optimal dynamic treatment regimes via smooth surrogate losses [Abstract 206]

Nilanjana LAHA, *Texas A&M University*
Aaron SONABEND, *Google research*
Rajarshi MUKHERJEE, *Harvard University*
Tianxi CAI, *Harvard University*

27.M2.I20 Optimizations

Venue: NSSC-B

Chair : Pritam RANJAN, pritamr@iimidr.ac.in
Organizer : Abhyuday MANDAL, University of Georgia

11:00 Modeling and Active Learning for Experiments with Quantitative-Sequence Factors [Abstract 220]

Abhyuday MANDAL, *University of Georgia*
Qian XIAO, *University of Georgia*
Xinwei DENG, *Virginia Tech*
Yaping WANG, *East China Normal University*

11:25 Optimal lifetesting plan for two sample censoring scheme [Abstract 203]

Debasis KUNDU, *Indian Institute of Technology Kanpur*

11:50 A smooth alternative to check loss in Deep inference tasks [Abstract 309]

Snehanshu SAHA, *snehanshus@goa.bits-pilani.ac.in*

27.M2.I21 Recent Advances in Survey Sampling *Venue:* NSSC-C

Chair and Organizer : Snigdhansu CHATTERJEE, University of Minnesota

11:00 On an Empirical Likelihood-Based Solution to Approximate Bayesian Computation Problem [Abstract 78]

Sanjay CHAUDHURI, *National University of Singapore*
Subhroshekhar GHOSH, *National University of Singapore*
Thi Kim Cuc PHAM, *National University of Singapore*

11:25 Multivariate Global-Local Priors for Small Area Estimation [Abstract 135]

Malay GHOSH, *University of Florida*
T GHOSH, *Citibank, Tampa*
J. J. MAPLES, *US Census Bureau*
X TANG, *University of Arizona*

11:50 Hierarchical Bayes estimation of small area proportions using statistical linkage of disparate data sources [Abstract 208]

Parthasarathi LAHIRI, *University of Maryland, College Park*
Soumojit DAS, *University of Maryland, College Park, USA*

27.M2.I22 Biostatistics *Venue:* MSSH

Chair : Yang NI,
Organizer : Pierpaolo DE BLASI, University of Torino and Collegio Carlo Alberto, Italy

11:00 Joint Modeling of Longitudinal Biomarker Data with Changepoint and Flexible Sigmoidal Response [Abstract 171]

Wesley JOHNSON, *UC Irvine*
Michelle NORRIS, *California State University Sacramento*
Edward BEDRICK, *U. of Arizona*

11:25 Bayesian nonparametric taxonomic Gibbs priors [Abstract 291]

Tommaso RIGON, *University of Milano-Bicocca*
Ching-Lung HSU, *Duke University*
David DUNSON,

11:50 Mixed Membership Models for Functional Data [Abstract 360]

Donatello TELESCA, *UCLA*
Nicholas MARCO, *UCLA*
Damla SENTURK, *UCLA*

27.M2.I23 Markov Chains and Urn Models *Venue:* MSCR-1

Chair : Sreekar VADLAMANI, TIFR CAM
Organizers : *Moumanti PODDER, Indian Institute of Science Education and Research (IISER) Pune* and *Parthanil ROY, Indian Statistical Institute*

11:00 Urn Models with Irreducible, but Asymptotically Reducible, Replacement Matrices [Abstract 228]

Krishanu MAULIK, *Indian Statistical Institute*
Rohan SARKAR, *University of Connecticut, Storrs*

11:25 Interacting Urns with Multiple drawings [Abstract 310]

Neeraja SAHASRABUDHE, *IISER Mohali*

11:50 **Adaptive schemes for sampling in infinite dimensions** [**Abstract 366**]

Sreekar VADLAMANI, *TIFR CAM*
Jonas WALLIN,
Sibsankar SINGHA,

27.M2.I24 Reliability Theory and its Applications *Venue:* MSCR-2

Chair and Organizer : Sanjeev SABNIS, IIT Bombay

11:00 **Test of exponentiality against NWBUE alternatives - An L-statistics approach** [**Abstract 140**]

Shyamal GHOSH, *Assistant Professor, Dept. of Mathematics, IIIT Guwahati*
Murari MITRA, *Professor, Dept. of Mathematics, IIEST Shibpur*

11:25 **Parameter Estimation for Multistate Series and Parallel Systems Under Different Observation Setting** [**Abstract 195**]

Leena KULKARNI, *Department of Statistics, School of Mathematics, Applied Statistics & Analytics, NMIMS Deemed-to-be University, Mumbai, India*
Sanjeev SABNIS, *Department of Mathematics, IIT Bombay, Mumbai, India*
Sujit GHOSH, *Department of Statistics, NC State University, Raleigh, USA*

11:50 **Preservation of Log-concavity under Multi-state Series and Multi-state Parallel Systems** [**Abstract 304**]

Sanjeev SABNIS, *IIT Bombay*
Priyanka MAJUMDER, *IIT Bombay*
Shyamal GHOSH, *IIIT Guwahati*

Lunch 12:20 - 14:00 Venue: NSSC

Plenary Lecture 2 Hao Helen Zhang *Venue:* JNTA

Chair : Snigdhansu CHATTERJEE, University of Minnesota

14:00 **Flexible and Interpretable Learning for High-Dimensional Complex Data** [**Abstract 385**]

Helen ZHANG, *University of Arizona*

Switching time 15:00 - 15:10

Panel Discussion 2 Leading to a better future with Diversity, Equity, and Inclusion in action

Venue: JNTA
Time : **15:10 - 16:30**
Organizer : Susmita DATTA, University of Florida

- **Carmen TEKWE**, Indiana University at Bloomington
- **Hiya BANERJEE**, Novartis Pharmaceuticals
- **Snigdhansu CHATTERJEE**, University of Minnesota
- **Subhashis GHOSHAL**, North Carolina State University

27.A1.I25 Different quantitative aspects in drug development

Venue: `NSSC-A`

Chair : Braja Gopal SAHOO, Novartis Healthcare Pvt. Ltd.
Organizer : Hiya BANERJEE, Novartis Pharmaceuticals

15:10 **Geometric approaches to assessing the numerical feasibility for conducting matching-adjusted indirect comparisons in clinical trials [Abstract 95]**

Kohinoor DASGUPTA, *Novartis Healthcare Pvt Ltd*

15:35 **Predictive role of neutrophil-lymphocyte ratio (NLR) in patients with metastatic melanoma: A post hoc exploratory analysis from phase 3 trial [Abstract 311]**

Braja Gopal SAHOO, *Novartis Healthcare Pvt. Ltd.*
Hiya BANERJEE,

16:00 **Regulatory Submissions with Real world data sources - Key challenges and opportunities [Abstract 65]**

Arunava CHAKRAVARTTY, *Novartis*

27.A1.I26 Advances in Statistical Learning Theory

Venue: `NSSC-B`

Chair : Dootika VATS, Indian Institute of Technology Kanpur
Organizer : Sakshi ARYA, Penn State University

15:10 **Spectral regularized kernel two-sample test [Abstract 353]**

Bharath SRIPERUMBUDUR, *Pennsylvania State University*

15:35 **Towards Proving the Capacity of Deep Operator Nets for Solving (Partial) Differential Equations [Abstract 241]**

Anirbit MUKHERJEE, *The University of Manchester*
Pulkit GOPALANI, *Indian Institute of Technology-Kanpur*
Sayar KARMAKAR, *UFlorida*

16:00 **Kernel Epsilon-Greedy Approach for Contextual Bandits [Abstract 8]**

Sakshi ARYA, *Penn State University*
Bharath SRIPERUMBUDUR, *Penn State University*

27.A1.I27 High dimensional Bayesian inference

Venue: `NSSC-C`

Chair and Organizer : Peter MUELLER, UT Austin

15:10 **A Bayesian Approach for Network Classification [Abstract 148]**

Sharmistha GUHA, *Texas A&M University*
Abel RODRIGUEZ, *University of Washington*

15:35 **Predictor-Informed Bayesian Approaches for Dynamic Functional Connectivity [Abstract 151]**

Michele GUINDANI, *University of California, Los Angeles*

16:00 **Bayesian Semiparametric Hidden Markov Tensor Partition Models for Longitudinal Data with Local Variable Selection [Abstract 317]**

Abhra SARKAR, *The University of Texas at Austin*
Giorgio PAULON, *Berry Consultants*
Peter MUELLER, *The University of Texas at Austin*

27.A1.I28 Stochastic Processes and Scaling Limits

Venue: MSSH

Chair : Srikanth IYER, Indian Institute of Science Bangalore

Organizers : *Moumanti PODDER, Indian Institute of Science Education and Research (IISER) Pune* and *Parthanil ROY, Indian Statistical Institute*

15:10 Asymmetric Super-Heston-rough volatility model with Zumbach effect as a scaling limit of quadratic Hawkes processes [Abstract 162]

Srikanth IYER, *Indian Institute of Science Bangalore*
Priyanka CHUDASAMA, *Indian Institute of Science Bangalore*

15:35 Branching Brownian motion seen as a Markov process and its invariant measures [Abstract 335]

Atul SHEKHAR, *TIFR-CAM*
Xinxin CHEN, *Beijing Normal University*
Christophe GARBAN, *University Claude Bernard Lyon 1,*

16:00 Convergence to the Brownian web for a perturbed Howard model [Abstract 308]

Kumarjit SAHA, *Ashoka University*

27.A1.I29 Quality and Reliability

Venue: MSCR-1

Chair and Organizer : Shovan CHOWDHURY, Indian Institute of Management Kozhikode

15:10 On the general δ-shock model [Abstract 159]

Nil Kamal HAZRA, *IIT Jodhpur*
Goyal DHEERAJ, *IIT Jodhpur*

15:35 Bootstrap Control Chart for Percentiles of Some Lifetime Distributions with Hybrid Censored Data [Abstract 202]

Amarjit KUNDU, *Professor, Department of Mathematics, Raiganj University, Raiganj*
Prof. Shovan CHOWDHURY, *Professor, Indian institute of Management Kozhikode, Kozhikode, India*
Bidhan MODOK, *Research Scholar, Department of Mathematics, Raiganj University*

16:00 Optimal Sampling Plan under Accelerated Life Test Setting [Abstract 84]

Shovan CHOWDHURY, *Indian Institute of Management Kozhikode*

27.A1.I30 Statistical techniques for using summary statistics in epidemiology/genetic epidemiology research

Venue: MSCR-2

Chair and Organizer : Debashree RAY, Johns Hopkins University

15:10 Quantifying portable genetic effects and improving cross-ancestry genetic prediction with GWAS summary statistics [Abstract 213]

Qiongshi LU, *University of Wisconsin-Madison*

15:35 A novel meta-analysis approach in the presence of disparate confounder information from studies [Abstract 289]

Debashree RAY, *Johns Hopkins University*
Alvaro MUOZ, *Johns Hopkins University*
Lisa JACOBSON, *Johns Hopkins University*
Bryan LAU, *Johns Hopkins University*

16:00 Bayesian mixture models for detecting protein-DNA footprints [Abstract 48]

Anushua BISWAS, *IISER Pune*
Leelavati NARLIKAR, *IISER Pune, Data Science Department*

Coffee 16:30 - 17:10 Venue: NSSC

27.A2.I31 Count data modelling and applications

Venue: JNTA

Chair : Thekke Variyam RAMANATHAN, Savitribai Phule Pune University
Organizer : Manik AWALE, Savitribai Phule Pune University

17:10 Urbanisation and Bird Diversity [Abstract 183]

Akanksha KASHIKAR, *Department of Statistics, Savitribai Phule Pune University*

17:35 A new look at Integer Valued Time Series models with Discrete distributions [Abstract 175]

Mariyamma K D, *Christ College (Autonomous) Irinjalakuda, Calicut University*

18:00 Seasonal INAR(1) model for overdispersed count time series data [Abstract 11]

Manik AWALE, *Savitribai Phule Pune University*
KASHIKAR AKANKSHA,

27.A2.I32 Extreme Value Analysis

Venue: NSSC-A

Chair and Organizer : Kaushik JANA, Ahmedabad University

17:10 Scoring predictions at extreme quantiles [Abstract 168]

Kaushik JANA, *Ahmedabad University*
Kaushik JANA, *Ahmedabad University*
Almut VERAART, *Imperial College London*
Axel GANDY, *Imperial College London*

17:35 Realistic and Fast Modeling of Spatial Extremes over Large Geographical Domains [Abstract 158]

Arnab HAZRA, *Indian Institute of Technology Kanpur*
Raphael HUSER, *King Abdullah University of Science and Technology*
David BOLIN, *King Abdullah University of Science and Technology*

18:00 Optimal pooling and distributed inference for the tail index and extreme quantiles [Abstract 355]

Gilles STUPFLER, *ENSAI & University of Angers*
Abdelaati DAOUIA, *Toulouse School of Economics*
Simone PADOAN, *Bocconi University*

27.A2.I33 Advances in Survival Analysis - 1

Venue: NSSC-B

Chair and Organizer : Arusharka SEN, Concordia University

17:10 Inferring causal effects using autopsy data [Abstract 56]

Marco CARONE, *University of Washington*
Eric MORENZ, *University of Washington*
Stijn VANSTEELANDT, *Ghent University + London School of Hygiene and Tropical Medicine*
Lianne SHEPPARD, *University of Washington*

17:35 Nonparametric Estimation for Combined Incident and Prevalent Cohort Data [Abstract 230]

James MCVITTIE, *University of Regina*

18:00 **Sparse estimation in the finite mixture of accelerated failure time regression models** [**Abstract 339**]

Farhad SHOKOOHI, *University of Nevada Las Vegas*
Masoud ASGHARIAN, *McGill University, Canada*
Abbas KHALILI, *McGill University, Canada*
Shili LIN, *Ohio State University, USA*

In Memorium: Saurabh Ghosh

Venue: NSSC-C

Chair and Organizer : Moulinath BANERJEE, University of Michigan

- **Krishanu MAULIK**, Indian Statistical Institute
- **Saonli BASU**, University of Minnesota
- **Debashree RAY**, Johns Hopkins University

27.A2.I34 Statistical Modeling and Inference in Complex Structured High-Dimensional Data

Venue: MSSH

Chair : Arvind RAO, Department of Computational Medicine and Bioinformatics, University of Michigan-Ann Arbor
Organizers : *Arvind RAO, Department of Computational Medicine and Bioinformatics, University of Michigan-Ann Arbor* and *Shariq MOHAMMED, Boston University School of Public Health*

17:10 **Approximate selective inference via maximum likelihood** [**Abstract 267**]

Snigdha PANIGRAHI, *University of Michigan*

17:35 **Bayesian Integrative Network Models for Multimodal and Spatial Omics Data** [**Abstract 3**]

Satwik ACHARYYA, *University of Michigan*

18:00 **A statistical framework to study the spatial variation in imaging phenotypes with applications to radiology and pathology** [**Abstract 237**]

Shariq MOHAMMED, *Boston University*
Santhoshi N. KRISHNAN, *Rice University and University of Michigan*
Timothy L. FRANKEL, *University of Michigan*
Arvind RAO, *University of Michigan*

27.A2.I35 Bayesian nonparametrics - foundation

Venue: MSCR-1

Chair : Raffaele ARGIENTO, Università degli studi di Bergamo
Organizer : Peter MUELLER, UT Austin

17:10 **Impact of the Prior in Bayesian Nonparametric Models** [**Abstract 57**]

Marta CATALANO, *University of Warwick*
Hugo LAVENANT,

17:35 **Random probability measures with fixed mean distributions** [**Abstract 122**]

FRANCESCO GAFFI, *Bocconi University*

27.A2.C6 Contributed Session 6

Venue: Math-LH1

Chair : Ronit JAISWAL, PhD Scholar, ICAR-INDIAN AGRICULTURAL RESEARCH INSTITUTE, New Delhi

17:10 **A Deterministic Approximation Approach To The Continuum Logit Dynamic With An Application To Supermodular Games** [Abstract 247]

Sayan MUKHERJEE, *Indian Statistical Institute, Kolkata*
Ratul LAHKAR, *Department of Economics, Ashoka University*
Souvik ROY, *Applied Statistics Unit, Indian Statistical Institute, Kolkata*

17:25 **Asymptotic Behaviour of Bonferroni, Holm and Benjamini-Hochberg Procedures under Correlated Normal** [Abstract 109]

Monitirtha DEY, *Indian Statistical Institute, Kolkata*
Subir Kumar BHANDARI, *Indian Statistical Institute, Kolkata*

17:40 **PARTIALLY BALANCED BIPARTITE BLOCK DESIGNS** [Abstract 372]

VINAYAKA, *PhD Scholar (Agricultural Statistics), ICAR-IASRI, New Delhi-110012*
Rajender PARSAD, *Principal Scientist, ICAR-IASRI, Pusa, New Delhi-110012*
B. N. MANDAL, *Senior Scientist, ICAR-IASRI, Pusa, New Delhi-110012*
Sukanta DASH, *Scientist, ICAR-IASRI, Pusa, New Delhi-110012*

17:55 **Classical estimation in log-logistic distribution using adaptive progressively type-II censored data** [Abstract 200]

Anita KUMARI, *Department of Statistics, Central University of Haryana, Mahendragarh India*
Kapil KUMAR, *Department of Statistics, Central University of Haryana, Mahendragarh, India*

18:10 **CNN-LSTM model for agricultural price forecasting using genetic algorithm optimization technique** [Abstract 166]

Ronit JAISWAL, *PhD Scholar, ICAR-INDIAN AGRICULTURAL RESEARCH INSTITUTE, New Delhi*
Girish Kumar JHA, *Principal Scientist, ICAR-IARI, New Delhi*

27.A2.C7 Contributed Session 7 *Venue:* Math-LH3

Chair : Priya YADAV, Department of Statistics, Central University of Haryana, Mahendergarh, India

17:10 **Statistical Inference for Generalized half-normal distribution for simple step stress case in the presence of hybrid censoring** [Abstract 380]

Priya YADAV, *Department of Statistics, Central University of Haryana, Mahendergarh, India*
Devendra KUMAR, *Department of Statistics, Central University of Haryana, India*
Sukhdev SINGH, *Department of Statistics, Amity University, Mohali, India*

17:25 **Bayesian group-shrinkage based estimation in panel VAR models with mixed frequency data** [Abstract 63]

Nilanjana CHAKRABORTY, *nchakraborty@ufl.edu*
Kshitij KHARE, *University of Florida*
George MICHAILIDIS, *University of Florida*

17:40 **Discriminating between generalized inverted half Rayleigh distribution and generalized inverted half logistic distribution.** [Abstract 113]

Bishal DIYALI, *Department of Statistics, Central University of Haryana*
Devendra KUMAR, *Department of Statistics, Central University of Haryana*

17:55 **EFFICIENT SEQUENTIAL THIRD ORDER RESPONSE SURFACE DESIGNS** [Abstract 371]

ANKITA VERMA, *STUDENT*
SEEMA JAGGI, *ADG (HRD) ,ICAR , New Delhi- 110012*

ELDHO VARGHESE, *Senior Scientist, ICAR-CMFRI, Kochi-682018*
ANINDITA DATTA, *Scientist, ICAR-IASRI, New Delhi-110012*

18:10 **A Bayesian Record Linkage Approach that Adjusts for Variables in One Dataset [Abstract 178]**

Gauri KAMAT, *Brown University*
Mingyang SHAN, *Eli Lilly and Company*
Roee GUTMAN, *Brown University*

27.A2.C8 Contributed Session 8

Venue: Math-LH4

Chair : Shiwani TIWARI, Banaras Hindu University

17:10 **A Novel Image Monitoring Procedure Using Jump Regression [Abstract 293]**

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

17:25 **A machine learning approach to predict vessel trajectories in real-time using AIS data [Abstract 184]**

Rohit KATTI, *S & P Global*
Jyotirmoy PATRA, *S & P Global*
Nidhi RAI, *S & P Global*
Sandip PAL, *S & P Global*

17:40 **Asymmetric Rotatable Designs for Fitting Response Surface Model Incorporating Neighbour Effects [Abstract 88]**

ASHUTOSH DALAL, *ICAR-INDIAN AGRICULTURAL STATISTICS RESEARCH INSTITUTE*
SEEMA JAGGI, *Indian Council of Agricultural Research, Education Division*
ELDHO VARGHESE, *ICAR-Central Marine Fisheries Research Institute*

17:55 **Likelihood-based inference for skewed responses in a crossover trial setup. [Abstract 269]**

Savita PAREEK, *IIT Bombay*
Siuli MUKHOPADHYAY, *IIT Bombay*
Kalyan DAS, *IIT Bombay*

18:10 **Joint Calibration Approach-Based Estimator for Dual Frame Surveys using Two Auxiliary Variables [Abstract 362]**

Shiwani TIWARI, *Banaras Hindu University*
Piyush RAI, *Department of Statistics, Banaras Hindu University, Varanasi, Uttar Pradesh, India*

Wednesday December 28

28.M1.I36 Modern methods for high-dimensional and distributed multivariate analysis *Venue:* `NSSC-A`

Chair : Samrat ROY, University of Pennsylvania
Organizer : Aaron MOLSTAD, University of Florida

9:00 **Gaussian Process Regression and Classification Using International Classification of Disease Codes as Covariates [Abstract 354]**

Sanvesh SRIVASTAVA, *The University of Iowa*
Sanvesh SRIVASTAVA, *University of Iowa*
Yunyi LI, *University of Iowa*
Stephanie GILBERTSON-WHITE, *University of Iowa*

9:25 **Regularized high dimension low tubal-rank tensor regression [Abstract 299]**

Samrat ROY, *University of Pennsylvania*
George MICHAILIDIS, *University of Florida*

9:50 **HODOR: A two-stage hold-out design for online controlled experiments on networks [Abstract 332]**

Srijan SENGUPTA, *North Carolina State University*
Nicholas LARSEN, *North Carolina State University*
Jonathan STALLRICH, *North Carolina State University*

28.M1.I37 Recent advances in Bayesian Computation for high-dimensional data *Venue:* `NSSC-B`

Chair : Srijata SAMANTA, M D Anderson Cancer Center
Organizer : Jyotishka DATTA, Virginia Tech

9:00 **Graphical Evidence [Abstract 33]**

Anindya BHADRA, *Purdue University*
Ksheera SAGAR, *Purdue University*
Sayantan BANERJEE, *IIM Indore*
Jyotishka DATTA, *Virginia Tech*

9:25 **Consistent Bayesian Variable Selection in High-Dimensional Hierarchical Regression [Abstract 316]**

Srijata SAMANTA, *M D Anderson Cancer Center*
Kshitij KHARE,
George MICHAILIDIS,

9:50 **New Directions in Bayesian Shrinkage for Sparse, Structured Data [Abstract 101]**

Jyotishka DATTA, *Virginia Tech*
Boss JONATHAN , *University of Michigan Biostatistics*
David DUNSON, *Duke University Statistics*
Matthew HEINER , *Brigham Young University*

28.M1.I38 Biostatistics *Venue:* `NSSC-C`

Chair : Karabi NANDY, University of Texas Southwestern Medical Center
Organizer : Subhashis GHOSHAL, North Carolina State University

9:00 **Scalable Data Integration in Genome-wide Association Studies (GWAS) through Generalized Method of Moments** [**Abstract 70**]

Nilanjan CHATTERJEE, *Johns Hopkins University*
Prosenjit KUNDU, *Johns Hopkins University*

9:25 **Model-based Clustering of Multiple Images Incorporating Covariates** [**Abstract 219**]

Amita MANATUNGA, *Emory University*
Ying CUI, *Emory University*
Jeong Hoon JANG, *Yonsei University*

9:50 **A Psychometric Analyses of the Concise Health Risk Tracking- Self-Report A Measure of Suicidal Risk: Performance in Adolescent Outpatients** [**Abstract 254**]

Karabi NANDY, *University of Texas Southwestern Medical Center*
Augustus John RUSH, *Curbstone consultant LLC*
Madhukar TRIVEDI, *University of Texas Southwestern Medical Center*
Thomas CARMODY, *University of Texas Southwestern Medical Center*

28.M1.I39 Dynamic Treatment Regimes and Reinforcement Learning *Venue:* `MSSH`

Chair and Organizer : Bibhas CHAKRABORTY, National University of Singapore & Duke University

9:00 **Adaptive Randomization in Sequential Multiple-Assignment Randomized Trial (SMART)** [**Abstract 136**]

Palash GHOSH, *Department of Mathematics, Indian Institute of Technology Guwahati, India*
Rik GHOSH, *Department of Mathematics, Indian Institute of Technology Guwahati India*
Bibhas CHAKRABORTY, *Duke-NUS Medical School, Singapore*
Megan E. PATRICK, *University of Michigan, USA*

9:25 **Designing Micro-randomized Trials for Mobile Health Studies via Thompson Sampling** [**Abstract 61**]

Bibhas CHAKRABORTY, *National University of Singapore & Duke University*
Xueqing LIU, *National University of Singapore*
Nina DELIU, *Sapienza University of Rome*

9:50 **Estimating heterogeneous treatment effects with right-censored data via causal survival forests** [**Abstract 87**]

Yifan CUI, *Zhejiang University*

28.M1.I40 Graphical models and dependence *Venue:* `MSCR-1`

Chair : Igor PRUENSTER, Bocconi University
Organizer : Pierpaolo DE BLASI, University of Torino and Collegio Carlo Alberto, Italy

9:00 **Clustering consistency with Dirichlet process mixtures** [**Abstract 290**]

Giovanni REBAUDO, *The University of Texas at Austin*
Filippo ASCOLANI, *Bocconi University*
Antonio LIJOI, *Bocconi University*
Giacomo ZANELLA, *Bocconi University*

9:25 **Hierarchical Modeling of Covariance Structure via Subspace Factor Analysis** [**Abstract 67**]

Noirrit Kiran CHANDRA, *The University of Texas at Dallas*
Jason XU, *Duke University*
David B. DUNSON, *Duke University*

9:50 **Bayesian nonparametric multilayer clustering of longitudinal data** [Abstract 120]

Beatrice FRANZOLINI, *franzolini@pm.me*
Maria DE IORIO, *University College London / National University of Singapore / Agency for Science Technology and Research*

28.M1.I41 Advances in Markov chain Monte Carlo *Venue:* MSCR-2

Chair and Organizer : Dootika VATS, Indian Institute of Technology Kanpur

9:00 **Lower Bounds on the Rate of Convergence for Metropolis-Hastings** [Abstract 172]

Galin JONES, *University of Minnesota*
Austin BROWN, *University of Minnesota*

9:25 **Stability of particle filters for chaotic, deterministic dynamical systems** [Abstract 6]

Amit APTE, *Indian Institute of Science Education and Research (IISER) Pune*
Anugu Sumith REDDY, *Indian Institute of Technology Bombay, Mumbai, India*
Pinal MANDAL, *International Centre for Theoretical Sciences TIFR, Bangalore, India*
Shashank Kumar ROY, *International Centre for Theoretical Sciences TIFR, Bangalore, India*

9:50 **Output Analysis for Parallel MCMC** [Abstract 369]

Dootika VATS, *Indian Institute of Technology Kanpur*
Agarwal MEDHA, *University of Washington, Seattle*
Kushagra GUPTA, *Stanford University*

Student Paper Competition 2 *Venue:* Math-LH1

Chair : Joyee GHOSH, The University of Iowa

9:00 **Model-based Personalized Synthetic Magnetic Resonance Imaging** [Abstract 264]

Subrata PAL, *Department of Statistics, Iowa State University*
Somak Dutta, Ranjan Maitra ,

9:15 **Bivariate DeepKriging for Large-scale Spatial Interpolation of Wind Fields** [Abstract 250]

Pratik NAG, *King Abdullah University of Science and Technology*
Ying sun, Brian J Reich ,

9:30 **A spatial copula interpolation in a random field with application in air pollution data** [Abstract 361]

Debjoy THAKUR, *Indian Institute of Technology, Tirupati*
Dr. Ishapathik Das, Ms. Shubhashree Chakravarty ,

9:45 **Efficient Model Compression for Sparse Bayesian Neural Networks** [Abstract 307]

Diptarka SAHA, *University of Illinois, Urbana - Champaign*
Zihe Liu, Feng Liang ,

10:00 **Model-based imputation methods for small area estimation** [Abstract 325]

Aditi SEN, *University of Maryland, College Park*
Dr. Partha Lahiri ,

28.M1.C9 Contributed Session 9 *Venue:* `Math-LH3`

Chair : Arkajyoti BHATTACHARYA, Post doctoral researcher, University of Groningen, The Netherlands

9:00 **The relationship of agricultural wage rate among states in India: A Vector Error Correction Model Analysis [Abstract 55]**

Yashavanth BS, *yashavanthbs@gmail.com*
Rakesh J, *Professor Jayashankar Telangana State Agricultural University*

9:15 **Multivariate strong invariance principle for Markov chain Monte Carlo [Abstract 22]**

Arka BANERJEE, *arkabee96@gmail.com*
Dootika VATS, *Indian Institution of Technology, Kanpur.*

9:30 **CONSTRUCTION OF STRUCTURALLY INCOMPLETE ROW-COLUMN DESIGNS [Abstract 225]**

B MANJUNATHA, *Indian Agricultural Statistics Research Institute-New Delhi*
Rajender PARSAD, *Principal Scientist, ICAR-IASRI, Pusa, New Delhi*
B.N. MANDAL, *Senior Scientist, ICAR-IASRI, Pusa, New Delhi*
Sukanta DASH, *Scientist, ICAR-IASRI, Pusa, New Delhi*

9:45 **Hybrid intelligence: An approach by combining human intelligence and Machine learning models for freight rate forecasting in the Dry bulk shipping market [Abstract 227]**

Irene MATHEWS, *S&P Global*
Rajeev PANDEY, *S&P Global*

10:00 **A divide-and-conquer approach to perform consensus independent component analysis on large scale datasets [Abstract 40]**

Arkajyoti BHATTACHARYA, *Post doctoral researcher, University of Groningen, The Netherlands*
Setareh REZAEE OSHTERNIAN, *PhD student, University of Groningen*
Mark FREDERIKS, *MSc*
Rudolf S. N. FEHRMANN, *Professor, MD, PhD, University of Groningen*

28.M1.C10 Contributed Session 10 *Venue:* `Math-LH4`

Chair : Abhishek BHATTACHARJEE, Pfizer

9:00 **Reliability Estimation for Bathtub Shaped Distribution under Progressive Type II Block Censoring [Abstract 201]**

Rani KUMARI, *National Institute of Technology Patna*
Rani KUMARI, *Department of Mathematics, National Institute of Technology Patna, India*
Yogesh Mani TRIPATHI, *Department of Mathematics, Indian Institute of Technology Patna, India*
Rajesh Kumar SINHA, *Department of Mathematics, National Institute of Technology Patna, India*

9:15 **Model-Based Clustering Using Finite-Mixture of Power-Normal Distribution [Abstract 176]**

Shiva Kumar K, *Pondicherry University*

9:30 **Probability Density-Based Clustering [Abstract 35]**

Shrikrishna BHAT K, *Pondicherry University*

9:45 **Death Process Approach for Modelling Changes in Marriage Probabilities [Abstract 152]**

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

10:00 **A weighted test to evaluate contribution of components in a randomized trial with combination therapy [Abstract 36]**

Abhishek BHATTACHARJEE, *Pfizer*

Coffee 10:20 - 11:00 Venue: NSSC

Special Invited Session 2 Bas Kleijn, R. V. Ramamoorthi

Venue: JNTA

Chair : Subhashis GHOSHAL, North Carolina State University

11:00 **Confidence sets in a sparse stochastic block model [Abstract 189]**

Bas KLEIJN, *b.j.k.kleijn@uva.nl*
Jan VAN WAAIJ, *Department of Mathematical Sciences, University of Copenhagen*

11:40 **Doob's theorem revisited [Abstract 283]**

R V RAMAMOORTHI, *ramamoor@msu.edu*

28.M2.I42 Modern Applied Statistics

Venue: NSSC-A

Chair : Marepalli RAO, University of Cincinnati
Organizer : IISA,

11:00 **Topological Data Analysis in Finance A Review of Applications [Abstract 205]**

Arnab LAHA, *Indian Institute of Management Ahmedabad*

11:25 **Data Arising from Destructive Samples: Some Informative Models: The Case of Spina Bifida [Abstract 288]**

Marepalli RAO, *University of Cincinnati*
Rigwed TATU, *Cincinnati Children's Hospital, Cincinnati*
Anand SETH, *SK Patent Associates, Columbus*
Shesh RAI, *Cancer Data Science Center, University of Cincinnati*

11:50 **Adjustment for the uncertainty of predicted expression in transcriptome-wide association study: an efficient approach based on measurement error theory and bootstrapping [Abstract 215]**

Arunabha MAJUMDAR, *Assistant Professor, Department of Mathematics, Indian Institute of Technology Hyderabad*
Arunabha MAJUMDAR, *Department of Mathematics, Indian Institute of Technology Hyderabad*
Tanushree HALDAR, *Department of Clinical Pharmacy, University of California San Francisco*

28.M2.I43 Understanding the failures of neural networks

Venue: NSSC-B

Chair and Organizer : Aditi RAGHUNATHAN, Carnegie Mellon University

11:00 **Pitfalls of Deep Learning [Abstract 258]**

Praneeth NETRAPALLI, *Google Research India*

11:25 **Understanding the failure modes of out-of-distribution generalization [Abstract 251]**

Vaishnavh NAGARAJAN, *Google*
Anders ANDREASSEN,
Behnam NEYSHABUR,

11:50 **Estimating and improving the performance of machine learning under natural distribution shifts [Abstract 281]**

Aditi RAGHUNATHAN, *Carnegie Mellon University*
Ananya KUMAR, *Stanford*
Christina BAEK, *Carnegie Mellon University*

28.M2.I44 Machine learning Venue: NSSC-C

Chair : Michele GUINDANI, University of California, Los Angeles
Organizer : Peter MUELLER, UT Austin

11:00 **Usage of Random Compression Matrices for Efficient Bayesian Computation [Abstract 150]**

Rajarshi GUHANIYOGI, *Texas A & M University*
Aaron SCHEFFLER, *UC San Francisco*
Rene GUTIERREZ, *Texas A & M University*

11:25 **DocLens: Digitizing hard-to-read handwritten healthcare documents [Abstract 275]**

Sujoy PAUL, *Google Research*

11:50 **Bayesian dimensionality reduction via the identifications of the data intrinsic dimensions [Abstract 232]**

Antonietta MIRA, *Universit della Svizzera italiana and University of Insubria*
Denti FRANCESCO, *Universit Cattolica Milano*
Di Noia ANTONIO, *Universit della Svizzera italiana*
Alessandro LAIO, *laio@sissa.it*

28.M2.I45 Recent developments in exploratory data analysis Venue: MSSH

Chair : Thekke Variyam RAMANATHAN, Savitribai Phule Pune University

11:00 **Focused Model Selection for Time Series Data [Abstract 285]**

Thekke Variyam RAMANATHAN, *Savitribai Phule Pune University*

11:25 **Managing the Biological Invasion by Invasive Alien Plant Species in India: Insights from the ILORA Database using Data Science Applications [Abstract 46]**

Amiya Ranjan BHOWMICK, *Institute of Chemical Technology, Mumbai*
Jyoti Jagdish PRAJAPATI, *Institute of Chemical Technology, Mumbai*
Achyut Kumar BANERJEE, *Sun Yat-sen University, Guangzhou, China*
Abhishek MUKHERJEE , *Indian Statistical Institute, Giridih Branch, Jharkhand*

11:50 **Geometric EDA for Random Objects [Abstract 114]**

Paromita DUBEY, *University of Southern California*

28.M2.I46 Challenges in the Analysis of Longitudinal Studies

Venue: MSCR-1

Chair and Organizer : Abdus SATTAR, Case Western Reserve University

11:00 **INFERENCE WITH JOINT MODELS UNDER MISSPECIFIED RANDOM EFFECTS DISTRIBUTIONS** [**Abstract 324**]

Abdus SATTAR, *Case Western Reserve University*
Sanjoy SINHA, *Carleton University*

11:25 **Constrained inference in mixed models for clustered data** [**Abstract 352**]

Sanjoy SINHA, *Carleton University, Ottawa, Canada*

11:50 **Adjusted Limited Dependent Variable Mixture Model for quality of health data** [**Abstract 23**]

Hiya BANERJEE, *Novartis Pharmaceuticals*
Elke-Christine ORTMANN, *Novartis Pharmaceuticals*
Minal JAIN, *Novartis Pharmaceuticals*

28.M2.I47 Inference for high-dimensional and functional data

Venue: MSCR-2

Chair : Soham SARKAR, Indian Statistical Institute, Delhi Centre
Organizer : Subhajit DUTTA, IIT Kanpur, India

11:00 **Multi-sample comparison using spatial signs for infinite dimensional data** [**Abstract 83**]

Joydeep CHOWDHURY, *King Abdullah University of Science and Technology*
Probal CHAUDHURI, *Indian Statistical Institute, Kolkata*

11:25 **Estimating the number of components in functional MRI (fMRI) data for Independent Component Analysis (ICA) using Entropy** [**Abstract 255**]

Rajesh NANDY, *UNT Health Science Center*

11:50 **High Dimensional Logistic Regression Under Network Dependence** [**Abstract 248**]

Somabha MUKHERJEE, *National University of Singapore*
Sagnik HALDER, *University of Florida*
Bhaswar BHATTACHARYA, *University of Pennsylvania*
George MICHAILIDIS, *University of Florida*

Lunch 12:20 - 14:00 Venue: NSSC

Plenary Lecture 3 Igor Pruenster

Venue: JNTA

Chair : Peter MUELLER, UT Austin

14:00 **Prediction rules, random partitions and borrowing of information: an invitation to Bayesian Nonparametrics** [**Abstract 279**]

Igor PRUENSTER, *Bocconi University*

Switching time 15:00 - 15:10

Panel Discussion 3 Women in Statistics and Data Science

Venue: JNTA

Organizers : *Pallavi CHITTURI, Temple University* and *Amarjot KAUR, Merck*

Moderator: Amarjot KAUR, Merck

- **Sayantee JANA**, IIT Hyderabad
- **Jayanti GUPTA**, NishKash Consulting
- **Amita MANATUNGA**, Emory University
- **Radha RAILKAR**, Merck
- **Pallavi CHITTURI**, Temple University

28.A1.I48 Some advances in Bayesian methods and computation for robust inference in complex data

Venue: NSSC-A

Chair : Giovanni REBAUDO, The University of Texas at Austin

Organizer : Abhra SARKAR, The University of Texas at Austin

15:10 **Bayesian nonparametric common atoms regression for generating synthetic matching populations** [**Abstract 240**]

Peter MUELLER, *UT Austin*
Noirrit CHANDRA, *UT Dallas*
Abhra SARKAR, *UT Austin*
Ying YUAN, *MDACC*

15:35 **Robust probabilistic inference via a constrained transport metric** [**Abstract 271**]

Debdeep PATI, *Texas A&M University*

16:00 **Generalized Bayesian Additive Regression Trees Models** [**Abstract 210**]

Antonio LINERO, *University of Texas at Austin*

28.A1.I49 Statistical Modeling of Covid-19 Pandemic

Venue: NSSC-B

Chair and Organizer : Debashis PAUL, University of California, Davis

15:10 **Bayesian Regularized Estimation of Latent Spatial Transmission of COVID-19.** [**Abstract 249**]

Ujjal Kumar MUKHERJEE, *College of Business, and Carle Illinois College of Medicine, UIUC*
Snigdhansu CHATTERJEE, *School of Statistics, UMN*

15:35 **Early assessment of potential endemic pattern of COVID19 in India** [**Abstract 37**]

Madhuchhanda BHATTACHARJEE, *University of Hyderabad, Hyderabad, India*
Arup BOSE, *Indian Statistical Institute, Kolkata, India*

16:00 **Time-dynamics of the COVID-19 Pandemic: Inference and Mitigation Strategy** [**Abstract 39**]

Satarupa BHATTACHARJEE, *UC Davis*
Shuting LIAO, *Graduate Group in BioStatistics, University of California, Davis, 95616, USA*
Debashis PAUL, *Department of Statistics, University of California, Davis, 95616, USA*
Sanjay CHAUDHURI, *Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117546, Singapore*

28.A1.I50 Random Growth and Games

Venue: NSSC-C

Chair : Riddhipratim BASU, International Centre for Theoretical Sciences

Organizers : *Moumanti PODDER, Indian Institute of Science Education and Research (IISER) Pune* and *Parthanil ROY, Indian Statistical Institute*

15:10 Space time correlations in models of random growth [Abstract 30]

Riddhipratim BASU, *International Centre for Theoretical Sciences*

15:35 Percolation models with correlations [Abstract 143]

Subhajit GOSWAMI, *Tata Institute of Fundamental Research*

16:00 A Game-Theoretic approach for disease propagation [Abstract 306]

Soumyarup SADHUKHAN, *Indian Institute of Technologuy Kanpur*
Sayar KARMAKAR , *Department of Statistics, University of Florida*
Moumanti PODDER , *Department of Mathematics, Indian Institute of Science Education and Research, Pune*
Souvik ROY, *Applied Statistics Unit, Indian Statistical Institute, Kolkata*

28.A1.I51 Modern methods time series analysis

Venue: MSSH

Chair : Junho YANG, Academia Sinica

Organizer : Suhasini SUBBARAO, suhasini@tamu.edu

15:10 Baxter-type convergence results for a block Toeplitz system under long memory with an application to linear prediction problem [Abstract 382]

Junho YANG, *Academia Sinica*
Akihiko INOUE,

15:35 Modeling Temporal Networks of Relational Events Data [Abstract 274]

Subhadeep PAUL, *The Ohio State University*

16:00 TBA [Abstract 218]

Tuhin MAJUMDER,

28.A1.I52 Large-scale Statistical Inference and applications

Venue: MSCR-1

Chair : Somya SINGHVI,

Organizer : Gourab MUKHERJEE, University of Southern California

15:10 A Scalable Dynamic Bayesian Mixture Model for Fine-grained Promotion Mix Analysis of Digital Coupons [Abstract 244]

Gourab MUKHERJEE, *University of Southern California*
Rashmi BHUYIAN, *University of Southern California*
Wreetabrata KAR, *Purdue University*

15:35 Estimation and Inference in Inhomogeneous Random Graph Models [Abstract 42]

Bhaswar BHATTACHARYA, *University of Pennsylvania*
Anirban CHATTERJEE , *University of Pennsylvania*
Soham DAN , *University of Pennsylvania*
Svante JANSON , *Uppsala University*

16:00 On Bayesian Methods for Change-point detection in High Dimension [Abstract 146]

Nilabja GUHA, *UMASS Lowell*
Jyotishka DATTA, *Virginia Tech*

28.A1.I53 Statistics of Financial Markets

Venue: MSCR-2

Chair and Organizer : Rituparna SEN, Indian Statistical Institute, Bangalore

15:10 **Simultaneous Confidence Intervals for Multi-way Clustered Stock Return Data [Abstract 191]**

Tamalika KOLEY, *Birla Institute of Technology Mesra*
Mrinal JANA, *Birla Institute of Technology Mesra*
Gopal BASAK, *Indian Statistical Institute, Kolkata*

15:35 **A lagged cross correction matrix model for Indian stock market data [Abstract 207]**

Ananya LAHIRI, *IIT Tirupati*
Nagavyshnavi GANTASALA, *IIT Tirupati*

16:00 **Option pricing using Hawkes Process [Abstract 342]**

Shubhangi SIKARIA, *Philips India Limited*
Rituparna SEN, *ISI Bangalore, India*

Coffee and Poster Setup 16:30 - 17:10 Venue: NSSC

Poster Session

Venue: NSSC

Time : **17:10 - 18:30**

- **Stochastic comparisons between two finite mixture models with exponentiated location-scale distributed components [Abstract 34]**

  RAJU BHAKTA, *National Institute of Technology Rourkela*
  Suchandan Kayal ,

- **Efficient Model Fitting and Two-Sample Testing for Large Networks via Subsampling [Abstract 62]**

  Kaustav CHAKRABORTY, *University of Illinois at Urbana-Champaign*
  Srijan Sengupta, Yuguo Chen ,

- **Transmuted Cure Models- A Competing Risk Perspective [Abstract 86]**

  Soorya C. S., *Department of Statistics, Cochin University of Science and Technology*
  Dr. Asha Gopalakrishnan ,

- **INTERMITTENT STATE WISE CANCER SURVIVAL RISK ESTIMATION USING MULTI-STATE FRAILTY MARKOV REGRESSION MODEL [Abstract 107]**

  DIVYA DENNIS, *Division of Cancer Epidemiology & Biostatistics, Regional Cancer Centre, Thiruvananthapuram, Kerala, India*
  Dr. Jagathnath Krishna K M, Dr. Preethi Sara George, Dr. Aleyamma Mathew ,

- **Bayesian and non-Bayesian inference of Weibull lifetime model based on partially observed competing risks data under unified hybrid censoring scheme [Abstract 116]**

  Subhankar DUTTA, *National Institute of Technology, Rourkela*
  Suchandan Kayal ,

- **Development of Model Confidence Set Based Ensemble Models Using CEEMDAN Decomposition and Machine Intelligence** [Abstract 124]

  Sandip GARAI, *Phd Scholar (Ag. Statistics)*
  Ranjit Kumar Paul ,

- **Nonparametric Screening and Selection in presence of dependence among predictors** [Abstract 139]

  Shanta GHOSH, *The University of Illinois at Chicago*
  Dr. Sanjib Basu ,

- **Efficient Approximation of Interval Specific Estimates for Biological Growth Curve Models using Localized Likelihood Maximization and Its Applications to Real Data Analysis** [Abstract 179]

  Md Aktar Ul KARIM, *Institute of Chemical Technology, Mumbai*
  Amiya Ranjan BHOWMICK, *Institute of Chemical Technology, Mumbai*

- **Individualized Inference in Bayesian Quantile Directed Acyclic Graphical Models** [Abstract 186]

  Ksheera Sagar KERALAPURA NAGARAJU, *Purdue University*
  Yang Ni, Veera Baladandayuthapani and Anindya Bhadra ,

- **Testing Quantiles of Several Logistic Populations with a Common Location** [Abstract 188]

  Habiba KHATUN, *National Institute of Technology Rourkela*
  Prof. Manas Ranjan Tripathy ,

- **Classification Rules for Two Inverse Gaussian Populations with Common Dispersion Parameter and Unequal Means** [Abstract 196]

  Pushkal KUMAR, *National Institute of Technolgy, Rourkela*
  Manas Ranjan Tripathy ,

- **Bivariate Distribution with Singular Components** [Abstract 198]

  Sanjay KUMAR, *PhD Student*
  NA ,

- **Data-Driven Selection of Fractional Differential Operator with Applications to Real Data** [Abstract 231]

  Dipali V. MESTRY, *Institute of Chemical Technology, Mumbai*
  Pratik Singh, Amiya Ranjan Bhowmick ,

- **Estimation methods for elementary chirp model parameters** [Abstract 235]

  Anjali MITTAL, *Indian Institute of Technology Kanpur*
  Prof. Rhythm Grover, Prof. Debasis Kundu, Prof. Amit Mitra ,

- **Convergence of Gibbs sampler in spatial location-scale mixture models under different centering and scaling parameterizations** [Abstract 242]

  Arghya MUKHERJEE, *Indian Institute of Technology, Kanpur*
  Arnab Hazra ,

- **A Latent Class Trajectory Model for Penile Microbiome: Mapping longitudinal heterogeneity in response to baseline behavioral factors and disease status** [Abstract 252]

  Debarghya NANDI, *Adjunct Professor*
  Supriya D. Mehta, Dulal Bhaumik ,

- **An Integrated Bayesian Framework for Multi-omics Prediction and Classification** [**Abstract 278**]

  Anupreet PORWAL, *University of Washington*
  Himel Mallick, Erina Paul, Satabdi Saha, Vladimir Svetnik ,

- **Some Inadmissibility Results on Estimating Ordered Scale Parameters of Two Exponential Populations with Different Locations using Progressive Type-II Censored Data** [**Abstract 318**]

  Mojammel Haque SARKAR, *National Institute of Technology Rourkela*
  Manas Ranjan Tripathy ,

- **Detection of Anomalous Behaviour of Aircraft's Engines for Predictive Maintenance** [**Abstract 320**]

  Shibajee SARKAR, *Ramakrishna Mission Vivekananda Educational and Research Institute*
  Srirup Mitra, Dr. Sudipta Das, Prof. Debasis Sengupta ,

- **A novel GEE type approach for analyzing zero-inflated clustered ordinal longitudinal data with an application to dental studies** [**Abstract 321**]

  Shoumi SARKAR, *University of Florida*
  Somnath Datta, Peihua Qiu, Jeremy Gaskins, Steven Levy ,

- **Inference of adaptive type-II progressively censored competing risks data for generalized inverted exponential distribution** [**Abstract 344**]

  Kundan SINGH, *Department of Mathematics, IIT Patna*
  Yogesh Mani Tripathi ,

- **Guided unsupervised outlier detection using domain knowledge** [**Abstract 363**]

  Vinay TIWARI, *Savitribai Phule Pune University*
  Dr. Akanksha Kashikar ,

- **An Application of Elliptical Spatial Linear Models** [**Abstract 365**]

  ASHIS RANJAN UDGATA, *ICAR-Indian Agricultural Statistics Research Institute*
  ANIL RAI, *ICAR-Indian Agricultural Statistics Research Institute*
  ANKUR BISWAS, *ICAR-Indian Agricultural Statistics Research Institute*
  NOBIN CHANDRA PAUL, *ICAR-Indian Agricultural Statistics Research Institute*

- **Modelling of sparse conditional spatial extremes processes subject to left-censoring** [**Abstract 381**]

  Rishikesh YADAV, *HEC Montreal*
  Leo R. Belzile, Nicholas R. Beck ,

28.A2.C11 Contributed Session 11

Venue: `Math-LH1`

Chair : Sagnik MONDAL, King Abdullah University of Science and Technology

17:10 BIDIRECTIONAL LSTM-CNN BASED GENERATIVE ADVERSARIAL NETWORKS FOR TIME SERIES FORECASTING [**Abstract 10**]

G AVINASH, *PhD Scholar at ICAR-Indian Agricultural Statistics Research Institute*
Ramasubramanian VAIDHYANATHAN, *Principal Scientist at ICAR-IASRI*
Harish NAYAK G.H, *PhD Scholar at ICAR-IASRI*
Mrinmoy RAY, *Scientist at ICAR-IASRI*

17:25 **GET Association Schemes and Designs** [Abstract 157]

Mohd HARUN, *Scientist, ICAR-Indian Agricultural Statistics Research Instiyute, New Delhi-12*
Cini VARGHESE, *Professor (Agricultural Statistics), ICAR-Indian Agricultural Statistics Research Institute, New Delhi-12*

17:40 **Designing Factorial Experiments for Cropping Sequence Experiments** [Abstract 98]

Sukanta DASH, *ICAR-Indian Agricultural Statistics Research Institute*
Ankit SINGH, *ICAR-Indian Agricultural Statistics Research Institute*
Rajender PARSAD, *ICAR-Indian Agricultural Statistics Research Institute*
Anil KUMAR, *ICAR-Indian Agricultural Statistics Research Institute*

17:55 **Cultural and Socio-demographic determinants affecting Female sterilization over other methods of Family planning in India: Evidence from Recent round of NFHS surveys.** [Abstract 346]

Sanjiv SINGH , *Department of Statistics, Institute of Science , Banaras Hindu University, Varanasi(U.P) India*
Dr. Kaushlendra Kumar SINGH, *Department of Statistics, Institute of Science Banaras Hindu University, Varanasi (221005)*

18:10 **Parallel Approximation of the Tukey g-and-h Likelihoods for Large-Scale Non-Gaussian Geostatistical Modeling** [Abstract 239]

Sagnik MONDAL, *King Abdullah University of Science and Technology*
Sameh ABDULAH, *King Abdullah University of Science and Technology*
Hatem LTAIEF, *King Abdullah University of Science and Technology*
Marc GENTON, *King Abdullah University of Science and Technology*

28.A2.C12 Contributed Session 12 *Venue:* Math-LH3

Chair : VEERSHETTY, PhD Scholar at ICAR-IASRI

17:10 **Tukey's Cause- Selecting Control Charts** [Abstract 121]

Shital GADEKAR, *Department of Statistics Savitribai Phule Pune University Pune*
Manik AWALE, *Department of Statistics Savitribai Phule Pune University,Pune*

17:25 **Asymptotic Bayes optimality under sparsity for equicorrelated multivariate normal test statistics** [Abstract 297]

Rahul ROY, *Indian Statistical Institute, Kolkata*
Subir Kumar BHANDARI, *Indian Statistical Institute, Kolkata*

17:40 **Infant Mortality Rate: A Demographic and Socio Economic Assessment of the Empowered Action Group (EAG) States of India** [Abstract 81]

Anuradha Rajkonwar CHETIYA, *Ramjas College, University of Delhi*
Vishal DEO, *Ramjas College, University of Delhi*

17:55 **Bayesian Estimation in Correlated Gamma Frailty Models** [Abstract 337]

CHAKRDHAR SHELAKE, *Savitribai Phule Pune University*
CHAKRDHAR SHELAKE, *Savitribai Phule Pune University*
RAOSAHEB LLATPATE, *Savitribai Phule Pune University*
DAVID HANAGAL , *Savitribai Phule Pune University*

18:10 **Hybrid Statistical and Deep Learning Technique Over Different Viable Approaches for Improved Predictive Performance of Rainfall** [Abstract 370]

VEERSHETTY, *PhD Scholar at ICAR-IASRI*
G AVINASH, *PhD Scholar at ICAR-IASRI*
Harish NAYAK. G.H., *PhD Scholar at ICAR-IASRI*
Moumita BAISHYA, *PhD Scholar at ICAR-IASRI*

28.A2.C13 Contributed Session 13 *Venue:* Math-LH4

Chair : Jing XU, Duke National University of Singapore

17:10 Forecasting Agricultural Price Series using Multilayer Perceptron Neural Network and Metaheuristic Algorithms [Abstract 257]

Harish NAYAK.G.H., *Ph.D. Scholar at ICAR-IASRI, New Delhi*
Mohammad Wasi ALAM, *Principal Scientist at ICAR-IASRI, New Delhi*
G AVINASH, *PhD Scholar at ICAR-IASRI*
Mrinmoy RAY, *Scientist at ICAR-IASRI*

17:25 Examining the factors influencing enrollment and grade appropriate learning outcomes in Indian rural schools [Abstract 349]

VIJAY SINGH, *Great Learning*
Srabashi BASU, *Great Learning, India*
Saathveenaa MOHAN, *Great Learning, India*
Yukta JOSHI, *Great Learning, India*

17:40 Calibration Approach-Based Estimator under Dual Frame Surveys: Revisited [Abstract 199]

Shubham KUMAR, *shubh.stat@bhu.ac.in*
Piyush Kant RAI, *Banaras Hindu University, Varanasi*

17:55 Estimation in Jointly Type-II Censored k Lindley Distributions [Abstract 193]

Hare KRISHNA, *Department of Statistics, Chaudhary Charan Singh University, Meerut (INDIA)*

18:10 A Flexible Micro-Randomized Trial Design and Sample Size Considerations [Abstract 379]

Jing XU, *Duke National University of Singapore*
Jing XU, *Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore*
Xiaoxi YAN, *Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore*
Bibhas CHAKRABORTY, *Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore*

Thursday December 29

29.M1.I54 Large scale Bayesian analyses

Venue: NSSC-A

Chair : Abhirup DATTA, Johns Hopkins University
Organizer : Snigdhansu CHATTERJEE, University of Minnesota

9:00 **MARS via LASSO** [**Abstract** 153]

Adityanand GUNTUBOYINA, *University of California Berkeley*
Dohyeong KI, *University of California Berkeley*
Billy FANG, *Google*

9:25 **Machine learning for spatial generalized mixed models** [**Abstract** 99]

Abhirup DATTA, *Johns Hopkins University*

9:50 **Bayesian Inference for Atmospheric CO2 Retrieval using emulators and a two-stage adaptive Monte Carlo method** [**Abstract** 238]

Anirban MONDAL, *Case Western Reserve University*
Otto LAMMINPAEAE,

29.M1.I55 Methods for Imaging and Omics Data Science: Advances, Applications, and Spatiotemporal Innovations

Venue: NSSC-B

Chair : Anupreet PORWAL, University of Washington
Organizer : Himel MALLICK, Merck Research Labs

9:00 **RONSeq: A Robust Non-Parametric Method for Detecting Differential Gene Expression in RNA-Sequencing Studies** [**Abstract** 75]

Suvo CHATTERJEE, *Indiana University, Bloomington*
Arindam FADIKAR, *Argonne National Laboratory*

9:25 **A Two-Part Tweedie Model for Differential Analysis of Omics Data** [**Abstract** 45]

Arinjita BHATTACHARYYA, *Merck*
Suvo CHATTERJEE, *Indiana University*
Himel MALLICK, *Merck*

9:50 **Evaluation of a Clinical Decision Support System using a Machine Learning Based Predictive Model: A Stepped-Wedge Cluster-Randomized Trial** [**Abstract** 229]

Madhu MAZUMDAR, *Icahn School of Medicine at Mount Sinai*
Chen YANG, *Icahn School of Medicine at Mount Sinai*
Asem BERKALIEVA, *Icahn School of Medicine at Mount Sinai*
Deukwoo KWON, *Icahn School of Medicine at Mount Sinai*

29.M1.I56 Pandemic forecasting: Lessons learned from COVID-19

Venue: NSSC-C

Chair and Organizer : Bhramar MUKHERJEE, University of Michigan

9:00 **Aspects of modelling CoVID-19: Understanding and quantifying the uncertainty** [**Abstract** 233]

Swapnil MISHRA, *University of Copenhagen*

9:25 **Predictions, Role of Interventions and the Crisis of Virus in India: A Data Science Call to Arms** [**Abstract** 243]

Bhramar MUKHERJEE, *University of Michigan*

9:50 **TBA** [**Abstract** 2]

Sam ABBOTT,

29.M1.I57 Recent Advancements in Clinical Trial Conduct, Design and Analyses *Venue:* `MSSH`

Chair : Jay MANDREKAR, MAYO CLINIC
Organizer : Saptarshi CHATTERJEE, Eli Lilly and Co

9:00 **Using data analytics to overcome challenges with clinical trials data and to improve patient engagement and experience [Abstract 222]**

Jay MANDREKAR, *MAYO CLINIC*

9:25 **The design and implementation of precision medicine trials in oncology [Abstract 223]**

Sumithra MANDREKAR, *Department of Quantitative Health Sciences, Mayo Clinic*

9:50 **Recent Advancements in Alzheimer's Clinical Trial Analyses [Abstract 72]**

Saptarshi CHATTERJEE, *Eli Lilly and Co.*

29.M1.I58 Reliability & Survival Modeling *Venue:* `MSCR-1`

Chair : Shyamsundar SAHOO, Associate Professor, Department of Statistics, Haldia Government College

Organizer : Sudipta DAS, Ramakrishna Mission Vivekananda Educational and Research Institute

9:00 **A semi-parametric model of ice accummulation rate based on temperature and age of ice for Antarctic ice core data [Abstract 330]**

Debasis SENGUPTA, *Indian Statistical Institute*
Radhendushka SRIVASTAVA, *IIT Bombay*
Sreelekshmy S, *IIT Bombay*

9:25 **A flexible model with linear approximation for left truncated right censored data [Abstract 123]**

Ayon GANGULY, *Indian Institute of Technology Guwahati*
D MITRA, *IIM Udaipur*
N BALAKRISHNAN, *McMaster University*
D KUNDU, *IIT Kanpur*

9:50 **On estimation of the covariate's scale effect on failure time using martingale residuals [Abstract 313]**

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

29.M1.I59 Probability and its Applications *Venue:* `MSCR-2`

Chair : Sanjay CHAUDHURI, National University of Singapore
Organizer : Arka GHOSH, Iowa State University

9:00 **Existence of MLEs for parameters of an adjusted linear preferential attachment model of random graphs with covariates [Abstract 131]**

ARKA GHOSH, *apghosh@iastate.edu*
Sungchan PARK, *Iowa State University*

9:25 **Analysis of co-integration in networks [Abstract 145]**

Apratim GUHA, *XLRI, Xavier School of Management*
Anindya S CHAKRABARTI,

9:50 **Co-existence in probability models [Abstract 298]**

Rishideep ROY, *IIM Bangalore*
Kumarjit SAHA,

29.M1.C14 Contributed Session 14

Venue: Math-LH1

Chair : Antik CHAKRABORTY, Purdue University

9:00 **Inference of dependent competing risks from Marshall-Olikin bivariate Burr-XII distribution under generalized progressive hybrid censoring** **[Abstract 68]**

PRAKASH CHANDRA, *Indian Institute of Technology, Patna*
Prakash CHANDRA, *Indian Institute of Technology, Patna*
Yogesh Mani TRIPATHI, *Indian Institute of Technology, Patna*

9:15 **Small Area Prediction under a Spatially Correlated Multivariate Area Level Model and its Application in Education Survey Data of India** **[Abstract 147]**

SAURAV GUHA, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*
Rajender PARSAD, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*
Saumyadipta PYNE, *Department of Statistics and Applied Probability, University of California, Santa Barbara, CA, USA*
Sudeep MARWAHA, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*

9:30 **Ordinal Non-Symmetric Correspondence Analysis for Four-way Contingency Table using Orthogonal Polynomials** **[Abstract 165]**

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

9:45 **A tensorial approach to partition the Asymmetric Measure of Association for Multiway Contingency Table** **[Abstract 177]**

Kirtee KAMALJA, *Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, Maharashtra*

10:00 **Bayesian inference on high-dimensional multivariate binary responses** **[Abstract 60]**

Antik CHAKRABORTY, *Purdue University*
Rihui OU, *Duke University*
David DUNSON, *Duke University*

29.M1.C15 Contributed Session 15

Venue: Math-LH3

Chair : ASHKAMINI, Banaras Hindu University

9:00 **Estimating Prevalence of Post-war Health Disorders based on Capture-recapture experiment with Heterogeneous Catchability and Behavioural Dependence** **[Abstract 69]**

Kiranmoy CHATTERJEE, *Department of Statistics, Bidhannagar College, Kolkata*
Prajamitra BHUYAN, *Assistant Professor, Indian Institute of Management Kolkata*
Kiranmoy CHATTERJEE, *Assistant Professor, Department of Statistics, Bidhannagar College, Kolkata*

9:15 **Regularized Estimation of Sparse Spectral Precision Matrices** **[Abstract 105]**

Navonil DEB, *Cornell University*
Amy KUCEYESKI, *Cornell University*
Sumanta BASU, *Cornell University*

9:30 **Bayesian modelling of volatility using R and Stan** **[Abstract 256]**

Muhammed NAVAS T, *Research scholar*
Athar Ali KHAN, *Professor*

9:45 **Multilevel Programming Model with Neutrosophic Approach for Supplier Selection Problem in a Fuzzy Environment.** [**Abstract 214**]

Anas M, *Research Scholar, D/o Statistics & Operations Research, Aligarh Muslim University, Aligarh*
Ahmad Yusuf ADHAMI, *Assistant Professor, D/o Statistics & Operations Research, Aligarh Muslim University, Aligarh*
Anas M, *Research Scholar, D/o Statistics & Operations Research, Aligarh Muslim University, Aligarh*
Firoz AHMAD, *Department of Management Studies, Indian Institute of Science (IISc), Bangalore*

10:00 **On the Analysis of One-Shot Device Testing Data Using Copula Models: A Bayesian Approach** [**Abstract 9**]

ASHKAMINI, *Banaras Hindu University*
Reema SHARMA, *BANARAS HINDU UNIVERSITY*
Satyanshu K UPADHYAY, *BANARAS HINDU UNIVERSITY*

29.M1.C16 Contributed Session 16

Venue: Math-LH4

Chair : Prashant DHAMALE, SVKM'S NMIMS Deemed to be University, Mumbai/ Savitribai Phule Pune University, Pune

9:00 **Limit Theorems for Entropic Maps** [**Abstract 305**]

Ritwik SADHU, *Cornell University*
Ziv GOLDFELD, *Cornell University*
Kengo KATO, *Cornell University*
Gabriel RIOUX, *Cornell University*

9:15 **GENERALIZED ROW-COLUMN DESIGNS WITH FACTORIAL TREATMENT STRUCTURE** [**Abstract 100**]

Anindita DATTA, *Scientist*
Seema JAGGI, *Additional Director General (HRD), ICAR*
Eldho VARGHESE, *Senior Scientist, ICAR-CMFRI*
Cini VARGHESE, *Principal Scientist, ICAR-IASRI*

9:30 **Sculpting Human Fertility Using Variance-Adjusted Logistic Family of Distributions** [**Abstract 348**]

SHAMBHAVI SINGH, *BANARAS HINDU UNIVERSITY*
AKANKSHA GUPTA, *BANARAS HINDU UNIVERSITY*
SATYANSHU K. UPADHYAY, *BANARAS HINDU UNIVERSITY*

9:45 **Spatio-temporal modelling of malaria mortality in India from the Million Death Study** [**Abstract 169**]

Sayantee JANA, *IIT Hyderabad*
Sze Hang FU, *Dalla Lana School of Public Health, Centre for Global Health Research, St. Michaels Hospital, University of Toronto, Toronto, ON, Canada*
Hellen GELBAND, *Dalla Lana School of Public Health, Centre for Global Health Research, St. Michaels Hospital, University of Toronto, Toronto, ON, Canada*
Patrick BROWN, *Dalla Lana School of Public Health, Centre for Global Health Research, St. Michaels Hospital, University of Toronto, Toronto, ON, Canada Department of Statistical Sciences, University of Toronto, Toronto, Canada*

10:00 **Spectral matrix shrinkage estimation using various loss functions.** [**Abstract 111**]

Prashant DHAMALE, *SVKM'S NMIMS Deemed to be University, Mumbai/ Savitribai Phule Pune University, Pune*
Akanksha KASHIKAR, *Savitribai Phule Pune University, Pune*

Coffee 10:20 - 11:00 Venue: NSSC

Special Invited Session 3 Kavita Ramanan, Bodhisattva Sen *Venue:* `JNTA`

Chair : Sanjay CHAUDHURI, National University of Singapore

11:00 **Multivariate Distribution-free Nonparametric Testing using Optimal Transport** [**Abstract 327**]

Bodhisattva SEN, *Columbia University*
Nabarun DEB, *University of British Columbia, Vancouver*
Bhaswar BHATTACHARYA, *University of Pennsylvania*

11:40 **Beyond mean-field limits for stochastic processes on networks** [**Abstract 284**]

Kavita RAMANAN, *Brown University*

29.M2.I60 Recent Advances in Online Learning *Venue:* `NSSC-A`

Chair and Organizer : Sabyasachi CHATTERJEE, University of Illinois at Urbana Champaign

11:00 **Constrained Stochastic Nonconvex Optimization with State-dependent Markov Data** [**Abstract 17**]

Krishnakumar BALASUBRAMANIAN, *University of California, Davis*

11:25 **Game-theoretic formulation of sequential nonparametric two-sample tests** [**Abstract 336**]

Shubhanshu SHEKHAR, *Department of Statistics and Data Science, Carnegie Mellon University*
Aaditya RAMDAS,

11:50 **Spatially Adaptive Online Prediction of Piecewise Regular Functions** [**Abstract 71**]

Sabyasachi CHATTERJEE, *Assistant Professor*
Subhajit GOSWAMI, *TIFR, Bombay*

29.M2.I61 Novel approaches to measurement error correction in wearable devices *Venue:* `NSSC-B`

Chair and Organizer : Carmen TEKWE, Indiana University at Bloomington

11:00 **Bayesian Semi-Parametric Scalar-On-Function Quantile Regression with Measurement Error using GAL** [**Abstract 386**]

Roger ZOH, *Indiana University*

11:25 **Quantile regression with a mixture of function-valued covariate prone to complex heteroscedastic measurement errors and a scalar-valued covariate prone to classical measurement error** [**Abstract 359**]

Carmen TEKWE, *Indiana University at Bloomington*
Xiwei CHEN, *Indiana University at Bloomington*
Yuanyuan LUAN, *Indiana University at Bloomington*
Roger ZOH, *Indiana University at Bloomington*

11:50 **A function-based approach to model the measurement error in wearable devices** [**Abstract 163**]

Sneha JADHAV, *Wake Forest University*
Carmen TEKWE, *Indiana University Bloomington*
Yuanyuan LUAN,

29.M2.I62 Recent Trends in Biostatistical Methods *Venue:* `NSSC-C`

Chair and Organizer : Shibasish DASGUPTA, Pfizer and Chennai Mathematical Institute (CMI)

11:00 Single Cell RNA-Seq Data Analysis and its Applications in Health Sciences [Abstract 103]

Susmita DATTA, *University of Florida*
Michael SEKULA, *University of Louisville*
Jeremy GASKINS, *University of Louisville*

11:25 A response adaptive allocation for ordinal response clinical trials [Abstract 43]

Rahul BHATTACHARYA, *Professor*

11:50 A Bayesian joint model for multivariate longitudinal and event time data with application to ALL maintenance studies [Abstract 91]

Kiranmoy DAS, *Indian Statistical Institute, Kolkata*

29.M2.I63 Statistical Inference on Complex Data *Venue:* `MSSH`

Chair : Sandipan ROY,
Organizer : Sharmodeep BHATTACHARYYA, Oregon State University

11:00 Exact Tests for Offline Change Detection in Multichannel Binary and Count Data [Abstract 104]

Shyamal Krishna DE, *Indian Statistical Institute Kolkata*
Soumendu Sundar MUKHERJEE, *Indian Statistocal Institute Kolkata*

11:25 Aspects of Community Detection For Networks In the Presence Of Complexities [Abstract 73]

Shirshendu CHATTERJEE, *City University of New York, Graduate Center & City College*

11:50 A framework for modelling multiplex networks [Abstract 66]

Swati CHANDNA, *Birkbeck, University of London*

29.M2.I64 Statistical methods in environmental research *Venue:* `MSCR-1`

Chair and Organizer : Soudeep DEB, Indian Institute of Management Bangalore

11:00 Comprehensive Simultaneous Inference on Trend-Cycle Model [Abstract 182]

Sayar KARMAKAR, *University of Florida*
Wei Biao WU,

11:25 Robust statistical modeling of monthly rainfall: The minimum density power divergence approach [Abstract 130]

Abhik GHOSH, *Indian Statistical Institute*
Arnab HAZRA, *IIT Kanpur*

11:50 EEMD based LSTM hybrid model for Agricultural Price Forecasting [Abstract 82]

Kapil CHOUDHARY, *ICAR-IASRI, New Delhi*
Girish Kumar JHA, *ICAR-IARI, New Delhi*
Ronit JAISWAL, *ICAR-IASRI, New Delhi*
Rajeev R. KUMAR, *ICAR-IASRI, New Delhi*

29.M2.I65 Statistical techniques for Hydroclimatic applications *Venue:* MSCR-2

Chair and Organizer : Arpita MONDAL, Indian Institute of Technology (IIT) Bombay

11:00 **Modelling, Detection and Attribution of Large-scale Compound Extreme Events using Probabilistic Graphical Models** [**Abstract 234**]

Adway MITRA, *Indian Institute of Technology Kharagpur*

11:25 **APPLICATIONS OF PROBABILISTIC MODELS IN HYDRO-METEOROLOGICAL DROUGHT MONITORING & FREQUENCY ANALYSIS** [**Abstract 282**]

Meenu RAMADAS, *Assistant Professor, School of Infrastructure, IIT Bhubaneswar*
Alok SAMANTARAY, *Student, School of Infrastructure, IIT Bhubaneswar*

11:50 **Modeling Hydrologic Extremes with Multivariate Extreme Value Models** [**Abstract 334**]

SHAILZA SHARMA, *IISc*

Lunch 12:20 - 14:00 Venue: NSSC

Bahadur memorial lecture Soumndranath Lahiri *Venue:* JNTA

Chair : Hira KOUL, Michigan State University

14:00 **The Central Limit Theorem and statistical inference in high dimensions** [**Abstract 209**]

Soumendra LAHIRI, *Washington University in St. Louis*

Switching time 15:00 - 15:10

Panel Discussion 4 Data, Models and Modelers during a Global Pandemic *Venue:* JNTA

Organizer : Bhramar MUKHERJEE, University of Michigan

- **Rukmini S**, Independent Data Journalist
- **Sam ABBOTT**,
- **Swapnil MISHRA**, University of Copenhagen

29.A1.I66 Recent advances in time series analysis *Venue:* NSSC-A

Chair : Soudeep DEB, Indian Institute of Management Bangalore
Organizer : Debashis MONDAL, Washington University in St Louis

15:10 **A Bayesian approach to identify changepoints in spatio-temporal ordered categorical data** [**Abstract 106**]

Soudeep DEB, *Indian Institute of Management Bangalore*
Siddharth RAWAT, *Indian Institute of Management Bangalore*
Candace BERRETT, *Brigham Young University*

15:35 **Wavelet Spatio-Temporal Change Detection on multi-temporal SAR images** [**Abstract 119**]

Rodney FONSECA, *Weizmann Institute of Science*
Rogerio NEGRI, *UNESP*
Aluisio PINHEIRO, *Unicamp*
Abdourrahmane ATTO, *Universite de Savoie*

16:00 **Polyspectral Mean Estimation of General Nonlinear Processes** [**Abstract 133**]

Dhrubajyoti GHOSH, *Duke University*
Tucker MCELROY, *US Census Bureau*
Soumendra LAHIRI, *Washington University in St. Louis*

29.A1.I67 Simulates and Surrogates in Complex Statistical Learning

Venue: NSSC-B

Chair : Jason XU, Duke University
Organizer : Surya TOKDAR, Duke University

15:10 **On the Robustness to Misspecification of -Posteriors and Their Variational Approximations** [**Abstract 303**]

Cynthia RUSH, *Columbia University*
Marco AVELLA MEDINA, *Columbia University*
Jos MONTIEL OLEA, *Cornell University*
Amilcar VALEZ, *Northeastern University*

15:35 **Likelihood-based Inference for Stochastic Epidemic Models via Data Augmentation** [**Abstract 378**]

Jason XU, *Duke University*

16:00 **Incorporating mechanistic knowledge in causal inference** [**Abstract 375**]

Alexander VOLFOVSKY, *Duke University*

29.A1.I68 Bayesian application in Drug Development

Venue: NSSC-C

Chair : Samopriyo MAITRA,
Organizer : Angshuman SARKAR, GSK India Global Service Private Ltd

15:10 **Borrowing of Historical Clinical trial data for Effiacy Analysis** [**Abstract 49**]

Arunangshu BISWAS, *GSK*

15:35 **Assessing Clinical Development Plans using Quantitative Decision Making Framework** [**Abstract 27**]

Rudrani BANERJEE, *GSK*

16:00 **Bayesian Approaches in Clinical Trials of Medical Devices** [**Abstract 154**]

Jayanti GUPTA, *NishKash Consulting*

29.A1.I69 Applied Probability and Heavy Tails

Venue: MSSH

Chair : Parthanil ROY, Indian Statistical Institute
Organizers : *Moumanti PODDER, Indian Institute of Science Education and Research (IISER) Pune* and *Parthanil ROY, Indian Statistical Institute*

15:10 **Persistence of heavy-tailed sample averages: principle of infinitely many big jumps** [**Abstract 41**]

Ayan BHATTACHARYA, *Assistant Professor, Department of Mathematics, IIT Bombay*
Zbigniew PALMOWSKI, *University of Wroclaw*
Bert ZWART, *CWI, Amsterdam*

15:35 Asymptotic analysis of stochastic differential equations with small random noise and fast periodic sampling [Abstract 261]

Chetan PAHLAJANI, *Indian Institute of Technology Gandhinagar*
Shivam DHAMA, *IIT Gandhinagar*

16:00 Stable random fields indexed by discrete amenable groups [Abstract 296]

Parthanil ROY, *Indian Statistical Institute*
Mahan MJ, *Tata Institute of Fundamental Research, Mumbai*
Sourav SARKAR, *University of Cambridge*

29.A1.I70 Applications of regression techniques in interdisciplinary research *Venue:* MSCR-1

Chair : Arnab LAHA, Indian Institute of Management Ahmedabad
Organizer : Partha Sarathi MUKHERJEE, Indian Statistical Institute, Kolkata

15:10 Image denoising and deblurring using jump regression and local pixel clustering [Abstract 245]

Partha Sarathi MUKHERJEE, *Indian Statistical Institute, Kolkata*

15:35 Causal Analysis at Extreme Quantiles with Application to London Traffic Flow Data [Abstract 47]

Prajamitra BHUYAN, *Indian Institute of Management, Calcutta*
Kaushik JANA, *Ahmedabad University*
Emma MCCOY, *Imperial College London*

16:00 Fully Bayesian estimation of virtual brain parameters with self-tuning Hamiltonian Monte Carlo [Abstract 170]

Jayant JHA, *Indian Statistical Institute*

29.A1.I71 Application of Statistics/ML/Data Science in Telecom domain *Venue:* MSCR-2

Chair and Organizer : Subhadip BANDYOPADHYAY, Principal Data Scientist, GAIA, Ericsson, Bangalore

15:10 AI solutions for traffic management in LTE networks [Abstract 160]

Ranjani H. G., *Principal Data Scientist, GAIA, Ericsson, Bangalore*

15:35 Supervised change point detection using non-stationary Hawkes process [Abstract 25]

Jyotirmoy BANERJEE, *Principal Data Scientist, GAIA, Ericsson, Bangalore*

16:00 Identification of Causal Dependencies amongst Telecom KPIs using Latent Convergent Cross Mapping [Abstract 20]

Subhadip BANDYOPADHYAY, *Principal Data Scientist, GAIA, Ericsson, Bangalore*
Sujoy Kumar ROY CHOWDHURY, *Principal Data Scientist, GAIA, Ericsson, Bangalore*

Coffee 16:30 - 17:10 Venue: NSSC

29.A2.I72 Recent advances in sequential methodologies *Venue:* NSSC-A

Chair and Organizer : Sudeep BAPAT, Indian Institute of Management Indore

17:10 Sequential Estimation for the Multiple Linear Regression Models with Balanced Loss Functions [Abstract 331]

Raghu Nandan SENGUPTA, *Indian Institute if Technology Kanpur*
Ren SCHILLING, *Institutfr Math. Stochastik, TU Dresden, GERMANY*
Sundeep R. BAPAT, *Indian Institute of Management Indore*
Neeraj JOSHI, *Indian Institute if Technology Kanpur*

17:35 Two-Stage Principal Component Analysis [Abstract 76]

Bhargab CHATTOPADHYAY, *Indian Institute of Technology Jodhpur*
Swarnali BANERJEE, *Loyola University Chicago*

18:00 Two-stage and purely sequential minimum risk point estimation of the scale parameter of a family of distributions under modified LINEX loss plus sampling cost [Abstract 174]

Neeraj JOSHI, *joshin@iitk.ac.in*
Sudeep R. BAPAT, *Indian Institute of Management Indore*
Raghu Nandan SENGUPTA, *Indian Institute of Technology Kanpur*

29.A2.I73 Advancements and contemporary applications of causal inference. *Venue:* NSSC-B

Chair and Organizer : Bhuvanesh PAREEK, IIM Bangalore

17:10 Bayesian Nonparametric Modeling of Restricted Mean Survival Time [Abstract 31]

Sanjib BASU, *University of Illinois Chicago*
Ruizhe CHEN, *Johns Hopkins University*

17:35 Balancing on the Leaves: Exploring Causal Roots of Decision Tree Algorithms [Abstract 268]

Bhuvanesh PAREEK, *IIM Bangalore*
Bikram KARMAKAR, *University of Florida*

18:00 A Double Machine Learning Approach to Combining Experimental and Observational Data [Abstract 270]

Harsh PARIKH, *Duke University*
Marco MORUCCI, *NYU*
Vittorio ORLANDI, *Duke University*
Alexander VOLFOVSKY, *Duke University*

29.A2.I74 Addressing healthcare challenges through data and machine learning *Venue:* NSSC-C

Chair : Sudipta DAS, Ramakrishna Mission Vivekananda Educational and Research Institute
Organizer : Subhashis GHOSHAL, North Carolina State University

17:10 AIM-AHEAD: Artificial Intelligence and Machine Learning to address health disparities and achieve health equity [Abstract 373]

Jamboor VISHWANATHA, *Univ. of North Texas Health Science Center*

17:35 Building the Infrastructure and Co-design of Data Sets for AI/ML applications to address health equity [Abstract 374]

Lavanya VISHWANATHA , *University of North Texas Health Science Center*

18:00 **SOFTWARE RELIABILITY BASED ON RENEWAL PROCESS MODELING FOR ERROR OCCURRENCE DUE TO EACH BUG WITH PERIODIC DEBUGGING SCHEDULE** [Abstract 94]

Sudipta DAS, *Ramakrishna Mission Vivekananda Educational and Research Institute*
ANUP DEWANJI, *Applied Statistics Unit, Indian Statistical Institute, Kolkata, India*
SUBRATA KUNDU, *George Washington University, Washington, DC, USA*

29.A2.I75 Statistical methods in Genetics, Genomics, Microbiome and Clinical Decision Making

Venue: `MSSH`

Chair and Organizer : Sanjay SHETE, UT MD Anderson Cancer Center

17:10 **Statistical methods for evaluating gene-exposure interactions** [Abstract 323]

Jaya SATAGOPAN, *Rutgers School of Public Health*

17:35 **Bayesian Meta-Analysis of Penetrance and its Application to Breast Cancer Risk among ATM Mutation Carriers** [Abstract 52]

Swati BISWAS, *University of Texas at Dallas*
Thanthirige Lakshika RUBERU, *University of Texas at Dallas*
Danielle BRAUN, *Harvard T.H. Chan School of Public Health*
Giovanni PARMIGIANI, *Dana Farber Cancer Institute*

18:00 **Sparse canonical correlation to identify breast cancer related genes regulated by copy number aberrations** [Abstract 115]

Diptavo DUTTA, *National Cancer Institute*
Ananda SEN, *University of Michigan*
Jaya SATAGOPAN, *Rutgers University*

29.A2.C17 Contributed Session 17

Venue: `Math-LH1`

Chair : BG MANJUNATH, University of Hyderabad

17:10 **Spatial Stochastic Frontier Model for Count Data** [Abstract 338]

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

17:25 **Multiple testing applied to Structured Hypotheses** [Abstract 253]

Shinjini NANDI, *Montana State University*

17:40 **RESCALING BOOTSTRAP WITHOUT REPLACEMENT TECHNIQUE OF VARIANCE ESTIMATION FOR ADAPTIVE CLUSTER SAMPLING** [Abstract 14]

MOUMITA BAISHYA, *Ph.D. Scholar*
Tauqueer AHMAD, *Principal Scientist & Head, Division of Sample Surveys, ICAR-IASRI*
Ankur BISWAS , *Scientist, ICAR-IASRI*

17:55 **A queueing-inventory control model with cross-trained servers** [Abstract 204]

INIYA L, *Madurai Kamaraj University*
SIVAKUMAR B, *Madurai Kamaraj University*

18:10 **Markov Chain model for extremal PM2.5** [Abstract 224]

BG MANJUNATH, *University of Hyderabad*

29.A2.C18 Contributed Session 18

Venue: `Math-LH3`

Chair : Harithasree GOPALAKRISHNAN, Great Learning

17:10 Bayesian Estimation from Pareto Type II Distribution with Record Values using Monte-Carlo Techniques [Abstract 345]

SABHIYA SINGH, *University of Lucknow*
Rajeev PANDEY, *University of Lucknow*

17:25 DEVELOPMENT OF EXPONENTIAL SMOOTHING MODEL USING A NEW APPROACH OF TIME VARYING SMOOTHING PARAMETER [Abstract 1]

Praveenkumar A, *praveenkumarupm@gmail.com*
Praveenkumar A, *Ph.D. Scholar*
Wasi ALAM, *Principal Scientist, ICAR-IASRI*
Kanchan SINHA, *Scientist, ICAR-IASRI*

17:40 Assessing the feasibility of feature selection algorithms for wheat yield forecasting in regression framework [Abstract 367]

Manoj VARMA, *PhD Scholar, ICAR-IASRI, New Delhi*
Kamalesh Narain SINGH, *Principal Scientist and Head (F and ASM), ICAR-IASRI, NEW DELHI*
Achal LAMA, *Scientist, ICAR_IASRI, New Delhi*

17:55 Deciding optimal number of segments in Detrended Correspondence Analysis [Abstract 112]

Shantaram DHUM, *Savitribai Phule Pune University, Pune*
Akanksha KASHIKAR, *Savitribai Phule Pune University, Pune*

18:10 Educational data mining on Indian states using clustering and ranking for targeted intervention [Abstract 142]

Harithasree GOPALAKRISHNAN, *Great Learning*
Suyash VERMA, *Great Learning*
Dr. Srabashi BASU, *Great Learning*
Natesan RAJAN, *Great Learning*

29.A2.C19 Contributed Session 19

Venue: `Math-LH4`

Chair : Richa SHARMA, Great Learning Mumbai India (Student)

17:10 Seasonal zero modified geometric INAR(1) process [Abstract 128]

Aishwarya GHODAKE, *Department of Statistics, Savitribai Phule Pune University, Pune*
Manik AWALE, *Department of Statistics , Savitribai Phule Pune University, Pune*

17:25 Dissimilarity Measure for Quantification of the Poetic/Prose Nature of a Text [Abstract 192]

Rutuja KORE, *Department of Statistics, Savitribai Phule Pune University*
Akanksha KASHIKAR, *Department of Statistics, Savitribai Phule Pune University, Pune*

17:40 Multiple Imputation for Directional Data [Abstract 13]

Sneha BABEL, *H.V.Desai College, Pune*
Akanksha KASHIKAR, *Department of Statistics, Savitribai Phule Pune University, Pune*

17:55 Method Based on State-Space Epidemiological Model for Cost-Effectiveness Analysis of Non-Medical Interventions for COVID-19 [Abstract 108]

Vishal DEO, *Ramjas College, University of Delhi*
Gurprit GROVER, *Department of Statistics, Faculty of Mathematical Sciences, University of Delhi*

18:10 **PREDICTING SUCCESS FOR A NEW SHOW ON A KIDS CHANNEL** [**Abstract 333**]

Richa SHARMA, *Great Learning Mumbai India (Student)*
MARVIN JOHN CARDOZA, *Great Learning Mumbai India*
RATNAKAR VIJAY ACHREKAR , *Great Learning Mumbai India*
UDAYAKUMAR D, *Great Learning Mumbai India*

Award ceremony and Banquet 19:00 - 22:00 Sheraton Grand, Brigade Gateway

Friday December 30

Short Course 2 Short Course on R (Part 1) *Venue:* NSSC-A

Time : **9:00 - 10:30**

- **Ayush PATEL**, University of Oxford
- **Heather TURNER**, University of Warwick

Short Course 3 Short Course on Deep learning (Part 1) *Venue:* NSSC-B

Time : **9:00 - 10:30**

- **Abhinanda SARKAR**, Great Learning

30.M1.I76 Statistics and Machine Learning *Venue:* NSSC-C

Chair : Ranjit Kumar PAUL, ICAR-Indian Agricultural Statistics Research Institute
Organizer : Sanjay CHAUDHURI, National University of Singapore

9:00 **Wavelets and Machine Learning Techniques for Forecasting Time Series [Abstract 273]**

Ranjit Kumar PAUL, *ICAR-Indian Agricultural Statistics Research Institute*

9:25 **Modeling and forecasting of crop yield using AI/MLT incorporating with weather variables [Abstract 343]**

Kamlesh SINGH, *ICAR-IASRI, New Delhi*
Bishal GURUNG, *ICAR-IASRI, New Delhi*
Achal LAMA, *ICAR-IASRI, New Delhi*
Md Ashraful HAQUE, *ICAR-IASRI, New Delhi*

9:50 **Modelling distribution of aquatic alien invasive weeds under future climate scenarios [Abstract 127]**

Yogita GHARDE, *ICAR-Directorate of Weed Research*
Sushil KUMAR, *ICAR-Directorate of Weed Research, Jabalpur*
JS MISHRA, *ICAR-Directorate of Weed Research, Jabalpur*

30.M1.I77 Copula and Survival Analysis *Venue:* MSSH

Chair and Organizer : Arusharka SEN, Concordia University

9:00 **Nonparametric smooth estimation of survival and density function for length biased data: An overview [Abstract 77]**

Yogendra CHAUBEY, *Concordia University*
Yogendra CHAUBEY,

9:25 **Copula-likelihood Maximization under Bivariate Random Censoring [Abstract 326]**

Arusharka SEN, *Concordia University*
Magloire LOUDEGUI DJIMDOU, *Concordia University*
Yogendra CHAUBEY, *Concordia University*

9:50 **Copula-based estimation of health concentration curves with an application to COVID-19 [Abstract 54]**

Taoufik BOUEZMARNI, *Universit de Sherbrooke*
Abderrahim Taamouti MOHAMED DOUKALI, *McGill University*
UNIVERSITY OF LIVERPPOL,

30.M1.I78 Applied Probability and Statistical Physics *Venue:* `MSCR-1`

Chair : Mathew JOSEPH, Indian Statistical Institute
Organizers : *Moumanti PODDER, Indian Institute of Science Education and Research (IISER) Pune* and *Parthanil ROY, Indian Statistical Institute*

9:00 **Spectral properties of random perturbations of non-self-adjoint operators [Abstract 28]**

Anirban BASAK, *ICTS-TIFR*
Elliot PAQUETTE,
Martin VOGEL,
Ofer ZEITOUNI,

9:25 **A random string in a medium of Poisson traps [Abstract 173]**

Mathew JOSEPH, *Indian Statistical Institute*
Siva ATHREYA, *Indian Statistical Institute*
Carl MUELLER, *University of Rochester*

9:50 **A multispecies totally asymmetric zero range process and Macdonald polynomials [Abstract 12]**

Arvind AYYER, *Indian Institute of Science*
James MARTIN, *Oxford University*
Olya MANDELSHTAM, *University of Waterloo*

Coffee 10:20 - 11:00 Venue: NSSC

Short Course 4 Short Course on R (Part 2) *Venue:* `NSSC-A`

Time : **11:00 - 12:30**

- **Ayush PATEL**, University of Oxford
- **Heather TURNER**, University of Warwick

Short Course 5 Short Course on Deep learning (Part 2) *Venue:* `NSSC-B`

Time : **11:00 - 12:30**

- **Abhinanda SARKAR**, Great Learning

Special Invited Session 4 Saonli Basu, Asha Gopalakrishnan *Venue:* `JNTA`

Chair : Subrata KUNDU, George Washington University

11:00 **Understanding the role of genomics in Human Health [Abstract 32]**

Saonli BASU, *University of Minnesota*

11:40 **Modeling and identifiability of non-homogenous Poisson process cure rate model. [Abstract 141]**

Asha GOPALAKRISHNAN, *asha@cusat.ac.in*
Anup DEWANJI, *ISI Calcutta, India*
C S SOORYA , *Cochin University of Science & Technology, India.*

30.M2.I79 Statistical Aspects of Robustness and Distribution-Shift *Venue:* `NSSC-C`

Organizer : Sivaraman BALAKRISHNAN, Carnegie Mellon University

11:00 Robust Universal Inference [Abstract 16]

Sivaraman BALAKRISHNAN, *Carnegie Mellon University*

11:25 Central Limit Theorem, Edgeworth expansion, and Neural Networks [Abstract 194]

Arun Kumar KUCHIBHOTLA , *Carnegie Mellon University*
Arisina BANERJEE, *Indian Statistical Institute*

11:50 Learning under Distribution Shift [Abstract 211]

Zachary LIPTON, *Carnegie Mellon University*

30.M2.I80 Design and Reliability *Venue:* `MSSH`

Chair : Debasis KUNDU, Indian Institute of Technology Kanpur
Organizer : Abhyuday MANDAL, University of Georgia

11:00 Stage life testing in presence of competing risks [Abstract 190]

Arnab KOLEY, *IIM Indore*
Sudeep BAPAT, *IIM Indore*

11:25 Bayesian Optimal Stepped Wedge Design [Abstract 347]

Satya Prakash SINGH, *Indian Institute of Technology Kanpur*

11:50 A-ComVar: A Flexible Extension of Common Variance Designs [Abstract 85]

Shrabanti CHOWDHURY, *Icahn School of Medicine at Mount Sinai*
Joshua LUKEMIRE, *Emory University*
Abhyuday MANDAL, *University of Georgia*

30.M2.I81 Random Sets and Graphs *Venue:* `MSCR-1`

Chair : Sanchayan SEN, Indian Institute of Science
Organizers : *Moumanti PODDER, Indian Institute of Science Education and Research (IISER) Pune* and *Parthanil ROY, Indian Statistical Institute*

11:00 Partial divisibility in random sets [Abstract 89]

Biltu DAN, *Indian Institute of Science*
Jnaneshwar BASLINGKER, *Indian Institute of Science*
Manjunath KRISHNAPUR , *Indian Institute of Science*

11:25 The cover time of a random walk on Linial-Meshulam Model. [Abstract 4]

Kartick ADHIKARI, *IISER Bhopal*

11:50 Scaling limits of random graph models at criticality: edge-weighted graphs converging to an L^3 graphon [Abstract 329]

Sanchayan SEN, *Indian Institute of Science*

30.M2.I82 Recent Advances in Functional Data Analysis *Venue:* MSCR-2

Chair and Organizer : Shyamal Krishna DE, Indian Statistical Institute Kolkata

11:00 **How to test for the rank of a stochastic process?** [**Abstract 59**]

Anirvan CHAKRABORTY, *IISER Kolkata*
Victor M. PANARETOS, *EPFL Switzerland*

11:25 **Covariance Networks for Functional Data on Multidimensional Domains** [**Abstract 322**]

Soham SARKAR, *Indian Statistical Institute, Delhi Centre*

11:50 **On Progressive Type-I Interval Censoring Schemes under Competing Risks** [**Abstract 300**]

Soumya ROY, *Indian Institute of Management Kozhikode*

Lunch 12:20 - 13:50 Venue: NSSC

Short Course 6 Short Course on R (Part 3) *Venue:* NSSC-A

Time : **13:50 - 17:00**

- **Ayush PATEL**, University of Oxford
- **Heather TURNER**, University of Warwick

Short Course 7 Short Course on Deep learning (Part 3) *Venue:* NSSC-B

Time : **13:50 - 17:00**

- **Abhinanda SARKAR**, Great Learning

COLORADOSCHOOLOFMINES
EARTH • ENERGY • ENVIRONMENT

IISA 2023 CONFERENCE

GOLDEN, COLORADO USA

June 1 (Thursday) – June 4 (Sunday), 2023
web: https://intindstat.org
Contact: iisa2023@intindstat.org

The conference brings together students, early career scientists, researchers and practitioners from academia, industry and government, connected through a common interest in statistics and data science.

1500 Illinois St., Golden, CO 80401

MINES.EDU

Abstracts

1. DEVELOPMENT OF EXPONENTIAL SMOOTHING MODEL USING A NEW APPROACH OF TIME VARYING SMOOTHING PARAMETER

[29.A2.C18, (page 49)]

Praveenkumar A, *praveenkumarupm@gmail.com*
Praveenkumar A, *Ph.D. Scholar*
Wasi ALAM, *Principal Scientist, ICAR-IASRI*
Kanchan SINHA, *Scientist, ICAR-IASRI*

DEVELOPMENT OF EXPONENTIAL SMOOTHING MODEL USING A NEW APPROACH OF TIME VARYING SMOOTHING PARAMETER Praveenkumar A1, Wasi Alam2 and Kanchan Sinha3 ICAR-Indian Agricultural Statistics Research Institute Library Avenue, New Delhi-110012, INDIA

Abstract. Time series modeling and forecasting is a vibrant research field that has attracted the interest of the scientific community in recent decades. Smoothing methods have been widely used in literature for modeling and forecasting of time series data. In moving average model, equal weights were allotted to all the observations of the time series. In exponential smoothing, weights are assigned in exponential manner as recent observations getting more weights than the older ones. So it doesnt capture entire pattern in the time series and the appropriate characteristics of time series cant be explained properly. Hence, exponential smoothing model with time varying parameter is developed with weights based on extreme values of the lagged observations. The monthly price series of Tea of Kolkata market has been used for this study. Based on forecast accuracy measures, it has been found that the proposed approach performs better than existing smoothing models. The proposed approach is quite novel and may be recommended for future study using price series of agricultural commodities. Key words. Agricultural price; Moving Average; Exponential Smoothing; Time varying parameter.

1Ph.D. Scholar, ICAR-IASRI, New Delhi (praveenkumarupm@gmail.com) 2Principal Scientist, ICAR-IASRI, New Delhi (mw.alam@icar.gov.in) 3Scientist, ICAR-IASRI, New Delhi (kanchan.sinha@icar.gov.in)

2. TBA

[29.M1.I56, (page 38)]

Sam ABBOTT,

TBA

3. Bayesian Integrative Network Models for Multimodal and Spatial Omics Data

[27.A2.I34, (page 21)]

Satwik ACHARYYA, *University of Michigan*

The development and clinical implementation of evidence-based precision medicine strategies has become a realistic possibility, primarily due to the rapid accumulation of large-scale multimodal genomics, clinical and phenotypic data from diverse disease systems. Network models are powerful tools to investigate complex dependence structures in such 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 heterogeneity (e.g. patient-specific) and additional structure (e.g. spatial) that such data often manifest. In this talk, I will discuss integrative network-based approaches to (a) decipher patient-specific networks integrating genomic and clinical phenotypic data; (b) integrate pan-omic data incorporating biological hierarchies among molecular-platforms to construct multi-layered networks, and (c) construct spatially-varying networks from spatial transcriptomics data. Bayesian formulations are used to admit sparse network recovery and are computationally tractable. These approaches will be illustrated using several case studies in oncology across multiple tumor types.

4. The cover time of a random walk on Linial-Meshulam Model.

[30.M2.I81, (page 53)]

Kartick ADHIKARI, *IISER Bhopal*

The cover time of a graph means the expected number step taken by a random walk on this graph to visit all the vertices of the graph. We consider a d-dimensional simplicial complex which contains the complete (d - 1)-skeleton and each d-simplex with probability p. In this talk, we shall discuss the cover time of a random walk on this complex, where the walk starts from a (d - 1)-simplex and each step walker moves uniformly to one of the neighbouring (d - 1)-simplex from its present position. This talk will be based on a joint work with Robert Adler, Omer

Bobrowski and Ron Rosenthal.

5. Construction Methods for Doubly Nested Block Designs

[26.A2.C2, (page 9)]

Nehatai AGASHE, *PGS IARI, New Delhi*
Cini VARGHESE, *IASRI, New Delhi*
Harun MOHD, *IASRI, New Delhi*
Vinay KUMAR, *PGS IARI, New Delhi*

Abstract Doubly nested block design is a block design which includes three systems of blockings, in which first system of blocking is nested within a second system and second system of blocking is nested within third system of blocking. This design is used to eliminate three sources of variation present in the experimental material of an experiment. New construction methods for proper doubly nested block design are obtained, by amalgamating different types of incomplete block designs at different layers. General form of information matrix pertaining to the treatment effects under doubly nested block design set up is derived. R code is written that generates these designs and computes information matrices for studying their characterization properties. Application potential of doubly nested block designs in obtaining mating plans for breeding trials is also illustrated. Keywords: breeding trials, doubly nested block design, incomplete block design, information matrix, R code

6. Stability of particle filters for chaotic, deterministic dynamical systems

[28.M1.I41, (page 26)]

Amit APTE, *Indian Institute of Science Education and Research (IISER) Pune*
Anugu Sumith REDDY, *Indian Institute of Technology Bombay, Mumbai, India*
Pinal MANDAL, *International Centre for Theoretical Sciences TIFR, Bangalore, India*
Shashank Kumar ROY, *International Centre for Theoretical Sciences TIFR, Bangalore, India*

The problem of stability of nonlinear filters and of particle filters has been studied extensively. Most results assume some form of controllability of the signal, which assumptions do not apply to deterministic dynamical systems. In this talk, I will discuss both theoretical as well as numerical challenges and methods to overcome these challenges in order to demonstrate stability in this context. I will argue that for such systems, no variant of particle filters approximate the true filter even in the limit of large sample size. (Based on works in collaboration with A.S.Reddy, S.Vadlamani, P.Mandal, and S.K.Roy)

7. Clustering grouped data via hierarchical normalized finite point process

[27.M1.I16, (page 12)]

Raffaele ARGIENTO, *Università degli studi di Bergamo*
Alessandro COLOMBI, *Universit Milano Bicocca*
Federico CAMERLENGHI, *Universit Milano Bicocca*
Lucia PACI, *Universit Cattolica del Sacro Cuore*

Almost surely discrete random probability measures have received close attention in the Bayesian nonparametric community. They have been used to model populations of individuals or latent parameters (in the mixture model setting) composed of unfixed species with unknown proportions. In this framework, data are usually assumed to be exchangeable. However, the latter assumption is not appropriate when data are divided in multiple groups which may share the same species. If so, partially exchangeability accommodates the dependence across populations.

8. Kernel Epsilon-Greedy Approach for Contextual Bandits

[27.A1.I26, (page 18)]

Sakshi ARYA, *Penn State University*
Bharath SRIPERUMBUDUR, *Penn State University*

Contextual bandit algorithms are popular for sequential decision making in several practical applications, ranging from online advertisement recommendations to mobile health. The goal of such problems is to maximize cumulative reward over time for a finite set of choices/arms, while taking covariate (or contextual) information into account. In this work, we assume that the similarities between the covariates and expected rewards can be modeled as arbitrary linear functions of the contexts' images in a certain reproducing kernel Hilbert space (RKHS). We propose KernelEpsilonGreedy, a kernelized epsilon-greedy algorithm, and establish its convergence rates for the estimation error and prediction error. The rates we achieve are closely tied to the intrinsic dimensionality of the RKHS.

9. On the Analysis of One-Shot Device Testing Data Using Copula Models: A Bayesian Approach

[29.M1.C15, (page 41)]

ASHKAMINI, *Banaras Hindu University*
Reema SHARMA, *BANARAS HINDU UNIVERSITY*
Satyanshu K UPADHYAY, *BANARAS HINDU UNIVERSITY*

Copula models are utilized to study the dependence structure among the random variables. Such models are capable of modelling the marginal behavior of the random variables and the dependence structure separately and thereby may be considered as popular tools in the areas of reliability engineering and survival analysis. The present paper utilizes three copula models for analyzing one shot device testing data with two correlated failure modes collected from a constant stress accelerated life test. The data on such devices are typical in the sense that one obtains data in binary forms instead of observing the actual failure times of the devices. The approach is throughout Bayesian and the posterior analysis is performed using the well-known Metropolis algorithm. Finally, a few important Bayesian model comparison tools are used to select the most appropriate model from the three proposed models for the considered data. A numerical study based on a real data illustration is provided that conveys the straightforwardness of our approach.

10 . BIDIRECTIONAL LSTM-CNN BASED GENERATIVE ADVERSARIAL NETWORKS FOR TIME SERIES FORECASTING

[28.A2.C11, (page 35)]

G AVINASH, *PhD Scholar at ICAR-Indian Agricultural Statistics Research Institute*
Ramasubramanian VAIDHYANATHAN, *Principal Scientist at ICAR-IASRI*
Harish NAYAK G.H, *PhD Scholar at ICAR-IASRI*
Mrinmoy RAY, *Scientist at ICAR-IASRI*

Indian edible oil and oilseeds business has several challenges in the newly liberalised trading environment. The challenge of reconciling the interests of several stakeholders in the oilseed is formidable. Edible oil prices witnessed a rise in recent months due to changing geopolitical scenario and exhibit more volatility. Such variation in edible stock price needs to be accounted for while modelling for its use in prediction. A study has been conducted in which the deep learning (DL) approach viz., Generative Adversarial Network (GAN) model has been used by employing separately Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) as the Generators to mine the data distributions of these stocks and Convolutional Neural Network (CNN) as the Discriminator discerning between the actual and generated stock data. To the best of our knowledge, an attempt has been made for the first time towards modeling Bidirectional LSTM-CNN based GANs for time series forecasting. For this, daily data on Ruchi Soya Ltd. stock prices relating to oilseeds based products have been considered and the daily closing prices have been predicted. The results revealed that the novel proposed GAN seem to have promising performance in closing price prediction on the real data compared with other DL models fitted on the same data using GRU, LSTM and Bi-LSTM.

11. Seasonal INAR(1) model for overdispersed count time series data

[27.A2.I31, (page 20)]

Manik AWALE, *Savitribai Phule Pune University*
KASHIKAR AKANKSHA,

In this paper a integer autoregressive model with seasonal period ‘s’ based on the negative binomial thinning operator and having negative binomial marginal distribution (SNBINAR(s) is proposed. The model is based on the negative binomial thinning operator. Yule- Walker, conditional least squares and conditional maximum likelihood estimation methods for the model parameters are studied. It has been shown that the models NGINAR(1) (2009), NBINAR(1) (2013), and SGINAR(s) (2020) are the special cases of the proposed model. Simulation study has been carried out to see the performance of the estimation methods and the forecasting performance of the model. A real data has been analyzed using the SNBINAR(s) model.

12 . A multispecies totally asymmetric zero range process and Macdonald polynomials

[30.M1.I78, (page 52)]

Arvind AYYER, *Indian Institute of Science*
James MARTIN, *Oxford University*
Olya MANDELSHTAM, *University of Waterloo*

Macdonald polynomials are a remarkable family of symmetric functions that are known to have connections to combinatorics, algebraic geometry and representation theory. Due to work of Corteel, Mandelshtam and Williams, it is known that they are

related to the asymmetric simple exclusion process (ASEP) on a ring.

The modified Macdonald polynomials are obtained from the Macdonald polynomials using an operation called plethysm. It is natural to ask whether the modified Macdonald polynomials are related to some other particle system. In this talk, we answer this question in the affirmative via a multispecies totally asymmetric zero-range process (TAZRP). We also present a Markov process on tableaux that projects to the TAZRP and derive formulas for stationary probabilities and certain correlations. We also prove a remarkable symmetry property for local correlations.

This is joint work with Olya Mandelshtam and James Martin (arXiv:2209.09859).

13. Multiple Imputation for Directional Data

[29.A2.C19, (page 49)]

Sneha BABEL, *H.V.Desai College, Pune*
Akanksha KASHIKAR, *Department of Statistics, Savitribai Phule Pune University, Pune*

Observations consisting of directions or angles are found across many areas of science, including ecology, earth sciences, environmental science, and medicine. Examples of such data are the angular movements of an animal relative to a food source or other attractor, wind directions, diurnal measurements of admission times to an intensive care unit, and departure directions of birds after release. Circular data arise whenever directions are measured, and are usually expressed as angles relative to some fixed reference point, such as Due North. Time data measured on a 24 h clock may also be converted to angular measurements, with 0:00 corresponding to 0 and 24:00 to 360. Like other datasets, missing values is a major problem in the case of circular data as well. In this paper, we discuss some imputation methods for missing values in circular data by using the technique of multiple imputation via chained equations. This ensures that, the relationships among these variables can be used for better imputation of missing values. We restrict our attention to bivariate datasets consisting of at least one circular variables. The performance of the method is assessed via comparison of the distribution of the original and imputed datasets through an extensive simulation study.

Keywords: Directional data, Missing data, Multiple imputation via chained equations, Simulations

14 . RESCALING BOOTSTRAP WITHOUT REPLACEMENT TECHNIQUE OF VARIANCE ESTIMATION FOR ADAPTIVE CLUSTER SAMPLING

[29.A2.C17, (page 48)]

MOUMITA BAISHYA, *Ph.D. Scholar*
Tauqueer AHMAD, *Principal Scientist & Head, Division of Sample Surveys, ICAR-IASRI*
Ankur BISWAS , *Scientist, ICAR-IASRI*

RESCALING BOOTSTRAP WITHOUT REPLACEMENT TECHNIQUE OF VARIANCE ESTIMATION FOR ADAPTIVE CLUSTER SAMPLING

15. The containment profile of hyperrecursive trees

[26.A1.I5, (page 5)]

Srinivasan BALAJI, *Department of Statistics, George Washington University*
Hosam MAHMOUD, *Department of Statistics, George Washington University*
Joshua SPARKS, *Department of Statistics, George Washington University*

We investigate vertex levels of containment in a random hypergraph grown in the spirit of a recursive tree. We consider a local profile tracking the evolution of the containment of a particular vertex over time, and a global profile concerned with counts of the number of vertices of a particular containment level.

For the local containment profile, we obtain the exact mean, variance, and probability distribution in terms of standard combinatorial quantities such as generalized harmonic numbers and Stirling numbers of the first kind. As for the global containment profile, we establish an asymptotically normal distribution for the number of vertices at the smallest containment level as well as their covariances with the number of vertices at the second smallest containment level and the variances of these numbers.

16. Robust Universal Inference

[30.M2.I79, (page 53)]

Sivaraman BALAKRISHNAN, *Carnegie Mellon University*

In this talk, I will discuss two methods – universal inference and a robust extension of it – which use sample-splitting to construct confidence sets which are valid under very weak regularity conditions. Universal inference uses sample-splitting to ease the construction of likelihood-ratio based confidence sets. Despite its generality it can fail catastrophically when the statistical model is misspecified. Building on the sample-splitting idea, I will then discuss a robust variant of universal inference, building on robust counterparts to the classical likelihood-ratio test. These methods are easy to apply, yield valid inference in a wide-range of challenging problems, and achieve these strong guarantees at a surprisingly small statistical price.

This is based on joint work with Beomjo Park, Aaditya Ramdas and Larry Wasserman.

17 . Constrained Stochastic Nonconvex Optimization with State-dependent Markov Data

[29.M2.I60, (page 42)]

Krishnakumar BALASUBRAMANIAN, *University of California, Davis*

We study online algorithms for constrained nonconvex stochastic optimization problems with streaming Markovian data. In particular, we focus on the case when the transition kernel of the Markov chain is state-dependent. Such stochastic optimization problems arise in various statistical machine learning problems including strategic classification and reinforcement learning. For this problem, we study both projection-based and projection-free algorithms. In both cases, we establish that the number of calls to the stochastic first-order oracle to obtain an appropriately defined e-stationary point is of the order $O(1/e^{2.5})$. In the projection-free setting we additionally establish that the number of calls to the linear minimization oracle is of order $O(1/e^{5.5})$.

18 . De-preferential attachment model through preferential attachment

[26.A1.I2, (page 4)]

Antar BANDYOPADHYAY, *Indian Statistical Institute, Delhi Centre*
Somak LAHA, *Indian Statistical Institute, Kolkata*

In this work, we will introduce a new model of a growing sequence of random graph, where vertices with high degree are preferred except a few higher degree vertices. The model may be described as follows: at every discrete time step, a fixed proportion or a fixed number of the existing vertices are chosen according to some random/deterministic rule preferring higher degree vertices which are then made ”taboo” vertices. Then a newly arrived vertex is only allowed to join to the non-tabooed vertices but via a (linear)-preferential attachment scheme. We will discus asymptotic properties of this novel model. In particular, we will present results on fixed degree asymptotic, asymptotic degree distribution and asymptotic distribution of the maximum degree and compare the findings with the classical Albert and Barabsi model.

[This is a joint work with Somak Laha]

19. Regridding uncertainty for statistical downscaling of solar radiation

[26.A2.I8, (page 6)]

Soutir BANDYOPADHYAY, *Colorado School of Mines*
Maggie BAILEY, *Colorado School of Mines*
Douglas NYCHKA, *Colorado School of Mines*
Manajit SENGUPTA, *National Renewable Energy Laboratory (NREL)*

As the photovoltaic (PV) industry moves to extend plant lifetimes to 50 years, the changing climate may have an effect on PV production and assumptions that current solar radiation patterns are representative of the future may not be appropriate. A key step in aiding the prediction of PV production is projecting solar radiation for future years based on a changing climate. This involves downscaling future climate projections for solar radiation to spatial and temporal resolutions that are useful for building PV plants. Initial steps in downscaling involve being able to closely predict observed data from regional climate models (RCMs). This prediction requires (1) regridding RCM output from their native grid on differing spatial resolutions to a common grid in order to be comparable to observed data and (2) bias correcting solar radiation data, via quantile mapping, for example, from climate model output. The uncertainty associated with (1) is not always considered for downstream operations in (2). This talk examines this uncertainty which is not often shown to the user of a regridded data product. This analysis is applied to data from the National Solar Radiation Database housed at the National Renewable Energy Lab and a case study of the mentioned methods in California is presented.

20. Identification of Causal Dependencies amongst Telecom KPIs using Latent Convergent Cross Mapping

[29.A1.I71, (page 46)]

Subhadip BANDYOPADHYAY, *Principal Data Scientist, GAIA, Ericsson, Bangalore*
Sujoy Kumar ROY CHOWDHURY, *Principal Data Scientist, GAIA, Ericsson, Bangalore*

Telecommunications networks operate on enormous amount of time-series data, and often exhibit anomalous trends in their behaviour. This is caused due to increased latency and reduced throughput in the network which inevitably leads to poor customer experience . One of the common problems in machine learning in the telecom domain is to predict anomalous behaviour ahead of time. Whilst this is a well-researched problem, though still open, there is far less work done in identifying causal structures from the temporal patterns of various Key Performance Indicators (KPI) in the telecom network. The ability to identify causal structures from anomalous behaviours would allow more effective intervention and generalisation of different environments and networks. The traditional methods of causal analysis are Granger causality and Judea Pearls Do-Why formulation. However, Granger causality is not very well suited for observations from industrial systems because of its limitations with respect to additive measurement errors, subsampling etc. Convergent cross mapping on the other hand is based on the system dynamics by considering the system to be a non-linear system. This is based on Takens Theorem which establishes a one-to-one correspondence between the manifold of a system and its shadow manifold obtained by delay embeddings of the state variables. Further, since data from real-world systems is inherently noisy and often has missing observations, takes it further to learn a latent representation using Neural Ordinary Differential Equations (ODE) using Gated Recurrent Networks . This approach is called Latent Convergent Cross Mapping. In this work, we look at 10 important KPIs in Telecom Networks which are relevant to understand the performance of the network. We use Convergent Cross Mapping to understand the relationships between these variables and learn from observational data the causal structure between the KPIs. The causal structure includes both the direction as well as the strength of the causal links. We then use domain knowledge and SME inputs to understand and validate the learnt causal structure and use this to show that except for one additional spurious causal relationship which it identified (with relatively low weight) the rest of the causal structure can be explained by an understanding of the basic signal processing and communications aspect of the network transmission.

21. Greater than the sum of its parts learning from piecewise data in spatial contexts

[26.A2.I11, (page 7)]

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

Collaborative statistical and machine learning (such as federated learning), are inference paradigm where many models are trained on piecewise data silos, while the inference parameters are combined using a central node. The unique feature of federated learning is that data is never shared between silos, maintaining privacy, thus keeping the model training decentralized. Meaningful inference in federated learning therefore depends critically on the combination step at the central server. We investigate the effect of silo heterogeneity on the resultant inference. While collaborative learning itself is a relatively new concept, there is limited literature on node heterogeneity and spatial covariance estimation, with most approaches assuming exchangeable data distributions. We study the bias when estimating federated learning models across inconsistent nodes. We present a novel methods to for weighting functions across piecewise learning nodes. Additionally, we also propose a statistical algorithm using MCMC for approximate inference in these heterogeneous silo settings. Finally, we demonstrate the performance of our proposed method in simulated and real data settings.

22. Multivariate strong invariance principle for Markov chain Monte Carlo

[28.M1.C9, (page 27)]

Arka BANERJEE, *arkabee96@gmail.com*
Dootika VATS, *Indian Institution of Technology, Kanpur.*

Strong invariance principle (SIP) for the partial sums of some mean-corrected random sequences holds if it can be approximated to a scaled Brownian motion in an almost everywhere sense. Over the years SIP rates have consistently improved for various dependent and independent structures of the random sequences. In this paper, we provide relatively tight SIP rates for polynomially and geometrically ergodic Markov chains. We use wide-sense regenerative structures of Harris ergodic Markov chains, bypassing the need to assume a one-step minorization. A key feature of the rates is that they are completely specified. This allows users of Markov chain Monte Carlo to verify key assumptions employed in output analysis.

23. Adjusted Limited Dependent Variable Mixture Model for quality of health data

[28.M2.I46, (page 30)]

Hiya BANERJEE, *Novartis Pharmaceuticals*
Elke-Christine ORTMANN, *Novartis Pharmaceuticals*
Minal JAIN, *Novartis Pharmaceuticals*

Health-related Utility data is important for populating cost-effectiveness analysis and inform health technology assessments and drug pricing prior to patient access. It assesses a patients perception of their own health described in terms of symptoms, function and well-being using numerical scoring systems and is increasingly used in clinical trials to test effectiveness. One common utility questionnaire is EQ-5D assessing health problems across five domains using ordinal Likert scales. Value sets or valuation functions which are country specific convert these profiles of ordinal measures into cardinal health-related utilities between 1 (perfect health) and minus infinity, where 0 represents death, and negative values represent health states worse than death. Health state utilities often show gaps between 1 and the next smaller utility in the value set. Health state utilities can be the consequence of multiple latent classes, or they can exhibit multi-modal marginal densities. The regular mixed model repeated measures (MMRM) are not the best method to use for these bounded, discontinued, multi-modal data. The alternative analyses, Adjusted Limited Dependent Variable Mixture Model (ADLVMM) can be used for such data. In our presentation, we will discuss the method and results using this method.

24. Minimax Nonparametric Estimation for Markov Decision Processes

[Student Paper Competition 1, (page 13)]

Imon BANERJEE, *Purdue University*
Vinayak Rao, Harsha Honnappa ,

Markov decision processes (MDPs) have wide applications in engineering and machine learning, forming a key component in many reinforcement learning problems. In this work, we consider the estimation of the transition probabilities of finite-state finite-control MDPs, and develop a minimax sample complexity bounds for nonparametric estimation of these transition probability matrices. Unlike most studies, which have been in the online setup, we consider offline MDPs. Our results are quite general since we do not assume anything specific about the logging policy. Instead, the dependence of our statistical bounds on the logging policy comes out in form of a natural mixing coefficient. We demonstrate an interesting tradeoff between stronger assumptions on mixing versus requiring more samples to achieve a particular PAC-bound. We finally demonstrate the validity of our results under various examples, like ergodic Markov chains, weakly ergodic inhomogenous Markov chains, and Markov decision processes with non-stationary Markov, episodic, and greedy (minorized) controls.

25. Supervised change point detection using non-stationary Hawkes process

[29.A1.I71, (page 46)]

Jyotirmoy BANERJEE, *Principal Data Scientist, GAIA, Ericsson, Bangalore*

In statistics change point detection tries to identify when the temporal data changes are significant than usual. The problem concerns both detecting whether or not a change has occurred and identifying the location of any such changes. Accurate detection of the change point helps in downstream tasks like root cause analysis, network monitoring etc. Hawkes process is a self-exciting point process. Time series can be modelled as stochastic processes. Hawkes process is a specific type of temporal point process and has become popular in machine learning due to its ability in handling non-stationary stochastic process. We propose an approach which uses Hawkes process to model the time series data. Additionally our supervised approach is able to handle the associated variability in the time series data present in real world scenario. The approach can rank the probable change points and the ranking threshold can be used to detect the change points. We evaluated our with state-of-the-art benchmarks and show promising results.

26. Transfer learning under posterior drift

[26.A1.I5, (page 5)]

Moulinath BANERJEE, *University of Michigan*
Subha MAITY, *University of Michigan*
Yuekai SUN, *University of Michigan*

I will speak on some of our recent work on transfer learning from a source to a target population in the presence of 'posterior drift': i.e. the regression function/Bayes classifier in the target population is different from that in the source. By modeling the posterior drift through a linear adjustment (on an appropriately transformed scale), we are able to learn the nature of the posterior drift using relatively few samples from the target population as compared to the source population, which usually provides an abundance of samples. I will present applications to mortality prediction from UK Biobank data as well as to image classification data using our new approach.

27 . Assessing Clinical Development Plans using Quantitative Decision Making Framework

[29.A1.I68, (page 45)]

Rudrani BANERJEE, *GSK*

Pharma organizations aim to make informed decision about developing a pipeline of medicinal drugs. This is to ensure the most potent drug makes its way to unmet needs of increasing patient populations. Discussions and decisions related to development of drugs are typically complex in nature with many folded risks. Clinical teams have now resorted to ways of quantifying the unknown risk of failures of a drug due to lack of efficacy and safety concerns based on historical or expert evidence. Quantitative Decision Making (QDM) means to describe in full the quantitative characteristics of the proposed study designs within a development plan, to transparently show the probability of being able to make a decision at the end of each trial and showing how the proposed studies de-risks the next stage of development. In this talk, we will describe an example of how QDM methods were used to assess different options of clinical drug development plans and enabled informed decision making.

28. Spectral properties of random perturbations of non-self-adjoint operators

[30.M1.I78, (page 52)]

Anirban BASAK, *ICTS-TIFR*
Elliot PAQUETTE,
Martin VOGEL,
Ofer ZEITOUNI,

Understanding spectral properties of non-self-adjoint operators are of significant importance as they arise in many problems such as scattering systems, open or damped quantum systems, and the analysis of the stability of solutions to nonlinear PDEs. Absence of suitable methods (e.g. variational methods) renders the study of the spectrum of such operators to be difficult. On the other hand, its high sensitivity to small perturbations leads to serious numerical errors. Motivated by problems in different fields such as numerical analysis, semiclassical analysis, fluid dynamics, and mathematical physics, during the last fifteen years there have been several works in understanding the spectral properties of random perturbations of non-self-adjoint operators. In this talk, we will focus on random perturbations of large dimensional non-self-adjoint Toeplitz matrices, and discuss (i) the limiting law for the empirical measure of its eigenvalues, (ii) limit law for outlier eigenvalues, and (iii) localization/delocalization of its eigenvectors, and the universality and non-universality of these features. I will also present some fun pictures and simulations. Based on joint works with Elliot Paquette, Martin Vogel, and Ofer Zeitouni.

29. An Empirical Bayes Approach to Controlling the False Discovery Exceedance

[26.A1.I3, (page 4)]

Pallavi BASU, *Indian School of Business*

In large-scale multiple hypothesis testing problems, the false discovery exceedance (FDX) provides a desirable alternative to the widely used false discovery rate (FDR) when the false discovery proportion (FDP) is highly variable. We develop an empirical Bayes approach to control the FDX. We show that, for independent hypotheses from a two-group model and dependent hypotheses from a Gaussian model fulfilling the exchangeability condition, an oracle decision rule based on ranking and thresholding the local false discovery rate (lfdr) is optimal in the sense that the power is maximized subject to the FDX constraint. We propose a data-driven FDX procedure that uses carefully designed computational shortcuts to emulate the oracle rule. We investigate the empirical performance of the proposed method using both simulated and real data and study the merits of FDX control through an application for identifying abnormal stock trading strategies.

30. Space time correlations in models of random growth

[28.A1.I50, (page 32)]

Riddhipratim BASU, *International Centre for Theoretical Sciences*

Last passage percolation is a canonical model of planar random growth believed to be in the Kardar-Parisi-Zhang (KPZ) universality class under mild conditions, but the rigorous progress is mostly limit to a handful of exactly solvable models including the exponential last passage percolation. In this talk we shall discuss a number of results regarding the decay of correlations in exponential last passage percolation in both the space and the time direction.

31. Bayesian Nonparametric Modeling of Restricted Mean Survival Time

[29.A2.I73, (page 47)]

Sanjib BASU, *University of Illinois Chicago*
Ruizhe CHEN, *Johns Hopkins University*

Restricted mean survival time (RMST) is increasingly being used in planning and analyzing time-to-event outcome in clinical, medical and health sciences. The concept of RMST is model independent and its popularity stems from its advantages as a summary in survival analysis over other conventional measures such as the hazard ratio. We develop a Bayesian nonparametric model for RMST and develop subject-level RMST inference as well as group-level causal inference of average treatment effect. We evaluate performance of the proposed Bayesian nonparametric approach and compare with non-Bayesian methods. We present an application of our proposed approach to analyze data from a metastatic colorectal cancer trial.

32. Understanding the role of genomics in Human Health

[Special Invited Session 4, (page 52)]

Saonli BASU, *University of Minnesota*

Recent advances in sequencing technologies have provided us with a wealth of genomic information on individuals. The current effort to generate genomic data on big cohorts such as the UK Biobank study and All of Us cohort provide us the opportunities to gain deeper insight into the genetic mechanism of a complex trait. I will discuss how statistical methodologies contribute to identifying the genetic risk factors and assist in a better understanding of gene and environment interplay. I will talk about the challenges in implementing these statistical tools in large-scale Biobank samples and in multiethnic cohorts to study the role of genetics in disease development. I will focus on some of the current strategies to perform genetic analyses in multi-ethnic Biobank samples and discuss the importance of reproducible research in for analyzing genomic data with such high dimensionality and complexity.

33. Graphical Evidence

[28.M1.I37, (page 24)]

Anindya BHADRA, *Purdue University*
Ksheera SAGAR, *Purdue University*
Sayantan BANERJEE, *IIM Indore*
Jyotishka DATTA, *Virginia Tech*

Marginal likelihood, also known as model evidence, is a fundamental quantity in Bayesian statistics. It is used for model selection using Bayes factors or for empirical Bayes tuning of prior hyper-parameters. Yet, the calculation of evidence has remained a longstanding open problem in Gaussian graphical models. Currently, the only feasible solutions that exist are for special cases such as the Wishart or G-Wishart, in moderate dimensions. We present an application of Chib's technique that is applicable to a very broad class of priors under mild requirements. Specifically, the requirements are: (a) the priors on the diagonal terms on the precision matrix can be written as gamma or scale mixtures of gamma random variables and (b) those on the off-diagonal terms can be represented as normal or scale mixtures of normal. This includes structured priors such as the Wishart or G-Wishart, and more recently introduced element-wise priors, such as the Bayesian graphical lasso and the graphical horseshoe. Among these, the true marginal is known in an analytically closed form for Wishart, providing a useful validation of our approach. For the general setting of the other three, and several more priors satisfying conditions (a) and (b) above, the calculation of evidence has remained an open question that this article resolves under a unifying framework.

34. Stochastic comparisons between two finite mixture models with exponentiated location-scale distributed components

[Poster Session, (page 33)]

RAJU BHAKTA, *National Institute of Technology Rourkela*

Suchandan Kayal ,

In this work, we provide sufficient conditions for stochastic comparisons between two finite mixture models with respect to usual stochastic order when the mixing components have exponentiated location-scale family of distributions. The sufficient conditions are based on various majorization orders between the vectors of mixing proportions and/or model parameters of the finite mixtures. Specifically, the stochastic comparisons are carried out while heterogeneity occurs in one model parameter or occurs in both mixing proportion and a model parameter. Furthermore, the usual stochastic order between two finite mixture models is also established based on the concept of unordered majorization order. To illustrate the theoretical findings, several relevant numerical examples and counterexamples are presented here.

35. Probability Density-Based Clustering

[28.M1.C10, (page 27)]

Shrikrishna BHAT K, *Pondicherry University*

The initialisation is one of the crucial steps in iterative algorithms, particularly for the EM algorithm. In the case of model-based clustering using finite mixture models, the results of the K-means algorithm are generally used for the initialisation of the EM algorithm. However, it only initialises the location and scale parameters of the cluster. Components in mixture models may have another parameter where K-means fail to initialise them. To deal with such an initialisation problem, a probability density-based clustering algorithm is introduced where individual components' probability density function is considered an asymmetric similarity proximity measure. The results are illustrated with simulated and real-life data sets and compared with the EM algorithm and Probabilistic density distance clustering algorithm. Keywords: EM algorithm, Asymmetric Similarity Proximity Measure, K-Means, Initialisation.

36. A weighted test to evaluate contribution of components in a randomized trial with combination therapy

[28.M1.C10, (page 28)]

Abhishek BHATTACHARJEE, *Pfizer*

There are some challenges in developing a combination therapy with two or more therapeutic agents. It is important for clinical trials with combination therapies to evaluate the contribution of components. As such, comparisons between the standard of care and each agent in the combination is required to justify the efficacy benefit of the combination against potential exposure to added toxicity. This presentation will propose the use of weighted linear contrasts to identify the active agents (factors) of a treatment combination in a concurrently randomized study utilizing a factorial design. In general, efficiency of factorial design is diminished in the presence of interaction effect for traditional main effects tests. However, there is very limited scope for testing an interaction effect in a factorial randomized trials due to sample size constraints. This new proposal to test and identify active factors based on weighted linear contrasts is more powerful in the presence of an interaction effect. Further, use of this weighted test for identifying an active therapeutic agent in a treatment combination may be used for designing efficient trials without compromising statistical power. This presentation will discuss several applications of the proposed method to phase 2 and phase 3 trials, even beyond trials with combination therapy.

37. Early assessment of potential endemic pattern of COVID19 in India

[28.A1.I49, (page 31)]

Madhuchhanda BHATTACHARJEE, *University of Hyderabad, Hyderabad, India*

Arup BOSE, *Indian Statistical Institute, Kolkata, India*

For a disease, within a given geographic region, endemic refers to the steady prevalence of the disease in the population. Public health experts believe that potentially COVID would stay in the population with a constant number of expected infections throughout the year with likely seasonal patterns and/or occasional spikes in incidences.

The traditional endemic models are based on various modified versions of the SIR models. A very few have attempted the more sophisticated spatially autocorrelated endemic-epidemic models (by Held et al 2005). However, the regional variations and apparent inharmonious occurrence of the disease in different parts of India causes difficulties in direct applications of such models to this subcontinent.

Since the beginning of the pandemic, we have been studying the paths COVID19 has taken across various parts of India and we have done so at various geographical scales. Our rigorous investigations into the spatio-temporal behaviour of this epidemic in India brought out salient commonality in the patterns observed across geographic locations and irrespective of epidemic magnitudes that could very well be the early indicators of the endemic pattern of COVID19 in India.

38. Asymptotics of Large Autocovariance Matrices

[26.A1.I3, (page 4)]

Monika BHATTACHARJEE, *Department of Mathematics, IIT Bombay*

We consider the high-dimensional moving average process and explore the asymptotics for eigenvalues of its sample autocovariance matrices. Under quite weak conditions, we prove, in a unified way, that the limiting spectral distribution (LSD) of any symmetric polynomial in the sample autocovariance matrices, after suitable centering and scaling, exists and is nondegenerate. We use methods from free probability in conjunction with the method of moments to establish our results. In addition, we are able to provide a general description of the limits in terms of some freely independent variables. We also establish asymptotic normality results for the traces of these matrices. We suggest statistical uses of these results in problems such as order determination of high-dimensional MA and AR processes and testing of hypotheses for coefficient matrices of such processes.

39. Time-dynamics of the COVID-19 Pandemic: Inference and Mitigation Strategy

[28.A1.I49, (page 31)]

Satarupa BHATTACHARJEE, *UC Davis*
Shuting LIAO, *Graduate Group in BioStatistics, University of California, Davis, 95616, USA*
Debashis PAUL, *Department of Statistics, University of California, Davis, 95616, USA*
Sanjay CHAUDHURI, *Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117546, Singapore*

Modeling the time-dynamic nature of the COVID-19 pandemic has been of great importance in understanding the course of the pandemic trajectories and hence implementing relevant intervention strategies accordingly to mitigate the spread of the disease.

In the first part of the talk, the evolution of the COVID-19 pandemic is described through a time-dependent stochastic dynamic model in discrete time. The proposed is multi-compartment and information on the social distancing measures and diagnostic testing rates are incorporated to characterize the dynamics of the various compartments of the model. A vital feature of the methodology is its ability to estimate latent unobservable compartments such as the number of asymptomatic but infected individuals who are known to be the critical vectors of COVID-19 spread. The methodology is applied to understand the true extent and dynamics of the pandemic in various states within the United States (US).

In the second part of the talk, we consider a more elaborate, network-based, mechanistic model to describe the joint dynamics to discuss the potential effects of different intervention options.. Intervention options consist of testing, vaccination, and partial lockdown. We propose a mechanism for determining an optimal lockdown scheme by optimizing an objective function that takes into account the economic cost associated with the policy. We conduct simulation studies to compare various possible scenarios across several population centers.

40. A divide-and-conquer approach to perform consensus independent component analysis on large scale datasets

[28.M1.C9, (page 27)]

Arkajyoti BHATTACHARYA, *Post doctoral researcher, University of Groningen, The Netherlands*
Setareh REZAEE OSHTERNIAN, *PhD student, University of Groningen*
Mark FREDERIKS, *MSc*
Rudolf S. N. FEHRMANN, *Professor, MD, PhD, University of Groningen*

Introduction: Currently consensus independent component analysis (c-ICA) is extensively used in transcriptomics to segregate complex mixture of signals present in mRNA expression profiles. However, in large set of mRNA expression profiles, users need to reduce a percentage of explained variance in the data to conduct c-ICA computationally feasibly. This reduction of variance is conducted by removing the principal components which explain least variance in the data. Due to this reduction of explained variance, a distinct set of biological signals from rare diseases with low number of samples, might not be present anymore in the c-ICA results. Hence, it is necessary to come up with a computationally feasible method which can maximize detecting independent signals without reducing explained variance. Method: We used a divide-and-conquer approach to conduct c-ICA (DCC-ICA) which can find biological signals that were previously missed. First, hierarchical clustering was used to obtain distinct clusters of homogenous mRNA expression profiles. Thereafter, c-ICA was conducted on the distinct clusters separately with 100% explained variance to identify statistically independent components in each cluster. Lastly, a consensus on independent components from all clusters was obtained by maximizing robustness and minimizing shared information between the components. Results: We simulated 2,000 mRNA expression profiles using random mixture of 10 independent components (source ICs) obtained from real mRNA expression profiles. With the aim of finding the 10 source ICs, we applied both regular c-ICA with 95% explained variance and DCC-ICA on the simulated dataset. DCC-ICA could find all the 10 source ICs in comparison to c-ICA finding 9 of them. Also, maximum used memory and computational time for DCC-ICA was lower than c-ICA. Application of DCC-ICA on 10,817 mRNA expression profiles of cancer patients from the cancer genome atlas (TCGA) resulted in finding numerous significantly biologically enriched ICs which were not present in the list of ICs obtained from c-ICA. Conclusion: In conclusion, we provide a computationally feasible approach to maximize insight into mRNA expression levels for large datasets.

41. Persistence of heavy-tailed sample averages: principle of infinitely many big jumps

[29.A1.I69, (page 45)]

Ayan BHATTACHARYA, *Assistant Professor, Department of Mathematics, IIT Bombay*
Zbigniew PALMOWSKI, *University of Wroclaw*
Bert ZWART, *CWI, Amsterdam*

We consider the sample average of a centered random walk in the d-dimensional Euclidean space with regularly varying step size distribution. For the first exit time from a compact convex set A not containing the origin, we show that its tail is of lognormal type. Moreover, we show that the typical way for a large exit time to occur is by having a number of jumps growing logarithmically in the scaling parameter. This is a joint work with Zbigniew Palmowski and Bert Zwart and available at https://arxiv.org/abs/1902.09922 .

42. Estimation and Inference in Inhomogeneous Random Graph Models

[28.A1.I52, (page 32)]

Bhaswar BHATTACHARYA, *University of Pennsylvania*
Anirban CHATTERJEE , *University of Pennsylvania*
Soham DAN , *University of Pennsylvania*
Svante JANSON , *Uppsala University*

Motifs (patterns of subgraphs), such as edges and triangles, encode important structural information about the geometry of a network. Consequently, counting motifs in a large network is an important statistical and computational problem. In this talk we will discuss the problem of estimating motif densities in an inhomogeneous random graph sampled from a graphon. We will show that the limiting distributions of subgraph counts can be Gaussian or non-Gaussian, depending on a notion of regularity of subgraphs with respect to the graphon. Using these results and a novel multiplier bootstrap for graphons, we will construct joint confidence sets for the motif densities and discuss applications in network two-sample testing.

(Joint work with Anirban Chatterjee, Soham Dan, and Svante Janson.)

43. A response adaptive allocation for ordinal response clinical trials

[29.M2.I62, (page 43)]

Rahul BHATTACHARYA, *Professor*

Ordinal categorical responses are natural outcomes in many disciplines of clinical trials like Orthopaedics and Ophthalmology. However, most of the clinical trial designs developed for such responses either club the categories to make the response binary or assign numerical scores to treat the response as continuous. Adaptive allocation designs are often considered in the context of clinical trials for their ability to favour the treatment doing better using all available data. Therefore, a response adaptive design is developed for ordinal categorical treatment outcome in a convenient way without losing any information. Properties of the proposed design are studied and the applicability is further illustrated using a real data.

44. Multiple inflated negative binomial regression for correlated multivariate count data

[27.M1.C5, (page 14)]

SUMANGAL BHATTACHARYA, *Department of Mathematics and Statistics, IIT Tirupati*
Joseph MATHEWS,
Sumen SEN, *Associate Professor, Austin Peay State Uni-*

versity
Ishapathik DAS, *Asst. Professor, IIT Tirupati*

This article aims to provide a method of regression for multivariate multiple inflated count responses assuming the responses follow a negative binomial distribution. Negative binomial regression models are common in the literature for modeling univariate and multivariate count data. However, two problems commonly arise in modeling such data: choice of the multivariate form of the underlying distribution and modeling the zero-inflated structure of the data. Copula functions have become a popular solution to the former problem because they can model the response variables dependence structure. The latter problem is often solved by modeling an assumed latent variable Z generating excess zero-valued counts in addition to the underlying distribution. However, despite their flexibility zero-inflation models do not account for the case of m additional inflated values at k1, k2, . . . , km. We propose a multivariate multiple-inflated negative binomial (MMINB) regression model for modeling such cases. Furthermore, we present an iterative procedure for estimating model parameters using maximum likelihood estimation (MLE). The multivariate distribution functions considering the dependence structure of the response vectors are found using copula methods. The proposed method is illustrated using simulated data and applied to real data.

45. A Two-Part Tweedie Model for Differential Analysis of Omics Data

[29.M1.I55, (page 38)]

Arinjita BHATTACHARYYA, *Merck*
Suvo CHATTERJEE, *Indiana University*
Himel MALLICK, *Merck*

One common objective in single-cell RNA-sequencing (scRNA-Seq) studies is to detect differentially expressed genes or genomic features across experimental conditions. Due to the nature of the associated data which is typically characterized by a large number of zero counts, most published methods employ two-part models to identify the effects of biological variation in single-cell RNA sequencing data: (i) a logistic sub-model to model the binarized expression profiles, and (ii) a count or continuous regression to model the non-zero values. While these methods are able to detect differences in the expression prevalence and the average expression level (among the expressed cells), they fail to provide an unconditional interpretation of covariate effects on the average gene expression, reducing their flexibility in practical applications. Here, we propose a two-part Tweedie regression model (TPCPLM) (with or without random effects) for testing the association between overall gene expression and clinical covariates for both individual-level and cell-level differential expression. The model includes a logistic regression component to model the binarized representation of the single-cell RNA sequencing data and a Tweedie regression component to model the overall gene expression, where each component may include a random effect to account for the repeated measurements. Simulation studies show that the TPCPLM model outperforms published methods in false discovery rate control while maintaining power. In real data, TPCPLM identifies uniquely detected genes not easily identified by published methods. TPCPLM is available as part of the open-source R package https://github.com/himelmallick/Tweedieverse.

46. Managing the Biological Invasion by Invasive Alien Plant Species in India: Insights from the ILORA Database using Data Science Applications

[28.M2.I45, (page 29)]

Amiya Ranjan BHOWMICK, *Institute of Chemical Technology, Mumbai*
Jyoti Jagdish PRAJAPATI, *Institute of Chemical Technology, Mumbai*
Achyut Kumar BANERJEE, *Sun Yat-sen University, Guangzhou, China*
Abhishek MUKHERJEE , *Indian Statistical Institute, Giridih Branch, Jharkhand*

Biological invasions represent one of the major environmental challenges for the conservation of global biodiversity and the continuation of ecosystem services and have a serious impact on human and animal health and economic development. Pant et al. (2021) and Prajapati et al. (2022) have created a comprehensive database, named, Indian Alien Flora Information (ILORA) (https://ilora2020.wixsite.com/ilora2020) which contains data for 14 variables related to ecology, biogeography, introduction pathway, socio-economy, and distribution of 1747 alien vascular plant species from 22 national and global sources. The data set is expected to assist a wide range of stakeholders involved in India's scientific research, policy formulation, and decision-making related to Invasive Alien Plant Species (IAPS). In recent years, several mod-

elling tools have been used to predict the consequences of climate change on the spatial distribution of animal species, weeds, and crops in India. However, most of the studies are confined to one or to a limited number of species. In this talk, we shall discuss how the ILORA database will be useful in implementing policy regulations with the aid of data science applications. In particular, the identification of the species which are expected to show significant range expansion in India is an important task which will provide management insights.

47. Causal Analysis at Extreme Quantiles with Application to London Traffic Flow Data

[29.A1.I70, (page 46)]

Prajamitra BHUYAN, *Indian Institute of Management, Calcutta*
Kaushik JANA, *Ahmedabad University*
Emma MCCOY, *Imperial College London*

Transport engineers employ various interventions to enhance traffic-network performance. Recent emphasises on cycling as a sustainable travel mode aims to reduce traffic congestion. Quantifying the impacts of Cycle Superhighways is complicated due to the non-random assignment of such intervention over the transport network and heavy-tailed distribution of traffic flow. Treatment effects on asymmetric and heavy tailed distributions are better reflected at extreme tails rather than at averages or intermediate quantiles. In such situations, standard methods for estimating quantile treatment effects at the extremes can provide misleading inferences due to the high variability of estimates. In this work, we propose a novel method which incorporates a heavy tailed component in the outcome distribution to estimate the extreme tails and simultaneously employs quantile regression to model the bulk part of the distribution utilising a state-of-the-art technique. Simulation results show the superiority of the proposed method over existing estimators for quantile causal effects at extremes in the case of heavy tailed distributions. The analysis of London transport data utilising the proposed method indicates that the traffic flow increased substantially after the Cycle Superhighway came into operation. The findings can assist government agencies in effective decision making to avoid high consequence events and improve network performance.

48. Bayesian mixture models for detecting protein-DNA footprints

[27.A1.I30, (page 19)]

Anushua BISWAS, *IISER Pune*
Leelavati NARLIKAR, *IISER Pune, Data Science Department*

Transcriptional regulation is governed by interactions between proteins called transcription factors (TFs) and their corresponding DNA binding sites. Identifying these binding sites is critical to understand their role in gene-regulation. High-throughput chromatin immunoprecipitation (ChIP) sequencing-based assays capture all the genomic regions that are associated with the TF. ChIP-exo is a modified protocol, which uses the lambda exonuclease enzyme to digest DNA close to the TF-DNA complex, thereby improving the positional resolution of the TF-DNA contact. Because the digestion occurs in the 5-3 orientation, the protocol produces directional footprints close to the complex, on both sides of the double stranded DNA. Like all ChIP-based methods, ChIP-exo reports a mixture of different regions associated with the TF: those binding directly to the TF as well as via intermediaries. The distribution of the footprints are likely to be indicative of the complex forming at the DNA. We present ExoDiversity[1], which uses a model-based framework to learn a joint distribution over footprints and motifs, thereby resolving the mixture of ChIP-exo footprints into diverse binding modes. It uses no prior motif or TF information and automatically learns the number of different modes from the data. The framework is written such that it is flexible to handle most discrete distributions for modeling the footprints. I will focus primarily on the Bernoulli distribution over the presence and absence of the footprint signals taken independently at each position in the window of interest around the motifs. We show its application on a wide range of TFs and organisms/cell-types. Because ExoDiversitys goal is to explain the complete set of reported regions, it is able to identify co-factor TF motifs that appear in a small fraction of the dataset. Further, ExoDiversity discovers small nucleotide variations within and outside canonical motifs, which co-occur with variations in footprints, suggesting that the TF-DNA structural configurations at those regions are likely to be different. Finally, we show that detected modes have specific DNA shape features and conservation signals, giving insights into the structure and function of the putative TF-DNA complexes. [1] Biswas, Anushua, and Leelavati Narlikar. ”Resolv-

ing diverse proteinDNA footprints from exonuclease-based ChIP experiments." Bioinformatics 37.Supplement_1 (2021): i367-i375.

49. Borrowing of Historical Clinical trial data for Effiacy Analysis

[29.A1.I68, (page 45)]

Arunangshu BISWAS, *GSK*

Pharmaceutical companies often leverage information from other sources to augment the current trial. This may involve using historical data, market data, prior elicitations, and others to inform on the behavior of the present drug under study. In this talk we discuss about the one such method, called Bayesian Dynamic Borrowing, that uses the Bayesian paradigm to choose a suitable prior from a subpopulation or another related population. We discuss the tipping point method that enables to choose the weights for different component of the prior. Finally, we discuss few examples of bayesian dynamic borrowing that have been used in the pharmaceutical industry.

50. Modeling of Load-sharing Systems using Piecewise Linear Approximation to Cumulative Hazard

[27.M1.C4, (page 14)]

Shilpi BISWAS, *Research Scholar, IIT Guwahati*
Ayon GANGULY, *Assistant Professor, IIT Guwahati*
Debanjan MITRA, *Assistant Professor, IIM Udaipur*

In this article, we consider load-sharing system consists of certain number of components connected in parallel. In such a systems, when component fails one by one, the total workload redistributes among the surviving components and hence loads on the remaining components increase. In this work, we describe a piecewise linear approximation (PLA) for the cumulative hazard in load-sharing systems with unknown load-share rules. This model is data-driven, and can be suitably tuned by choosing the number of linear pieces for the PLA at each stage. Maximum likelihood estimates of the model parameters are obtained. Percentile bootstrap and bias adjusted bootstrap methods are used to construct confidence intervals for model parameters. Monte Carlo simulations are performed to study the performance of inferential methods. The robustness of the model is also studied through detailed Monte Carlo simulations. The mean time to failure, mean residual time and reliability at a mission time of the system for the proposed model are discussed. A real data on two-component systems is thoroughly studied for illustrative purpose.

51. Estimation of Spectral Risk Measure for Left Truncated and Right Censored Data

[26.A2.I12, (page 7)]

Suparna BISWAS, *Indian Statistical Institute, Bangalore*
Rituparna SEN,

Left truncated and right censored data are encountered frequently in insurance loss data due to deductibles and policy limits. Risk estimation is an important task in insurance as it helps in evaluating deductible credits and price for higher limits. 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 spectral risk measure using the product limit estimator. Consistency and asymptotic normality for the estimator of spectral risk measure are derived. 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 Norwegian fire claims data set.

52. Bayesian Meta-Analysis of Penetrance and its Application to Breast Cancer Risk among ATM Mutation Carriers

[29.A2.I75, (page 48)]

Swati BISWAS, *University of Texas at Dallas*
Thanthirige Lakshika RUBERU, *University of Texas at Dallas*
Danielle BRAUN, *Harvard T.H. Chan School of Public Health*
Giovanni PARMIGIANI, *Dana Farber Cancer Institute*

Multi-gene panel testing allows testing of many cancer susceptibility genes quickly at a lower cost making such testing accessible to a broader population. Thus, more patients carrying pathogenic germline mutations in various cancer-susceptibility genes are being identified. This creates a great opportunity, as well as an urgent need, to counsel these patients about appropriate risk reducing management strategies. Counseling hinges on accurate estimates of age-specific risks of developing various cancers associated with mutations in a specific gene, i.e., penetrance estimation. We propose a meta-analysis approach based on a Bayesian hierarchical random-effects model to obtain penetrance estimates by integrating studies reporting different types of risk measures (e.g., penetrance, relative risk, odds ratio) while accounting for the associated uncertainties. After estimating posterior distributions of the parameters via a Markov chain Monte Carlo algorithm, we estimate penetrance and credible intervals. We investigate the proposed method and compare with an existing approach via simulations based on studies reporting risks for two moderate-risk breast cancer susceptibility genes, ATM and PALB2. Our proposed method is far superior in terms of coverage probability of credible intervals and mean square error of estimates. Finally, we apply our method to estimate the penetrance of breast cancer among carriers of pathogenic mutations in the ATM gene.

53. Spectrum of High Dimensional Sample Covariance and Related Matrices: A Selective Review

[27.M1.I15, (page 12)]

Arup BOSE, *Indian Statistical Institute*

We provide a ‘selective review on the behaviour of the high dimensional sample covariance matrix S, the most important random matrix in high dimensional statistics, and some related matrices. The asymptotic distribution of the empirical spectrum of (centered and scaled) S is either the semi-circular law or the Marchenko-Pastur law, depending on how the dimensions n and p grow. From a non-commutative probability point of view, independent copies of S converge algebraically to freely independent semi-circular or compound free Poisson variables. We also discuss the asymptotic normality of the linear spectral statistics, and the convergence of the maximum eigenvalue to a Tracy-Widom law. Some related matrices are also discussed. These include the separable, generalised, and cross-covariance matrices, Kendall’s tau and Spearman’s rho matrices, and the autocovariance matrices in one and high dimensions. Finally, we present some statistical applications.

54. Copula-based estimation of health concentration curves with an application to COVID-19

[30.M1.I77, (page 51)]

Taoufik BOUEZMARNI, *Universit de Sherbrooke*
Abderrahim Taamouti MOHAMED DOUKALI, *McGill University*
UNIVERSITY OF LIVERPPOL,

COVID-19 has created an unprecedented global health crisis that caused millions of infections and deaths worldwide. Many, however, argue that pre-existing social inequalities have led to inequalities in infection and death rates across social classes, with the most-deprived classes are worst hit. In this paper, we derive semi/non-parametric estimators of Health Concentration Curve (HC) that can quantify inequalities in COVID-19 infections and deaths and help identify the social classes that are most at risk of infection and dying from the virus. We express HC in terms of copula function that we use to build our estimators of HC. For the semi-parametric estimator, a parametric copula is used to model the dependence between health and socio-economic variables. The copula function is estimated using maximum pseudo-likelihood estimator after replacing the cumulative distribution of health vari- able by its empirical analogue. For the non-parametric estimator, we replace the copula function by a Bernstein copula estimator. Furthermore, we use the above estimators of HC to derive copula-based estimators of health Gini coefficient. We establish the consistency and the asymptotic normality of HCs estimators. Using different data-generating processes and sample sizes, a Monte-Carlo simulation exercise shows that the semiparametric estimator outperforms the smoothed nonparametric estima- tor, and that the latter does better than the empirical estimator in terms of Integrated Mean Squared Error. Finally, we run an extensive empirical study to illustrate the importance of HCs estimators for investigating inequality in COVID-19 infections and deaths in the U.S. The empirical results show that the inequalities in states socio-economic variables like poverty, race/ethnicity, and economic prosperity are behind the observed inequalities in the U.S.s COVID-19 infections and deaths.

55. The relationship of agricultural wage rate among states in India: A Vector Error Correction Model Analysis

[28.M1.C9, (page 27)]

Yashavanth BS, *yashavanthbs@gmail.com*
Rakesh J, *Professor Jayashankar Telangana State Agricultural University*

Indian agriculture is labour intensive and serves as the source of livelihood for majority of the population. Often, due to variations in the wage rate among neighbouring states, agricultural labourers are found to be migrating from one state to another looking for better wage rates. This study attempted to explore the interactive relations among agricultural labour wage rates in five neighbouring Indian states viz., Andhra Pradesh, Karnataka, Tamilnadu Telangana and Chhattisgarh using monthly time series data of 2005-2020. The objective of this study was to examine the degree of integration among wage rates of agricultural labourers in neighbouring states. We built vector error correction model (VECM) by conducting stationarity test, Granger causality test and cointegration test. For building the VEC Model, the complete data set (180 data points) was split into training (160 data points) and testing (20 data points) data sets. The nonstationarity of the data was established by the Augmented Dickey Fuller test. The Granger causality test indicated that the wage rates in different states are influenced by each other. Results from the Johansen cointegration test proved that the wage rates among different states are cointegrated suggesting the long run equilibrium. For the purpose of forecasting, VECM (1) was built and tested for goodness of fit using Mean Absolute Percentage Error (MAPE) which were found to be ¡ 10% for all the states suggesting good fit of the VECM model. The impulse response functions were used to understand the influence of shocks in wage rate of one state on the others. The results revealed that fluctuations in wage rates are transmitted across the states, indicating that changes in wage rate in one state are consistently related and influenced by the wage rate changes in other states. Keywords: Agriculture; Cointegration; Labour wage; VECM; Forecasting

56. Inferring causal effects using autopsy data

[27.A2.I33, (page 20)]

Marco CARONE, *University of Washington*
Eric MORENZ, *University of Washington*
Stijn VANSTEELANDT, *Ghent University + London School of Hygiene and Tropical Medicine*
Lianne SHEPPARD, *University of Washington*

It is often of interest to characterize how an exposure affects a particular disease biomarker. For biomarkers that can only be measured at autopsy, naive comparisons of observed biomarker values can be highly misleading when the exposure under consideration affects survival. In such cases, harmful exposures can even be found to be protective. In this work, we study this problem from a causal inference perspective, and provide a framework for studying the effect of a point-exposure on a time-varying biomarker process that can only be sampled at death. We identify relevant causal estimands and describe how to perform nonparametric efficient inference under several practical study designs, characterized by various forms of censoring and truncation. We discuss how our methods can be used to better understand drivers of dementia biomarkers measured in brain tissue.

57. Impact of the Prior in Bayesian Nonparametric Models

[27.A2.I35, (page 21)]

Marta CATALANO, *University of Warwick*
Hugo LAVENANT,

The Dirichlet process has been a cornerstone in the development of Bayesian Nonparametrics, allowing one to learn the law of an exchangeable sequence through closed form expressions. Still, its learning rule is often too restrictive and many generalizations have been proposed to increase its flexibility, such as the class of normalized completely random measures. Here we investigate a simple yet fundamental matter: will a different nonparametric prior lead to a different learning outcome? To this end we define a tractable distance between the induced exchangeable sequences, which relies on optimal transport and allows one to perform comparisons between models both a priori and a posteriori. We put very mild assumptions on the data generating process.

58. Inhomogeneous Erds-Rnyi random graphs: bulk and edge of the spectrum

[26.A2.I7, (page 5)]

Arijit CHAKRABARTY, *Indian Statistical Institute, Kolkata*

The talk is on inhomogeneous Erds-Rnyi random graphs in the non-dense regime. The eigenvalues of the adjacency matrix of the graph are studied. The empirical spectral distribution of the matrix after suitable scaling and centering is shown to have a deterministic limit in probability. 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. Assum-

ing 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.

The talk is based on joint works with Sukrit Chakraborty, Rajat S. Hazra, Frank den Hollander and Matteo Sfragara.

59. How to test for the rank of a stochastic process?

[30.M2.I82, (page 54)]

Anirvan CHAKRABORTY, *IISER Kolkata*
Victor M. PANARETOS, *EPFL Switzerland*

How can we determine whether a mean-square continuous stochastic process has a finite rank, and if so, what this precise rank is? And how can we do so at a given level of confidence? This question is central to a great deal of methods for functional data, which require low-dimensional representations whether by functional PCA or other methods. The difficulty is that the determination is to be made on the basis of iid replications of the process observed discretely and with measurement error contamination. This adds a ridge to the empirical covariance, obfuscating the underlying dimension. We build a matrix-completion inspired test statistic that circumvents this issue by measuring the best possible least square fit of the empirical covariance's off-diagonal elements, optimized over covariances of given finite rank. For a fixed grid of sufficient size, we determine the statistic's asymptotic null distribution as the number of replications grows. We then use it to construct a bootstrap implementation of a step-wise testing procedure controlling the family-wise error rate corresponding to the collection of hypothesis formalizing the question at hand. Under minimal regularity assumptions we show that the procedure is consistent and that its bootstrap implementation is valid. The procedure involves no tuning parameters or pre-smoothing, is indifferent to the homoskedasticity or lack of it in the measurement errors, and does not assume a low-noise regime. An extensive study of the procedure's finite-sample accuracy demonstrates remarkable performance in both real and simulated data.

This work is joint with Victor M. Panaretos (EPFL, Switzerland)

60 . Bayesian inference on high-dimensional multivariate binary responses

[29.M1.C14, (page 40)]

Antik CHAKRABORTY, *Purdue University*
Rihui OU, *Duke University*
David DUNSON, *Duke University*

It has become increasingly common to collect high-dimensional binary response data; for example, with the emergence of new sampling techniques in ecology. In smaller dimensions, multivariate probit (MVP) models are routinely used for inferences. However, algorithms for fitting such models face issues in scaling up to high dimensions due to the intractability of the likelihood, involving an integral over a multivariate normal distribution having no analytic form. Although a variety of algorithms have been proposed to approximate this intractable integral, these approaches are difficult to implement and/or inaccurate in high dimensions. Our main focus is in accommodating high-dimensional binary response data with a small to moderate number of covariates. We propose a two-stage approach for inference on model parameters while taking care of uncertainty propagation between the stages. We use the special structure of latent Gaussian models to reduce the highly expensive computation involved in joint parameter estimation to focus inference on marginal distributions of model parameters. This essentially makes the method embarrassingly parallel for both stages. We illustrate performance in simulations and applications to joint species distribution modeling in ecology.

61. Designing Micro-randomized Trials for Mobile Health Studies via Thompson Sampling

[28.M1.I39, (page 25)]

Bibhas CHAKRABORTY, *National University of Singapore & Duke University*
Xueqing LIU, *National University of Singapore*
Nina DELIU, *Sapienza University of Rome*

Mobile health (mHealth) or digital health interventions are becoming increasingly common in tandem with advances in technology. In this talk, we will present an innovative trial design arising in mHealth, namely, the micro-randomized trial (MRT) that involves sequential, within-person randomization over many instances. MRTs often aim to estimate the proximal effects of push-type mHealth interventions, e.g., motivational text-messages to promote physical activity or other healthy behaviors. The basic MRT design can be further improved to make it 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 is often operationalized via contextual bandit algorithms. Specifically, we will discuss the role of a particular contextual bandit algorithm, namely, Thompson sampling, in designing adaptive MRTs with count-type outcomes. Simulation results will be shown to validate the proposed design approach. mHealth case studies will be discussed in detail.

62. Efficient Model Fitting and Two-Sample Testing for Large Networks via Subsampling

[Poster Session, (page 33)]

Kaustav CHAKRABORTY, *University of Illinois at Urbana-Champaign*
Srijan Sengupta, Yuguo Chen ,

Big data in many areas of science are now routinely represented in the form of large networks. Analysis of such networks, e.g. network model fitting and two-sample hypothesis testing are topics of considerable interest in the statistics community. However, most of the methods proposed in the literature are unsuitable for very large networks due to their computational cost. Here we present new techniques based on subsampling strategies in the Random Dot Product Graph Model (RDPG) framework, in order to reduce the time complexity. The key idea for subsampling is to divide the whole network into several sub-networks with an overlap region and applying the corresponding method on each of the sub-networks. We develop a step-by-step procedure to perform the subsampling for estimating parameters and two algorithms using parametric bootstrap and multiple testing to perform the two-sample hypothesis testing for network models. The time complexity for a network can be reduced by a divisive factor of the order k^2 by using subsampling, where k is the number of subnetworks. Detailed time complexity analysis, consistency theorems for parameter estimation and hypothesis testing in RDPG framework are provided. Additionally, we extend all the methods for a class of network models containing RDPG and develop the corresponding consistency results. Finally, we demonstrate the validity of our methods by experimenting on various simulation setups as well as real data examples. The experiments indicate considerable runtime improvements for the subsampling based methods over the original method.

63. Bayesian group-shrinkage based estimation in panel VAR models with mixed frequency data

[27.A2.C7, (page 22)]

Nilanjana CHAKRABORTY, *nchakraborty@ufl.edu*
Kshitij KHARE, *University of Florida*
George MICHAILIDIS, *University of Florida*

Panel vector auto-regressive models are popular and effective tools for modeling the evolution of multivariate time series (with an identical set of variables) across different regions/sub-populations. A key objective in this setting is to link the region-specific VAR models through appropriate homogeneity restrictions on their transition matrices to borrow strength from the common features, but at the same time providing enough flexibility for regional idiosyncrasies. For macroeconomic data, this challenge is often further enhanced by the fact that some variables are observed at a different frequency than others, and panel VAR literature for this mixed frequency setting is very sparse. In this paper, we develop a Bayesian approach for mixed frequency panel VAR models that uses group shrinkage to borrow strength across regions and to tackle parameter proliferation. Existing Bayesian approaches for linking region-specific VARs try to fuse relevant coefficients of regional VAR transition matrices to a common value, while we employ a less stringent generalized hierarchical group-lasso prior based on groups of coefficients. This prior distribution allows us to incorporate and learn the interdependence structure between various regions through an inter-region covariance (matrix) parameter. Extensive performance evaluation on simulated data and on a motivating Eurozone macroeconomic dataset illustrates the efficacy of the proposed methodology.

64. Recent Advances in Data Integration and Distributed Computing

[27.M1.I14, (page 11)]

Sounak CHAKRABORTY, *University of Missouri*
Tanujit DEY, *Harvard University*
Anjishnu Banerjee BANERJEE, *Medical College of Wisconsin*
Yilun HUANG, *University of Missouri*

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.

65. Regulatory Submissions with Real world data sources - Key challenges and opportunities

[27.A1.I25, (page 18)]

Arunava CHAKRAVARTTY, *Novartis*

Real-world evidence (RWE), derived from data from real-world sources is an increasingly important source of evidence that holds great potential to increase efficiency and improve clinical development and life cycle management of medical products. Regulatory agencies, have considered the use of the RWD both as a primary source of data for a new investigational product or as a source of information for contextualization and comparison as an external control. Furthermore additional considerations apply when the real world data relates to time to event variables such as differences in assessment frequencies and determination of time zero as these impact the statistical interpretation of the evidence.

In this talk we two case studies that outline some the most commonly encountered challenges and methods for the use of RWE for medical product development specific to these settings.

66. A framework for modelling multiplex networks

[29.M2.I63, (page 43)]

Swati CHANDNA, *Birkbeck, University of London*

Consider the setting where multiple networks are observed on the same set of nodes, also known as multiplex networks. Such data may arise in the form of networks observed over time, e.g., friendship networks, or from different subjects at a given point in time, e.g. structural brain networks from multiple individuals. Given such data, it is crucial to quantify dependence between pairs of networks and have a framework which allows for the empirical description of correlated networks. We describe an approach to modelling multiplex networks with two layers and show how it leads to simpler models that generate correlated networks. This is joint work with S. Janson and S.C. Olhede.

67. Hierarchical Modeling of Covariance Structure via Subspace Factor Analysis

[28.M1.I40, (page 25)]

Noirrit Kiran CHANDRA, *The University of Texas at Dallas*
Jason XU, *Duke University*
David B. DUNSON, *Duke University*

Factor analysis provides a canonical framework for modeling of lower-dimensional structure in high-dimensional data, inducing a parsimonious covariance structure model. In many contexts, high-dimensional data are collected under different conditions, and there is interest in modeling shared versus group-specific structure in the data. With this motivation, various hierarchical extensions of factor analysis have been proposed, but these approaches face practical issues including identifiability of the shared versus group-specific components. To address these problems, we propose a class of SUbspace Factor Analysis (SUFA) models, which characterize variation across groups at the level of a lower-dimensional subspace. We prove that the proposed class of SUFA models lead to identifiability of the shared versus group-specific components of the covariance. In addition, taking a Bayesian approach to inference, we develop efficient posterior computational algorithms with complexity not dependent on sample size. We illustrate the methods through application to integration of multiple gene expression datasets relevant to immunology.

68 . Inference of dependent competing risks from Marshall-Olikin bivariate Burr-XII distribution under generalized progressive hybrid censoring

[29.M1.C14, (page 40)]

PRAKASH CHANDRA, *Indian Institute of Technology, Patna*
Prakash CHANDRA, *Indian Institute of Technology, Patna*
Yogesh Mani TRIPATHI, *Indian Institute of Technology, Patna*

Competing risk models have attracted much attention among researchers. Such models deal with more than one failure causes involved in product life cycle. Inference under dependent competing risk models has found wide applications. The Marshal-Olikin bivariate life distribution introduced by Marshall and Olikin is widely used in such inference. Samanta and Kundu derived inference for a competing risk model based on the Marshall-Olikin bivariate Weibull distribution. We have considered Marshal-Olikin bivariate Burr XII distribution (MOBBXII) as introduced by Tain and Gui and obtain useful inferences under generalized progressive hybrid censoring scheme. We estimate unknown model parameters of the MOBBXII distribution using frequentist and Bayesian procedures. The highest posterior density intervals are discussed as well. Monte Carlo simulations are implemented to compare different methods, and finally a real data set representing the blindness time of eyes during a diabetic retinopathy study is analysed.

69. Estimating Prevalence of Post-war Health Disorders based on Capture-recapture experiment with Heterogeneous Catchability and Behavioural Dependence

[29.M1.C15, (page 40)]

Kiranmoy CHATTERJEE, *Department of Statistics, Bidhannagar College, Kolkata*
Prajamitra BHUYAN, *Assistant Professor, Indian Institute of Management Kolkata*
Kiranmoy CHATTERJEE, *Assistant Professor, Department of Statistics, Bidhannagar College, Kolkata*

Effective surveillance on the long-term public health impact due to war or terrorist attacks remains limited. Such health issues are commonly under-reported, specifically for a large group of individuals. For this purpose, efficient estimation of the size of the population under the risk of physical and mental health hazards is of utmost necessity. In this context, multiple system estimation is a potential strategy that has recently been applied to quantify under-reported events allowing heterogeneity among the individuals and dependence between the sources of information. To model such complex phenomena, a novel trivariate Bernoulli model is developed, and an estimation methodology using Monte Carlo based EM algorithm is proposed and it successfully overcomes the identifiability issue present in the model. Simulation results show superiority of the performance of the proposed method over existing competitors and robustness under model mis-specifications. The method is applied to analyze real case studies on the Gulf War and 9/11 Terrorist Attack at World Trade Center, US. Estimates of the incident rate and survival rate are computed by adjusting the under-count estimates for an unbiased evaluation of the post-war syndromes. The results provide interesting insights that can assist in effective decision-making and policy formulation for monitoring the health status of post-war survivors.

70 . Scalable Data Integration in Genome-wide Association Studies (GWAS) through Generalized Method of Moments

[28.M1.I38, (page 25)]

Nilanjan CHATTERJEE, *Johns Hopkins University*
Prosenjit KUNDU, *Johns Hopkins University*

We have shown earlier (Kundu et al., Biometrika, 2019) that generalized method of moments (GMM) can provide a unified framework for building complex models through integration of information across multiple disparate data sources. In this work, we consider data integration in the context of large-scale genetic association studies which requires parallel analysis of hundreds of thousands or even millions of genetic variants. We now develop score-tests and one-step optimization techniques within the GMM framework for scalable hypothesis testing and fast parameter estimation for the analysis of GWAS scale datasets. We will show how this general framework can be useful for data integration across individual level data and external GWAS summary-statistics to carry out a whole set of cutting-edge applications which requires incorporation of non-genetic covariate into the underlying models for data analysis.

71. Spatially Adaptive Online Prediction of Piecewise Regular Functions

[29.M2.I60, (page 42)]

Sabyasachi CHATTERJEE, *Assistant Professor*

Subhajit GOSWAMI, *TIFR, Bombay*

We consider the problem of estimating piecewise regular functions in an online setting, i.e., the data arrive sequentially and at any round our task is to predict the value of the true function at the next revealed point using the available data from past predictions. We propose a suitably modified version of a recently developed online learning algorithm called the sleeping experts aggregation algorithm. We show that this estimator satisfies oracle risk bounds simultaneously for all local regions of the domain. As concrete instantiations of the expert aggregation algorithm proposed here, we study an online mean aggregation and an online linear regression aggregation algorithm where experts correspond to the set of dyadic subrectangles of the domain. The resulting algorithms are near linear time computable in the sample size. We specifically focus on the performance of these online algorithms in the context of estimating piecewise polynomial and bounded variation function classes in the fixed design setup. The simultaneous oracle risk bounds we obtain for these estimators in this context provide new and improved (in certain aspects) guarantees even in the batch setting and are not available for the state of the art batch learning estimators.

72 . Recent Advancements in Alzheimer's Clinical Trial Analyses

[29.M1.I57, (page 39)]

Saptarshi CHATTERJEE, *Eli Lilly and Co.*

The mixed model for repeated measures (MMRM) [1] is the traditional approach for analyzing the longitudinal outcomes of clinical trials in progressive diseases like Alzheimers. With MMRM, the scheduled visit time is treated as categorical and with minimal assumptions of the mean trajectory it offers a flexible estimation of the longitudinal profile for the treatment groups over time. However, due to a saturated specification, the model requires many parameters that often entails conservative estimation of the treatment effect. Furthermore, MMRM typically focuses on the treatment contrast at the last visit to quantify the treatment effect that does not depend on the previous follow-ups and in situations like a progressive disease this quantification might not be appropriate. The primary focus of this talk is to present recent advancements on alternative models that can be utilized to overcome the limitations of MMRM. A version of the disease progression model (DPM) [2] will be presented that can estimate the overall slowing in disease progression if a drug has cumulative effect over time resulting in proportional efficacy. The natural cubic spline model [3] is another alternative to MMRM that treats time as continuous and can be used to flexibly model the longitudinal trend without sacrificing the robust specification of the variance covariance matrix. A few simulation studies will be presented to compare these models and highlight the benefits and limitations of the alternative longitudinal models over MMRM from the perspective of the trial operating characteristics.

References:

1. Mallinckrodt, C. H., Clark, W. S., Carroll, R. J., Molenberghs, G. (2003). Assessing response profiles from incomplete longitudinal clinical trial data under regulatory considerations. Journal of biopharmaceutical statistics,13(2), 179-190.

2. Wang, G., Berry, S., Xiong, C., Hassenstab, J., Quintana, M., McDade, E. M., ... Dominantly Inherited Alzheimer Network Trials Unit. (2018). A novel cognitive disease progression model for clinical trials in autosomaldominant Alzheimer's disease. Statistics in medicine, 37(21), 3047-3055.

3. Donohue, M. C., Langford, O., Insel, P., van Dyck, C. H., Petersen, R. C., Craft, S., ... Aisen, P. S. (2022). Natural cubic splines for the analysis of Alzheimer's clinical trials. arXiv preprint arXiv:2208.08077.

73 . Aspects of Community Detection For Networks In the Presence Of Complexities

[29.M2.I63, (page 43)]

Shirshendu CHATTERJEE, *City University of New York, Graduate Center & City College*

We consider the problem of identifying the common communitystructures for multilayer networks. Several complexities (e.g., correlation among different layers, assortativity patterns of the communitystructures, sparsity) can make the task of accurate community detection quite tricky.We consider certain extensions and modifications of the spectral clustering methods and show that such methods, unlike many other existing methods, performconsistently for multilayerstochastic block models and multilayer degree-corrected block models in the presence of many such complexities.

74. Nonparametric Hypothesis Testing in High Dimensions and GWAS Applications

[27.M1.I17, (page 12)]

Snigdhansu CHATTERJEE, *University of Minnesota*

High-dimensional data, where the dimension of the feature space is much larger than sample size, arise in a number of statistical applications. In this context, we present the generalized multivariate sign transformation, defined as a vector divided by its norm. For different choices of the norm function, the resulting transformed vector adapts to certain geometrical features of the data distribution. We obtain one-sample and two-sample testing procedures for mean vectors of high-dimensional data using these generalized sign vectors. These tests are based on U-statistics using kernel inner products, do not require prohibitive assumptions, and are amenable to a fast randomization-based implementation. Theoretical developments, simulated data and real data examples are discussed. A genome wide association study (GWAS) on the Minnesota Center for Twin and Family Research data is reported.

75. RONSeq: A Robust Non-Parametric Method for Detecting Differential Gene Expression in RNA-Sequencing Studies

[29.M1.I55, (page 38)]

Suvo CHATTERJEE, *Indiana University, Bloomington*
Arindam FADIKAR, *Argonne National Laboratory*

The current practice to detect differential expressed (DE) genes in bulk RNA-Seq studies involve applying parametric models, typically based on the negative binomial (NB) distribution. Accumulating evidence suggests that sequence count data is inadequately fit by a NB model. Many DE models also assume that genes have equal variances across experimental conditions, which is often not valid in practice. To address this, we introduce RONSeq, a fully nonparametric model that allows flexible modeling of genes when homoscedasticity is violated. Our formulation is based on i) M-estimators to robustly estimate fold changes in gene expression; ii) sandwich estimators to robustly estimate standard errors under heteroscedasticity; iii) a nonparametric resampling scheme to detect DE genes. Through synthetic benchmarking and real data analyses, we show that RONSeq leads to efficient FDR control, yielding substantial power gain as compared to published methods in detecting biologically relevant genes. RONSeq can be applied to other similar data types such as microbiome and single-cell, thus providing a unified framework for diverse omics data types. Our open-source R package is publicly available.

76 . Two-Stage Principal Component Analysis

[29.A2.I72, (page 47)]

Bhargab CHATTOPADHYAY, *Indian Institute of Technology Jodhpur*
Swarnali BANERJEE, *Loyola University Chicago*

Online Principal Component Analysis is a popular multivariate analytic tool that can be used for dimension reduction without losing much information for sequentially arriving data. Existing online PCA research works revolve around proposing efficient, scalable updating algorithms focusing on compression loss only and do not take into account the sample size. It is well known that the sample size impacts the compression loss - the smaller the sample size, the larger the compression loss, and vice versa. However, reducing compression loss by increasing sample size will increase the total sampling cost. This work focuses on optimizing a cost-compression loss that considers the compression loss and data collection cost using a two-stage PCA algorithm. The resulting two-stage algorithm is a fast and efficient alternative to sequential PCA and exhibits attractive convergence properties with no assumption on specific data distribution.

77. Nonparametric smooth estimation of survival and density function for length biased data: An overview

[30.M1.I77, (page 51)]

Yogendra CHAUBEY, *Concordia University*
Yogendra CHAUBEY,

Length biased data often arise in observational studies, when the observations are selected from the population with probability proportional to their length (Cox and Miller, 1965; Zelen and Feinlab, 1969; Vardi 1982). Such data entail constraints that many traditional nonparametric density estimators may not satisfy. Here we consider non-parametric density estimator based on the smoothing of the Cox's (1969) estimator. This talk will highlight some recent work in the area of nonparametric density estimation for length biased data.

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

[27.M2.I21, (page 16)]

Sanjay CHAUDHURI, *National University of Singapore*
Subhroshekhar GHOSH, *National University of Singapore*
Thi Kim Cuc PHAM, *National University of Singapore*

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

79 . Imputation-Based Q-Learning for Optimizing Dynamic Treatment Regimes with Right-Censored Survival Outcome

[26.A2.I9, (page 6)]

Yu CHENG, *University of Pittsburgh*
Lingyun LYU, *University of Pittsburgh*
Abdus WAHED, *University of Pittsburgh*

Q-learning has been one of the most commonly used methods for optimizing dynamic treatment regimes (DTRs) in multi-stage decision making. Right-censored survival outcome poses a significant challenge to Q-Learning due to its reliance on parametric models for counterfactual estimation, which are subject to misspecification and sensitive to missing covariates. Limited works reported for estimating optimal DTRs for survival outcomes with Q-learning mostly used parametric models. In this talk, we propose an imputation-based Q-learning (IQ-learning) where flexible nonparametric or semiparametric models are employed to estimate optimal treatment rules for each stage, and then weighted hot-deck multiple imputation (MI) and direct-draw MI are used to predict optimal potential survival times. Missing data are handled using inverse probability weighting and MI, and the non-random treatment assignment among the observed is accounted for using a propensity-score approach. We investigate the performance of IQ-learning via extensive simulations and show that it is more robust to model misspecification than existing Q-Learning methods, imputes only plausible potential survival times contrary to parametric models, and provides more flexibility in terms of baseline hazard shape. We demonstrate IQ-Learning by developing an optimal DTR for leukemia treatment based on a randomized trial with observational follow-up.

80 . Statistical Planning for Non-Proportional Hazards in Design and Analysis of Cancer Clinical Trials

[27.M1.C5, (page 14)]

Suresh CHENJI, *Eli Lilly*

Commonly cancer drug trials have time-to-event endpoints overall survival (OS) and progression-free survival (PFS) as primary outcome measures, designed, and analysed based on key assumption of proportional hazards (PH), the event hazard rate can change over time, but that the ratio of event hazards between two individuals is constant. The log-rank test, the Kaplan Meier survival plot and the Cox-proportional hazards model are standard statistical approaches for these time-to-event endpoint analyses depend on the PH assumption.

In recent years, cancer trials particularly immuno-oncology trials often encounter non-proportional hazards (NPH) scenarios where the proportionality assumption does not hold true, which may include the examples of crossing hazards, diminishing treatment effect and delayed treatment effect. In these situations, standard statistical methods do not provide valid treatment effect estimates and summary results.

Duke-FDA and cross-pharma working group (2018) discussed various alternative statistical methods for evaluating treatment effects of time-to-event endpoints under different NPH scenarios, like three categories of testing methods, including weighted log-rank tests, KaplanMeier curve-based tests (including weighted KaplanMeier and restricted mean survival time), and combination tests (including Breslow test, Lees combo test, and MaxCombo test). Under non-proportional hazards loss of power may be an issue and for the treatment effect changes over time, multiple approaches may need to be used to comprehend the overall treatment effect. For this talk we will use relevant case examples and new simulations to discuss various recommendations for design (power, sample size and follow-up time), primary and sensitivity approaches which can be prespecified when NPH is plausible

81. Infant Mortality Rate: A Demographic and Socio Economic Assessment of the Empowered Action Group (EAG) States of India

[28.A2.C12, (page 36)]

Anuradha Rajkonwar CHETIYA, *Ramjas College, University of Delhi*
Vishal DEO, *Ramjas College, University of Delhi*

Infant Mortality Rate (IMR) has been on a decline in India over the past decades. However, not only has the rate of decline been much lower than anticipated, but there exists a striking disparity in the IMR of different states even in 2020, with some states performing significantly better than others. Kerala, for example, has had a consistently low IMR ranging from 12 in 1000 live births in 1997 to 6 in 1000 live births in 2020. On the other hand, states like Madhya Pradesh and Uttar Pradesh have not fared very well in terms of reducing their IMRs. Madhya Pradesh had the highest IMR in 1997 at 94 per 1000 live births and continues to record the highest IMR in 2020 as well at 43 per 1000 live births.

Statistics after almost 45 years of implementation of various policies and programs clearly indicate that attainment of an equitable distribution in public health policies still remain a distant target as far as the IMR is concerned. Recognizing this disparity, eight states of India that come under the umbrella of the Empowered Action Group (EAG) have been given special focus in government health policies owning to the high IMR in these states.

The objective of this study is to assess the effects of different demographic and socio economic factors and their direct and spatial impact on child mortality in a cross sectional set up with respect to the EAG states. A Generalized Linear Model (GLM) has been used to assess the impact of factors, like female literacy (percentage of females who have completed 15 years of education), shortage of health centres and the per capita state domestic products (SDP), on IMR. Results indicate significant impact of these factors on IMR of the EAG states.

82. EEMD based LSTM hybrid model for Agricultural Price Forecasting

[29.M2.I64, (page 43)]

Kapil CHOUDHARY, *ICAR-IASRI, New Delhi*
Girish Kumar JHA, *ICAR-IARI, New Delhi*
Ronit JAISWAL, *ICAR-IASRI, New Delhi*
Rajeev R. KUMAR, *ICAR-IASRI, New Delhi*

Abstract Agricultural commodities prices are very unpredictable and complex, and thus, forecasting these prices is one of the research hotspots. In this paper, a data-adaptive decomposition method i.e., ensemble empirical mode decomposition (EEMD) is integrated with an efficient forecasting tool i.e., long short-term memory (LSTM) to develop a unique hybrid EEMD-LSTM model intending to increase the precision of agricultural price forecasting. Specifically, EEMD disintegrates a price series into a set of stable subseries that are independently modelled and forecasted using LSTM models. Finally, the actual forecast of the original series is obtained by merging the predictions of each stable series. EEMD overcomes the limitation of the mode mixing and end effect problems of the empirical mode decomposition (EMD). The prediction ability of the proposed model is compared with LSTM, and EMD-based hybrid LSTM using groundnut oil and soybean price series in terms of root mean square error, mean absolute percentage error, mean absolute error and, directional prediction statistics. Additionally, the Diebold-Mariano test is used to assess the performance of the models whose results confirm that the proposed hybrid model improves the prediction accuracy significantly. Keywords: Agricultural price forecasting; Ensemble empirical mode decomposition; Intrinsic mode function; Long short-term memory

83. Multi-sample comparison using spatial signs for infinite dimensional data

[28.M2.I47, (page 30)]

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

Probal CHAUDHURI, *Indian Statistical Institute, Kolkata*

We consider an analysis of variance type problem, where the sample observations are random elements in an infinite dimensional space. This scenario covers the case, where the observations are random functions. For such a problem, we propose a test based on spatial signs. We develop an asymptotic implementation as well as a bootstrap implementation and a permutation implementation of this test and investigate their size and power properties. We compare the performance of our test with that of several mean based tests of analysis of variance for functional data studied in the literature. Interestingly, our test not only outperforms the mean based tests in several non-Gaussian models with heavy tails or skewed distributions, but in some Gaussian models also.

84. Optimal Sampling Plan under Accelerated Life Test Setting

[27.A1.I29, (page 19)]

Shovan CHOWDHURY, *Indian Institute of Management Kozhikode*

.

85. A-ComVar: A Flexible Extension of Common Variance Designs

[30.M2.I80, (page 53)]

Shrabanti CHOWDHURY, *Icahn School of Medicine at Mount Sinai*
Joshua LUKEMIRE, *Emory University*
Abhyuday MANDAL, *University of Georgia*

We consider non-regular fractions of factorial experiments for a class of linear models. These models have a common general mean and main effects; however, they may have different 2-factor interactions. Here we assume for simplicity that 3-factor and higher-order interactions are negligible. In the absence of a priori knowledge about which interactions are important, it is reasonable to prefer a design that results in equal variance for the estimates of all interaction effects to aid in model discrimination. Such designs are called common variance designs and can be quite challenging to identify without performing an exhaustive search of possible designs. In this work, we introduce an extension of common variance designs called approximate common variance or A-ComVar designs. We develop a numerical approach to finding A-ComVar designs that is much more efficient than an exhaustive search. We present the types of A-ComVar designs that can be found for different number of factors, runs, and interactions. We further demonstrate the competitive performance of both common variance and A-ComVar designs using several comparisons to other popular designs in the literature.

86. Transmuted Cure Models- A Competing Risk Perspective

[Poster Session, (page 33)]

Soorya C. S., *Department of Statistics, Cochin University of Science and Technology*
Dr. Asha Gopalakrishnan ,

A cure rate model under a competing risk scenario where the number of competing causes follow a shifted binomial distribution with parameter 'p' is proposed. Interestingly, the resulting distribution is the well-studied transmuted class of distribution. Few existing cure rate models are shown to be special cases of the proposed model. The identifiability issues of the model are studied in detail. Further properties of the model are investigated, and we discuss the maximum likelihood estimation of the parameter. The performance is confirmed through a simulation study using a defective Gompertz baseline and with competing causes. The Bayesian approach to the estimation of the parameter is adopted. The complexity of the likelihood function is handled through the Metropolis-Hastings algorithm. We analyse the data consisting of 8966 patients who have undergone bone marrow transplantation at the European Society for Blood and Marrow Transplantation (EBMT). The validation of the estimation algorithm is confirmed using the bootstrap technique.

87 . Estimating heterogeneous treatment effects with right-censored data via causal survival forests

[28.M1.I39, (page 25)]

Yifan CUI, *Zhejiang University*

Forest-based methods have recently gained in popularity for non-parametric treatment effect estimation. Building on this line of work, we introduce causal survival forests, which can be used to estimate heterogeneous treatment effects in a survival and observational setting where outcomes may be right-censored. Our approach relies on orthogonal estimating equations to robustly adjust for both censoring and selection effects under unconfoundedness. In our experiments, we find our approach to perform well relative to a number of baselines.

88. Asymmetric Rotatable Designs for Fitting Response Surface Model Incorporating Neighbour Effects

[27.A2.C8, (page 23)]

ASHUTOSH DALAL, *ICAR-INDIAN AGRICULTURAL STATISTICS RESEARCH INSTITUTE*
SEEMA JAGGI, *Indian Council of Agricultural Research, Education Division*
ELDHO VARGHESE, *ICAR-Central Marine Fisheries Research Institute*

Response Surface Methodology (RSM) is an experimental approach for investigating the relationship between the response variable and the input variables and for creating a suitable relationship between them. It is employed when many input variables have an impact on a process's performance metric or quality attribute. When experimental units are set up linearly with no spacing in an RSM trial, the independence of observations cannot be guaranteed. For instance, in field experiments, if a chemical spray is put into one plot, the wind may cause it to spread to the adjacent plots. As a result, when fitting RSM, it is necessary to account for the influence that effects from nearby plots may have on a given plot's response. This paper describes the response surface model for factors with asymmetric levels where experimental units are affected by the nearby neighbouring units. Different methods of constructing asymmetric response surface designs have been discussed, incorporating neighbour effects into the model, which satisfies the rotatability conditions. These designs result in more precise estimates of the parameters of the response surface model. An R package named "rsdNE" has also been developed for generating these types of designs.

89. Partial divisibility in random sets

[30.M2.I81, (page 53)]

Biltu DAN, *Indian Institute of Science*
Jnaneshwar BASLINGKER, *Indian Institute of Science*
Manjunath KRISHNAPUR , *Indian Institute of Science*

The distribution of a random closed set is determined by its void functional. We study the set of non-negative α such that the α-th power of a void functional is again a void functional. This is in parallel with the study of Hadamard powers of positive semi-definite matrices. Also, we study the problem of approximating n-divisible random closed sets by infinitely divisible random closed sets. This is in parallel with the study of approximating sums of i.i.d. random variables by infinitely divisible random variables.

This talk is based on an ongoing joint work with J. Baslingker and M. Krishnapur.

90. Milestones in Personalized Medicine An overview on subgroup identification and some potential directions

[Plenary Lecture 1, (page 3)]

Kalyan DAS, *kalyanstat@gmail.com*
Ujjwal DAS, *OM,QM & IS Area IIM Udaipur*

Milestones in Personalized Medicine An overview on subgroup identification and some potential directions

Kalyan Das and Ujjwal Das

In the fight against hard-to treat diseases like cancer, it is often difficult to find a new treatment that benefits all patients. Scientific advances in the past two decades have transformed the treatment of several types of cancers and led to the design of targeted therapies based on individualized genomic profiles for lung cancer, breast cancer, and melanoma. The development of tailored therapies, often referred to as personalized or stratified medicine, is all set to revolutionize the drug development. The emergence of personalized strategies has embraced a wider range of biomedical fields and cross-domain expertise, bridging the bench-to-bed gap toward the fulfilment of integrated patient profiling, disease modelling and targeted drug design. With the central goal to customize medical care and cures, the focus of Personalized medicine is to provide right drug at the right dose to the right patient based on a large number of patient characteristics like several clinical features and biomarkers. The statistical techniques are essential to provide effective tools for construction and evaluation of evidence-based personalized intervention strategies. The focus of this article is to provide a review of this vibrant area of statistical research on personalized medicine.

In the present discussion, we stress upon the development in the direction of tracking the right group of patients given a specific treatment. This leads to appropriate subgroup identification. Alternatively, one can also look at the progress on finding the right treatment– often referred to as optimal treatment regimes (OTR) for a specific patient.

Following the review, we propose a subgroup iden-

tification technique for clustered data based on some supervised learning approach. The performance of the proposed technique has been assessed using simulation studies.

91. A Bayesian joint model for multivariate longitudinal and event time data with application to ALL maintenance studies

[29.M2.I62, (page 43)]

Kiranmoy DAS, *Indian Statistical Institute, Kolkata*

The most common type of cancer diagnosed among children is the Acute Lymphocytic Leukemia (ALL). A study was conducted by Tata Translational Cancer Research Center (TTCRC) Kolkata, in which 236 children (diagnosed as ALL patients) were treated for the first two years (approximately) with two standard drugs (6MP and MTx), and were then followed nearly for the next three years. The goal is to identify the longitudinal biomarkers that are associated with time to relapse, and also to assess the effectiveness of the drugs. We develop a Bayesian joint model in which a linear mixed model is used to jointly model three biomarkers, i.e. white blood cell count, neutrophil count, and platelet count; and a semi-parametric proportional hazards model is used to model relapse time. Our proposed joint model can assess the effects of different covariates on the biomarkers, and the effects of the biomarkers (and the covariates) on relapse time. The proposed model addresses some clinically important research questions. Specifically, we infer that white blood cell count is not associated with relapse time, but neutrophil count and platelet count are significantly associated. We also infer that a lower dose of 6MP and a higher dose of MTx jointly result in a lower relapse probability in the follow-up period. Interestingly, we find that relapse probability is the lowest for the patients classified into "high-risk" group at presentation. The effectiveness of the proposed joint model is assessed through extensive simulation studies.

92. Estimating the optimal linear combination of predictors using spherically constrained optimization

[27.M1.I14, (page 11)]

Priyam DAS, *Department of Biostatistics, Virginia Commonwealth University*
Debsurya DE, *Indian Statistical Institute*
Raju MAITI, *Indian Statistical Institute*
Christine PETERSON, *University of Texas MD Anderson Cancer Center*

In the context of a binary classification problem, the optimal linear combination of continuous predictors can be estimated by maximizing the area under the receiver operating characteristic curve. For ordinal responses, the optimal predictor combination can similarly be obtained by maximization of the hypervolume under the manifold (HUM). Since the empirical HUM is discontinuous, non-differentiable, and possibly multi-modal, solving this maximization problem requires a global optimization technique. Estimation of the optimal coefficient vector using existing global optimization techniques is computationally expensive, becoming prohibitive as the number of predictors and the number of outcome categories increases. We propose an efficient derivative-free black-box optimization technique based on pattern search to solve this problem, which we refer to as Spherically Constrained Optimization Routine (SCOR). Through extensive simulation studies, we demonstrate that the proposed method achieves better performance than existing methods including the step-down algorithm. Finally, we illustrate the proposed method to predict the severity of swallowing difficulty after radiation therapy for oropharyngeal cancer based on radiation dose to various structures in the head and neck. Our proposed method addresses an important challenge in combining multiple biomarkers to predict an ordinal outcome. This problem is particularly relevant to medical research, where it may be of interest to diagnose a disease with various stages of progression or a toxicity with multiple grades of severity.

93. On Distance to Default for Unlisted Firm using Hierarchical Bayesian Models

[26.A2.I12, (page 7)]

Sourish DAS, *CMI*
Susan THOMAS, *XKDR Research Forum*

Merton's Distance to Default (aka. DtD) works very well for firms listed with stock exchanges. The DtD is a good measure of the health of the firm. If available, the DtD is a better estimate of the firm's health than the probability of default (PD). However, based on data from the Center for Monitoring Indian Economy (CMIE), more than 97% of all Indian firms are not listed with any stock exchanges. Therefore for most of the firms, we cannot estimate the DtD.

This paper presents a methodology to estimate the DtD for unlisted firms using the hierarchical Bayesian models. We present the case study of Air India vs Jet Airways.

94 . SOFTWARE RELIABILITY BASED ON RENEWAL PROCESS MODELING FOR ERROR OCCURRENCE DUE TO EACH BUG WITH PERIODIC DEBUGGING SCHEDULE

[29.A2.I74, (page 48)]

Sudipta DAS, *Ramakrishna Mission Vivekananda Educational and Research Institute*
ANUP DEWANJI, *Applied Statistics Unit, Indian Statistical Institute, Kolkata, India*
SUBRATA KUNDU, *George Washington University, Washington, DC, USA*

The process of software testing usually involves the correction of a detected bug immedi- ately upon detection. In this article, in contrast, we discuss continuous time testing of a software with periodic debugging in which bugs are corrected, instead of at the instants of their detection, at some pre-specified time points. Under the assumption of renewal distribution for the time between successive occurrence of a bug, maximum-likelihood estimation of the initial number of bugs in the software is considered, when the renewal distribution belongs to any general parametric family or is arbitrary. The asymptotic prop- erties of the estimated model parameters are also discussed. Finally, we investigate the finite sample properties of the estimators, specially that of the number of initial number of bugs, through simulation.

95. Geometric approaches to assessing the numerical feasibility for conducting matching-adjusted indirect comparisons in clinical trials

[27.A1.I25, (page 18)]

Kohinoor DASGUPTA, *Novartis Healthcare Pvt Ltd*

Matching-adjusted indirect comparison (MAIC) is a propensity score matching method. It leverages individual patient data (IPD) from one or more internal clinical studies and adjusts their average patient characteristics to match to published results (or aggregated data (AD)) from external studies for which IPD are not available. Each patient in each of the IPD studies is assigned a weight, measuring the patient's propensity of being in the AD study. One key assumption is that the IPD and AD studies are from the same target patient population.

We discuss how to handle MAIC from a data analyst's perspective for assessing the common population assumption. We introduce several multivariate data analysis methods that focus on comparing the baseline variables used in the matching between the IPD and the AD studies. The methods identify situations when no numerical solutions are possible with the MAIC method.

Moreover, we identify an intrinsic property of MAIC where the estimated propensity weights increase along the steepest ascent of the multidimensional space of the matching variables. This has led us to an alternative way of assigning weights to the IPD data.

We demonstrate the approaches with motivating examples. The methods are straightforward to implement especially with the companion R package maicChecks. These methods help data analysts to best utilize the strength of MAIC, and to avoid generating misleading results when the method should not be used.

96. Bayesian validation framework for dynamic epidemic models

[27.M1.I18, (page 13)]

Sayan DASGUPTA, *Fred Hutchinson Cancer Center*
Mia MOORE, *Fred Hutchinson Cancer Center*
Dobromir DIMITROV, *Fred Hutchinson Cancer Center*
James HUGHES, *University of Washington, Seattle*

Complex models of infectious diseases are used to understand the transmission dynamics of the disease, project the course of an epidemic, predict the effect of interventions and/or provide information for power calculations of community level intervention studies. However, there have been relatively few opportunities to rigorously evaluate the predictions of such models till now. Indeed, while there is a large literature on calibration (fitting model parameters) and validation (comparing model outputs to data) of complex models based on empirical data, the lack of uniformity in accepted criteria for such procedures for models of infectious diseases has led to simple procedures being prevalent for such steps. However, recently, several community level randomized trials of combination HIV intervention have been planned and/or initiated, and in each case, significant epidemic modeling efforts were conducted during trial planning which were integral to the design of these trials. The existence of these models and the (antic-

ipated) availability of results from the related trials, provide a unique opportunity to evaluate the models and their usefulness in trial design. In this talk, we outline a framework for evaluating the predictions of complex epidemiological models and describe experiments that can be used to test their predictions.

97. Application of Statistical Machine Learning in Biomarker Selection

[26.A2.I10, (page 7)]

Shibasish DASGUPTA, *Pfizer and Chennai Mathematical Institute (CMI)*

With the increasing availability of large-scale biomedical data, researchers are offered with unprecedented opportunities to discover novel biomarkers for clinical outcomes. At the same time, they are also faced with great challenges to accurately identify important biomarkers from numerous candidates, specifically in high-dimensional settings. Many novel statistical machine learning methodologies have been developed to tackle these challenges in the last couple of decades. When the clinical outcome is time-to-event data, special statistical machine learning algorithms are needed to analyze this type of data due to the presence of censoring. This presentation will focus on some of the most used modern statistical machine learning methodologies for variable selection for time-to-event data with respect to a real data based on a clinical trial.

98. Designing Factorial Experiments for Cropping Sequence Experiments

[28.A2.C11, (page 36)]

Sukanta DASH, *ICAR-Indian Agricultural Statistics Research Institute*
Ankit SINGH, *ICAR-Indian Agricultural Statistics Research Institute*
Rajender PARSAD, *ICAR-Indian Agricultural Statistics Research Institute*
Anil KUMAR, *ICAR-Indian Agricultural Statistics Research Institute*

Cropping Systems Research often consider crop sequences comprising of two or more crops grown in the respective cropping seasons instead of a mono crop. In these types of experiment, not only estimation of main effects and interaction effects are important but also residual effect of treatment of one season have important and significant effect on crop of next upcoming season. This article proposes different methods for constructing block designs for factorial experiments with block size two and unblock designs for fractional factorial experiments for estimation of all main effects and consecutive two factor interactions. Several authors worked in this direction for estimation of all main effects and two factor interactions applied in different experimental situations. While considering cropping sequence experiments generally it is preferred to estimate only two factor interactions between consecutive factors with all main effects but, in these situations no such general method of construction available. Hence, in the present study a general method of construction has been developed with two special situations i.e. (i) consecutive two factor interactions estimated with full efficiency (ii) all the main effects and two factor interactions having equal efficiency. The developed construction methods are easy to implement, capable of estimating all main effects and selected two factor interaction accommodating any number of factors, and with a smaller number of replications comparing to available designs in literature.

99. Machine learning for spatial generalized mixed models

[29.M1.I54, (page 38)]

Abhirup DATTA, *Johns Hopkins University*

Spatial generalized linear mixed-models, consisting of a linear covariate effect and a Gaussian Process (GP) distributed spatial random effect, are widely used for analyses of geospatial data. We consider the setting where the covariate effect is non-linear and propose modeling it using a flexible machine learning algorithm like random forests or deep neural networks. We propose well-principled extensions of these methods, for estimating non-linear covariate effects in spatial mixed models where the spatial correlation is still modeled using GP. The basic principle is guided by how ordinary least squares extends to generalized least squares for linear models to account for dependence. We demonstrate how the same extension can be done for these machine learning approaches like random forests and neural networks. We provide extensive theoretical and empirical support for the methods and show how they fare better than nave or brute-force approaches to use machine learning algorithms for spatially correlated data. We demonstrate the RandomForestsGLS R-package that implements this extension for random forests.

100. GENERALIZED ROW-COLUMN DESIGNS WITH FACTORIAL TREATMENT STRUCTURE

[29.M1.C16, (page 41)]

Anindita DATTA, *Scientist*
Seema JAGGI, *Additional Director General (HRD), ICAR*
Eldho VARGHESE, *Senior Scientist, ICAR-CMFRI*
Cini VARGHESE, *Principal Scientist, ICAR-IASRI*

A generalized row-column design with factorial treatment structure is an arrangement of s^n factorial combinations in p rows and q columns such that the intersection of each row and column receives s^k $(k < n)$ treatment combinations. These designs are used when the number of treatment combinations is substantially large with limited number of replicates. Here, Generalized Confounded Row-Column (GCRC) designs have been introduced and methods of constructing GCRC designs have been discussed. Fractionally replicated GCRC designs have also been constructed. The designs obtained ensure balancing with respect to estimable effects.

101. New Directions in Bayesian Shrinkage for Sparse, Structured Data

[28.M1.I37, (page 24)]

Jyotishka DATTA, *Virginia Tech*
Boss JONATHAN , *University of Michigan Biostatistics*
David DUNSON, *Duke University Statistics*
Matthew HEINER , *Brigham Young University*

Sparse signal recovery remains an important challenge in large scale data analysis and global-local (G-L) shrinkage priors have undergone an explosive development in the last decade in both theory and methodology. In this talk, I will present two recent developments in Bayesian inference using G-L priors for sparse and structured data. In the first half of my talk, I will present a a new prior (called GIGG) for sparse Bayesian linear regression models designed explicitly for predictors with bi-level grouping structure, which generalizes normal beta prime and horseshoe shrinkage. In the second half, I consider high-dimensional compositional data with sparsity or complex dependence structure as routinely observed in many areas including ecology, microbiomics, geology, environmetrics and social sciences, and propose a new prior distribution, specially designed to enable scaling to data with many categories. I will discuss the methodological challenges associated with each of these problems, and briefly review and discuss these recent developments. I will provide some theoretical support for the proposed methods and show improved performance in simulation settings and application to environmentrics and microbiome data.

102. Regression Analysis of a Future State Entry Time Distribution Conditional on a Past State Occupation in a Progressive Multistate Model

[27.M1.I13, (page 11)]

Somnath DATTA, *University of Florida*

We present a nonparametric method for estimating the conditional future state entry probabilities and distributions of state entry time conditional on a past state visit when data are subject to dependent censorings in a progressive multistate model where Markovity of the system is not assumed. These estimators are constructed using the competing risk techniques with risk sets consisting of fractional observations and inverse probability of censoring weights. The fractional observations correspond to estimates of the numbers of persons who ultimately enter a state from which the future state in question can be reached in one step. We then address the corresponding regression problem by combining these marginal estimators with the pseudo-value approach. Performance of our regression scheme is studied using a comprehensive simulation study. An analysis of an existing data on graft-versus-host disease for bone marrow transplant individuals is presented using our novel methodology. A second analysis of another well known data set on burn patients is also included.

103. Single Cell RNA-Seq Data Analysis and its Applications in Health Sciences

[29.M2.I62, (page 43)]

Susmita DATTA, *University of Florida*
Michael SEKULA, *University of Louisville*
Jeremy GASKINS, *University of Louisville*

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

104 . Exact Tests for Offline Change Detection in Multichannel Binary and Count Data

[29.M2.I63, (page 43)]

Shyamal Krishna DE, *Indian Statistical Institute Kolkata*
Soumendu Sundar MUKHERJEE, *Indian Statistocal Institute Kolkata*

In this talk, we first consider offline detection of a single changepoint in binary and Poisson time series. We discuss various approaches such as cumulative sum (CUSUM) and likelihood ratio (LR) based exact tests as well as a new proposal that combines exact conditional two-sample tests with multiplicity correction. We demonstrate empirically that the exact tests are much more powerful than the standard asymptotic tests based on the Brownian bridge approximation of CUSUM statistic, especially when (1) sample size is small, (2) changepoint occurs near boundary, and (3) the data is generated from a distribution with sparse parametric setting. Further, we consider a multi-channel changepoint problem where each channel consists of a binary or Poisson time series. Adopting a False Discovery Rate controlling mechanism, we detect possible changes in multiple channels and show that this local approach is not only more informative but more powerful than multivariate global testing approaches when the number of channels with changepoints is much smaller than the total number of channels. As an application of the proposed methodologies, we detect changes in network-valued time-series with (1) edges as binary channels and (2) node-degrees as count (Poisson) channels.

105. Regularized Estimation of Sparse Spectral Precision Matrices

[29.M1.C15, (page 40)]

Navonil DEB, *Cornell University*
Amy KUCEYESKI, *Cornell University*
Sumanta BASU, *Cornell University*

Spectral precision matrix, the inverse of a spectral density matrix, is an object of central interest in frequency-domain multivariate time series analysis. Estimation of spectral precision matrix is a key step in calculating partial coherency and graphical model selection of stationary time series. When the dimension of a multivariate time series is moderate to large, traditional estimators of spectral density matrices such as averaged periodograms tend to be severely ill-conditioned, and one needs to resort to suitable regularization strategies. In this work, we propose complex graphical lasso (cglasso), an L1-penalized estimator of spectral precision matrix based on local Whittle likelihood maximization. We develop fast pathwise coordinate descent algorithms to implement cglasso for large dimensional time series. We also present a complete non-asymptotic theory of our proposed estimator which shows that consistent estimation is possible in high-dimensional regime as long as the underlying spectral precision matrix is suitably sparse. We illustrate the advantage of cglasso over competing alternatives using extensive numerical experiments on simulated data sets.

106 . A Bayesian approach to identify changepoints in spatio-temporal ordered categorical data

[29.A1.I66, (page 44)]

Soudeep DEB, *Indian Institute of Management Bangalore*
Siddharth RAWAT, *Indian Institute of Management Bangalore*
Candace BERRETT, *Brigham Young University*

Although there is substantial literature on identifying structural changes for continuous spatio-temporal processes, the same is not true for categorical spatio-temporal data. This work bridges that gap and proposes a novel spatio-temporal model to identify changepoints in ordered categorical data. The model leverages an additive mean structure with separable gaussian space-time processes. Our proposed technique is defined in such a way that it can detect a shift in the mean structure as well as in the covariance structures in both the spatial and temporal associations. Our approach's capability to handle ordinal categorical data provides an added advantage from an application perspective. We implement the model through a Bayesian framework, which gives a computational edge over a classical method. For application, we use county-wise COVID-19 data from New York by categorizing the daily cases according to CDC guidelines. Our model is able to identify changepoints in the data and helps in providing interesting insights about the "waves" encountered during the pandemic.

107. INTERMITTENT STATE WISE CANCER SURVIVAL RISK ESTIMATION USING MULTI-STATE FRAILTY MARKOV REGRESSION MODEL

[Poster Session, (page 33)]

DIVYA DENNIS, *Division of Cancer Epidemiology & Biostatistics, Regional Cancer Centre, Thiruvananthapuram, Kerala, India*
Dr. Jagathnath Krishna K M, Dr. Preethi Sara George, Dr. Aleyamma Mathew ,

In cancer survival studies the patients may passes through intermittent states of transition such as recurrences, metastasis, stable disease and death, which affects the survival differently and hence the hazard needs to be estimated simultaneously. Also there may exist hidden random heterogeneity which affects the simultaneous risk prediction. Hence a multi-state frailty model (MFM) is more useful. In the present study we derived a four state multi-state model (MSM) and corresponding MFM. We assumed frailty parameters to follow gamma or log normal distribution. The model parameters were estimated using penalized partial likelihood method. The model identification were done using Akaike information criteria (AIC), Bayesian information criteria (BIC), likelihood cross-validation (LCV) and predictability of the model were accessed using survival concordance (C-index). These derived models were illustrated using breast cancer survival data (covariate considered: composite stage, age at diagnosis, laterality, lymph node involvement, history of metastasis, comorbidity and tumor size at diagnosis). From the present study we observed that, there exists a high risk from state of recurrence to death followed by diagnosed as disease to death for most of the covariates. The multi-state model with log-normal frailty was found to be a better predictive model than the other two models. Key words: Multi-state model (MSM), multi-state frailty models (MFM), Penalized partial likelihood, breast cancer

108 . Method Based on State-Space Epidemiological Model for Cost-Effectiveness Analysis of Non-Medical Interventions for COVID-19

[29.A2.C19, (page 49)]

Vishal DEO, *Ramjas College, University of Delhi*
Gurprit GROVER, *Department of Statistics, Faculty of Mathematical Sciences, University of Delhi*

Non-medical containment measures, like quarantine, lockdown, travel restrictions, physical distancing etc., are paramount towards containing the spread of a novel epidemic, especially at its initial stage when little is known about its transmission dynamics. For these containment measures to be effective, timely identification of infectives through clinical testing is essential. To stress upon the importance of extensive random testing for breaking the chains of transmissions, we have designed a detailed framework for carrying out cost-effectiveness analysis (CEA) of extensive random testing in comparison to targeted testing (the testing policy followed by most countries). This framework can be easily extended to CEA of any other non-medical or even medical interventions for containing epidemics.

A new dynamic version of the basic susceptible-infected-removed (SIR) compartmental model, called the susceptible-infected (quarantined/ free) - recovered- deceased [SI(Q/F)RD] model, has been developed to assimilate the impact of the time-varying proportion of undetected cases on the transmission dynamics of the epidemic. Further, we have presented a Dirichlet-Beta state-space formulation of the SI(Q/F)RD model for the estimation of its parameters using posterior realizations from the Gibbs sampling procedure. The state-space SI(Q/F)RD model, which enables predictions of transmission dynamics in the presence of undetected cases, has been employed to forecast epidemiological parameters under the two scenarios being compared. The health outcomes have been measured in terms of the estimates of total number of deaths, and number of infections prevented because of the intervention. Since long-term generic health state measurement is not involved in this study, utility scores are not required for evaluating health benefits.

As a demonstration, the proposed methodology is applied to the COVID-19 data of California and Florida to carry out CEA of extensive random testing over targeted testing for containing the spread of the epidemic. During the period of the study, these two states were among the worst affected states in the USA, and also had very high percentages of positivity of COVID-19 tests, which indicated possibility of inadequate testing capacity.

Results of the CEA conclude that extensive random testing would be cost-effective over targeted testing if the states of California and Florida were willing to conduct around 9 extra tests and 2 extra tests, respectively, for saving one additional person from getting infected, and 375 more tests and 200 more tests, respectively, for saving one more infected person from dying.

109 . Asymptotic Behaviour of Bonferroni, Holm and Benjamini-Hochberg Procedures under Correlated Normal

[27.A2.C6, (page 22)]

Monitirtha DEY, *Indian Statistical Institute, Kolkata*
Subir Kumar BHANDARI, *Indian Statistical Institute, Kolkata*

Large-scale multiple testing problems in various disciplines routinely involve dependent observations. Tackling this dependence among observations has been one of the pertinent problems in simultaneous inference. However, very little literature is there that elucidates the effect of correlation on different testing procedures. This talk considers the multiple testing problem under the positively equicorrelated multivariate normality assumption. Here we focus on the family-wise error rate (FWER) and the false discovery rate (FDR), two of the most widely considered frequentist approaches to multiple testing.

Defined as the probability of erroneously rejecting at least one true null hypothesis, FWER is the most natural generalization of type I error rate to simultaneous testing problems. We prove that under the equicorrelated normality assumption, the FWER of Bonferroni's procedure asymptotically goes to zero when the number of hypotheses approaches infinity. We extend this result to the general positively dependent normal setup and generalized family-wise error rates. We also discuss the asymptotic power of Bonferroni's procedure. We also establish similar asymptotic zero behavior for Holm's FWER controlling method.

We also study the Benjamini-Hochberg (BH) procedure under correlated normality that controls the FDR, i.e., the expected proportion of false discoveries. It turns out that the BH method can retain positive FDR asymptotically, unlike Bonferroni or Holm's procedure. Specifically, we show that when the number of tests increases to infinity, then BH with a prespecified FDR level controls FDR at some strictly positive quantity which is a function of the chosen FDR level and the equicorrelation.

Our technique thus facilitates a broad class of error rate criteria and testing procedures. We also empirically demonstrate our results through simulations.

(This is joint work with Prof. Subir Kumar Bhandari, Indian Statistical Institute, Kolkata).

110. Bayesian Learning for disparities from Integrated Data Sources Related to the Rare Events

[27.M1.I14, (page 11)]

Tanujit DEY, *Harvard Medical School*
Anjishnu BANERJEE, *Medical College of Wisconsin*
Sounak CHAKRABORTY, *University of Missouri*

Health disparities are a key area of interest at the regional, and national levels. A dearth of statistical methods, which can distill out factors at the hospital levels are among the key contributors to the lack of understanding. In this work, innovative learning methods are proposed which can lead to better understanding of such factors.

111. Spectral matrix shrinkage estimation using various loss functions.

[29.M1.C16, (page 41)]

Prashant DHAMALE, *SVKM'S NMIMS Deemed to be University, Mumbai/ Savitribai Phule Pune University, Pune*
Akanksha KASHIKAR, *Savitribai Phule Pune University, Pune*

In frequency domain analysis of multivariate time series, Partial Coherence (PC) is one of the important quantities. It is used in many application areas such as seismology, meteorology, neuroscience etc. It is the frequency domain analogue of partial correlation. It is derived from spectral or precision matrices, which are poorly conditioned if the complex degrees of freedom slightly exceed the dimension of the multivariate time series. Previously, shrinkage estimators have been proposed using quadratic loss function and Hilbert-Schmidt loss function. However, it is observed through simulations that some estimated PCs are high when the actual PC value equals zero, with shrinkage estimation under quadratic loss. To resolve this issue, we propose shrinkage estimators of spectral / precision matrices under Absolute loss and Huber loss. In this paper, the performance of each method is studied and evaluated extensively using simulations. The performance is measured based on the percentage relative improvement in the squared error of the proposed estimator compared to the raw estimator of Partial Coherence.

Keywords: Multivariate time series, Spectral Matrix estimation, Loss function, Partial coherence.

112. Deciding optimal number of segments in Detrended Correspondence Analysis

[29.A2.C18, (page 49)]

Shantaram DHUM, *Savitribai Phule Pune University, Pune*
Akanksha KASHIKAR, *Savitribai Phule Pune University, Pune*

In Ecology, Correspondence Analysis (CA) and Detrended Correspondence Analysis (DCA) are commonly used techniques in the ordination of species * sites (abundance data). Species X Sites data is analogous to contingency table data and data in each cell of the table are frequencies i.e., population of particular species at the specified location. 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, arch effect or horseshoe effect is more serious. To eliminate arch/ horseshoe effect, two methods of detrending, namely I) Detrending by segments, II) Detrending by polynomials have been suggested in literature. This gives birth to DCA. In detrending by segments, first correspondence axis is divided into arbitrary number of segments and within each segment, the mean of scores along second axis is made equal to zero. In this method, numbers of segments are arbitrary and different segmentations lead to different ordination along the second correspondence axis. Twenty six is the recommended number of segments and no theoretical justification is provided for this number. We propose a method to find the optimal number of segments by bivariate clustering of the first two correspondence axes.

113. Discriminating between generalized inverted half Rayleigh distribution and generalized inverted half logistic distribution.

[27.A2.C7, (page 22)]

Bishal DIYALI, *Department of Statistics, Central University of Haryana*
Devendra KUMAR, *Department of Statistics, Central University of Haryana*

The generalized inverted half Rayleigh (GIHR) and generalized inverted half logistic (GIHL) distributions are two of the most recently introduced distributions very flexible for studying positively skewed lifetime data with unimodal hazard rates and they share a number of interesting properties. In this study, we used the problem of discriminating between the two distributions. First, we have considered the case of complete data and also it is assumed that the data are type II censored and come from either GIHR or GIHL distributions. To discriminate between the two distribution functions, we use the difference of the maximized log-likelihood functions. The asymptotic distribution of the discrimination statistic is obtained for both the complete and the censored cases. Some simulation studies have been performed for determining the probability of correct selection (PCS), to check the performance of the discrimination procedure under all considered situations. Also, the asymptotic results are used to calculate PCS for varying sample sizes and shape parameter for the complete data case and different censoring proportions in the type II censored case. The asymptotic results are found to perform well in almost all the situations. In addition, we propose a method for finding the minimum sample size required for discrimination for some pre-specified PCS. For illustration purposes, two real data sets are analysed.

114. Geometric EDA for Random Objects

[28.M2.I45, (page 29)]

Paromita DUBEY, *University of Southern California*

In this talk I will propose new tools for the exploratory data analysis of data objects taking values in a general separable metric space. First, I will introduce depth profiles, where the depth profile of a point in the metric space refers to the distribution of the distances between and the data objects. I will describe how depth profiles can be harnessed to define transport ranks, which capture the centrality of each element in the metric space with respect to the data cloud. Next, I will discuss the properties of transport ranks and show how they can be an effective device for detecting and visualizing patterns in samples of random objects. Together with practical illustrations I will establish the theoretical guarantees for the estimation of the depth profiles and the transport ranks for a wide class of metric spaces. I will demonstrate the efficacy of this new approach on distributional data comprising of a sample of age-at-death distributions for various countries and on fMRI data. This talk is based on joint work with Yaqing Chen and Hans-Georg Müller.

115 . Sparse canonical correlation to identify breast cancer related genes regulated by copy number aberrations

[29.A2.I75, (page 48)]

Diptavo DUTTA, *National Cancer Institute*
Ananda SEN, *University of Michigan*
Jaya SATAGOPAN, *Rutgers University*

Copy number aberrations (CNAs) in cancer affect disease outcomes by regulating molecular phenotypes, such as gene expressions, that drive important biological processes. To gain comprehensive insights into molecular biomarkers for cancer, it is critical to identify key groups of CNAs, the associated gene networks, regulatory modules, and their downstream effect on outcomes. In this talk, we will demonstrate an innovative use of sparse canonical correlation analysis (sCCA) to effectively identify the ensemble of CNAs, gene networks and regulatory modules in the context of binary and censored disease endpoints. Our approach detects potentially orthogonal gene expression modules which are highly correlated with sets of CNA and then identifies the genes within these modules that are associated with the outcome. Analyzing clinical and genomic data on 1,904 breast cancer patients from the METABRIC study, we found 14 gene modules to be regulated by groups of proximally located CNA sites. We validated this finding using an independent set of 1,077 breast invasive carcinoma samples from The Cancer Genome Atlas (TCGA). Our analysis on 7 clinical endpoints identified several novel and interpretable regulatory associations, highlighting the role of CNAs in key biological pathways and processes for breast cancer. Genes significantly associated with the outcomes were enriched for early estrogen response pathway, DNA repair pathways as well as targets of transcription factors such as E2F4, MYC and ETS1 that have recognized roles in tumor characteristics and survival. Subsequent meta-analysis across the endpoints further identified several genes through aggregation of weaker associations. Our findings suggest that sCCA analysis can aggregate weaker associations to identify interpretable and important genes, networks and pathways that are clinically consequential.

116. Bayesian and non-Bayesian inference of Weibull lifetime model based on partially observed competing risks data under unified hybrid censoring scheme

[Poster Session, (page 33)]

Subhankar DUTTA, *National Institute of Technology, Rourkela*
Suchandan Kayal ,

In this paper, a competing risk model is analyzed based on unified hybrid censoring scheme. It is assumed that the latent failure times follow Weibull distributions with a common shape and different scale parameters. The maximum likelihood estimates (MLEs) and approximate confidence intervals (ACIs) of the distributional parameters are obtained. Sufficient conditions under which the MLEs exist (uniquely) have been studied. Further, the Bayes estimates are obtained with respect to symmetric and asymmetric loss functions. The non-informative and informative prior distributions are considered for computing the approximate Bayes estimates. Furthermore, the highest posterior density (HPD) credible intervals have been obtained by using Markov Chain Monte Carlo (MCMC) method with the Gibbs sampling technique. The coverage probabilities for each confidence intervals are computed. Then, hypothesis testing has been implemented using likelihood ratio statistic. A Monte Carlo simulation study is carried out to compare the performance of the proposed estimates. Finally, two real data sets are provided to illustrate the estimates.

117 . Semiparametric Transformation Cure model for Current Status data under Competing risks.

[26.A1.I1, (page 3)]

Sreedevi E P., *Maharaja's College , Ernakulam*
Sudheesh KATTUMANNIL, *ISI, Chennai*
Sankaran P. G., *Cochin University of Science and Technology, Cochin*

Current status data occurs if each patient is observed only once at a random monitoring time. In lifetime studies, a proportion of subjects/individuals may not exposed to the event of interest even after a long period of observation. Accordingly, it is important to incorporate long-term survivors in the analysis of competing risks data when current status information is available. In the present paper, we propose a new method for analyzing current status competing risks data with long term survivors. The new method enable us to estimate the cure fraction from the proposed model itself. We formulate the effects of covariates on cumulative incidence functions using linear transformation model. The regression estimators are obtained using counting process based estimating equation. The asymptotic properties of the estimators are studied using martingale representation of the regression estimators. An extensive Monte Carlo simulation study is carried out to assess the finite sample performance of the proposed estimators. Finally, we illustrate our method using a real data set on prostate cancer.

118. High-degree vertices in weighted recursive trees

[26.A2.I7, (page 6)]

Laura ESLAVA, *IIMAS-UNAM*
Bas LODEWIJKS, *Institut Camille Jordan*
Marcel ORTGIESE, *University of Bath*

We will present the behavior of high-degree vertices of weighted recursive trees; which are a generalization of uniform recursive trees. Contrary to the qualitative behavior of linear preferential attachment trees where high-degree vertices become established from the early stages of the process, high-degree vertices in weighted recursive trees keep changing throughout the process. We provide a description of both the order of the degree and the height of such high-degree vertices for a wide class of weighted recursive trees and provide more specific properties for a class containing the uniform recursive tree. This is joint work with Bas Lodewijks and Marcel Ortgiese.

119. Wavelet Spatio-Temporal Change Detection on multi-temporal SAR images

[29.A1.I66, (page 45)]

Rodney FONSECA, *Weizmann Institute of Science*
Rogerio NEGRI, *UNESP*
Aluisio PINHEIRO, *Unicamp*
Abdourrahmane ATTO, *Universite de Savoie*

We introduce WECS (Wavelet Energies Correlation Screening), an unsupervised procedure to detect spatio-temporal change points on multi-temporal SAR images. The procedure is based on wavelet approximation for the multi-temporal images, wavelet energy apportionment, and ultra-high dimensional correlation screening for the wavelet coefficients. We show WECS performance on simulated multi-temporal image data. We also evaluate the proposed method on a time series of 85 satellite images in a forest region at the border of Brazil and the French Guiana. The proposed method displays good results in covering change regions, with the additional benefit of having simple and fast computation.

120. Bayesian nonparametric multilayer clustering of longitudinal data

[28.M1.I40, (page 26)]

Beatrice FRANZOLINI, *franzolini@pm.me*
Maria DE IORIO, *University College London / National University of Singapore / Agency for Science Technology and Research*

We introduce a new class of Bayesian nonparametric models to make inference on an ordered collection of partitions of the same objects, namely, a multilayer partition. The class is suited to analyze panel/longitudinal data, where repeated observations are collected over time for the same observational units. The core of our proposal is a conditional partial exchangeable structure, which we argue is a natural and general modeling strategy for this context. The resulting class of Bayesian models guarantees analytical and computation tractability both in terms of the clustering structure and the underlying random probabilities measures. It allows predictions for any new number of observations and may, ultimately, constitute a powerful reference framework -currently missing in the literature- for the development of tailored Bayesian nonparametric models for panel data. Here, we further explore in detail two specific models within this class: one based on a novel prior and another employing the well-known hierarchical Dirichlet process as building block.

121. Tukey’s Cause- Selecting Control Charts

[28.A2.C12, (page 36)]

Shital GADEKAR, *Department of Statistics Savitribai Phule Pune University Pune*
Manik AWALE, *Department of Statistics Savitribai Phule Pune University,Pune*

Tukeys Cause-Selecting Control Charts

The cause - selecting control chart is constructed for an output variable only after it has been adjusted for the effect of input quality characteristics. In phase I, the parameters of cause - selecting control chart are not known and should be estimated from the historical data. In phase II, the estimated parameters in phase I are used to set up control limits for monitoring process performance in future. The cause-selecting control charts are effective in monitoring and diagnosing the processes with cascade property. A cascade process is multistage process in which the output characteristics of next stage depends upon the output characteristics of previous stage. In this paper, we proposed Tukeys cause-selecting control charts in which Box plot principle is applied to establish control limits. Average run length (ARL) performance of Tukeys cause - selecting control chart is evaluated for normal and non-normal processes. ARL performance of proposed charts is compared with Zhangs cause-selecting control charts through simulation.The simulation-based study reveals that Tukeys cause - selecting control charts outperform Zhangs control charts for normal and non-normal processes in terms of ARL criterion significantly.

122. Random probability measures with fixed mean distributions

[27.A2.I35, (page 21)]

FRANCESCO GAFFI, *Bocconi University*

Linear functionals, or means, of discrete random probability measures are a natural probabilistic object and the investigation of their properties have a long and rich history. They appear in several areas of mathematics, including statistics, combinatorics, special functions, excursions of stochastic processes and financial mathematics, among others. Most contributions have aimed at determining their distribution starting from a fully specified random probability. This work addresses the inverse problem: the identification of the base measure of a discrete random probability measure yielding a specific mean distribution. The only available results stem from the combinatorial literature and concern the very specific case of a Dirichlet process with unit concentration parameter. Here we address the problem in much greater generality and our results cover the Dirichlet process with concentration parameter not necessarily equal to 1, the normalized stable process and the Pitman–Yor process. In addition to their theoretical interest, the results are of practical relevance to Bayesian nonparametric inference, where the law of a random probability measure acts as a prior distribution: often pre-experimental information is available about a finite-dimensional projection of the data generating distribution, such as the mean, rather than about an infinite-dimensional parameter. We further extend our findings to mixture models, ubiquitous in Statistics and Machine Learning.

123. A flexible model with linear approximation for left truncated right censored data

[29.M1.I58, (page 39)]

Ayon GANGULY, *Indian Institute of Technology Guwahati*
D MITRA, *IIM Udaipur*
N BALAKRISHNAN, *McMaster University*
D KUNDU, *IIT Kanpur*

Left truncation and right censoring are two common forms of incompleteness in many reliability and survival data. In this talk, a flexible model for left truncated right censored data based on linear approximation will be discussed. The likelihood estimation of the model parameters will be addressed. The robustness of the model is studied through an extensive simulation. A real data analysis will be provided for illustration purpose.

124. Development of Model Confidence Set Based Ensemble Models Using CEEMDAN Decomposition and Machine Intelligence

[Poster Session, (page 34)]

Sandip GARAI, *Phd Scholar (Ag. Statistics)*
Ranjit Kumar Paul ,

Stock market price prediction is highly demanding to people from all over the World as it is a source of high income for investors, traders and companies. In this paper stock price data has been predicted using several state-of-the-art methodologies such as stochastic forecasting models, machine learning programs, and deep learning algorithms. A complaisant decomposition method resonating with these machine intelligence models entrenched with boosting ensemble method and finally a Model Confidence Set (MCS) based mustering maneuver has been premised for forecasting stock price data. Coalition of forecasts observed from various methodologies is a good approach for improving the result. Despite optimizing the weights for the combination of all the models, a heuristic MCS based snuffing of the least important models prior averaging is conceded as a potent approach. MCS rescinds insignificant models based on the out-of-sample forecasting or in-sample prediction performance prior to equally average the superior models. The proposed methodologies will be compared to the existing standalone techniques using several validation measures to check the improvement in predicting stock price data along with an automated system of predicting a time series data will be premised.

125. Shrinkage Estimators for order restricted Parameters under the Generalized Pitman Nearness Criterion

[27.M1.C4, (page 14)]

Naresh GARG, *IIT Kanpur*
Neeraj MISRA, *IIT Kanpur*

We consider component-wise estimation of two order restricted location/scale parameters of a general bivariate location/scale distribution. We consider a general class of equivariant estimators and use the generalized Pitman nearness (GPN) criterion to derive improved estimators over arbitrary location/scale equivariant estimators. Under the unrestricted parameter space, the Pitman nearest equivariant estimator (PNEE) and the best location/scale equivariant estimator (BLEE/BSEE) are also derived. Then we used the GPN criterion, under the restricted parameter space, over PNEE/BLEE/BSEE and obtained an estimator that performs better than PNEE/BLEE/BSEE in the sense of the GPN criterion. These estimators do not depend on the form the loss function provided loss function is strictly monotone. Best of our knowledge, the GPN criterion is not yet used in the literature to compare and obtain improved estimators for order-restricted parameters. The usefulness of the obtained results is illustrated through various specific probability models. A simulation study has been considered to compare how well different estimators work under the GPN criterion.

126. Random slopes

[26.A2.C1, (page 8)]

Disha GHANDWANI, *Stanford University*
Trevor HASTIE, *Stanford University*
Swarnadip GHOSH, *Radix Trading*
Art OWEN, *Stanford University*

The crossed random-effects model is popular in applied statistics with applications in longitudinal studies, e-commerce, and recommender systems, to name a few. However, these models suffer from scalability issues with computational time running in (n1.5) where n stands for the number of data points. We draw our inspiration from the recommender system used by Stitch Fix; we have a dataset with more than 700,000 clients and 6,000 items, with the number of data points being in the range of 5,000,000. In such cases, fitting crossed random effects using the maximum likelihood approach could be highly inefficient and it could take up to several days to obtain the fit. There has been previous work on the scalability issues by Swarnadip Ghosh, Trevor Hastie, and Art Owen. They studied linear and logistic regression using client and item variables as fixed-effect features, with random intercept terms for clients and items. We work on a generalization of this problem where we allow effect sizes to be random too. It helps us in capturing the variability of effect size among clients as well as items. We have developed a scalable solution to the above problem, i.e., computational time of our algorithm is O(n) and we obtained empirical consistency of our estimates, i.e., as the number of data points increases, our estimates converge to the true parameters.

127. Modelling distribution of aquatic alien invasive weeds under future climate scenarios

[30.M1.I76, (page 51)]

Yogita GHARDE, *ICAR-Directorate of Weed Research*
Sushil KUMAR, *ICAR-Directorate of Weed Research, Jabalpur*
JS MISHRA, *ICAR-Directorate of Weed Research, Jabalpur*

Alien species are non-native species that occur outside their natural adapted ranges and with dispersal potential. Sometimes, these alien species become invasive when they are introduced deliberately or unintentionally outside their natural habitats into new areas where they express the capability to establish, invade and outcompete native species. These alien invasive species are considered as second biggest threat to biodiversity after habitat destruction. Invasive species cause loss of biodiversity including species extinctions, and changes in hydrology and ecosystem function. Many freshwater species such as Alternanthera philoxeroides, Salvinia molesta etc. have become a nuisance for aquatic ecosystems in India. Alligator weed, (Alternanthera philoxeroides (Mart.) Griseb. (Amaranthaceae) originated in the Parana

River region of South America. It has spread to other areas of South America and to the continents of North America, Asia and Australia and some of the adjacent island countries. It has invaded many parts of India including Northern, Eastern and Southern parts. The Salvinia molesta, commonly known as Water Fern is an aggressive and a fast growing Alien Invasive Aquatic Weed of the South-Eastern Brazil origin. During the past 60 Years, it has been spread widely throughout the world and recently added on to the list of the Worlds 100 Most Invasive Species.

Environmental niche modeling (ENM) is commonly used to evaluate the expansion risk of alien invasive species into non-native regions based on their climatic niche. Hence in the present study two aquatic species viz. Alternanthera philoxeroides and Salvinia molesta were studied for their invasion potential using ENM models. Their current geographical extent along with habitat suitability under RCP 4.5 and 8.5 for 2050 and 2070 were obtained using MaxEnt program. Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) was considered to check the model accuracy and validation using test data in MaxEnt. AUC values for Alternanthera philoxeroides and Salvinia molesta were obtained as 0.791 and 0.933 respectively. Jackknife test was employed to check the relative significance of different environmental/input variables. Results revealed that in case of Salvinia molesta, the environmental variable with highest gain when used in isolation is temperature seasonality, followed by Min temperature of coldest month. However, in case of Alternanthera philoxeroides, Annual precipitation and elevation played important role in explaining the suitability of the area. Current and future distribution maps of two species were obtained and prepared in four classes of climate suitability i.e. not suitable (0-0.25), low (0.25-0.50), moderately (0.50-0.75) and highly suitable (¿0.75) categories., An overall decrease in climate suitability for Alternanthera philoxeroides was found under future climate scenarios for 2050 and 2070 under RCP 4.5 and 8.5. However, increase in climatic suitability was obtained in case of Salvinia molesta.

128. Seasonal zero modified geometric INAR(1) process

[29.A2.C19, (page 49)]

Aishwarya GHODAKE, *Department of Statistics, Savitribai Phule Pune University, Pune*
Manik AWALE, *Department of Statistics , Savitribai Phule Pune University, Pune*

Non negative Integer-valued auto regressive (INAR) models have been widely used for modeling the count time series data. These models have shown promising applicability in various fields such as health, insurance, and marketing etc. Number of daily/weekly cases of a disease, weekly number of insurance claims, number of items sold per day of a particular product are some of the examples of the count time series which can be modeled using INAR models. In this paper we propose a Seasonal zero modified geometric INAR model of order one or SZMGINAR(1) with geometric marginal distribution based on negative binomial thinning operator. The proposed model is also capable of capturing under and over dispersion in presence of seasonality, which may be caused by deflation or inflation of zeros. We have derived the Yule-Walker and conditional least squares estimators. The conditional maximum likelihood estimates are obtained using the numerical optimization. We have carried out extensive simulation study to evaluate the asymptotic performance of these estimators. Simulation study reveals that the estimators obtained using these three methods are consistent. We have used prediction root mean squared error (PRMSE), prediction mean absolute error (PMAE) and percentage of true prediction (PTP) as accuracy measures to study the forecast performance of the proposed model. The forecast performance of the model is also compared with its competitive models. A real data set has been analyzed using the proposed model.

129 . An overview of application of Bayesian statistics in early drug development

[27.M1.I17, (page 12)]

Nairita GHOSAL, *Merck & Co., Inc., Rahway, NJ, USA*

Early drug development is a complex, strategy-driven process which spans over Phase I and Phase IIa. Phase I studies are used to evaluate safety, tolerability, and pharmacokinetic parameters in small number of healthy volunteers. Phase IIa studies evaluates a drugs preliminary efficacy, safety, tolerability in relatively higher number (approximately 50 -200) of subjects. Bayesian statistics is gaining traction in drug development since last decades and been extensively applied in early stage of clinical trials. A Bayesian posterior probability-based framework facilitate intuitive Go/No-Go decision making with small number of samples in Phase I studies. A brief description of application of Bayesian statistics in early drug development will be illustrated using simulated examples.

Keywords: Early drug development, Go/No-Go decision, Bayesian inference

130 . Robust statistical modeling of monthly rainfall: The minimum density power divergence approach

[29.M2.I64, (page 43)]

Abhik GHOSH, *Indian Statistical Institute*
Arnab HAZRA, *IIT Kanpur*

Statistical modeling of rainfall is an important challenge in meteorology, particularly from the perspective of rainfed agriculture where a proper assessment of the future availability of rainwater is necessary. The probability models mostly used for this purpose are exponential, gamma, Weibull, and lognormal distributions, where the unknown model parameters are routinely estimated using the maximum likelihood estimator (MLE). However, the presence of outliers or extreme observations is quite common in rainfall data and the MLEs being highly sensitive to them often leads to spurious inference. In this paper, we discuss a robust parameter estimation approach based on the minimum density power divergence estimators (MDPDEs) which provides a class of estimates through a tuning parameter including the MLE as a special case. The underlying tuning parameter controls the trade-offs between efficiency and robustness of the resulting inference; we also discuss a procedure for data-driven optimal selection of this tuning parameter as well as a robust selection of an appropriate model that provides the best fit to some specific rainfall data. We fit the above four parametric models to the areally-weighted monthly rainfall data from the 36 meteorological subdivisions of India for the years 1951-2014 and compare the fits based on the MLE and the proposed optimum MDPDE; the superior performances of the MDPDE based approach are illustrated for several cases. For all month-subdivision combinations, the best-fit models and the estimated median rainfall amounts are provided.

131. Existence of MLEs for parameters of an adjusted linear preferential attachment model of random graphs with covariates

[29.M1.I59, (page 39)]

ARKA GHOSH, *apghosh@iastate.edu*
Sungchan PARK, *Iowa State University*

In this talk, we consider a variation on preferential attachment model (PAM) of random graphs with covariates. It is similar to the usual PAM but the probability of existence of an edge depends on the node/vertex level covariates. For this parametric model, the likelihood function becomes complex and hard to solve analytically. In this work, we show the conditions under which the maximum likelihood estimates (MLEs) are guaranteed to exist.

132. Fractile Graphical Analysis in comparing conditional distributions: Applications in Financial Economics

[26.A2.I10, (page 7)]

Aurobindo GHOSH, *Singapore Management University*
ANIL BERA,

Fractile Graphical Analysis (FGA) was proposed by Mahalanobis (1961) as a method for comparing two distributions at two dierent points (of time or space) controlling for the rank of a covariate through fractile groups. We use bootstrap techniques to formalize the heuristic method used by Mahalanobis for approximating the standard error of the dependent variable using fractile graphs from two independently selected "interpenetrating network of subsamples." We highlight the potential and revisit this underutilized technique of FGA with a historical perspective. We explore a new non-parametric regression method called Fractile Regression where we condition on the ranks of the covariate, and compare it with existing regression techniques. The method thus developed us applied to a financial economic and policy questions.

133. Polyspectral Mean Estimation of General Nonlinear Processes

[29.A1.I66, (page 45)]

Dhrubajyoti GHOSH, *Duke University*
Tucker MCELROY, *US Census Bureau*
Soumendra LAHIRI, *Washington University in St. Louis*

Higher-order spectra (or polyspectra), defined as the Fourier Transform of a stationary process autocumulants, are useful in the analysis of nonlinear and non-Gaussian processes. Polyspectral means are weighted averages over Fourier frequencies of the polyspectra, and estimators can be constructed from analogous weighted averages of the higher-order periodogram (a statistic computed from the data samples

discrete Fourier Transform). We derive the asymptotic distribution of a class of polyspectral mean estimators, obtaining an exact expression for the limit distribution that depends on both the given weighting function as well as on higher-order spectra. Secondly, we use bispectral means to define a new test of the linear process hypothesis. Simulations document the finite sample properties of the asymptotic results. Two applications illustrate our results utility: we test the linear process hypothesis for a Sunspot time series, and for the Gross Domestic Product we conduct a clustering exercise based on bispectral means with different weight functions.

134. Bayesian Multivariate Regression Models for Tropical Cyclone Activity

[26.A1.I6, (page 5)]

Joyee GHOSH, *The University of Iowa*
Xun LI,
Gabriele VILLARINI,

We develop Bayesian regression models for multiple correlated response variables that measure the severity of a hurricane season. Some covariates have a few missing values, and the covariates differ in their predictive performance. This motivates us to develop multivariate regression models that can accommodate missing covariates, as well as variable selection uncertainty. We develop a model that allows dependency in the error terms, and we propose a model space prior that promotes a shared pattern of important/unimportant covariates across response variables. One of the aims of this multivariate analysis is to assess how much of the correlation between the response variables can be attributed to the shared dependence on the covariates. Another important goal of this work is to answer the general question: in moderate dimensional problems with a sparse regression coefficient matrix with many zeros, do Bayesian multivariate regression models improve prediction compared to their univariate counterparts? The answer is not immediately clear from the existing literature, and in this work, we try to fill this gap. To answer this question, we empirically study the properties of Bayesian multivariate regression models and compare them with those of univariate regression models via extensive simulation studies under many scenarios.

135. Multivariate Global-Local Priors for Small Area Estimation

[27.M2.I21, (page 16)]

Malay GHOSH, *University of Florida*
T GHOSH, *Citibank, Tampa*
J. J. MAPLES, *US Census Bureau*
X TANG, *University of Arizona*

It is now widely recognized that small area estimation (SAE) needs to be model-based. Global-local (GL) shrinkage priors for random effects are important in sparse situations where many area level effects do not have significant impact on the reponse beyond what is offered by covariates. We propose in this paper a hierarchical multivariate model with GL priors. We prove propriety of the posterior density when the regression coefficient matrix has a improper uniform prior. Some concentration inequalities are derived for the tail probabilities of the shrinkage estimators. The proposed method is illustrated both via data analysis and simulations.

136 . Adaptive Randomization in Sequential Multiple-Assignment Randomized Trial (SMART)

[28.M1.I39, (page 25)]

Palash GHOSH, *Department of Mathematics, Indian Institute of Technology Guwahati, India*
Rik GHOSH, *Department of Mathematics, Indian Institute of Technology Guwahati India*
Bibhas CHAKRABORTY, *Duke-NUS Medical School, Singapore*
Megan E. PATRICK, *University of Michigan, USA*

In a sequential multiple-assignment randomized trial (SMART), a sequence of treatments is given to a patient over multiple stages of the trial. There are multiple randomization procedures that assign treatments in a SMART across different stages. In spite of SMART designs becoming popular among clinical researchers, the methodologies for adaptive randomization at different stages of a SMART are very few and not sophisticated enough to handle the complexity of optimal allocation of treatments at every stage of a trial. Lack of optimal allocation methodologies can raise serious concerns about SMART designs from an ethical point of view, particularly when, in a specific stage, one treatment is already proven to be better in a (single stage) randomized controlled trial (RCT). In this article, we develop an optimal adaptive allocation procedure for a SMART with the binary primary outcome. Issues related to optimal adaptive allocations are explored theoretically with supporting simulations. Practical issues are discussed.

137. Application of Statistics in the Cell and Gene Therapy modality

[27.M1.C5, (page 14)]

Rochishnu GHOSH, *GSK India Global Services Pvt. Ltd.*

Genes can change as they go through inherited mutations, as they age, or by being altered or damaged by chemicals and radiations which could lead to inherited or acquired diseases. Gene therapy is a technique that uses genetic material (DNA) to manipulate a patient's cells for the treatment of an inherited or acquired disease. Cell therapy is defined as the infusion or transplantation of whole living cells which have been manipulated ex-vivo into a patient for the treatment of an inherited or acquired disease. Together, the two areas combined are known as the Cell and Gene Therapy. At Chemistry Manufacturing and Controls Statistics, we support the Cell and Gene Therapy analytical development process within GSK. The presentation aims at summarising the application of Statistics at every stage of the analytical development process, starting from the time when the blood sample is collected from the patient, also known as leukapheresis, transduced with a viral vector to the point when the final drug product is administered back into the patient.

138. Some aspects of SMART design: Methodological Developments with an Application in Non-inferiority Trial

[26.A2.I9, (page 6)]

Samiran GHOSH, *University of Texas School of Public Health*

Sequential multiple assignment randomized trial (SMART) design is becoming increasingly ubiquitous in the field of precision medicine. This design allows comparisons of sequences of adaptive interventions (AIs) tailored to the individual patient. Superiority testing is usually the initial goal in order to determine which embedded AI yields the best primary outcome on average. When direct superiority is not evident, yet an AI pose other benefits, then non-inferiority testing is warranted. Non-inferiority testing in the SMART setup is rather new and involves specification of non-inferiority margin and other important assumptions that are often unverifiable internally. These challenges are not specific to SMART and apply to two-arm non-inferiority trials that do not include a standard-of-care (SOC or placebo) arm. To address some of these challenges, three-arm non-inferiority trials that include SOC arm are proposed. However, methods developed so far for three-arm non-inferiority trials are not SMART specific. This is because apart from embedded AIs, SMART typically does not include a third SOC arm. In this paper we consider a three-arm SMART from a NIH-funded study of symptom management strategies among people undergoing cancer treatment. Motivated by that example, we for the first time propose a novel data analytic method for non-inferiority testing in the framework of three-arm SMART. Sample size and power considerations are discussed through extensive simulation study to elucidate our method.

139 . Nonparametric Screening and Selection in presence of dependence among predictors

[Poster Session, (page 34)]

Shanta GHOSH, *The University of Illinois at Chicago*
Dr. Sanjib Basu ,

Variable selection under multicollinearity is well studied but remains challenging. Our motivating application arises in environmental epidemiology where individuals are exposed simultaneously to a multitude of pollutants in the environmental mixture that potentially interact and present a health risk. The pollutant measures are often highly correlated at levels that are generally not seen in other areas of science. We develop a model-free screening and selection method using distance correlation, which is a non-parametric measure of dependence in arbitrary dimensions. We compare the performance of our proposed method via simulations with existing methods under linear, non-linear and gam data generating models. In presence of non-linear dependencies, we observed that our proposed method outperforms the traditional methods. We apply the proposed method to environmental mixtures data in NHANES involving many strongly correlated persistent organic pollutants.

140 . Test of exponentiality against NWBUE alternatives - An L-statistics approach

[27.M2.I24, (page 17)]

Shyamal GHOSH, *Assistant Professor, Dept. of Mathematics, IIIT Guwahati*
Murari MITRA, *Professor, Dept. of Mathematics, IIEST Shibpur*

In this work, we develop a class of Hollander-Proschan type tests for testing exponentiality against NWBUE alternatives. We derive the exact distribution of our statistics under the null hypothesis and asymptotic distributions under both null and alternative hypotheses exploiting the L-statistics theory. The performance of the tests for different sample sizes has also been studied in various alternative scenarios. We analyze two well-known data sets in the light of our result.

141. Modeling and identifiability of non-homogenous Poisson process cure rate model.

[Special Invited Session 4, (page 52)]

Asha GOPALAKRISHNAN, *asha@cusat.ac.in*
Anup DEWANJI, *ISI Calcutta, India*
C S SOORYA , *Cochin University of Science & Technology, India.*

Modelling time to event data, when there is always a proportion of the individuals, commonly referred to as immunes who do not experience the event of interest, is of importance in many biomedical studies. Improper distributions are used to model these situations and they are generally referred to as cure rate models. The mixture model is a popular cure rate model.The promotion time cure rate models or bounded cumulative hazards model (BCH) was proposed as an alternative to the mixture cure rate models. This model is modified to provide a class of cure rate models based on a non-homogeneous Poisson process. When censored observations are present, distinguishing censored individuals from the cured group lead to identifiability issues in the members of this class. These identifiability issues are investigated along with other characteristics of the proposed model.

142. Educational data mining on Indian states using clustering and ranking for targeted intervention

[29.A2.C18, (page 49)]

Harithasree GOPALAKRISHNAN, *Great Learning*
Suyash VERMA, *Great Learning*
Dr. Srabashi BASU, *Great Learning*
Natesan RAJAN, *Great Learning*

Annual Status of Education Report (Rural) 2018 (ASER 2018) has found that status of school education is not uniform among the Indian states. Whereas in a few states, e.g. West Bengal and Tripura overall school enrolment is above 85%, in other states, e.g. Manipur, it is below 30%. Early foundational skills in reading and arithmetic are necessary for any developing country to reap benefits of demographic dividend. ASER 2018 includes a total of 15,280 school visits across 26 states in India for information on various infrastructural and quality of teaching indicators. These variables include provision of mid-day meals, ensuring recommended pupil-teacher ratio, construction of toilets, gender-wise distribution of enrolment in government schools etc. The data also includes summary results of household surveys to assess children to understand their basic reading and arithmetic abilities. To our knowledge no comprehensive attempt has been made to discover pattern in the ASER data though states have been ranked according to their performances separately on reading and arithmetic ability of school children. In this study, states have been clustered based on the underlying factors extracted from the set of infrastructure and quality of teaching features. Preliminary results from factor analysis suggest that four latent factors may be obtained, namely school infrastructure, teacher adequacy, space adequacy and enrolment. Based on the above factors, states have been grouped into five clusters. Within the clusters, the states have been ranked based on the learning outcomes of reading and arithmetic abilities at grade-appropriate level. Distribution of the states at grade-appropriate learning has been considered to assign scores for ranking of states within the clusters. This analysis is expected to help the respective state governments to identify the gaps and perform targeted interventions to improve the effectiveness of primary education in India.

Keywords - learning outcomes, foundational skills, primary education, ASER data, enrolment, school infrastructure, reading and arithmetic abilities, gender inclusive education

143 . Percolation models with correlations

[28.A1.I50, (page 32)]

Subhajit GOSWAMI, *Tata Institute of Fundamental Research*

In this talk we will report on some recent exciting developments on a robust technique for answering some classical questions in percolation theory in non-classical contexts. The cases in focus would be the level sets of Gaussian free field on a hypercubic lattice in dimensions 3 and higher, random walks on tours and random interlacements.

144. A general class of shock models with dependent inter-arrival times

[26.A2.C3, (page 9)]

Dheeraj GOYAL, *Department of Mathematics, Indian Institute of Technology Jodhpur, 342037, Karwar, Rajasthan, India*
Nil Kamal HAZRA, *Department of Mathematics, Indian Institute of Technology Jodhpur, 342037, Karwar, Rajasthan, India*
Maxim FINKELSTEIN, *Department of Mathematical Statistics and Actuarial Science, University of the Free State, 339 Bloemfontein 9300, South Africa*

We introduce and study a general class of shock models with dependent inter-arrival times of shocks that occur according to the homogeneous Poisson generalized gamma process. A lifetime of a system affected by a shock process from this class is represented by the convolution of inter-arrival times of shocks. This class contains many popular shock models, namely, the extreme shock model, the generalized extreme shock model, the run shock model, the generalized run shock model, specific mixed shock models, etc. For systems operating under shocks, we derive and discuss the main reliability characteristics (namely, the survival function, the failure rate function, the mean residual lifetime function and the mean lifetime) and study relevant stochastic comparisons. Finally, we provide some numerical examples and illustrate our findings by the application that considers an optimal mission duration policy.

145. Analysis of co-integration in networks

[29.M1.I59, (page 39)]

Apratim GUHA, *XLRI, Xavier School of Management*
Anindya S CHAKRABARTI,

In this work, we explore the problem of portfolio allocation by decomposition of the network financial securities into core and periphery. Considering a financial market as a network where the securities are taken as nodes and existence of co-integration between two nodes as an edge between them, we attempt to create a well-diversified portfolio that can earn higher returns, or reduce risks, as the choice may be. The division in core and periphery, as common in literature, is done through maximisation of depth estimates. We explore how the choice of core and periphery changes with time, and investigate whether there exists correlation with business cycles.

146. On Bayesian Methods for Change-point detection in High Dimension

[28.A1.I52, (page 32)]

Nilabja GUHA, *UMASS Lowell*
Jyotishka DATTA, *Virginia Tech*

We present a Bayesian framework of change point estimation for high-dimensional observations. Such high dimensional observations may appear in many practical applications, where the high dimensional mean parameter changes with time. The dimension of the observations can be much larger than the number of observations. A lower dimensional embedding is proposed based on random projection. We establish change point estimation consistency and convergence rates. Results are shown under both known and unknown covariance structure, and related examples and applications are explored.

147. Small Area Prediction under a Spatially Correlated Multivariate Area Level Model and its Application in Education Survey Data of India

[29.M1.C14, (page 40)]

SAURAV GUHA, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*
Rajender PARSAD, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*
Saumyadipta PYNE, *Department of Statistics and Applied Probability, University of California, Santa Barbara, CA, USA*
Sudeep MARWAHA, *ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India*

The model based small area estimation (SAE) methods based on either implicit or explicit models are extensively applied to produce precise estimates using additional information from related areas. The widely used area-level FayHerriot (FH) model in SAE assumes that area-specific direct survey estimates follow an area-level linear mixed model with area as uncorrelated random effects. However, in practice the boundaries that define a small area are arbitrarily set and it is rational enough to anticipate that the effects of adjacent areas defined by a contiguity criterion are spatially correlated. Unlike the FH model, more than one variable of interest is modelled via multivariate FayHerriot model (MFH) by allowing for different covariance structure between the vector of the variable of interest and the random effects. A spatial version

of MFH model is introduced and small area predictor under this model is proposed. The residual maximum likelihood (REML) is employed for estimating the parameter of the proposed model. Analytical and bootstrap approaches for estimating the mean squared error (MSE) of the proposed small area predictor are also developed. The empirical results from the simulation studies clearly show that the proposed small area predictor outperforms the existing predictors. Both the proposed MSE estimators track the actual value of MSE reasonably well, with confidence intervals based on them achieving close to nominal coverage. An application to measure the district-level spatial distribution of proportion of households having access to internet connectivity and availability of smart devices for online education purposes using the recent available 2017-18 Household Social Consumption: Education in India survey is also presented. In the era of post-pandemic online education system in India, it is expected to offer key insights to the policymakers to identify the areas required more attention.

Keywords: Multivariate linear mixed model; Fay-Herriot model; Spatial correlation; SAR model; Mean squared error; Parametric bootstrap

148. A Bayesian Approach for Network Classification

[27.A1.I27, (page 18)]

Sharmistha GUHA, *Texas A&M University*
Abel RODRIGUEZ, *University of Washington*

We present a novel Bayesian binary classification framework for networks with labeled nodes. Our approach is motivated by applications in brain connectome studies, where the overarching goal is to identify both regions of interest (ROIs) in the brain and connections between ROIs that influence how study subjects are classified. We propose a novel binary logistic regression framework with the network as the predictor, and model the associated network coefficient using a novel class of global-local network shrinkage priors. We perform a theoretical analysis of a member of this class of priors (which we call the Network Lasso Prior) and show asymptotically correct classification of networks even when the number of network edges grows faster than the sample size. Our approach is implemented using an efficient Markov Chain Monte Carlo algorithm, and empirically evaluated through simulation studies and the analysis of a real brain connectome dataset.

149. A zero-inflated Bayesian nonparametric approach for identifying differentially abundant taxa in multigroup microbiome data with covariates

[27.M1.I13, (page 11)]

Subharup GUHA, *University of Florida*
Archie SACHDEVA, *University of Florida*
Somnath DATTA, *University of Florida*
Subharup GUHA, *University of Florida*

Scientific studies in the last two decades have established the central role of the microbiome in disease and health. Differential abundance analysis aims to identify microbial taxa associated with two or more sample groups defined by attributes such as disease subtype, geography, or environmental condition. The results, in turn, help clinical practitioners and researchers diagnose disease and develop new treatments more effectively. However, detecting differential abundance is uniquely challenging due to the high dimensionality, collinearity, sparsity, and compositionality of microbiome data. Further, there is a critical need for unified statistical approaches that can directly compare more than two groups and appropriately adjust for covariates. We develop a zero-inflated Bayesian nonparametric (ZIBNP) methodology that meets the multipronged challenges posed by microbiome data and identifies differentially abundant taxa in two or more groups, while also accounting for sample-specific covariates. The proposed hierarchical model flexibly adapts to unique data characteristics, casts the typically high proportion of zeros in a missing data framework, and mitigates high dimensionality and collinearity issues by utilizing the dimension reducing property of the semiparametric Chinese restaurant process. The approach relates the microbiome sampling depths to inferential precision and conforms with the compositional nature of microbiome data. In simulation studies and in the analyses of the CAnine Microbiome during Parasitism (CAMP) dataset on infected and uninfected dogs and a Global Gut microbiome dataset on human subjects belonging to three geographical regions, we compare ZIBNP with established statistical methods for differential abundance analysis in the presence of covariates.

150. Usage of Random Compression Matrices for Efficient Bayesian Computation

[28.M2.I44, (page 29)]

Rajarshi GUHANIYOGI, *Texas A & M University*

Aaron SCHEFFLER, *UC San Francisco*
Rene GUTIERREZ, *Texas A & M University*

Bayesian computation of high dimensional linear regression models with a popular Gaussian scale mixture prior distribution using Markov Chain Monte Carlo (MCMC) or its variants can be extremely slow or completely prohibitive due to the heavy computational cost that grows in the cubic order of p, with p as the number of features. Although a few recently developed algorithms make the computation efficient in presence of a small to moderately large sample size (with the complexity growing in the cubic order of n), the computation becomes intractable when sample size n is also large. In this talk I will discuss the data sketching approach to compress the n original samples by a random linear transformation to m samples in p dimensions, and compute Bayesian regression with Gaussian scale mixture prior distributions with the randomly compressed response vector and feature matrix. Our proposed approach yields computational complexity growing in the cubic order of m. Another important motivation for this compression procedure is that it anonymizes the data by revealing little information about the original data in the course of analysis. Our detailed empirical investigation with the Horseshoe prior from the class of Gaussian scale mixture priors shows closely similar inference and a massive reduction in per iteration computation time of the proposed approach compared to the regression with the full sample. I will also discuss the posterior contraction rate for high dimensional predictor coefficient with a general class of shrinkage priors on them under data compression/sketching. In particular, we characterize the dimension of the compressed response vector m as a function of the sample size, number of predictors and sparsity in the regression to guarantee accurate estimation of predictor coefficients asymptotically, even after data compression.

151. Predictor-Informed Bayesian Approaches for Dynamic Functional Connectivity

[27.A1.I27, (page 18)]

Michele GUINDANI, *University of California, Los Angeles*

Time-Varying Functional Connectivity investigates how the interactions among brain regions vary over the course of an fMRI experiment. The transitions between different individual connectivity states can be modulated by changes in underlying physiological mechanisms that drive functional network dynamics, e.g., changes in attention or cognitive effort as measured by concurrent measurements. In this talk, I will describe Bayesian approaches for estimating dynamic functional networks as a function of time-varying exogenous physiological covariates that are simultaneously recordedin subjects during the fMRI experiment. We apply our modeling framework on resting-state experiments where fMRI data have been collectedconcurrently with pupillometry measurements, leading us to assess the heterogeneity of the effects of changes in pupil dilation on the subjects' propensity to change connectivity states.

152. Death Process Approach for Modelling Changes in Marriage Probabilities

[28.M1.C10, (page 28)]

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

We model the NFHS-IV data on marriages as a continuous-time death process observed over discrete equidistant time points, where marriage is considered as an exit from the marriageable population and hence is considered as a death in the death process terminology. A death process with a change point has been proposed as a model to assess the changes in marriage rates over time. We derive the maximum likelihood estimators for the parameters and discuss their asymptotic distribution. The asymptotic distribution of the estimators is then used to propose a test for examining the presence/absence of a change point. Parametric bootstrap has been proposed as an alternative approach to derive the rejection region for the test statistic. For the estimation of the change point we have also used expectation and maximization (EM) method to provide confidence interval for change point using parametric bootstrap. In case of two change points, we consider that initially marriage rate is linearly increasing, then it is constant for some time points and then it is linearly decreasing. For the model with two change points, constrained numerical maximization of the likelihood has been carried out for the estimation of the parameters. The linear regression method is used for the estimation of initial values of the parameters. An extensive simulation study is carried out to examine the performance of the proposed procedures. NFHS-IV data is modelled via these approaches to study the changes in male and female marriage rates over time.

153. MARS via LASSO

[29.M1.I54, (page 38)]

Adityanand GUNTUBOYINA, *University of California Berkeley*
Dohyeong KI, *University of California Berkeley*
Billy FANG, *Google*

MARS is a popular method for nonparametric regression introduced by Friedman in 1991. MARS fits simple nonlinear and non-additive functions to regression data. We propose and study a natural LASSO variant of the MARS method. Our method is based on least squares estimation over a convex class of functions obtained by considering infinite-dimensional linear combinations of functions in the MARS basis and imposing a variation based complexity constraint. We show that our estimator can be computed via finite-dimensional convex optimization and that it is naturally connected to nonparametric function estimation techniques based on smoothness constraints. Under a simple design assumption, we prove that our estimator achieves a rate of convergence that depends only logarithmically on dimension and thus avoids the usual curse of dimensionality to some extent. We implement our method with a cross-validation scheme for the selection of the involved tuning parameter and show that it has favorable performance compared to the usual MARS method in simulation and real data settings.

154. Bayesian Approaches in Clinical Trials of Medical Devices

[29.A1.I68, (page 45)]

Jayanti GUPTA, *NishKash Consulting*

Adoption of Bayesian methods in clinical development has seen a gradual increase over the years. The pace of adoption has been better for medical devices relative to drugs. This is due to some key differences between devices and drugs and their clinical pathways. In this talk we will discuss these along with some examples of implementation of Bayesian methods in successful regulatory filings of devices.

155 . Brief description of the key methodological aspects of a vaccine trials with an emphasis on the Covid-19 vaccine study in Pfizer

[Special Invited Session 1, (page 15)]

Pritam GUPTA, *Pfizer Healthcare India Pvt Ltd*

Vaccines are complex biological products which are administered to healthy individuals. Safety is therefore paramount; vaccine development often entails large, time-consuming, and resource-intensive studies to detect rare safety issues and to establish vaccine efficacy. Before a vaccine is licensed and brought to the market, it undergoes a long and rigorous process of research, followed by many years of clinical testing. This talk will focus on introducing some of the key concepts and terminologies related to vaccine trials, discussing on common statistical methodologies related to analyzing trial data followed by the detailed description of the seamless development paradigm used to develop an mRNA vaccine for Covid-19. The talk will include key statistical aspects and regulatory challenges of the designs with an emphasis on the Bayesian framework implemented, which enabled us to obtain efficient designs using decision criteria based on the probability of benefit or harm. Publicly disclosed results will be summarized and discussed. Finally, the talk will conclude by describing my role as a blinded safety statistician involved with quality checks and review of multiple safety outputs on a real time basis.

156. A Novel Approach to Solve Fixed Charge Interval Transportation Problems

[26.A2.C2, (page 9)]

Ummey HABIBA, *Integral University, Lucknow*
Ummey HABIBA, *Integral University, Lucknow*
Abdul QUDDOOS,

This paper presents a new solution approach to solve fixed charge interval transportation problems, in which associated costs, fixed costs, supplies and demands are interval in nature. Objective is to minimize the total interval costs under deterministic decision variables. Firstly, an equivalent bi-objective fixed charge transportation problem is constructed from the given fixed charge interval transportation problem then the newly constructed equivalent crisp problem is solved using neutrosophic programming technique. To demonstrate the solution procedure, two existing numerical examples of Safi and Razmjoo [1] are coded and solved in Lingo 19.0. The obtained results are then compared with two other methods to show the superiority of the proposed solution technique.

Reference [1] Safi, M. R., & Razmjoo, A., (2013): Solving fixed charge transportation problem with interval parameters , Applied Mathematical Modelling , 37, 8341-8347.

157. GET Association Schemes and Designs

[28.A2.C11, (page 36)]

Mohd HARUN, *Scientist, ICAR-Indian Agricultural Statistics Research Instiyute, New Delhi-12*
Cini VARGHESE, *Professor (Agricultural Statistics), ICAR-Indian Agricultural Statistics Research Institute, New Delhi-12*

A new m-class association scheme viz. generalised extended triangular (GET) association scheme has been introduced. Method of constructing partially balanced incomplete block (PBIB) designs based on GET association scheme has been developed. It has been found that the well-known classes of triangular designs and extended triangular designs are special cases of GET designs. These designs also find wide application potential in designing mating experiments.

158. Realistic and Fast Modeling of Spatial Extremes over Large Geographical Domains

[27.A2.I32, (page 20)]

Arnab HAZRA, *Indian Institute of Technology Kanpur*
Raphael HUSER, *King Abdullah University of Science and Technology*
David BOLIN, *King Abdullah University of Science and Technology*

Various natural phenomena, such as precipitation, generally exhibit spatial extremal dependence at short distances only, while the dependence usually fades away as the distance between sites increases arbitrarily. However, the available models proposed in the literature for spatial extremes, which are based on max-stable or Pareto processes or comparatively less computationally demanding sub-asymptotic models based on Gaussian location and/or scale mixtures, generally assume that spatial extremal dependence persists across the entire spatial domain. This is a clear limitation when modeling extremes over large geographical domains, but surprisingly, it has been mostly overlooked in the literature. In this paper, we develop a more realistic Bayesian framework based on a novel Gaussian scale mixture model, where the Gaussian process component is driven by a stochastic partial differential equation that yields a sparse precision matrix, and the random scale component is modeled as a low-rank Pareto-tailed or Weibull-tailed spatial process determined by compactly supported basis functions. We show that our proposed model is approximately tail-stationary despite its non-stationary construction in terms of basis functions, and we demonstrate that it can capture a wide range of extremal dependence structures as a function of distance. Furthermore, the inherently sparse structure of our spatial model allows fast Bayesian computations, even in high spatial dimensions, based on a customized Markov chain Monte Carlo algorithm, which prioritizes calibration in the tail. In our application, we fit our model to analyze heavy monsoon rainfall data in Bangladesh. A simulation and a cross-validation study indicate that the proposed model outperforms some natural alternatives and that the model fits precipitation extremes satisfactorily well. Finally, we use the fitted model to draw inferences on long-term return levels for marginal precipitation at each site, and for spatial aggregates.

159. On the general δ-shock model

[27.A1.I29, (page 19)]

Nil Kamal HAZRA, *IIT Jodhpur*
Goyal DHEERAJ, *IIT Jodhpur*

The -shock model is one of the basic shock models which has a wide range of applications in reliability, finance and related fields. In existing literature, it is assumed that the recovery time of a system from the damage induced by a shock is constant as well as the shocks magnitude. However, as technical systems gradually deteriorate with time, it takes more time to recover from this damage, whereas the larger magnitude of a shock also results in the same effect. Therefore, we introduce a general -shock model when the recovery time depends on both the arrival times and the magnitudes of shocks. Moreover, we also consider a more general and flexible shock process, namely, the Poisson generalized gamma process. It includes the homogeneous Poisson process, the nonhomogeneous Poisson process, the Plya process and the generalized Plya process as the particular cases. For the defined survival model, we derive the relationships for the survival function and the mean lifetime and study some relevant stochastic properties. As an application, an example of the corresponding optimal replacement policy is discussed.

160. AI solutions for traffic management in LTE networks

[29.A1.I71, (page 46)]

Ranjani H. G., *Principal Data Scientist, GAIA, Ericsson, Bangalore*

Traffic in LTE networks is increasing at a rapid pace due to popularity and accessibility of smartphones and tablets. However, spectral resources are limited. To meet the dynamic demand for high bandwidth and capacity, it is important to optimally load the network cells. In this talk, we discuss challenges and some AI/ML based approaches in managing cellular congestion, in order to optimally distribute cellular load across various network cells. We also discuss scalability aspects of such solutions, real-time deployment challenges and life cycle management aspects involved in such a system. Title: Supervised change point detection using non-stationary Hawkes process

161. Fay-Herriot Area Level Model with Arc-Sin transformation for Sample Proportions

[26.A2.I8, (page 6)]

Masayo HIROSE, *Institute of Mathematics for Industry, Kyushu University*
Malay GHOSH, *Department of Statistics, University of Florida*
Tamal GHOSH, *Citibank, N.A*

An empirical best linear unbiased predictor can contribute to more efficiency, especially when the sample size within each area is not large enough to make reliable direct estimates. Sometimes it is essential to transform it back properly to the original scale to conclude. However, the natural back transformation could produce a bias, especially when the sample size within an area is not large enough. In this study, we find explicit empirical Bayes estimators for arc-sin transformed data that correct biases asymptotically. Moreover, we explicitly obtain the second-order unbiased estimators of these mean squared prediction errors, maintaining strict positivity. We also apply the proposed method to the positive rate in PCR testing for COVID-19.

162. Asymmetric Super-Heston-rough volatility model with Zumbach effect as a scaling limit of quadratic Hawkes processes

[27.A1.I28, (page 19)]

Srikanth IYER, *Indian Institute of Science Bangalore*
Priyanka CHUDASAMA, *Indian Institute of Science Bangalore*

Modeling price variation has always been of interest, from options pricing to risk management. It has been observed that the high-frequency financial market is highly volatile, and the volatility is rough. Moreover, we have the Zumbach effect, which means that past trends in the price process convey important information on future volatility. Microscopic price models based on the univariate quadratic Hawkes (hereafter QHawkes) process can capture the Zumbach effect and the rough volatility behavior at the macroscopic scale. But they fail to capture the asymmetry in the upward and downward movement of the price process. To incorporate asymmetry in price movement at the micro-scale and rough volatility and the Zumbach effect at the macroscale, we introduce the bivariate Modified-QHawkes process for upward and downward price movement. After suitable scaling and shifting, we show that the limit of the price process in the behaves as the so-called Super-Heston-rough model with the Zumbach effect.

163 . A function-based approach to model the measurement error in wearable devices

[29.M2.I61, (page 42)]

Sneha JADHAV, *Wake Forest University*
Carmen TEKWE, *Indiana University Bloomington*
Yuanyuan LUAN,

Physical activity (PA) is an important risk factor for many health outcomes. Wearable-devices such as accelerometers are increasingly used in biomedical studies to understand the associations between PA and health outcomes. Statistical analyses involving accelerometer data are challenging due to the following three characteristics: (i) high-dimensionality, (ii) temporal dependence, and (iii) measurement error. To address these challenges we treat accelerometer-based measures of PA as a single function-valued covariate prone to measurement error. Specifically, in order to determine the relationship between PA and a health outcome of interest, we propose a regression model with a functional covariate that accounts for measurement error. Using regression calibration, we develop a two-step estimation method for the model parameters and establish their consistency. A test is also proposed to test the significance of the estimated model parameters. Simulation studies are conducted to compare the proposed methods with existing alternative approaches under varying scenarios. Finally, the developed methods are used to assess the relationship between PA intensity and BMI obtained from the National Health and Nutrition Examination Survey data.

164. Reverse Experience Replay: An Efficient Way to Learn with Dependent Data.

[Special Invited Session 1, (page 15)]

Prateek JAIN, *Google Research India*
Dheeraj NAGARAJ, *Google Research India*
Praneeth NETRAPALLI, *Google Research India*
Naman AGARWAL, *Google Research Princeton*

Stochastic gradient descent (SGD) is the workhorse of modern machine learning. While SGD has been thoroughly analyzed for independent data and tight finite time guarantees are known, its finite sample performance with dependent data has not been as thoroughly analyzed. In this talk, we will consider SGD-style algorithms for two problems where the data is not independent but rather comes from a Markov chain: learning dynamical systems and Q-learning for reinforcement learning. While vanilla SGD is biased and does not converge to the correct solution for these problems, we show that SGD along with "reverse experience replay" can efficiently find the optimal solutions. Joint work with Naman Agarwal, Syomantak Chaudhuri, Praneeth Netrapalli, Suhas Kowshik and Dheeraj Nagaraj.

165 . Ordinal Non-Symmetric Correspondence Analysis for Four-way Contingency Table using Orthogonal Polynomials

[29.M1.C14, (page 40)]

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

Non-symmetric correspondence analysis (NSCA) allows the analyst to explore, interpret and portray the nature of asymmetric association between two or more nominal or ordinal categorical variables in an exploratory way. We extend the theory of ordinal three-way NSCA to four-way contingency tables (CT) using orthogonal polynomials. We propose to use tensorial approach to decompose the centered-profile array instead of kronecker product. The array of generalised correlations obtained using tensor operations for four-way CT can be easily generalized to multi-way CT. The coordinates for interactively-coded predictor isometric Biplot are also calculated. The Matlab codes for performing four-way NSCA are developed and the theoretical developments are demonstrated with an example.

166 . CNN-LSTM model for agricultural price forecasting using genetic algorithm optimization technique

[27.A2.C6, (page 22)]

Ronit JAISWAL, *PhD Scholar, ICAR-INDIAN AGRICULTURAL RESEARCH INSTITUTE, New Delhi*
Girish Kumar JHA, *Principal Scientist, ICAR-IARI, New Delhi*

Agricultural price series possess the characteristics of nonlinearity, non-stationarity and randomness which makes the task of accurate price prediction challenging. Convolutional neural network (CNN), a very prominent feature extracting deep learning technique and long short-term memory (LSTM), an efficient recurrent neural network forecasting technique, are integrated in this study to create a hybrid called CNN-LSTM model that can forecast any complex agricultural price series. The study uses four international monthly price series of palm oil, soybean oil, soybean, and maize for the comparison of the price prediction capabilities of the CNN-LSTM model with those of the LSTM, time-delay neural network (TDNN), and ARIMA models. The genetic algorithm (GA) technique is utilized to examine and choose various hyperparameters of the hybrid model to develop the final optimized CNN-LSTM model for each series. Several assessment metrics, including root mean square error, mean absolute percentage error, mean absolute deviation, and directional statistics indicate that the GA optimized CNN-LSTM model outperforms other models in the experimental data. In addition, the accuracy of the forecast is also tested for significant improvement using the Friedman test and Diebold-Mariano test, whose results validate that the CNN-LSTM model outperforms the others. Keywords: Agricultural price prediction; Convolutional neural network; Deep learning; Genetic algorithm; Long short-term memory

167 . A nonparametric test of cospectrality of networks

[27.M1.I15, (page 12)]

Indrajit JANA, *IIT Bhubaneswar*
Srijan SENGUPTA, *North Carolina State University*
Chetkar JHA, *University of Pennsylvania*

We live in an interconnected world where network valued data arises in many domains, and, fittingly, statistical network analysis has emerged as an active area in the literature. However, the topic of hypothesis testing in networks has received relatively less attention. In this work we consider the problem where one is given two networks, and the goal is to test whether the given networks are cospectral, i.e., they have the same non-zero eigenvalues.

Cospectral graphs have been well studied in graph theory and computer science. Cospectrality is relevant in real-world networks since it implies that the two networks share several important path-based properties, such as the same number of closed walks of any given length, the same epidemic threshold, etc. However, to the extent of our knowledge, there has not been any formal statistical inference work on this topic.

We propose a non-parametric test of cospectrality by leveraging some recent developments in random matrix theory. We establish theoretical results for the proposed test, and demonstrate its empirical accuracy using synthetic networks sampled from a wide variety of models as well as several well-known real-world network datasets.

This work is in collaboration with Chetkar Jha (University of Pennsylvania) and Srijan Sengupta (North Carolina State University).

168 . Scoring predictions at extreme quantiles

[27.A2.I32, (page 20)]

Kaushik JANA, *Ahmedabad University*
Kaushik JANA, *Ahmedabad University*
Almut VERAART, *Imperial College London*
Axel GANDY, *Imperial College London*

Prediction of quantiles at extreme tails is of interest in numerous applications. Extreme value modelling provides various competing predictors for this point prediction problem. A common method of assessment of a set of competing predictors is to evaluate their predictive performance in a given situation. However, due to the extreme nature of this inference problem, it can be possible that the predicted quantiles are not seen in the historical records, particularly when the sample size is small. This situation poses a problem to the validation of the prediction with its realization. In this article, we propose two non-parametric scoring approaches to assess extreme quantile prediction mechanisms. The proposed assessment methods are based on predicting a sequence of equally extreme quantiles on different parts of the data. We then use the quantile scoring function to evaluate the competing predictors. The performance of the scoring methods is compared with the conventional scoring method and the superiority of the former methods are demonstrated in a simulation study. The methods are then applied to analyze cyber Netflow data from Los Alamos National Laboratory and daily precipitation data at a station in California available from Global Historical Climatology Network.

169 . Spatio-temporal modelling of malaria mortality in India from the Million Death Study

[29.M1.C16, (page 41)]

Sayantee JANA, *IIT Hyderabad*
Sze Hang FU, *Dalla Lana School of Public Health, Centre for Global Health Research, St. Michaels Hospital, University of Toronto, Toronto, ON, Canada*
Hellen GELBAND, *Dalla Lana School of Public Health, Centre for Global Health Research, St. Michaels Hospital, University of Toronto, Toronto, ON, Canada*
Patrick BROWN, *Dalla Lana School of Public Health, Centre for Global Health Research, St. Michaels Hospital, University of Toronto, Toronto, ON, Canada Department of Statistical Sciences, University of Toronto, Toronto, Canada*

Background India has a substantial burden of malaria, concentrated in specific areas and population groups. Spatio-temporal modelling of deaths due to malaria in India is a critical tool for identifying high-risk groups for effective resource allocation and disease control policy-making, and subsequently for the countrys progress towards United Nations 2030 Sustainable Development Goals.

Methods In this study, a spatio-temporal model with the objective of understanding the spatial distribution of malaria mortality rates and the rate of temporal decline, across the country, has been constructed. A spatio-temporal random slope model was used, with malaria risk depending on a spatial relative risk surface and a linear time effect with a spatially-varying coefficient. The models were adjusted for urban/rural status (residence of the deceased) and Normalized Difference Vegetation Index (NDVI), using 200413 data from the Million Death Study (MDS) (the most recent data available), with nationwide geographic coverage. Previous studies based on MDS had focused only on aggregated analyses.

Results The rural population had twice the risk of death due to malaria compared to the urban population. Malaria mortality in some of the highest-risk regions, namely the states of Odisha and Jharkhand, are declining faster than other areas; however, the rate of decline was not uniformly correlated with the level of risk. The overall decline was faster after 2010.

Conclusion The results suggest a need for increased attention in high-risk rural populations, which already face challenges like inadequate infrastructure, inaccessibility to health care facilities, awareness, and education around malaria mortality and prevalence. It also points to the urgent need to restart the MDS to document changes since 2013, to develop appropriate malaria control measures.

170. Fully Bayesian estimation of virtual brain parameters with self-tuning Hamiltonian Monte Carlo

[29.A1.I70, (page 46)]

Jayant JHA, *Indian Statistical Institute*

Virtual brain models are data-driven patient-specific brain models integrating individual brain imaging data with neural mass modeling in a single computational framework, capable of autonomously generating brain activity and its associated brain imaging signals. Along the example of epilepsy, we develop an efficient and accurate Bayesian methodology estimating the parameters linked to the extent of the epileptogenic zone. State-of-the-art advances in Bayesian inference using Hamiltonian Monte Carlo (HMC) algorithms have remained elusive for large-scale differential-equations based models due to their slow convergence. We propose appropriate priors and a novel reparameterization to facilitate efficient exploration of the posterior distribution in terms of computational time and convergence diagnostics. The methodology is illustrated for in-silico dataset and then, applied to infer the personalized model parameters based on the empirical stereotactic electroencephalography (SEEG) recordings of retrospective patients. This improved methodology may pave the way to render HMC methods sufficiently easy and efficient to use, thus applicable in personalized medicine.

171 . Joint Modeling of Longitudinal Biomarker Data with Changepoint and Flexible Sigmoidal Response

[27.M2.I22, (page 16)]

Wesley JOHNSON, *UC Irvine*
Michelle NORRIS, *California State University Sacramento*
Edward BEDRICK, *U. of Arizona*

We develop a joint model for longitudinal bivariate diagnostic outcome data; one dichotomous and one continuous. In veterinary medicine, perfect diagnostic methods rarely exist. For the situation we consider, the continuous outcome is used as a biomarker for a specific infection/disease, and in practice, a cutoff is selected whereby if the outcome exceeds that cutoff, the individual is diagnosed as having the infection. That outcome could be a false positive, and when the cutoff is not achieved, and they are declared negative, it could be a false negative. Similarly, the binary outcome, which directly indicates either positive or negative, could be false. Our data involve n pairs sampled through time; timing and number of times vary by individual. The actual time of infection for those who become infected is unknown, but our changepoint model allows for its imputation. As a function of time, the continuous biomarker outcome is modeled with a flexible sigmoid that allows for clusters of individuals with different rates of increase and/or different curve heights. This is accomplished using a Dirichlet Process Mixture of sigmoids. We analyze data on Johnes disease in cattle.

172. Lower Bounds on the Rate of Convergence for Metropolis-Hastings

[28.M1.I41, (page 26)]

Galin JONES, *University of Minnesota*
Austin BROWN, *University of Minnesota*

Practitioners are often left tuning Metropolis-Hastings algorithms by trial and error or using optimal scaling guidelines to avoid poor empirical performance. We develop general lower bounds on the convergence rates of Metropolis-Hastings algorithms to study their computational complexity. If the target density concentrates with a parameter n (e.g. Bayesian posterior concentration, Laplace approximations), we show the convergence rate can tend to

1 exponentially fast if the tuning parameters do not depend carefully on n. We show this is the case for random-walk Metropolis in Bayesian logistic regression with Zellner's g-prior. We focus on more general target densities using a special class of Metropolis-Hastings algorithms with a Gaussian proposal (e.g. random walk and Metropolis-adjusted Langevin algorithms) where we give more general conditions. An application to flat prior Bayesian logistic regression is given.

173. A random string in a medium of Poisson traps

[30.M1.I78, (page 52)]

Mathew JOSEPH, *Indian Statistical Institute*
Siva ATHREYA, *Indian Statistical Institute*
Carl MUELLER, *University of Rochester*

We consider a random string modeled by the additive stochastic heat equation on a torus. The string evolves in R^n among obstacles of finite width centered around points coming from a Poisson point process. The string dies when any part of it hits an obstacle. We discuss the probability that the string survives up to time T. This talk is based on work with Siva Athreya and Carl Mueller.

174. Two-stage and purely sequential minimum risk point estimation of the scale parameter of a family of distributions under modified LINEX loss plus sampling cost

[29.A2.I72, (page 47)]

Neeraj JOSHI, *joshin@iitk.ac.in*
Sudeep R. BAPAT, *Indian Institute of Management Indore*
Raghu Nandan SENGUPTA, *Indian Institute of Technology Kanpur*

In this research, we present two-stage and purely sequential methodologies for estimating the scale parameter of the Moore and Bilikam family of lifetime distributions (see Moore and Bilikam, 1978). We propose our methodologies under the minimum risk point estimation setup, whereby we consider the modified LINEX loss function plus non linear sampling cost. We study some interesting exact distributional properties associated with our stopping rules. We also present simulation analyses using Weibull and Pareto distributions (special cases of the Moore and Bilikam family) to check the performance of our two-stage and purely sequential procedures. Finally, we provide two real data sets, namely, COVID-19 data and shock measurement data and analyze them using the Weibull and Pareto models in support of the practical utility of our proposed two-stage methodology.

References: Moore, A.H. and Bilikam, J.E. (1978). Bayesian Estimation of Parameters of Life Distributions and Reliability from Type II Censored Samples. IEEE Transactions on Reliability, 27: 64-67.

175. A new look at Integer Valued Time Series models with Discrete distributions

[27.A2.I31, (page 20)]

Mariyamma K D, *Christ College (Autonomous) Irinjalakuda, Calicut University*

The work is concerned with the discrete distributions and integer- valued autoregressive processes. It illustrates the specific convolutions of non-Gaussian distributions, and to extend the unification to larger classes. It also focuses on studying various generalizations of discrete distributions like discrete Mittag-Leffler, discrete stable-Linnik, geometric discrete semi stable-Linnik, Luders Formel I, Delaporte, discrete Poisson-Laplace, Katz Family of distributions, etc. Characteristic properties of the extended models are investigated, the advantage of generalized model over the base model is established and finally various application aspects of the newly developed models are envisaged. Key words: Binomial thinning; Infinite divisibility; Discrete self-decomposability; Integer valued autoregressive processes; Statistical modelling. Autoregressive processes; Count data models.

176 . Model-Based Clustering Using Finite-Mixture of Power-Normal Distribution

[28.M1.C10, (page 27)]

Shiva Kumar K, *Pondicherry University*

Clustering is a technique for segregating observations into similar groups, namely clusters. In recent times, finite-mixture models have grabbed much attention in model-based clustering, where each component density is considered as a cluster that arises from a probability distribution of a mixture population. While the finite mixture of normal and t models handle the symmetric data, the scale mixture of skew-normal (SMSN) models deal with skewed data. Here, the focus is on skewed data and its clustering using a finite mixture of power normal distribution as an alternative to the finite mixture of skew-normal distribution. The model parameters are derived and estimated through the EM algorithm. The model is illustrated with simulated and real-life data sets. The clustering results are compared with normal and skew-normal finite mixture models. Keywords: Model-Based Clustering, Finite-Mixture Model, EM Algorithm, Finite Mixture of Power Normal Distribution.

177. A tensorial approach to partition the Asymmetric Measure of Association for Multiway Contingency Table

[29.M1.C14, (page 40)]

Kirtee KAMALJA, *Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, Maharashtra*

When there is one-way or asymmetric relationship between the categorical variables associated with two-way contingency table (CT), GoodmanKruskal - index is a more suitable measure of association instead of Pearson chi-squared statistic. When three variables are under study, the multivariate extensions of -index such as Gray-Williams index, Marcotorchinos Index, Delta Index are used depending on number of response variables in the CT. We propose a generalization of Marcotorchino index as a measure of asymmetric association for a multiway CT. Further we partition this index into partial measures of predictability of the response variable given each of the predictor variables, a scalar multiple of symmetric association between all predictors and a residual term. We propose an algorithm based on a tensorial approach to partition the index which is computationally efficient than that of the traditional one. A demonstration with an example confirms the results.

178. A Bayesian Record Linkage Approach that Adjusts for Variables in One Dataset

[27.A2.C7, (page 23)]

Gauri KAMAT, *Brown University*
Mingyang SHAN, *Eli Lilly and Company*
Roee GUTMAN, *Brown University*

In many healthcare applications, information about units is dispersed across multiple datasets. Linking records across datasets becomes necessary when the goal is to estimate associations among variables exclusively appearing in each dataset. Record linkage is a statistical technique that identifies records representing the same entity across multiple datasets when unique identifiers are absent. Common Bayesian record linkage algorithms rely on similarities between variables recorded in both datasets, and do not adjust for relationships between variables that are exclusive to each dataset. We extend existing Bayesian record linkage methods to integrate associations between variables that are exclusive to one dataset. We show analytically, and using simulations, that our method improves the linking process and results in accurate estimates when identifying information is limited. We apply our method to link Meals on Wheels recipients to Medicare Enrollment records, and examine the association between activities of daily living and healthcare utilization among Meals on Wheels recipients.

179 . Efficient Approximation of Interval Specific Estimates for Biological Growth Curve Models using Localized Likelihood Maximization and Its Applications to Real Data Analysis

[Poster Session, (page 34)]

Md Aktar Ul KARIM, *Institute of Chemical Technology, Mumbai*
Amiya Ranjan BHOWMICK, *Institute of Chemical Technology, Mumbai*

Biological growth curves are used to make predictions about natural growth phenomena and have an excellent record of success in several disciplines. Typical approach is to treat the growth equations as mechanistic nonlinear models and fit them to real data sets by nonlinear least squares or maximum likelihood method. Bhowmick et al.[1]s method explicitly utilized the functional form of growth equations and obtained interval-specific estimates of parameters. These estimates have been found to perform better in estimating relative growth rate (RGR) and have the potential for selecting the best model from a set of competitive nonlinear models.

Indeed the model-specific estimates of RGR have already been used in growth studies for model selection purposes. However, its computation involves a lengthy mathematical calculation, and many times explicit expressions for the interval-specific estimators are not available that are a function of data alone. For highly nonlinear models, where the parameters are nonlinearly related, the computation of interval estimates is almost impossible and may suffer from significant approximation error. We propose a computational approach to obtain interval estimates of parameters based on the maximum likelihood estimation method. The likelihood function is maximized using the data on smaller intervals. We found that the proposed strategy provides stable estimates as compared to the existing approach which involves several steps of mathematical calculations. It is also found to be robust when the process error has a large variance. In addition, it does not require manual mathematical manipulation, instead, it uses optimizers in software packages efficiently, therefore, giving more accessibility to the practitioners. The validity of the proposed estimation process has been carried out by both simulation studies and real data sets.

180. An approximation algorithm for blocking of an experimental design

[26.A1.I3, (page 4)]

Bikram KARMAKAR, *University of Florida*

Blocked randomized designs are used to improve the precision of treatment effect estimates compared to a completely randomized design. A block is a set of units that are relatively homogeneous and consequently would tend to produce relatively similar outcomes if the treatment had no effect. The problem of finding the optimal blocking of the units into equal sized blocks of any given size larger than two is known to be a difficult problemthere is no polynomial time method guaranteed to find the optimal blocking. All available methods to solve the problem are heuristic methods. We propose methods that run in polynomial time and guarantee a blocking that is provably close to the optimal blocking. In all our simulation studies, the proposed methods perform better, create better homogeneous blocks, compared with the existing methods. Our blocking method aims to minimize the maximum of all pairwise differences of units in the same block. We show that bounding this maximum difference ensures that the error in the average treatment effect estimate is similarly bounded for all treatment assignments. In contrast, if the blocking bounds the average or sum of these differences, the error in the average treatment effect estimate can still be large in several treatment assignments.

181. Structurally incomplete 2-part designs involving treatment combinations

[26.A2.C2, (page 8)]

Sayantani KARMAKAR, *ICAR-Indian Agricultural Research Institute*
Cini VARGHESE, *ICAR-Indian Agricultural Statistics Research Institute*
Mohd HARUN, *ICAR-Indian Agricultural Statistics Research Institute*
Vinaykumar L.N., *ICAR-Indian Agricultural Research Institute*

2-part designs involve two groups of treatments such that the blocks are incomplete with respect to both the groups and each pair of treatments occurs together a constant number of times within each group (11 and 22 for first and second group respectively) and between groups (12). Fusion of two Incomplete Block Designs (IBD) in a systematic manner can yield 2-part row-column designs for investigating a set of treatment combinations. Sometimes, due to unavoidable constraints, the same treatment may have to be allotted to the entire row, thereby, making the row-blocking ineffective. This gives rise to 2-part block designs, forming a particular case of 2-part row-column designs. Further, in situations where certain experimental units are not available for the application of treatments, structurally incomplete (SI) designs viz. 2-part SI row-column designs and 2-part SI block designs prove to be helpful. General expressions of the information matrices have been derived to study the properties of these 2-part SI designs. Canonical efficiency factors have also been calculated.

182. Comprehensive Simultaneous Inference on Trend-Cycle Model

[29.M2.I64, (page 43)]

Sayar KARMAKAR, *University of Florida*
Wei Biao WU,

Cyclostationary process is a special type of nonstationary processes showing a periodic pattern. Motivated from various weather related data such as temperature, rainfall, etc. we analyze the popular trend-cycle model in presence of a very general nonstationary error and build simultaneous inference on the trend function post the estimation of the period length and the periodic component. Our simultaneous confidence bands are comprehensive in the sense it does not suffer from the boundary problem and thus it could allow a more general ARMA-GARCH type of noise processes that essentially requires larger bandwidth to estimate. We also discuss how to jointly study multiple such similar series and test whether the periods are synchronized or not. We conclude by providing some simulations and analysis of temperature anomaly data.

183. Urbanisation and Bird Diversity

[27.A2.I31, (page 20)]

Akanksha KASHIKAR, *Department of Statistics, Savitribai Phule Pune University*

Since the second half of the 20th century, the rapid growth of urban areas has been considered the primary driver of land-use changes and a significant threat to biodiversity worldwide. Although there are many causes of habitat loss, urbanisation is one of the most damaging factors in terms of the number of species lost or threatened. Therefore, understanding the impact of urbanisation on biodiversity becomes imperative from the point of view of conservation and planning of sustainable cities. Our current study deals with the effects of urbanisation on bird diversity in different states of India. The dataset consists of different urbanisation and biodiversity indicators (i.e., number of vehicles registered, urban population, species count, forest and tree cover areas etc.) for 28 states and seven union territories in India from 2000-2017. The primary purpose of the analysis is to check whether urbanisation affects biodiversity and to identify the states / UTs greatly affected by urbanisation. Techniques such as Regression analysis, Spearman rank correlation test, and TOPSIS are used for this analysis. We observe that urbanisation negatively affects the diversity of birds. We also attempt to identify regions greatly affected by urbanisation and need immediate attention.

184. A machine learning approach to predict vessel trajectories in real-time using AIS data

[27.A2.C8, (page 23)]

Rohit KATTI, *S & P Global*
Jyotirmoy PATRA, *S & P Global*
Nidhi RAI, *S & P Global*
Sandip PAL, *S & P Global*

In recent years, real-time vessel monitoring, and vessel management systems have become popular research topics in the maritime industry to measure factors such as marine pollution, port congestion management, intelligent ship navigation, ship collision risk estimation, etc. In this paper, we present an Artificial Neural Network (ANN)-based framework for predicting the course and destination of ships of a particular size class. Given the geospatial positions of the previous journey on the same voyage, the proposed framework can be utilized to estimate the trajectory for the remainder of the journey and the destination port at any time during a voyage. The key to this framework is dividing the entire sea surface into hexagonal grids of a certain resolution (based on Uber's H3 system) and mapping each of the Automatic Identification System (AIS) coordinates to specific hexagonal grid cells. We've converted trajectories into a series of geo-hexagonal indices using these hexagonal grid cells. This paper introduces a deep learning framework for predicting vessel trajectories and destination ports using Long Short-Term Memory (LSTM) and Gate Recurrent Unit (GRU) models. Our method can also be utilized for proactive collision prevention, anomaly detection, and route planning.

185 . Semiparametric transformation model for competing risks data with cure fraction

[26.A1.I1, (page 3)]

Sudheesh KATTUMANNIL, *Indian Statistical Institute*

We propose a new method for the analysis of competing risks data with long term survivors. The proposed method enables us to estimate the overall survival probability and cure fraction simultaneously. We formulate the effect of covariates on cumulative incidence functions using linear transformation models. Estimating equations based on counting process are developed to estimate regression coefficients. The asymptotic properties of the estimators are studied using martingale theory. An extensive Monte Carlo simulation study is carried out to assess the finite sample performance of the proposed estimators. Finally, we illustrate our method using a real data set.

186 . Individualized Inference in Bayesian Quantile Directed Acyclic Graphical Models

[Poster Session, (page 34)]

Ksheera Sagar KERALAPURA NAGARAJU, *Purdue University*
Yang Ni, Veera Baladandayuthapani and Anindya Bhadra ,

We propose an approach termed "qDAGx" for Bayesian covariate-dependent quantile directed acyclic graphs (DAGs) where these DAGs are individualized, in the sense that they depend on individual-specific covariates. A key distinguishing feature of the proposed approach is that the individualized DAG structure can be uniquely identified at any given quantile based on purely observational data without strong assumptions such as a known topological ordering. For scaling the proposed method to a large number of variables and covariates, we propose for the model parameters a novel parameter expanded horseshoe prior that affords a number of attractive theoretical and computational benefits to our approach. By modeling the conditional quantiles, qDAGx overcomes the common limitations of mean regression for DAGs, which can be sensitive to the choice of likelihood, e.g., an assumption of multivariate normality, as well as to the choice of priors. We demonstrate the performance of qDAGx through extensive numerical simulations and via an application in precision medicine by inferring patient-specific protein–protein interaction networks in lung cancer.

187. Near Optimal inference in Adaptive Linear Regression

[27.M2.I19, (page 15)]

Koulik KHAMARU, *Rutgers University*
Yash DESHPANDE, *Voleon*
Tor LATTIMORE , *Deep Mind*
Lester MACKEY, *Microsoft Research*

When data is collected in an adaptive manner, even simple methods like ordinary least squares can exhibit non-normal asymptotic behavior. As an undesirable consequence, hypothesis tests and confidence intervals based on asymptotic normality can lead to erroneous results. We propose an online debiasing estimator to correct these distributional anomalies in least squares estimation. Our proposed method takes advantage of the covariance structure present in the dataset and provides sharper estimates in directions for which more information has accrued. We establish an asymptotic normality property for our proposed online debiasing estimator under mild conditions on the data collection process, and provide asymptotically exact confidence intervals. We additionally prove a minimax lower bound for the adaptive linear regression problem, thereby providing a baseline by which to compare estimators. There are various conditions under which our proposed estimator achieves the minimax lower bound up to logarithmic factors. We demonstrate the usefulness of our theory via applications to multi-armed bandit, autoregressive time series estimation, and active learning with exploration.

188. Testing Quantiles of Several Logistic Populations with a Common Location

[Poster Session, (page 34)]

Habiba KHATUN, *National Institute of Technology Rourkela*
Prof. Manas Ranjan Tripathy ,

In this study, we consider the problem of hypothesis testing for the quantiles when there are several logistic populations available with a common location parameter and possibly unknown and unequal scale parameters. The traditional asymptotic likelihood ratio test (ALRT) is derived. Because of its asymptotic nature, it does not attain the nominal level of size of a test for small samples. Hence, we apply two of its modified versions, such as the standardized likelihood ratio test (call it SLRT) and the parametric bootstrap likelihood ratio test (call it PLRT). Taking advantage of the advanced computational facilities available nowadays, we apply a computational approach test (call it CAT1) and a modified version of it (call it CAT2) using the MLEs of the parameters. Using a comprehensive simulation study, we numerically compare the power and size of all the proposed tests. In the end, a real-life example has been considered to demonstrate the proposed test procedures.

189. Confidence sets in a sparse stochastic block model

[Special Invited Session 2, (page 28)]

Bas KLEIJN, *b.j.k.kleijn@uva.nl*
Jan VAN WAAIJ, *Department of Mathematical Sciences, University of Copenhagen*

In a sparse stochastic block model with two communities of unequal sizes we derive two posterior concentration inequalities, for (1) posterior (almost-)exact recovery of the community structure; (2) construction of confidence sets for the community assignment from credible sets with finite graph sizes, enabling exact frequentist uncertain quantification with MCMC-simulated credible sets at non-asymptotic graph sizes. We consider the enlargement of credible sets methodologically in greater generality and illustrate with computation of confidence sets derived from simulated graph data. It is also argued that a form of early stopping applies to MCMC sampling of the posterior, to enable the computation of confidence sets at larger graph sizes. (Based on joint work with J. van Waaij.)

190 . Stage life testing in presence of competing risks

[30.M2.I80, (page 53)]

Arnab KOLEY, *IIM Indore*
Sudeep BAPAT, *IIM Indore*

Life testing analysis of stage life testing in presence of competing risk is performed in this study. We take two independent competing risks following Weibull distribution with same shape but different scales in the initial stage, whereas two independent competing risks following one parameter exponential distributions with different scales in the second stage. The experiment stops when all the units fail. The whole analysis has been conducted under both the classical and Bayesian framework.

191. Simultaneous Confidence Intervals for Multi-way Clustered Stock Return Data

[28.A1.I53, (page 33)]

Tamalika KOLEY, *Birla Institute of Technology Mesra*
Mrinal JANA, *Birla Institute of Technology Mesra*
Gopal BASAK, *Indian Statistical Institute, Kolkata*

Return of a particular stock is usually affected by different months of a year and hence, it is necessary to consider any type of correlation if present in the computation of standard errors for the estimated stock returns. An exploratory analysis of the monthly stock returns data collected from the Bombay Stock Exchange in the period of January 2012 - December 2021 is performed to study the impact of a month and a year on the stock returns. Next simultaneous confidence intervals for the stock return considering different types of clustering in the data have been constructed in order to validate the results obtained from the initial analysis. In order to capture the market uncertainty, the stock returns are taken in the form of closed intervals instead of real values.

192. Dissimilarity Measure for Quantification of the Poetic/Prose Nature of a Text

[29.A2.C19, (page 49)]

Rutuja KORE, *Department of Statistics, Savitribai Phule Pune University*
Akanksha KASHIKAR, *Department of Statistics, Savitribai Phule Pune University, Pune*

Sanskrit is considered to be one of the oldest languages. One of the characteristic of the Sanskrit language is that word orders can be changed without much effect on the meaning conveyed. Poets tend to shuffle the words to comply with the poetic meter. It is speculated that the ancient Indian text, Bhagvadgita has been updated by three different authors. According to Khair (1969), Author 2 of the Bhagvadgita tends to write in a prose style, while other two writers follow poetic style. In this paper, we construct a measure for quantifying the poetic/prose nature of the text, especially in the context of the Bhagvadgita. We verify that the proposed measure satisfies the mathematical properties of a distance function. The proposed metric captures the difference between linguistic features of a prose and poetic text. The difference in the word sequence in the original verse and the verse obtained after anvaya (shuffled verse as per the prose order of the verse) has been calculated using k-grams. No. of words which have different positions in original and shuffled verse also contributes to the quantification of the dissimilarity. We compute this measure for the verses supposed to be written by every author and examine if there is a statistically significant difference between the written styles of the three authors.

193. Estimation in Jointly Type-II Censored k Lindley Distributions

[28.A2.C13, (page 37)]

Hare KRISHNA, *Department of Statistics, Chaudhary Charan Singh University, Meerut (INDIA)*

In engineering, reliability theory, and life testing experiments, products and units are collected from various production lines. In the statistical literature, the joint censoring scheme is proposed to perform comparative lifetime tests on products which are from two or more production lines under the same survival conditions. This research paper deals with the inferences when a joint type-II censored sample is taken from k Lindley populations. The maximum likelihood estimators of the unknown parameters are calculated. The construction of asymptotic confidence intervals and the log-transformed asymptotic confidence intervals of the unknown model parameters are given. Additionally, the boot-p and boot-t confidence intervals are constructed. In the Bayesian framework, the approximate Bayes estimates of the parameters are evaluated using the importance sampling and Gibbs sampling techniques under the balanced loss function assuming informative and non-informative priors. Also, a Monte Carlo simulation study is carried out to evaluate the performance of the various estimators. Finally, a real data analysis is performed to illustrate all the developed estimation methods.

194. Central Limit Theorem, Edgeworth expansion, and Neural Networks

[30.M2.I79, (page 53)]

Arun Kumar KUCHIBHOTLA , *Carnegie Mellon University*
Arisina BANERJEE, *Indian Statistical Institute*

Quantitative central limit theorems via Berry–Esseen bounds consider bounding the distance between probabilities of various events under the distribution of the normalized sum of random variables and under the corresponding Gaussian distribution. Investigations into quantitative CLT bounds in terms of expectations of functions were also considered in the literature, but with no explicit dependence on the dimension or the "adaptive" notion of smoothness of the functions. In this talk, we make use of function spaces commonly studied in neural networks and obtain quantitative CLTs that show explicit dimension dependence.

195. Parameter Estimation for Multi-state Series and Parallel Systems Under Different Observation Setting

[27.M2.I24, (page 17)]

Leena KULKARNI, *Department of Statistics, School of Mathematics, Applied Statistics & Analytics, NMIMS Deemed-to-be University, Mumbai, India*
Sanjeev SABNIS, *Department of Mathematics, IIT Bombay, Mumbai, India*
Sujit GHOSH, *Department of Statistics, NC State University, Raleigh, USA*

Consider multi-state series and multi-state parallel systems consisting of N independent components each. It is assumed that (i) each component and both the systems take values in the set $\{0, 1, 2\}$, (ii) each system and each component start out in state 2, the perfect state, and they make the transition to state 1, depending upon system configuration, and, eventually each system enters state 0, failed state. This multi-state nature of components and systems leads to N scenarios under which each of the systems makes the transition from state 2 to state 1, and eventually to state 0. The joint probability function for times spent in state 2 and state 1 is obtained based on these N scenarios for each of the systems. It is interesting to note that by merely changing set $\{0, 1\}$ of a standard binary series (parallel) system to a set $\{0, 1, 2\}$ of a multi-state series (multi-state parallel) system, renders expression of the joint probability function of system spending times in state 2 and state 1 of a multi-state series (multi-state parallel) system quite complex as compared to the univariate survival probability of the binary series (parallel) system being in the functioning state. Finally, as a proof of concept, graphical comparison between these analytical joint probability functions and joint empirical probability functions for each of the multi-state series and multi-state parallel systems based on Farlie-Gumbel-Morgernsten (FGM) distribution is made and it is found that they compare very well. The graphical comparison between corresponding univariate cumulative distribution functions is also found to be very good. Further, we obtain the parameter estimates under this observation setup for the FGM distribution.

196. Classification Rules for Two Inverse Gaussian Populations with Common Dispersion Parameter and Unequal Means

[Poster Session, (page 34)]

Pushkal KUMAR, *National Institute of Technolgy, Rourkela*
Manas Ranjan Tripathy ,

The classification problem for two inverse Gaussian populations with a common dispersion parameter and unequal means is considered. To develop the new classification rules using the proposed model, we derive the Bayesian estimators of model parameters and compare their performances with the existing classical estimators. The proposed estimators perform better than some existing ones in terms of bias and mean squared error (MSE) in most of the parameter spaces. Further, utilizing these estimators, we have constructed several plug-in type classification rules. We have numerically compared the performances of all the proposed rules, including the existing one, in terms of the expected probability of correct classification (EPC). The proposed plug-in type classification rules perform better than some existing ones. Finally, a real-life example illustrates the current model problem's importance.

197. DUG-MLES: An efficient hybrid machine learning algorithm to classify Dual Prong Goods measure the compliance risk

[27.M1.C4, (page 14)]

Rajat KUMAR, *S&P Global*
Suman SENAPATI, *S&P Global*
Byron MCKINNEY, *S&P Global*

Trade finance is cross-border in nature, requiring the shipment of goods from one port to another or through high-risk jurisdictions. In addition, trade transactions necessitate the manual processing of non-standardized documentation to verify shipment compliance. Due to these factors, trade transactions may pose a greater degree of compliance risks, and the variety of goods shipped may expose trade financiers to anti-money laundering, terrorist financing, and reputational risks associated with military equipment. The ability to detect dual prong/use goods is highly subjective and highly dependent on the trade operator's training and personal experience. Current industry solutions provide a low level of dual-use product and brand items, generally viewed to be below 10,000 items (with a strong possibility of missing military and dual-usage products), off-the-shelf search functionality (not tailored to the trade finance industry) and no prior model input from real-life goods descriptions (no base-line benchmarking strategy). Market solutions today are keyword lookup tools only, many of them require users to cut and paste the goods name from wider text (minus the packaging, incoterms, and port information) prior to activating a search for dual-use goods identification. Keyword lists alone provide a poor level of accuracy and return high numbers of false positives (roughly in the region of 40-50%) result. To compensate for current industry solutions which have a low baseline set of product keywords to work from, our objective of the dual use model is to automate the Level 0 maker and if appropriate, checker review conducted as part of the real-time transaction process used in the trade finance business. The model will provide system recommendations if further action, or no action is required in the dual use risk category. If action is required, the reviewer will use the output to enhance due diligence further. The model was developed through a hybrid application of a) SGD machine learning algorithm and b) elastic keyword search. A feedback module is incorporated into the model in order to facilitate ongoing output precision, thereby enhancing its overall predictive capabilities. The dual use module aims to provide a globally standardised control and reduce the risk associated with identifying these items. Therefore, whenever a good description is provided, the dual use model will analyse and return insightful results in order to process the transaction.

Keywords: Compliance risk, Machine Learning, Elastic Search, Dual Prong Goods, Hybrid Model.

198. Bivariate Distribution with Singular Components

[Poster Session, (page 34)]

Sanjay KUMAR, *PhD Student*
NA ,

Real-life data sets with ties arise quite commonly in reliability and survival analysis. We attempt to model such types of data sets, first using bivariate distributions with singular components, and then, explore further extensions to multivariate distributions. For this purpose, we consider mainly two types of approaches, the "Minimization approach" and the "Maximization approach". The Marshall-Olkin bivariate and multivariate extensions of the exponential distribution have been discussed in detail. The Marshall-Olkin bivariate exponential (MOBE) distribution has both an absolutely continuous part and a singular part. We discuss the physical interpretation and properties of the MOBE distribution. The existence and uniqueness of maximum likelihood estimates (MLEs) of parameters of the MOBE distribution have been discussed, and obtained using the EM algorithm. We also obtain Bayes estimates for parameters of the MOBE distribution with the Dirichlet-Gamma prior under the squared error loss function. Due to some limitations of the MOBE distribution, we consider the more flexible Marshall-Olkin bivariate Weibull (MOBW) distribution. MLEs of parameters of the MOBW distribution, using the EM algorithm are obtained under the random censoring scheme. We also obtain Bayes estimates of parameters for the MOBW distribution under type-I censored data. Further, MLEs of parameters of dependent competing risk models using MOBW distribution under type-II progressive hybrid censoring have been derived. Finally, we obtain MLEs of parameters of the more general bivariate modified Weibull (BMW) distribution. MOBE and MOBW distributions are special cases of the BMW distribution. We fit the BMW model under real life censored data. Some numerical exercises have been carried out for

illustration purpose.

199. Calibration Approach-Based Estimator under Dual Frame Surveys: Revisited

[28.A2.C13, (page 37)]

Shubham KUMAR, *shubh.stat@bhu.ac.in*
Piyush Kant RAI, *Banaras Hindu University, Varanasi*

Abstract To decrease the sampling costs or reduce the frame under-coverage errors occurring due to a single sampling frame, the use of multiple-frame surveys has gained significance in the last decade. However, these multiple frame estimators are difficult to deal with (in practice) but they are distinguished theoretically and may have smaller variances as compared to other estimators which are in general relevant. Also, a variety of auxiliary information available in some cases can obviously be used to provide a more efficient estimator. In the present article, a Joint Calibration Estimator (JCE) for dual frame survey is considered which already exists in sampling literature and we discussed some of its aspects and limitations. Also, we proposed JCE under the Hellinger distance function for the estimation of the population total. The fundamental concept of calibration approach-based estimation is to adjust the original design weights in such a way that the value of the given distance function is minimum and at the same time it satisfies a set of given constraints for the auxiliary information. We have derived and explored some of the properties of the proposed estimator theoretically and a numerical illustration is provided for the clarity and efficiency comparison of the developed estimator. Keyword: Dual Frame survey, Joint Calibration Estimator, Distance Functions.

200. Classical estimation in log-logistic distribution using adaptive progressively type-II censored data

[27.A2.C6, (page 22)]

Anita KUMARI, *Department of Statistics, Central University of Haryana, Mahendragarh India*
Kapil KUMAR, *Department of Statistics, Central University of Haryana, Mahendragarh, India*

This article deals with the classical estimation of the parameters of log-logistic distribution under adaptive progressively type-II censored data. The maximum likelihood and maximum product spacing estimation methods are used to obtain the point estimates of the unknown parameters with their corresponding asymptotic confidence interval as the interval estimates of the parameters. A simulation study is carried out to check the performance of the estimates. The illustration of the study is shown by analyzing a real-life problem.

201. Reliability Estimation for Bathtub Shaped Distribution under Progressive Type II Block Censoring

[28.M1.C10, (page 27)]

Rani KUMARI, *National Institute of Technology Patna*
Rani KUMARI, *Department of Mathematics, National Institute of Technology Patna, India*
Yogesh Mani TRIPATHI, *Department of Mathematics, Indian Institute of Technology Patna, India*
Rajesh Kumar SINHA, *Department of Mathematics, National Institute of Technology Patna, India*

In this article, we derive estimates for survival characteristics such as reliability function, hazard rate and median time to failure as well as differences in different test facilities under block progressive type-II censored data. Estimates are obtained under non-hierarchical as well as hierarchical frameworks when the lifetime of test units follow a bathtub shaped distribution. In non-hierarchical framework, maximum likelihood estimators of unknown quantities are derived and their existence and uniqueness properties are studied. Further Bayes estimators are derived using a hybrid Metropolis-Hasting sampling procedure. The performance of proposed approaches are compared via simulation study. We observed that hierarchical approach under non informative priors produce better estimation results. Finally, a real data is analysed to illustrate the proposed estimates.

202. Bootstrap Control Chart for Percentiles of Some Lifetime Distributions with Hybrid Censored Data

[27.A1.I29, (page 19)]

Amarjit KUNDU, *Professor, Department of Mathematics, Raiganj University, Raiganj*
Prof. Shovan CHOWDHURY, *Professor, Indian institute of Management Kozhikode, Kozhikode, India*
Bidhan MODOK, *Research Scholar, Department of Mathematics, Raiganj University*

Bootstrap Control Chart for Percentiles of Some Lifetime Distributions with Hybrid Censored Data Prof. Amarjit Kundu* Department of Mathe-

matics, Raiganj University, Raiganj, India Prof. Shovan Chowdhury Indian institute of Management Kozhikode, Kozhikode, India, and Bidhan Modok Department of Mathematics, Raiganj University, Raiganj, India

Abstract: Parametric bootstrap control charts are proposed for process monitoring of percentiles of some lifetime distributions for type-I hybrid censored data assuming in-control parameters to be unknown. Similar charts can be derived for type-I and type-II censored data as a special case of the proposed censoring scheme. Monte Carlo simulations are carried out for various combinations of percentiles, false-alarm rates and sample sizes to evaluate the in-control performances of the proposed charts in terms of average run lengths. The out-of-control behavior and performance of the chart is thoroughly investigated for several choices of shifts in the parameters of the distribution. Conventional Shewhart-type chart is also proposed under the same set-up asymptotically and compared with bootstrap chart using a skewed data set. The chart under hybrid censoring scheme is found to be more effective than the same under type-I and type-II censoring schemes in terms of magnitude and speed of detection of out-of-control signals. An application of the proposed chart is shown from clinical practice.

203 . Optimal lifetesting plan for two sample censoring scheme

[27.M2.I20, (page 15)]

Debasis KUNDU, *Indian Institute of Technology Kanpur*

Optimal life-testing plan for two sample censoring scheme 1 Debasis Kundu Department of Mathematics and Statistics Indian Institute of Technology Kanpur Kanpur, Pin-208016 Abstract Joint progressive censoring schemes are quite useful to conduct comparative life-testing experiment of different competing products. Recently, the balanced joint progressive censoring (BJPC) scheme has been intro- duced in the literature. Optimal planning of such progressive censoring scheme is an important issue to the experimenter. This article consid- ers optimal life-testing plan under the BJPC scheme using the Bayesian precision and D-optimality criteria, assuming that the lifetimes follow Weibull distribution. In order to obtain the optimal BJPC life-testing plans, one needs to carry out an exhaustive search within the set of all admissible plans under the BJPC scheme. However, for large sample size, determination of the optimal life-testing plan is difficult by exhaus- tive search technique. A meta-heuristic algorithm based on the variable neighborhood search method is employed for computation of the optimal life- testing plan. Optimal plans are provided under different scenarios.

204 . A queueing-inventory control model with cross-trained servers

[29.A2.C17, (page 48)]

INIYA L, *Madurai Kamaraj University*
SIVAKUMAR B, *Madurai Kamaraj University*

In this article, we consider a queueing-inventory system with ser- vice facility. The service facility has multiple servers and all the servers are cross-trained so that they can be assigned to either front room or back room. We assume the customers arrive according to a Poisson process and request service at the front room and demands one unit of item. The demanded item is delivered to them after performing a random service which is distributed as exponential. The inventory is replenished according to a (s,Q) policy. The lead times for the orders are assumed to have exponential distribution. The service facility has finite waiting hall and customers who arrive during the waiting hall is full are considered to be lost. By switching the servers from the front room to back room or vice versa, depending on the queue size and the on-hand inventory level, management may be able to mini- mize the total operating cost. The problem is modelled as a Markov decision problem and the problem is solved using Linear programming method. A numerical study is conducted to analyze the impact of various parameters on the performance of the servers.

205 . Topological Data Analysis in Finance A Review of Applications

[28.M2.I42, (page 28)]

Arnab LAHA, *Indian Institute of Management Ahmedabad*

Topological Data Analysis (TDA) deals with applications of computational algebraic topology for analysing real world data sets. Unlike other well-known statistical methods, TDA techniques attempt to understand the shape of data and use that information to draw conclusions. Over the last two decades TDA has found innovative applications in a variety of disciplines such as genomics, chemical engineering, astrophysics, evolution, image analysis, medicine etc.

Recently TDA has been applied to problems of finance such as clustering and classification of financial time series, investment decisions, financial crisis, corporate failure etc. Interesting insights that are of significant importance have been revealed through use of TDA techniques. In this paper after a brief review of TDA, I discuss some of these recent innovative applications of TDA techniques to analysis of financial data.

206 . Optimal dynamic treatment regimes via smooth surrogate losses

[27.M2.I19, (page 15)]

Nilanjana LAHA, *Texas A&M University*
Aaron SONABEND, *Google research*
Rajarshi MUKHERJEE, *Harvard University*
Tianxi CAI, *Harvard University*

: Large health care data repositories such as electronic health records (EHR) open new opportunities to derive individualized treatment strategies for complicated diseases such as sepsis. In this talk, I will discuss the problem of estimating sequential treatment policies tailored to a patient's individual characteristics, often referred to as dynamic treatment regimes (DTRs). Our main objective will be to find the optimal DTR that maximizes a discontinuous value function through direct maximization of Fisher consistent surrogate loss functions. In this regard, we demonstrate that a large class of concave surrogates fails to be Fisher consistent – a behavior that differs from the classical binary classification problems. We further characterize a non-concave family of Fisher consistent smooth surrogate functions, which can be optimized via gradient descent using off-the-shelf machine learning algorithms. Compared to the existing direct search approach under the support vector machine framework (Zhao et al., 2015), our proposed DTR estimation via surrogate loss optimization (DTRESLO) method is more computationally scalable to large sample sizes and allows for broader functional classes for treatment policies. We establish theoretical properties for our proposed DTR estimator and obtain a sharp upper bound on the regret corresponding to our DTRESLO method. The finite sample performance of our proposed estimator is evaluated through extensive simulations. Finally, we illustrate the working principles and benefits of our method for estimating an optimal DTR for treating sepsis using EHR data from sepsis patients admitted to intensive care units.

207. A lagged cross correction matrix model for Indian stock market data

[28.A1.I53, (page 33)]

Ananya LAHIRI, *IIT Tirupati*
Nagavyshnavi GANTASALA, *IIT Tirupati*

We propose a model for lagged cross correction functions from Indian stock return data. We study the corresponding cross correction matrices and its empirical spectral distributions. We will see how good the model describes the actual data. This is a joint work with PhD student Gantasala Naga Vyshnavi.

208. Hierarchical Bayes estimation of small area proportions using statistical linkage of disparate data sources

[27.M2.I21, (page 16)]

Parthasarathi LAHIRI, *University of Maryland, College Park*
Soumojit DAS, *University of Maryland, College Park, USA*

We propose a Bayesian approach to estimate finite population proportions for small areas. The proposed methodology improves on the traditional sample survey method because, unlike the traditional methods, our proposed method borrows strength from multiple data sources. Our approach is fundamentally different from the existing small area Bayesian approach to the finite population sampling, which typically assumes a hierarchical model for all units of the finite population. We assume such model only for the units of the finite population with outcome variable observed because for these units assumed model can be checked using existing statistical tools. Modeling unobserved units of the finite population is challenging because assumed model cannot be checked in the absence of data on the outcome variable. To make reasonable modeling assumptions, we propose to form a large number of cells for each small area using factors that potentially influence the binary outcome variable of interest. This strategy is expected to bring some degree of homogeneity within a given cell and also among cells from different small areas that are constructed with the same factor level combination. Instead of assuming an elaborate model for the true probabilities for individual unobserved units, we assume that population means of cells with the same combination of factor levels are identical across small areas and the population mean of true probabilities for a cell is identical to the mean of true values for the

observed units in that cell. We apply our proposed methodology to a real-life COVID-19 survey, linking information from multiple disparate data sources to estimate vaccine-hesitancy rates (proportions) for 50 US states and Washington, D.C. (small areas). We also provide practical ways of model selection that can be applied to a wider class of models under similar setting but for a diverse range of scientific problems.

209. The Central Limit Theorem and statistical inference in high dimensions

[Plenary Lecture 3, (page 44)]

Soumendra LAHIRI, *Washington University in St. Louis*

The Central Limit Theorem (CLT) plays an indispensable role in classical statistical inference for finite dimensional parameters, including calibration of confidence sets and statistical tests. However, validity of the CLT in high dimensional problems where the dimension (p) of the observations diverges to infinity with the sample size (n) is no longer guaranteed. There is extensive recent work on the problem, following the seminal paper by Chernozhukov, Chetverikov and Kato (2013; Annals of Statistics), that attempts to establish the CLT (or Gaussian Approximation) in high dimensions under various growth conditions on the dimension p. In this talk, we present some new results on Gaussian Approximation for different classes of sets, providing insights into specific distributional characteristics of the underlying high dimensional random vectors that determine the optimal growth rates. We also consider some extensions and applications of Gaussian Approximation to high dimensional statistical inference problems.

210. Generalized Bayesian Additive Regression Trees Models

[28.A1.I48, (page 31)]

Antonio LINERO, *University of Texas at Austin*

Bayesian additive regression trees have seen increased interest in recent years due to their ability to combine machine learning techniques with principled uncertainty quantification. The Bayesian backfitting algorithm used to fit BART models, however, limits their application to a small class of models for which conditional conjugacy exists. In this article, we greatly expand the domain of applicability of BART to arbitrary generalized BART models by introducing a very simple, tuning-parameter-free, reversible jump Markov chain Monte Carlo algorithm. Our algorithm requires only that the user be able to compute the likelihood and (optionally) its gradient and Fisher information. The potential applications are very broad; we consider examples in survival analysis, structured heteroskedastic regression, and gamma shape regression.

211. Learning under Distribution Shift

[30.M2.I79, (page 53)]

Zachary LIPTON, *Carnegie Mellon University*

Methods for prediction can fail catastrophically when faced with data with a different distribution at test time. A broad goal of my recent research has been to develop methods which have rigorous guarantees under various types of constrained distribution shift. I will present some vignettes in the domain adaptation setting, where we can access unlabeled target samples, and can use these unlabeled samples to adapt our source predictor. I will first present results for various generalizations of the label shift problem, where shifts can occur in the label marginals, but the class conditional distributions are stable. Then I will present some results on domain adaptation under missingness-shift where source and target distributions are identical, but are corrupted by different missingness mechanisms.

This is based on joint work with Saurabh Garg, Helen Zhou and Sivaraman Balakrishnan.

212. Optimal PBIB designs in two-replicates

[26.A2.C2, (page 9)]

Vinaykumar L.N., *ICAR - Indian Agricultural Research Institute*
Cini VARGHESE, *ICAR - Indian Agricultural Statistics Research Institute*
Harun MOHD, *ICAR - Indian Agricultural Statistics Research Institute*
Sayantani KARMAKAR , *ICAR - Indian Agricultural Research Institute*

A new method of constructing a series of four associate class partially balanced incomplete block (PBIB) designs in two replicates having blocks of unequal sizes has been proposed. The underlying association scheme of these designs is defined. A second series of three associate class PBIB designs with two replications and equal block sizes has been obtained,

as a special case, using the same method. The efficiency factors of the designs are observed to be quite high when compared to an orthogonal block design. Also, both series of designs are affine resolvable and hence are optimal in the competing class of resolvable incomplete block designs. As treatments require minimum number of replications in these designs, they can be advantageously used when there is a scarcity in resources.

213 . Quantifying portable genetic effects and improving cross-ancestry genetic prediction with GWAS summary statistics

[27.A1.I30, (page 19)]

Qiongshi LU, *University of Wisconsin-Madison*

Polygenic risk scores (PRS) calculated from genome-wide association studies (GWAS) of Europeans are known to have substantially reduced predictive accuracy in non-European populations, limiting its clinical utility and raising concerns about health disparities across ancestral populations. However, existing methods to address this problem have several critical limitations (e.g., most methods require some individual-level data that are independent from input GWAS; these data rarely exist in practice), and the improvement in predictive performance has been incremental. Here, we present a novel statistical framework named X-Wing with three main innovations to improve PRS performance in ancestrally diverse populations. First, we introduce and estimate cross-population local genetic correlation which directly quantifies correlated (portable) genetic effects between multiple ancestral populations. Second, we introduce a novel Bayesian method to incorporate local genetic correlation annotation into multi-population PRS modeling, where annotation-dependent statistical shrinkage amplifies the effects of annotated variants (i.e., variants with correlated effects between populations). Finally, we propose an innovative strategy to combine multiple PRS trained in various populations into an omnibus score with improved prediction accuracy using GWAS summary data alone as input. The entire X-Wing procedure only requires GWAS summary data, which is a major advance compared to existing approaches. Through extensive benchmarking using numerous GWAS datasets including UK Biobank, Biobank Japan, and Population Architecture using Genomics and Epidemiology Consortium study, we demonstrate that X-Wing pinpoints portable genetic effects and substantially improves PRS performance in non-European populations. Applied to 31 traits in East Asians, cross-population genetic correlations using X-Wing-annotated SNPs are substantially higher than the genome-wide genetic correlation estimates, while correlations in the remaining genome are consistently lower. X-Wing PRS achieves 31.0%-96.1% gain in predictive R2 compared to state-of-the-art methods based on GWAS summary statistics. Followed up on 13 traits in admixed Americans, X-Wing also shows a substantial improvement in prediction accuracy, with the R2 increase ranging from 18.7%-122.1%. Overall, X-Wing introduces several methodological innovations that will likely have broad and impactful applications and provides an accurate and privacy-preserving method to construct PRS in ancestrally diverse populations.

214 . Multilevel Programming Model with Neutrosophic Approach for Supplier Selection Problem in a Fuzzy Environment.

[29.M1.C15, (page 41)]

Anas M, *Research Scholar, D/o Statistics & Operations Research, Aligarh Muslim University, Aligarh*
Ahmad Yusuf ADHAMI, *Assistant Professor, D/o Statistics & Operations Research, Aligarh Muslim University, Aligarh*
Anas M, *Research Scholar, D/o Statistics & Operations Research, Aligarh Muslim University, Aligarh*
Firoz AHMAD, *Department of Management Studies, Indian Institute of Science (IISc), Bangalore*

Abstract: In supply chain management, the selection of suppliers is of vital importance and plays a valuable role in the performance of organizations. A hierarchical structure, with different levels in the selection of suppliers, is employed, wherein sequential decisions are made from the top level to the bottom level. Decision variables, called controlling factors, are partitioned among different levels. In the decision-making process, often because of the lack of confidence or uncertainty, It becomes hard for decision-makers to give explicit/crisp values to any parameter, resulting in uncertainty in the problem. In this paper, A multilevel supplier selection problem with fuzzy demand and supply is considered. To avoid decision conflicts, superior or upper-level decision makers provide possible relaxations in the form of tolerances. Thus, the problem is employed with fuzzy constraints. Based on a neutrosophic decision set, the neutrosophic compromise programming ap-

proach (NCPA) is used as a solution technique with the idea of indeterminacy degree as well as membership and non-membership degree of different objectives. Membership functions (Linear-type) are used to develop satisfactory solutions by fuzzily describing objective functions and controlling factors. A numerical example has been provided to demonstrate the validity and appropriateness of NCPA.

215. Adjustment for the uncertainty of predicted expression in transcriptome-wide association study: an efficient approach based on measurement error theory and bootstrapping

[28.M2.I42, (page 28)]

Arunabha MAJUMDAR, *Assistant Professor, Department of Mathematics, Indian Institute of Technology Hyderabad*
Arunabha MAJUMDAR, *Department of Mathematics, Indian Institute of Technology Hyderabad*
Tanushree HALDAR, *Department of Clinical Pharmacy, University of California San Francisco*

Transcriptome-wide association study (TWAS) is a promising approach to identifying novel genes associated with complex traits. It provides better biological insights into the gene-trait association than a genome-wide association study (GWAS). The method commonly used in a standard TWAS approach combines two subsequent regressions. In the first-step regression (in general penalized), we train a prediction model to estimate the genetic component of gene expression based on an external panel of expression data. Next, the prediction model is used to estimate the genetic component of expression (predicted expression) in separate GWAS data. Finally, in the GWAS data, we regress a trait on the predicted expression in the second step to test for an association between the gene and the trait. A significant criticism of the standard TWAS approach is that the uncertainty of the predicted expression is ignored in the second-step regression. It can lead the final inference on the gene trait association subject to an uncontrolled rate of false positives, reducing the reliability of the detected association signals. We propose a novel approach to adjust for the uncertainty of predicted expression in the TWAS approach. We borrow techniques from measurement error theory to derive the adjustment factor that needs to be incorporated in the test statistic obtained from the standard TWAS approach. Next, we apply bootstrapping techniques for penalized regression to estimate the adjustment factor. We use simulations to show that the standard TWAS approach inflates the type 1 error rate, whereas our adjusted TWAS approach adequately controls it. Due to an inflated false positive rate, the standard TWAS produces a higher power than the adjusted TWAS approach. In summary, we propose a TWAS approach which is a valid test for detecting gene-trait association while controlling for the false positive rate appropriately, contrary to the standard TWAS approach.

216. Stochastic Correlation modelling with von Mises Process

[Student Paper Competition 1, (page 13)]

Sourav MAJUMDAR, *Indian Institute of Management Ahmedabad*
Arnab Kumar Laha ,

The von Mises process is an Ornstein-Uhlenbeck type process on the circle, whose equilibrium distribution is the von-Mises distribution on the circle. We obtain an approximate analytical form of the transition density of this process. We consider estimation of the parameters of the diffusion from discretely observed von Mises process. We propose two methods for estimation using maximum likelihood and score matching. We consider a novel application of the von Mises process to stochastic correlation modelling in finance. We propose a stochastic correlation model and consider it's estimation. We illustrate the application of our model to examine the Indian FX and equity markets during the onset of the COVID-19 pandemic.

217. Statistical power and sample size requirements to detect an intervention by time interaction in four-level longitudinal cluster randomized trials

[26.A2.C3, (page 9)]

Priyanka MAJUMDER, *Postdoctoral Fellow*
Samiran GHOSH, *Professor, Department of Family Medicine and Public Health Sciences, Wayne State University, Detroit, MI, USA.*
Siuli MUKHOPADHYAY, *Professor, Department of Mathematics, Indian Institute of Technology Bombay, Mumbai, India*
Bo WANG, *Professor, Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts, USA*

Cluster/group randomized controlled trials

(CRTs) have a long history in the study of health sciences. CRT is a special type of intervention trial in which a complete group is randomly assigned to a study condition (or intervention). It is typically performed when individual randomization is difficult/impossible without substantial risk of contamination across study arms or prohibitive from the cost or group dynamics point of view. In this work, the aim is to design and analyze four-level longitudinal cluster randomized trials. The main interest here is to study the difference between treatment groups over time for such a four-level hierarchical data structure. This work is motivated by a real-life study for education-based HIV prevention. A random intercept mixed-effects linear regression, including a time by intervention interaction is used for modelling. Closed form expression of the power function to detect the interaction effect is determined. Sample size equations depend on the correlation among schools but not on correlations among classes or students, while the power function depends on the product of the number of units at different levels. Optimal allocation of units under a fixed cost by minimizing the expected standardized variance is also determined and is shown to be independent of correlations among units at any level. Results of detailed simulation studies find the theoretical power estimates based on the derived formulae close to the empirical estimates.

218. TBA

[28.A1.I51, (page 32)]

Tuhin MAJUMDER,

TBA

219. Model-based Clustering of Multiple Images Incorporating Covariates

[28.M1.I38, (page 25)]

Amita MANATUNGA, *Emory University*
Ying CUI, *Emory University*
Jeong Hoon JANG, *Yonsei University*

This work is motivated by the need to develop a non-invasive tool for screening anemia with large populations in under-developed world.The core idea is to leverage color intensity pixel matrices extracted from a photo of patients fingernails. To delineate the subpopulation structures of extracted color intensity pixel matrices according to varying anemia risks, we introduce a new method for clustering two-dimensional functional data (images). The proposed clustering method is based on a latent class functional mixed model that leverages the structural information of two-dimensional functional data using an appropriate basis system while effectively borrowing information across different subjects by placing random effects on basis coefficients. The method is particularly useful in modern data-rich applications, as it can simultaneously cluster multiple images while controlling for effects of available covariates (e.g., image metadata) on cluster membership. An efficient EM algorithm is derived to estimate the model parameters and cluster memberships. We further introduce a data-driven approach for choosing the appropriate number of clusters based on a distortion function adapted to our setting. We demonstrate the practical utility of the proposed method via extensive simulation studies and an analysis on the patient-sourced fingernail data.

220 . Modeling and Active Learning for Experiments with Quantitative-Sequence Factors

[27.M2.I20, (page 15)]

Abhyuday MANDAL, *University of Georgia*
Qian XIAO, *University of Georgia*
Xinwei DENG, *Virginia Tech*
Yaping WANG, *East China Normal University*

A new type of experiment that aims to determine the optimal quantities of a sequence of factors is eliciting considerable attention in medical science, bioengineering, and many other disciplines. Such studies require the simultaneous optimization of both quantities and the sequence orders of several components which are called quantitative-sequence (QS) factors. Given the large and semi-discrete solution spaces in such experiments, efficiently identifying optimal or near-optimal solutions by using a small number of experimental trials is a nontrivial task. To address this challenge, we propose a novel active learning approach, called QS-learning, to enable effective modeling and efficient optimization for experiments with QS factors. QS-learning consists of three parts: a novel mapping-based additive Gaussian process (MaGP) model, an efficient global optimization scheme (QS-EGO), and a new class of optimal designs (QS-design). The theoretical properties of the proposed method are investigated, and optimization techniques using analytical gradients are developed. The performance of the proposed method is demonstrated via a real drug experiment on lymphoma treatment and several simulation studies.

221. Assembling multiple data sources and machine learners to assess ambient PM2.5 in India and linking it with cardiovascular health

[27.M1.I18, (page 13)]

Siddhartha MANDAL, *Center for Chronic Disease Control, New Delhi, India*
Poornima PRABHAKARAN, *Public Health Foundation of India*
Petter LJUNGMAN, *Karolinska Institute*
Joel SCHWARTZ, *Harvard TH Chan School of Public Health*

India experiences both high levels of ambient air pollution and a high burden of cardiometabolic disease. However, there is a lack of detailed exposure assessment due to a sparse ground monitoring network, mostly concentrated in urban areas, and unavailability of longitudinal studies on health impacts of ambient PM2.5. Thus it is important to develop models that provide exposure estimates at high spatiotemporal resolution across India both retrospectively and prospectively, and study their links with health outcomes. In this talk, we will describe the air pollution epidemiology work carried out as part of the GEOHealth Hub (funded by Fogarty International Center, NIH) and CHAIR (funded by the Swedish Research Council) projects in India. To assess daily ambient PM2.5, we used a machine-learning-based approach by ensemble averaging across four different learners to develop a model calibrated against ground monitoring-based data from 1060 stations spanning 2008-2020. To counter the sparsity of available PM2.5, we also implemented a calibration model for the ratio of PM2.5 and PM10. Predictors encompassing all major areas such as meteorology, land-use patterns, satellite observations, emissions inventories and chemical transport models were utilized. Overall cross-validation (CV) was conducted by leaving out monitors and performance was assessed at the daily level.The overall daily CV-R2 was 0.86 ranging between 0.73-0.92 across years, while daily RMSE ranged between 14.1-25.4 g/m3. Among the learners, tree-based algorithms (gradient boosting, extreme gradient boosting and random forests) performed better than deeplearning and ensemble averaging further improved the predictive performance. Among predictors, meteorology, reanalysis based PM2.5 and elevation were more predictive by the tree based learners while urban-rural classification and road density were high-rank predictors by the deeplearning algorithm. Long-term exposure to PM2.5 assessed from the ensemble averaging based models have been linked to incident hypertension and type 2 diabetes in an ongoing cardiometabolic surveillance cohort in Chennai and Delhi. We observed a 1.53 (95% CI: 1.19-1.96) times increase and 1.22 (95% CI:1.09-1.36) times increased risk of developing hypertension and incident T2DM respectively, against 10 g/m3 difference in annual exposure to PM2.5, which clearly indicates the meaningful clinical and public health benefits of mitigating ambient PM2.5 in India. In summary, we have developed a valuable resource for air pollution epidemiology in a LMIC scenario with high levels of pollution as well as high burden of disease. This presents an opportunity to leverage existing, ongoing and future health studies in both urban and rural India to accurately assess the burden of ambient PM2.5 on a multitude of health outcomes.

222. Using data analytics to overcome challenges with clinical trials data and to improve patient engagement and experience

[29.M1.I57, (page 39)]

Jay MANDREKAR, *MAYO CLINIC*

This talk will focus on two aspects. First part will discuss some of the challenges encountered while conducting clinical trials. These will include setting up databases, recruitment of participants, missing data issues on key variables of interest. Some examples, on overcoming some of these challenges using novel data analytic techniques will be discussed from a Data Scientists perspective. Second part will illustrate how one can use novel data analytic techniques such as exploratory factor analysis and logistic regression for improving patient participation, engagement and experience in clinical research studies.

223. The design and implementation of precision medicine trials in oncology

[29.M1.I57, (page 39)]

Sumithra MANDREKAR, *Department of Quantitative Health Sciences, Mayo Clinic*

With the launch of the National Cancer Institutes Precision Medicine Initiative in 2015, there has been a shift in trial designs that tailor health care solutions to individual patients using a screening platform and move away from the one-trial / one biomarker at a time approach. Basket, and umbrella designs represent a class of novel designs for testing targeted therapeutics, and individualizing treatment in oncology. Umbrella trials include a central infrastructure for screening and identification of patients, and focus on a single tumor type or histology with multiple sub trials, each testing a targeted therapy within a molecularly defined subset. Basket trial designs offer the possibility to include multiple molecularly defined subpopulations, often across histology or tumor types, but included in one cohesive design to evaluate the targeted therapy in question. I will discuss the unique characteristics of these designs as well as the challenges encountered in the setting of an umbrella master protocol design using experiences from the ALCHEMIST study in lung cancer. Specifically, I will present the design considerations, as well as logistical details pertaining to the collection of clinical and correlative data for the master protocol and the sub-studies.

224. Markov Chain model for extremal PM2.5

[29.A2.C17, (page 48)]

BG MANJUNATH, *University of Hyderabad*

Extreme value theory for environmental process or meteorological data includes problems of short term temporal dependences. However, in extremal analysis of a such data often required computation (prediction) of the probabilities of events that are rarer than those observed and also prediction of its return periods. In the literature, modeling extremal pattern of the environmental process, such as air quality indices, wind speed, heavy rainfall and thereon, focues on the development of techniques that make adequate use of the information in the data and more robust to the extrapolation. In the current note we will be merely interested in modelling extremal pattern of the particulate matter (PM2.5) of size 2.5 micrometers, which is a key variable in predicting the air quality index(AQI). We will be considering Markovian approach of modelling threshold exceeds of PM2.5, which is unlike with the classical annaul maxima approach. We also exam the power of predicting return-value and return-period of the extreme events which are severe than observed under the Markovian modelling.

225. CONSTRUCTION OF STRUCTURALLY INCOMPLETE ROW-COLUMN DESIGNS

[28.M1.C9, (page 27)]

B MANJUNATHA, *Indian Agricultural Statistics Research Institute-New Delhi*
Rajender PARSAD, *Principal Scientist, ICAR-IASRI, Pusa, New Delhi*
B.N. MANDAL, *Senior Scientist, ICAR-IASRI, Pusa, New Delhi*
Sukanta DASH, *Scientist, ICAR-IASRI, Pusa, New Delhi*

Structurally Incomplete row-column (SIRC) are useful for two-way elimination heterogeneity settings wherein empty nodes do not represent the underutilized or wasted experimental units or two stage experiments wherein the blocking criteria are implemented in sequences. A subclass of SIRC designs known as Balanced Incomplete Latin Square (BILS) designs (number of rows, columns and treatments are equal) are useful for the situations wherein it may not be possible to accommodate all treatments in rows and columns and in which number of rows and columns are equal to number of treatments and there is equal number of empty nodes in each row and each column. Most of the BILS designs available in literature do not satisfy the property of pairwise balance and /or variance balance. Also in some experimental situations, the number of rows or columns may be more or less than the number of treatments and experimenter is interested in making all possible pairwise treatment comparisons with same variance. For such experimental situations, some methods of construction for obtaining pairwise and/or variance balanced SIRC/BILS designs have been obtained using symmetric BIB designs, union of two variance balanced SIRC/BILS designs or adding treatments in existing variance balanced SIRC designs.

226. Estimation of the Selected Treatment Mean in Two Stage Drop-the-Losers Design

[26.A2.C1, (page 8)]

MASIHUDDIN, *INDIAN INSTITUTE OF TECHNOLOGY KANPUR*
Neeraj MISRA,

We consider the two stage drop-the-losers design in which multiple subjects are further administered the adjudged promising treatment in the second stage of the design. In order to study performances of various estimators, we obtain some admissibility and minimaxity results for estimating the mean effect of the selected promising treatment, under the mean squared error criterion. The maximum likelihood estimator, which is the weighted average of the first and second stage sample means is shown to be minimax and admissible. We show that the uniformly minimum variance conditionally unbiased estimator (UMVCUE) of the selected treatment mean is inadmissible and obtain an improved estimator. In this process, we also derive a sufficient condition for inadmissibility of an arbitrary location and permutation equivariant estimator and provide dominating estimators in cases, where this sufficient condition is satisfied. The mean squared error and absolute bias performances of various competing estimators are compared via a simulation study. A real data example is also provided for illustration purpose.

227. Hybrid intelligence: An approach by combining human intelligence and Machine learning models for freight rate forecasting in the Dry bulk shipping market

[28.M1.C9, (page 27)]

Irene MATHEWS, *S&P Global*
Rajeev PANDEY, *S&P Global*

Machines, with their high computing speed and efficiency, and human intelligence, with its creativity and dynamic nature, have unique and, in some cases, mutually exclusive strengths and weaknesses in making predictions. When dealing with a highly volatile industry, such as the shipping market, that is influenced by demand for commodities, availability of vessels, climatic changes, geopolitical risks, variability in fuel costs, market sentiments, etc., many existing models (or combinations of models) may give forecasts based on their very complex understanding of the data but may not be effective in explaining the market and the influencing drivers at certain periods. At the opposite end of this complexity paradigm are very simple linear models, which offer a high degree of explanation but are poor in terms of accuracy. In this paper, we use a hybrid approach that combines human intelligence and various machine learning models to determine the dry bulk freight rate, describing how each of them is used sequentially so as to meet both the accuracy and the explanation criteria. For this study, we use the Baltic exchange index, BDI (Baltic dry index). It provides a benchmark for the prices (freight rates) of moving the major raw materials by sea. Freight rate forecasting is not only a matter of obtaining the future prices but also the ability of the model to explain the market in terms of drivers, seasonality, trends, sentiments, and other dynamic aspects. We are concentrating primarily on two frameworks, a monthly model and a weekly model process flow, which provide the BDI forecast for the next three years and three months, respectively. The paper describes the workflow, from assembling the key drivers for forecasting, developing models, and selecting results from all the outputs based on domain knowledge and market fluctuations. Each process unveils the form of intelligence that is being used. This would provide a clear understanding of why a hybrid framework, a two-way mutual learning system, is necessary while dealing with data and forecasting in volatile markets. On the one hand, a model is developed based on a thorough understanding of the domain and close observation of the market. On the other hand, the output derived from the model explains the major drivers, trends, and seasonality captured. This provides the right balance of accuracy and quantitative story telling in a volatile market with many moving parts.

228 . Urn Models with Irreducible, but Asymptotically Reducible, Replacement Matrices

[27.M2.I23, (page 16)]

Krishanu MAULIK, *Indian Statistical Institute*
Rohan SARKAR, *University of Connecticut, Storrs*

The urn models with irreducible replacement matrices are well understood. In this presentation, we consider a sequence of replacement matrices which are irreducible at every stage but converges to a matrix which is reducible. In particular, when there are two colors, the limiting replacement matrix can be taken as upper triangular. In this case, the count of the balls of the second color will continue to grow linearly. We provide a phase transition for the rate of growth of the count of the balls of the first color. We shall also indicate models with higher numbers of colors.

Joint work with Rohan Sarkar, University of Connecticut.

229 . Evaluation of a Clinical Decision Support System using a Machine Learning Based Predictive Model: A Stepped-Wedge Cluster-Randomized Trial

[29.M1.I55, (page 38)]

Madhu MAZUMDAR, *Icahn School of Medicine at*

Mount Sinai
Chen YANG, *Icahn School of Medicine at Mount Sinai*
Asem BERKALIEVA, *Icahn School of Medicine at Mount Sinai*
Deukwoo KWON, *Icahn School of Medicine at Mount Sinai*

Stepped-Wedge Cluster-Randomized Trials (SW-CRTs) are increasingly utilized for evaluating healthcare delivery interventions. Appealing features of SW-CRTs include having each cluster acting as their own control and not needing to withhold the intervention from any patient. However, the design and analysis of SW-CRT is complex and methodology is not available for detecting interaction effects. Detecting interaction effect is important for diversity research to examine differential effect between minority and nonminority groups. We present a power calculation formulation and apply to a SW-CRT evaluating a clinical decision support system intervention that incorporates mortality predictive tool data to prompt goals of care conversations for cancer patients at high risk of short-term mortality. The composite outcome is documentation of goals of care conversations (Yes/No) and level of aggressive care utilization at the end-of-life (High/Low). We hypothesize that minority patients will have differentially increased gain in outcome compared to the non-minority group. Sample size calculation needs to be performed on detection of interaction between minority status and intervention.

230 . Nonparametric Estimation for Combined Incident and Prevalent Cohort Data

[27.A2.I33, (page 20)]

James MCVITTIE, *University of Regina*

An electronic health records database can comprise a large number of variables that have been sampled using a variety of different techniques. One such example is time-to-event data that have been captured by an observation window yielding a combination of failure/censoring times comprised of ”prevalent cases” (times captured cross-sectionally at the start of the window) and ”incident cases” (times which began inside the window). Although these data are based on the same underlying time-to-event variable, they have been sampled using different mechanisms. We address three questions in this talk: (i) Are we justified in combining these independent data sets to model the underlying failure time survival function? (ii) If the data can be combined, how do we nonparametrically estimate the failure time survival function?, and (iii) Are there statistical benefits to combining the incident and prevalent cases into a single estimator? We examine the theory to address these questions and discuss a real-world example where this modelling procedure has be applied.

231. Data-Driven Selection of Fractional Differential Operator with Applications to Real Data

[Poster Session, (page 34)]

Dipali V. MESTRY, *Institute of Chemical Technology, Mumbai*
Pratik Singh, Amiya Ranjan Bhowmick ,

Fractional order differential equations (FDE) have been found to be a useful tool in many research areas such as biology, chemistry, ecology, epidemiology, etc. Multiple definitions of fractional differential operator (FDO) are available in the literature and the solution of a differential equation varies under different choices of definitions. Therefore, an appropriate selection of FDO is necessary for real data analysis. In this manuscript, we adopt a Bayesian model selection framework to select the appropriate definition for a given data set. We utilized the Gibbs sampler and Reversible Jump MCMC for computing the posterior model probabilities. For the simulation study, we consider fractional logistic differential equations under six different FDOs. The best definition is the one having the highest posterior model probability. We checked the robustness of the analysis using a simulation study under different parameter setups. The method has also been applied to analyze real data sets for predicting growth trajectories. All the computations have been carried out using R software.

232. Bayesian dimensionality reduction via the identifications of the data intrinsic dimensions

[28.M2.I44, (page 29)]

Antonietta MIRA, *Universit della Svizzera italiana and University of Insubria*
Denti FRANCESCO, *Universit Cattolica Milano*
Di Noia ANTONIO, *Universit della Svizzera italiana*
Alessandro LAIO, *laio@sissa.it*

Even if they are defined on a space with a large dimension, data points usually lie onto a hypersur-

face, or manifold with a much smaller intrinsic dimension (ID). The recent TWO-NN method (Facco et al., 2017, Scientific Report) allows estimating the ID when all points lie onto a single manifold. TWO-NN makes only a fairly weak assumption, that the density of points is approximately constant in a small neighborhood around each point. Under this hypothesis, the ratio of the distances of a point from its first and second neighbour follows a Pareto distribution that depends parametrically only on the ID, allowing for an immediate estimation of the latter. We have extended the TWO-NN model to the case in which the data lie onto several manifolds with different ID. While the idea behind the extension is simple (the Pareto distribution is just replaced by a mixture of K Pareto distributions), a non-trivial Bayesian scheme is required for correctly estimating the model and assigning each point to the correct manifold. Applying this method, which we dub Hidalgo (heterogeneous intrinsic dimension algorithm), we uncover a surprising ID variability in several real-world datasets. The Hidalgo model obtains remarkable results, but its limitation consists in fixing a priori the number of component in the mixture. To adopt a fully Bayesian approach, a possible extension would be the specification of a prior distribution for the parameter K. Instead, with even greater flexibility, we let K go to infinity, using a Bayesian Nonparametric approach and model the data with a Dirichlet Process Mixture Model as an infinite mixture of Pareto distributions. This approach, the same time, takes into account the uncertainty relative to the number of mixture components. Since the posterior distribution has no closed form, to sample from it we rely on the Slice Sampler algorithm. From preliminary analyses performed on simulated data, the model provides promising results.

233 . Aspects of modelling CoVID-19: Understanding and quantifying the uncertainty

[29.M1.I56, (page 38)]

Swapnil MISHRA, *University of Copenhagen*

Despite trends in modern medicine and epidemiological control, the risk for novel outbreaks and previously existing pathogens is currently greater than ever. Indeed, the current outbreak of SARS-CoV-2 has exposed the need for precise, robust, and principled mathematical modelling of disease outbreaks that can perform well with noisy and potentially biased data. To tackle these challenges, I will present a unifying view of modelling infectious diseases that contributes to the new understanding of the spread of the diseases and their epidemiological properties. The unified framework allows flexible probablistic models that are capable of fitting complex and noisy data from different sources. I will touch upon how the new unified framework, built using Stan (numpyro), has helped us to characterize the initial spread of SARS-CoV-2 and quantify the altered epidemiological characteristics of various 'variants of concerns (VOCs)'.

234. Modelling, Detection and Attribution of Large-scale Compound Extreme Events using Probabilistic Graphical Models

[29.M2.I65, (page 44)]

Adway MITRA, *Indian Institute of Technology Kharagpur*

As the threat of climate change looms all over the world, extreme climatic events such as droughts, extreme rainfall and heatwaves are expected to become more frequent and more intense. These extreme events often tend to occur concurrently, such as hot/dry extremes during summer, resulting in compound events with additional social damage. In such a situation, it is of special interest to study the nature of these extreme events and understand their changing characteristics such as location, spatial extents and durations, intensity and causes. Much of the statistical analysis of extreme events in the climate and hydrology communities is done location-wise, by setting thresholds like quantiles, return periods and standardized indices for specific meteorological variables like precipitation and temperature. However, these extreme events often cover wide areas and involve multiple variables. Moreover, one extreme event may trigger off others. Region-wise threshold-based analysis may not be suitable to identify such properties. We consider the use of Probabilistic Graphical Models which define binary latent variables, indicating whether or not there is an extreme event at a given spatio-temporal point, taking into account not only the measurements of different meteorological variables there, but also the presence/absence of extreme events in the neighbourhood, and relations between different types of extreme events. This approach not only allows us to avoid setting thresholds, but inferring the most likely values of these binary variables help us to identify the spatial and temporal extents of each event and relationships between different events. However, this also gives rise to significant computational challenges, especially if we are considering a large region. To circumvent these challenges, we need parallelized variational inference techniques. This approach can be especially useful to identify and localize previously undetected extreme events, and also study the properties of extreme events as simulated by different CMIP6 models under various future scenarios.

235. Estimation methods for elementary chirp model parameters

[Poster Session, (page 34)]

Anjali MITTAL, *Indian Institute of Technology Kanpur*
Prof. Rhythm Grover, Prof. Debasis Kundu, Prof. Amit Mitra ,

We propose some estimation methods to estimate the elementary chirp model parameters, which are encountered in sonar, radar, acoustics, and other areas. We derive asymptotic theoretical properties of least squares estimators (LSEs) and approximate least squares estimators (ALSEs) for the one-component elementary chirp model. It is proved that the proposed estimators are strongly consistent and follow the normal distribution asymptotically. We also suggest how to obtain proper initial values for these methods. The problem of finding initial values is a difficult problem when the number of components in the model is large, or when the signal-to-noise ratio is low, or when two frequency rates are close to each other. We propose sequential procedures to estimate the multiple-component elementary chirp model parameters. We prove that asymptotic properties of the sequential LSEs and sequential ALSEs coincide with the LSEs and ALSEs, respectively. To evaluate the performance of the proposed estimators, numerical experiments are performed. It is observed that the proposed sequential estimators perform well even in situations where LSEs do not perform well. We also compare the performance of these estimators with the estimators based on cubic phase function (CPF), product CPF (PCPF) and dechirping method. Real data analysis on a bat echolocation data is performed, illustrating the performance of the proposed sequential algorithm and validating the capability of the model to capture such real datasets.

236. A bootstrap control chart for some well known finite range distributions.

[26.A2.C1, (page 8)]

BIDHAN MODOK, *RAIGANJ UNIVERSITY*
Shovan CHOWDHURY, *Indian institute of Management Kozhikode, Kozhikode, India.*
Amarjit KUNDU, *Raiganj University, Raiganj, India*

Abstract: Modern day products are highly reliable due to advancements in technology and processes. It is difficult for any process monitoring scheme to observe large proportion of non-conforming items, which makes the use of average proportion not so relevant to set up the control limits for proportion data. For such instances, certain percentiles of proportions are worth investigating over the average of proportions. Here, a bootstrap control chart, known as bootstrap truncated beta control chart (BTBC chart) is set up for monitoring percentiles of proportion data following truncated beta distribution with unknown parameters. Extensive Monte Carlo simulations are carried out with various combinations of percentiles, false-alarm rates and sample sizes to evaluate in-control performance of the chart in terms of the average run length. The out-of-control performance of the chart is thoroughly investigated for several choices of shifts in the parameters of the distribution and is compared with beta distributed process. The proposed chart is applied to two skewed data sets for illustration and one application is provided from healthcare.

237. A statistical framework to study the spatial variation in imaging phenotypes with applications to radiology and pathology

[27.A2.I34, (page 21)]

Shariq MOHAMMED, *Boston University*
Santhoshi N. KRISHNAN, *Rice University and University of Michigan*
Timothy L. FRANKEL, *University of Michigan*
Arvind RAO, *University of Michigan*

We develop a statistical modeling framework to estimate the spatial interactions in imaging phenotypes through a geographically weighted regression (GWR). We leverage the model estimates/residuals from the GWR to construct a probability density function (serving as a signature quantifying spatial interactions). These density functions are then analyzed using a Riemannian-geometric framework. We develop permutation-based hypothesis tests for the density functions and classification models with the density functions as predictors. We present our methodology with applications to radiology and pathology imaging phenotypes. We employ our framework to quantify the T2-FLAIR mismatch signature, an important diagnostic sign in gliomas, and use it to predict the molecular status of lower-grade gliomas. We observe statistically significant differences between the averages of density signatures for different tumor categories. We will also present another application of our framework with pathology images in the context of pancreatic cancer.

238 . Bayesian Inference for Atmospheric CO2 Retrieval using emulators and a two-stage adaptive Monte Carlo method

[29.M1.I54, (page 38)]

Anirban MONDAL, *Case Western Reserve University*
Otto LAMMINPAEAE,

NASA's Orbiting Carbon Observatory-2 collects tens of thousands of observations of reflected sunlight daily, and the mission's retrieval algorithm processes these indirect measurements into estimates of atmospheric CO2 and other states. The retrieval is an inverse problem and consists of a physical forward model for the transfer of radiation through the atmosphere. Here we use a Bayesian approach, which casts this inverse solution as the posterior distribution of the state given the observed spectra. The posterior distribution is intractable hence MCMC based sampling method is used for Bayesian inference, which requires a large number of evaluations of the expensive forward model. Here we focus on the emulator approach where a statistical representation of the forward model is built based on some simulation runs of the forward model using the nearest neighbor Gaussian process. We use the functional principal component analysis approach to reduce the dimension of the high-dimensional functional radiance data and the active subspace approach to reduce the dimension of the input state vector. The emulator runs instantaneously resulting in a computationally efficient retrieval algorithm based on MCMC. Moreover, we use an efficient two-stage adaptive Monte Carlo method for the sampling-based inference, where the emulator is used in the first stage and the full physics-based forward model in the second stage. While the adaptive nature of the algorithm guarantees faster convergence of the chain and very good mixing properties, the two-stage approach helps in rejecting the bad proposals in the inexpensive first stage, making the algorithm computationally efficient. As the proposals are dependent on the previous states the chain loses its Markov property, but we prove that it retains the desired ergodicity property.

239 . Parallel Approximation of the Tukey g-and-h Likelihoods for Large-Scale Non-Gaussian Geostatistical Modeling

[28.A2.C11, (page 36)]

Sagnik MONDAL, *King Abdullah University of Science and Technology*
Sameh ABDULAH, *King Abdullah University of Science and Technology*
Hatem LTAIEF, *King Abdullah University of Science and Technology*
Marc GENTON, *King Abdullah University of Science and Technology*

Maximum likelihood estimation is an essential tool in the procedure to impute missing data in climate/weather applications. By defining a particular statistical model, the maximum likelihood estimation can be used to understand the underlying structure of given geospatial data. The Gaussian random field is one of the most popular models that has been widely used to describe geospatial data under the hood of maximum likelihood estimation. However, computation of Gaussian log-likelihood demands operations on a dense symmetric positive definite matrix, often parameterized by the Matern correlation function. This computation of the log-likelihood requires $O(n^2)$ storage and $O(n^3)$ operations, which can be a huge task considering that the number of geographical locations n can be enormous for some problems. However, despite its appealing theoretical properties, the assumptions of Gaussianity are unrealistic since real data usually show signs of skewness or have some extreme values. Herein, we consider the Tukey g-and-h (TGH) random field as an example of a non-Gaussian random field that shows more robustness in modeling geospatial data by including two more parameters to incorporate skewness and heavy tail features in the model. This work provides the first HPC implementation of the TGH random field's inference on parallel hardware architectures. Using task-based programming models associated with dynamic runtime systems, our implementation leverages high concurrency of current parallel systems. This permits to run the exact log-likelihood evaluation of the Tukey g-and-h (TGH) random fields for a decent number of geospatial locations. To tackle large-scale problems, we provide additionally an implementation of the given model using two different low-rank approximations. We compress the aforementioned positive-definite symmetric matrix for computing the log-likelihood and rely on the Tile Low-Rank (TLR) and the Hierarchical off-Diagonal Low-Rank (HODLR) matrix approximations. We assess the performance and accuracy of the proposed implementations using synthetic datasets up to 800K and a 300K precipitation data of Germany to demonstrate the advantage of using non-Gaussian over Gaussian random fields. Moreover, by relying on TLR/HODLR matrix computations, we can now solve for larger matrix sizes, while preserving the required accuracy for prediction. We show the performance superiority of TLR over HODLR matrix computations when calculating the TGH likelihoods and predictions. Our TLR-based approximation shows a speedup up to 7.29X and 2.96X on shared-memory and distributed-memory systems, respectively, compared to the exact implementation.

240. Bayesian nonparametric common atoms regression for generating synthetic matching populations

[28.A1.I48, (page 31)]

Peter MUELLER, *UT Austin*
Noirrit CHANDRA, *UT Dallas*
Abhra SARKAR, *UT Austin*
Ying YUAN, *MDACC*

We propose a model-based approach using nonparametric Bayesian common atoms models to create synthetic matching populations from available data. The model and method are developed in the context of single-arm, treatment only clinical trials. The single-arm cohort is complemented by a synthetic control arm that is created from readily available external data in the form of electronic health records (EHR). Although randomized clinical trials (RCT) remain the gold standard for approvals by regulatory agencies, the increasing availability of such real world data has opened opportunities to supplement increasingly expensive and difficult to carry out RCT's with evidence from readily available real world data.

We propose a novel nonparametric Bayesian common atoms mixture model that allows us to find equivalent population strata in the EHR and the treatment arm and then resample the EHR data to create equivalent patient populations under both, the single arm trial and the resampled EHR. Resampling is implemented as a density-free importance sampling scheme. Using the synthetic control arm inference for the treatment effect can then be carried out using any method available for RCTs. Alternatively the proposed nonparametric Bayesian model allows straightforward model-based inference. In simulation experiments, the proposed method vastly outperforms alternative methods. We apply the method to supplement single arm treatment-only glioblastoma studies with a synthetic control arm based on historical trials.

241. Towards Proving the Capacity of Deep Operator Nets for Solving (Partial) Differential Equations

[27.A1.I26, (page 18)]

Anirbit MUKHERJEE, *The University of Manchester*
Pulkit GOPALANI, *Indian Institute of Technology-Kanpur*
Sayar KARMAKAR, *UFlorida*

One of the core techniques of modelling nature is via (P)DEs and trying to understand their solutions goes to the heart of many of the hardest challenges in mathematics. "Physics Informed Machine Learning" largely revolves around using neural nets for numerically solving differential equations. Some of the recent methodological advances in this field include Physics Inspired Neural Nets (PINNs), Fourier Neural Operators (FNO), Deep Ritz Method (DRM), Deep Galerkin Method (DGM) - and maybe most interestingly Deep Operator Nets (DONs). DONs are a significant departure from conventional neural nets because their architecture involves taking an inner-product of two nets - and this surprisingly enables it to approximate maps between infinite dimensions. There is increasing evidence of the power of DONs in solving differential equations and in this work we start to advance the theory of generalization error bounds for DONs.

Our key contribution in our recent work is to uncover the first nearly size independent upperbound for the Rademacher complexity of DONs. Along the way we prove a new kind of a Talagrand contraction lemma which might be of independent mathematical interest - and we shall end with delineating certain open questions about empirical processes that our research leads us to ask.

This talk is based off our work in, https://arxiv.org/abs/2205.11359, with my undergrad intern Pulkit Gopalani (IIT-Kanpur) and Sayar Karmakar (UFlorida). An extended abstract of this work appeared at DeepMath 2022.

242. Convergence of Gibbs sampler in spatial location-scale mixture models under different centering and scaling parameterizations

[Poster Session, (page 34)]

Arghya MUKHERJEE, *Indian Institute of Technology, Kanpur*
Arnab Hazra ,

While generalized extreme value distribution is the only possible limit of re-normalized univariate block maxima, for point-wise block maxima of re-normalized stochastic processes, the only possible limit is called a max-stable process. Despite being theoretically sound, fitting max-stable processes in large-scale real data applications is challenging due to the high computation burden; the order of computations grows super-exponentially with the dimension of the problem. On the other hand, Gaussian processes (GPs) have been used widely for geostatistical modeling due to their several attractive theoretical and computational properties, and several variants of GPs are available in the literature that allow very high-dimensional spatial data analysis. However, GPs are often criticized for modeling spatial extremes due to their thin-tailed finite-dimensional marginal distributions. As a solution, spatial location and/or scale mixtures of GPs have been recently proposed by several researchers for modeling spatial extremal dependence, and most of those models are Bayesian that require Gibbs sampling. The layers of the hierarchical models are chosen arbitrarily in those studies, and the convergence diagnostics of the Gibbs sampling are mostly ignored. In this poster, we intend to demonstrate some theoretical and numerical results about the convergence of the Gibbs sampler for such hierarchical Bayesian location/scale mixtures of GPs. For location-mixture models, we study the differences between the centering (CP) and non-centering parameterizations (NCP), while for the scale-mixture models, differences between scaling (SP) and non-scaling parameterizations (NSP) are studied. Under certain realistic assumptions, we derive the analytic expressions of the Gaussian transition kernel of the conditional posteriors. Our results show substantial improvements in the convergence of the Gibbs sampler under CP and SP, in comparison with NCP and NSP.

243. Predictions, Role of Interventions and the Crisis of Virus in India: A Data Science Call to Arms

[29.M1.I56, (page 38)]

Bhramar MUKHERJEE, *University of Michigan*

India, world's largest democracy, had three very different surges of SARS-CoV-2 in 2020,2021 and 2022 corresponding to the transmission of the ancestral strain, the rise of the Delta variant and the final Omicron wave. The human behavior and public health intervention strategies were also very different during these three waves. In this presentation, we provide a brief chronicle of the modeling experience of our study team over the last two years, looking at the data from India, leading to the development of a tiered data-driven framework for public health interventions towards pandemic resilience. Through mathematical modeling we study the timing and duration of public health interventions with intervention effects estimated from actual data. We illustrate that early and sustained interventions can help us avoid harsh lockdowns and reduce COVID mortality drastically. We also quantify the estimated number of missing COVID-deaths in India which are orders of magnitude larger than reported COVID-deaths. This is joint work with many, with all supporting research materials and products available at covind19.org.

244. A Scalable Dynamic Bayesian Mixture Model for Fine-grained Promotion Mix Analysis of Digital Coupons

[28.A1.I52, (page 32)]

Gourab MUKHERJEE, *University of Southern California*
Rashmi BHUYIAN, *University of Southern California*
Wreetabrata KAR, *Purdue University*

We develop a novel dynamic mixture model for analyzing the effects of varied marketing components in a digital promotion campaign that uses online coupons. A key feature of the proposed model is that it segments customers based on their purchase history and provides fine-grained estimates of the heterogeneous effects that marketing mix variables have on the different consumer segments. The proposed model not only captures long-term heterogeneous segments in the customer pools but also tracks short term changes in customer engagement through dynamic indices that tabulates stocks of unresponded recent coupons. We conduct Bayesian estimation of the model parameters by using a novel Gibbs algorithm which is highly scalable due to the usage of Polya-Gamma distributions based data-augmentation strategy in handling Binomial likelihoods of customer responses to promotional coupons. Finally, through a path-algorithm we provide an integrated framework for providing fine-grained analysis of the marketing component effects at various levels of heterogeneity. We establish large-sample properties on the operational characteristics of the developed algorithm. We apply the proposed model on recent consumer responses data from the apparel industry and obtain encouraging results.

245. Image denoising and deblurring using jump regression and local pixel clustering

[29.A1.I70, (page 46)]

Partha Sarathi MUKHERJEE, *Indian Statistical Institute, Kolkata*

Image denoising and image deblurring are important tasks to be performed so that subsequent image analyses are reliable. Researchers from diverse backgrounds have been working on these for long. However, we still do not have good enough solutions so that we can stop working on such problems. In this talk, the primary emphasis will be on preservation of important image features such as edge-structures, other fine details, and so on. In this regard, an overview of many such attempts using jump regression analysis and local pixel clustering will be provided. Using the strengths of both, a lot of image details can be preserved while removing noise and blur from images.

246. Adaptive Estimation of Nonparametric Functionals

[27.M2.I19, (page 15)]

Rajarshi MUKHERJEE, *Harvard University*
Lin LIU, *Shanghai Jiao Tong University*
James ROBINS, *Harvard University*
Eric TCHETGEN TCHETGEN, *University of Pennsylvania*

We provide general adaptive upper bounds for estimating nonparametric functionals based on second-order U-statistics arising from finite-dimensional approximation of the infinite-dimensional models. We then provide examples of functionals for which the theory produces rate optimally matching adaptive upper and lower bounds. Our results are automatically adaptive in both parametric and nonparametric regimes of estimation and are automatically adaptive and semiparametric efficient in the regime of parametric convergence rate.

247 . A Deterministic Approximation Approach To The Continuum Logit Dynamic With An Application To Supermodular Games

[27.A2.C6, (page 22)]

Sayan MUKHERJEE, *Indian Statistical Institute, Kolkata*
Ratul LAHKAR, *Department of Economics, Ashoka University*
Souvik ROY, *Applied Statistics Unit, Indian Statistical Institute, Kolkata*

We consider the logit dynamic in a large population game with a continuum of strategies. The deterministic approximation approach requires us to derive this dynamic as the finite horizon limit of a stochastic process in a game with a finite but large number of strategies and players. We first establish the closeness of this dynamic with a step–wise approximation. We then show that the logit stochastic process is close to the step–wise logit dynamic in a discrete approximation of the original game. Combining the two results, we obtain our deterministic approximation result. We apply the result to large population supermodular games with a continuum of strategies. Over finite but sufficiently long time horizons, the logit stochastic process converges to logit equilibria in a discrete approximation of the supermodular game. By the deterministic approximation approach, so does the logit dynamic in the continuum supermodular game.

248. High Dimensional Logistic Regression Under Network Dependence

[28.M2.I47, (page 30)]

Somabha MUKHERJEE, *National University of Singapore*
Sagnik HALDER, *University of Florida*
Bhaswar BHATTACHARYA, *University of Pennsylvania*
George MICHAILIDIS, *University of Florida*

The classical formulation of logistic regression relies on the independent sampling assumption, which is often violated when the outcomes interact through an underlying network structure, such as over a temporal/spatial domain or on a social network. This necessitates the development of models that can simultaneously handle both the network peer-effect (arising from neighbourhood interactions) and the effect of (possibly) high-dimensional covariates. In this talk, I will describe a framework for incorporating such dependencies in a high-dimensional logistic regression model by introducing a quadratic interaction term, as in the Ising model, designed to capture the pairwise interactions from the underlying network. The resulting model can also be viewed as an Ising model, where the node-dependent external fields linearly encode the high-dimensional covariates. We use a penalized maximum pseudo-likelihood method for estimating the network peer-effect and the effect of the covariates (the regression coefficients), which, in addition to handling the high-dimensionality of the parameters, conveniently avoids the computational intractability of the maximum likelihood approach. Our results imply that even under network dependence it is possible to consistently estimate the model parameters at the same rate as in classical (independent) logistic regression, when the true parameter is sparse and the underlying network is not too dense. Towards the end, I will talk about the rates of consistency of our proposed estimator for various natural graph ensembles, such as bounded degree graphs, sparse Erdos-Renyi random graphs, and stochastic block models, which follow as a consequence of our general results. This is a joint work with Ziang Niu, Sagnik Halder, Bhaswar Bhattacharya and George Michailidis.

249. Bayesian Regularized Estimation of Latent Spatial Transmission of COVID-19.

[28.A1.I49, (page 31)]

Ujjal Kumar MUKHERJEE, *College of Business, and Carle Illinois College of Medicine, UIUC*
Snigdhansu CHATTERJEE, *School of Statistics, UMN*

Context: A distinguishing characteristic of COVID-19 from other epidemics such as Ebola and SARS is the presence of a relatively larger number of asymptomatic individuals leading to predominantly latent transmission of the disease. The detection and prediction of the latent spatial transmission of COVID-19 over a network of locations have been a key challenge to healthcare delivery organizations, and policymakers.

Research objective: In this paper, we propose a Bayesian graphical method with regularization for the estimation of the latent spatial transmission of COVID-19 over a network of interconnected locations. We model the latent spatial covariance of infection transmission as a Gaussian process on a graph with the interdependence of COVID-19 transmission across locations. Additionally, we use a Bayesian regularization of the parameters based on the horseshoe priors (half-Cauchy priors) to estimate the dominant paths of infection transmission. The parameters of the proposed model are estimated using a Markov Chain Monte Carlo (MCMC) method. We demonstrate the application of the proposed method for policies related to isolation and lockdown of locations to dampen the spread of the COVID-19 pandemic.

Contributions: Our primary contribution is the Bayesian methodological approach to an important policy problem related to COVID-19 management and the adoption of data-driven mitigation actions. The methodology proposed, and the application of the methodology on spatial isolation policies will help policymakers mitigate the geographic transmission of COVID-19. This study is intended to provide a data-driven interpretable and rational basis for such strategies in order to minimize the cost of mitigation strategies while dampening the spread of the infection. Unlike most black-box machine learning algorithms, we model the infection dynamics explicitly, which makes the methodological approach data-driven as well as interpretable. The proposed methodology has several operational and policy-level implications such as testing and vaccination-related resource allocation policies.

250. Bivariate DeepKriging for Large-scale Spatial Interpolation of Wind Fields

[Student Paper Competition 2, (page 26)]

Pratik NAG, *King Abdullah University of Science and Technology*
Ying sun, Brian J Reich ,

Wind energy, a well-known green energy source, contributes to the reduction of pollutants, which benefits public health. Setting up wind energy requires a proper spatial interpolation of the wind fields. A difficult task is however, large-scale spatial interpolation or downscaling of bivariate wind fields with velocity in two dimensions because wind data typically exhibits high spatial heterogeneity and non-Gaussianity. In spatial statistics, cokriging, which is commonly used for predicting bivariate spatial fields, is not optimal except for Gaussian processes. Additionally, cokriging is computationally prohibitive for large datasets. In this paper, we propose a method, called bivariate DeepKriging, which is a spatially dependent deep neural network (DNN) with an embedding layer constructed by spatial radial basis functions for bivariate spatial data prediction. Then, using ensemble DNN and bootstrap, we create a distribution-free uncertainty quantification approach. We use the bivariate DeepKriging approach to the wind data over the Middle East region at 506,771 points. The prediction performance of the proposed method is superior to the cokriging predictors and dramatically reduces computation time.

251. Understanding the failure modes of out-of-distribution generalization

[28.M2.I43, (page 29)]

Vaishnavh NAGARAJAN, *Google*
Anders ANDREASSEN,
Behnam NEYSHABUR,

Classifiers often rely on features like the background that may be spuriously correlated with the label. In practice, this results in poor test-time accuracy as the classifier may be deployed in an environment where these spurious correlations no longer hold. While many algorithms have been developed to heuristically tackle this challenge, in this talk, we will take a step back to ask: why do classifiers rely on spurious correlations in the first place? We'll see that the answer to this is not as straightforward as one might think. We'll see two distinct mechanisms by which classifiers fall into this trap, one "statistical" in nature and one "geometric" in nature. We'll wrap up the talk with some future directions about how we could use these insights to improve or propose new algorithms.

252. A Latent Class Trajectory Model for Penile Microbiome: Mapping longitudinal heterogeneity in response to baseline behavioral factors and disease status

[Poster Session, (page 34)]

Debarghya NANDI, *Adjunct Professor*
Supriya D. Mehta, Dulal Bhaumik ,

Several studies demonstrate the association of the penile microbiome with Bacterial vaginosis (BV) in female sex partners, but there is minimal evaluation of what shapes the penile microbiome. In this study, we characterize the longitudinal variation of penile microbiome and identified groups of individuals with similar trajectories, and baseline factors associated with varying trajectories. Our study uses data from a prospective cohort of 218 heterosexual Kenyan couples with 740 observations over a period of 1 year. To identify men with similar trajectories, we conducted univariate latent class trajectory analysis for the top 4 most abundant taxa using a zero inflated Poisson regression. For Corynebacterium, 75.5% of men exhibited a stable RA trajectory, while 24.5% of men exhibited a high RA trajectory that decreased over time. Men grouped in the high, decreasing RA trajectory were found to be more likely to be HIV infected (OR: 4.35) and less likely to be HSV-2 seropositive (OR: 0.42). We then applied a multivariate trajectory analysis to identify clusters of individual trajectories across multiple taxa. EM algorithm was used for parameter estimation and posterior probabilities were used to predict latent group memberships. Understanding of latent group trajectories and factors influencing them may lead to therapeutic avenues for better sexual health in men and their female sex partners.

253. Multiple testing applied to Structured Hypotheses

[29.A2.C17, (page 48)]

Shinjini NANDI, *Montana State University*

Modern statistical applications often involve large sets of hypotheses. These hypotheses may have an underlying structure, often implied by the experimental setup. Such structures provide important information regarding the distribution of signals, the dependence among the tests, order in which testing of the hypotheses is meaningful, etc. Evidently, use of such structural information aid the development of new multiple testing procedures that are more precise than pre-established procedures, both in terms of control on false discoveries and power. Complex classification of hypotheses can be realized in GWAS studies, neuroimaging data etc. where hypotheses can be simultaneously grouped by multiple criteria. The groups can occur hierarchically or simultaneously, or both. Utilizing such structural information can be challenging, nevertheless, vital in the search of improved multiple testing procedures. In this presentation, I will present an overview of my research in grouped structures of hypotheses and some ideas of my ongoing research involving other structured hypotheses.

254. A Psychometric Analyses of the Concise Health Risk Tracking- Self-Report A Measure of Suicidal Risk: Performance in Adolescent Outpatients

[28.M1.I38, (page 25)]

Karabi NANDY, *University of Texas Southwestern Medical Center*
Augustus John RUSH, *Curbstone consultant LLC*
Madhukar TRIVEDI, *University of Texas Southwestern Medical Center*
Thomas CARMODY, *University of Texas Southwestern Medical Center*

Objectives: The Concise Health Risk Tracking Self-Report (CHRT-SR) is a self-reported survey instrument that is meant to measure the risk of suicidal behavior. In this work, we explore its psychometric properties in a representative sample of adolescent outpatients.

Methods: A sample (n=657) of adolescents (¡ 18 years of age) in primary or psychiatric care completed the 14-item version of the CHRT-SR at baseline and again within three months. To identify an optimal, briefer solution for the scale, we evaluated the factor structure of CHRT-SR using multigroup confirmatory factor analysis, and testing measurement invariance across age and gender. The item response theory (IRT) and classical test theory (CTT) characteristics of the optimal solution were further evaluated. Concurrent validity (both cross-sectional and as a change measure over time) of the optimal solution was assessed by comparing it to an independent suicide measure, the PHQ-9 suicide item # 9.

Results: Confirmatory factor analysis and measurement invariance analyses identified the 9-item CHRT-SR as the optimal solution. Spearman-Brown coefficient was 0.80 at first visit and 0.86 at second visit. CTT revealed corrected item-total correlations between 0.58 and 0.80. IRT analyses revealed excellent item performance and a unifactorial instrument. Cross-sectional and change over time concurrent validity analyses that compared the CHRT-SR9 with responses to item 9 of the PHQ-9 revealed that it can measure both the improvement and worsening of suicidality over time.

Conclusion: The CHRT-SR9 is a brief self-report with excellent psychometric properties in a sample of adolescents that is sensitive to changes in suicidality over time. Its performance in other populations and ability to predict future suicidal events deserves study.

255. Estimating the number of components in functional MRI (fMRI) data for Independent Component Analysis (ICA) using Entropy

[28.M2.I47, (page 30)]

Rajesh NANDY, *UNT Health Science Center*

Since its introduction, ICA has become an extremely popular approach for analyzing fMRI data, as it is a data driven approach which does not require the a priori definition of the response in a regression type model and is able to capture both spatial and temporal inter-subject variability. However, in fMRI data, the observed data typically has a dimension higher than the number of independent components (ICs) and it has to be estimated using information theoretic criteria. Unfortunately, such an approach is valid only when the noise is weak or white. Neither is usually true for real data and existing methods generally perform poorly. In this work, a entropy-based approach is proposed that offers a more accurate estimation of the true dimension of the observed data.

256. Bayesian modelling of volatility using R and Stan

[29.M1.C15, (page 40)]

Muhammed NAVAS T, *Research scholar*
Athar Ali KHAN, *Professor*

This study uses modelling and model comparison to compare three widely used GARCH models with their stochastic volatility (SV) counterparts in modelling the dynamics of inflation rates using the Bayesian approach. BRICS country consumer price index (CPI) data is used to assess these models. We discover that the stochastic volatility models generally outperform the GARCH models. The stochastic volatility in mean (SV-L) model is also demonstrated to be the most effective for the BRICS nations that we took into consideration. The article also looks at which model attributes are crucial in simulating inflation rates. It turns out that when modelling inflation rates, inflation volatility feedback is an important component to take into account. For each of the five countries we took into consideration, the SV-L outperforms all other models. The study was done in Rstan, which is a programming language for statistical inference, and the simulation uses the Hamiltonian Monte Carlo (HMC) algorithm of Markov Chain Monte Carlo (MCMC) to sample from the posterior distribution.

257. Forecasting Agricultural Price Series using Multilayer Perceptron Neu-

ral Network and Metaheuristic Algorithms

[28.A2.C13, (page 37)]

Harish NAYAK.G.H., *Ph.D. Scholar at ICAR-IASRI, New Delhi*
Mohammad Wasi ALAM, *Principal Scientist at ICAR-IASRI, New Delhi*
G AVINASH, *PhD Scholar at ICAR-IASRI*
Mrinmoy RAY, *Scientist at ICAR-IASRI*

Food security management in agriculture-dominated developing countries like India requires efficient and reliable agricultural commodity price forecasting models more than ever before. Hence, forecasts of agricultural commodity prices are intended to help farmers, policymakers, and agribusiness industry. Statistical models showed several drawbacks in modeling accurately the complex patterns and volatile behaviours of agricultural prices driven by numerous factors. This can be solved by some extent using Machine Learning (ML) models but they experience problems during training process such as feature extraction, over-fitting or trapping in searching for global optima. To overcome these limitations, the current study attempts to hybridize the recently developed nature-inspired swarm-based metaheuristic algorithms (MHAs) such as Artificial Bee Colony (ABC), Ant Colony Optimization (ACO), Firefly Algorithm (FFA), and Bat Algorithm (BA) with Multi-layer Perceptron (MLP). The accuracy of these models will be inspected to solve the agriculture price series forecasting problem where the performance of proposed models will be compared with two traditional neural network models viz., Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM), and a linear model namely Autoregressive Integrated Moving Average (ARIMA) model. Result revealed that the newly proposed hybrid models show greater accuracy than the other competitive models. The study concludes that augmenting MHAs with MLP can be a reliable tool in forecasting the agriculture price series data with a high level of precision, fast convergence, and high constancy level.

258. Pitfalls of Deep Learning

[28.M2.I43, (page 28)]

Praneeth NETRAPALLI, *Google Research India*

While deep neural networks have achieved large gains in performance on benchmark datasets, their performance often degrades drastically with changes in data distribution encountered during real-world deployment. In this work, through systematic experiments and theoretical analysis, we attempt to understand the key reasons behind such brittleness of neural networks in real-world settings and why fixing these issues is exciting but challenging.

259 . Optimal Multivariate Crossover Designs

[27.M1.C5, (page 14)]

Shubham NIPHADKAR, *niphadkarshubham@gmail.com*
Siuli MUKHOPADHYAY, *Indian Institute of Technology Bombay*

In this work, we study optimal crossover designs with multiple responses measured in each period. This work is motivated by a 33 crossover trial studying the effect of two single doses of oral prednisone with placebo on biomarkers of mucosal inflammation. To model the outcomes of multiple variables in a single period, a multivariate fixed effect model is used. Three different types of dispersion matrices are considered for the error terms. In the first type, we assume that the error terms are homoscedastic and uncorrelated. In the second type of error dispersion matrix, we assume that the error terms corresponding to the observations only on the same response variable are homoscedastic. The response variables are assumed to be uncorrelated. In the third type, we assume that corresponding to the observations on the same response variable, the error terms within each subject are correlated. However, the error terms corresponding to the observations on any two different response variables are uncorrelated. Corresponding to these three types of dispersion matrix of the error terms, the joint information matrix for the direct and carryover effects, the information matrix for the direct effects and the information matrix for the carryover effects are obtained. Various results regarding these information matrices are studied. For designs having t treatments and p periods with p=t3, we study A-optimal, D-optimal and E-optimal balanced uniform designs for the first two dispersion matrix structures, and for the third structure in which the error terms are correlated, we study A-optimality, D-optimality and E-optimality of designs represented by Type I orthogonal array. It is noted that the proposed multivariate model gives rise to different design optimality results as compared to the univariate model case.

260. A Conditional Bayesian Approach with Valid Inference for High Dimensional Logistic Regression

[Student Paper Competition 1, (page 13)]

Abhishek OJHA, *University of Illinois, Urbana-Champaign*
Naveen N. Narisetty ,

We consider the problem of performing inference for a continuous treatment effect on a binary outcome variable while controlling for high dimensional baseline covariates. We propose a novel Bayesian framework for performing inference for the desired low dimensional parameter in a high dimensional logistic regression model. While it is relatively easier to address this problem in linear regression, the nonlinearity of the logistic regression poses additional challenges that make it difficult to orthogonalize the effect of the treatment variable from the nuisance variables. Our proposed approach provides the first Bayesian alternative to the recent frequentist developments and can incorporate available prior information on the parameters of interest, which plays a crucial role in practical applications. In addition, the proposed approach takes advantage of an adaptive orthogonalization as opposed to a fixed orthogonalization of frequentist methods to improve performance. We provide uniform convergence results that show the validity of credible intervals resulting from the posterior. Our method has competitive empirical performance when compared with state-of-the-art methods.

261. Asymptotic analysis of stochastic differential equations with small random noise and fast periodic sampling

[29.A1.I69, (page 46)]

Chetan PAHLAJANI, *Indian Institute of Technology Gandhinagar*
Shivam DHAMA, *IIT Gandhinagar*

We study the dynamics of an ordinary differential equation (ODE) in the presence of small Brownian perturbations and fast periodic sampling. By the latter, we mean a scenario where the vector field governing the ODE depends not just on the instantaneous value of the state of the system, but also on its most recent measurement. For the resulting continuous-time stochastic process indexed by two small parameters, we obtain effective ordinary and stochastic differential equations (SDE) describing the mean behavior and the typical fluctuations about the mean in the fast sampling, small noise limit. The limiting fluctuation process is found to vary depending on the relative rates at which the two small parameters go to zero. When the two parameters are comparable, the limiting SDE is found to have an effective drift term which captures the cumulative effect of the fast sampling. In this regime, our results yield a time-inhomogeneous Markov process which provides a strong (pathwise) approximation of the original non-Markovian process. We also discuss connections of these calculations with the performance analysis of feedback control systems with sampling effects. This is joint work with my student Shivam Dhama.

262. Quantitative Decision Making in Early Stages of Drug Development

[26.A2.I10, (page 7)]

RUPAM PAL, *GSK Pharmaceutical*

The pharmaceutical companies are currently experiencing huge challenges to fund their research and development due the increased cost and increased numbers of programs failing at the later stage of the development. As a result, it has become extremely important to take key decisions on the progress of a molecule at different stages of the development based on evidence and science. To address this issue, Quantitative decision making (QDM) has become an integrated part of the decision- making strategy for these companies. QDM enables development plans of a molecule and clinical trials to be designed optimally by quantifying and minimizing the risk and reducing the confirmation bias. The most popular approach to implement the QDM framework is through predictive inference using Bayesian paradigm. But using this methodology can be challenging at an early stage of the development, when we dont have a lot of clinical data or knowledge on the molecule. The focus of this presentation is how to use predictive inference in such scenarios to develop the QDM framework for the molecule to help the decision-making bodies take the best possible decision.

263. Preventive Maintenance of a Machine: A statistical approach to model and analysis of recurrent events from machine logs to predict failures

[26.A1.I4, (page 4)]

Sandip PAL, *S&P Global*
Pritam RANJAN, *IIM Indore*
Arnab KOLEY, *IIM Indore*

Debasis KUNDU, *IIT Kanpur*

In recent years, the requirement for real-time understanding of machine behavior has become an important objective in industrial sectors to reduce the cost of unscheduled downtime and to maximise production with expected quality. The vast majority of high-end machines are equipped with a number of sensors that can record event logs over time. This machine log data is accessible and stored in a database in real time, allowing for further analysis. In this article, we have analyzed machine log data with multiple discrete event states. Failure of the machine necessitates maintenance, after which it becomes operational again. We have designated maintenance time as machine downtime. Every maintenance state is followed by three distinct states: ”Running with issues,” ”Running without issues,” and ”Failure.” The sensors are also capable of capturing the corresponding operational conditions of the machine as well as the defined states of events. These operational conditions are external factors, or covariates, of the machine. Several reliability studies have been conducted in order to construct a failure prediction model using this type of dataset. However, very little research has been conducted to comprehend the statistical distribution of event sequences. This paper presents a model for predicting the time to failure based on different distributional assumptions for these three states. We've also developed a Cox-Proportional hazard model to predict the machine's failure time. The proposed method can help reduce downtime costs by controlling operational parameters in advance to prevent failures from occurring too soon.

Keywords: Cox proportional hazard model, principal component analyses , maximum likelihood estimate, discrete events sequence, failure event prediction.

264 . Model-based Personalized Synthetic Magnetic Resonance Imaging

[Student Paper Competition 2, (page 26)]

Subrata PAL, *Department of Statistics, Iowa State University*
Somak Dutta, Ranjan Maitra ,

Synthetic Magnetic Resonance (MR) imaging predicts images at new design parameter settings from a few observed MR scans. Model-based methods, or methods that use both the physical and statistical properties underlying the MR signal and its acquisition, can predict images at any setting from as few as three scans, allowing it to be used in individualized patient and anatomy-specific contexts. However, the estimation problem in model-based synthetic MR imaging is ill-posed and so regularization, in the form of correlated Gaussian Markov Random Fields, is imposed on the voxel-wise spin-lattice relaxation time, spin-spin relaxation time and the proton density underlying the MR image. We develop theoretically sound but computationally practical matrix-free estimation methods for synthetic MR imaging. Our evaluations show excellent ability of our methods to synthetize MR images in a clinical framework. An added strength of our model-based approach, also developed and illustrated here, is the accurate estimation of standard errors of regional means in the predicted images.

265. Inflated binomial and Beta Binomial Model to study the Distribution of Rural Out migrants

[26.A2.C1, (page 8)]

Richa PANCHGAUR, *Research Scholar*
Alok KUMAR,

Although migration from the home is a significant aspect of demography, it is less well-studied than fertility and death. Migrationboth domestic and internationalhas become a bigger issue for researchers and other sociologists as a result of the rise in family and individual needs. Migration from rural regions has been a key topic of interest for social scientists as well as program designers in developing nations, notably in India where 70% of the population still resides in villages (Census 2011). India's migration speed and pattern have evolved during the past few decades. The difference between rural and urban areas, particularly in the previous two decades, has progressively grown and is more pronounced in some regions and states. To understand the pattern of risk associated with adult outmigration, an attempt has been made in the current study to examine trends in rural adult outmigration at the household level for rural regions. Inflated binomial and beta-binomial distribution has been used to explain the occurrence and it has been applied to the observed distribution of migrants from their homes. It was discovered that inflated binomial and beta-binomial distributions may properly describe the migratory trend under specific assumptions. Planners and policymakers may find the findings useful in developing more equal and successful rural and urban programs.

266 . Generating Optimal Designs for Mixture Experiments using Genetic Algorithm

[27.M1.C4, (page 13)]

Mahesh Kumar PANDA, *Associate Professor of Statistics, Department of Statistics, Ravenshaw University, Cuttack, Odisha*

Abstract: Mixture experiments play a vital role in many fields including industrial processes such as chemical and pharmaceutical drug formulation, textile fiber production, food processing, etc. In a standard mixture experiment, the factors are the components of a mixture, and the response is assumed to depend only on the relative proportion of the ingredients present but not on the total amount of the mixture. For a q-component mixture, the ingredient proportions (xi, i = 1, 2, q) must be lies between 0 and 1 and sum to unity. In the present work, we propose and develop a genetic algorithm for generating A- and R-optimal designs where the experimental region is an irregularly shaped polyhedral region. The proposed method does not require the selection of points from a user-defined candidate set of mixtures and is even flexible for obtaining A-and R-optimal designs in a continuous region that includes highly constrained mixture regions.

Keywords: Mixture experiment, A-optimal design, R-optimal design, Genetic algorithm, Polyhedral region.

267. Approximate selective inference via maximum likelihood

[27.A2.I34, (page 21)]

Snigdha PANIGRAHI, *University of Michigan*

Querying the data has become a fairly common practice for anyone who wishes to learn a model from a range of different candidates. Naively using the same data twice, first to learn a model and then infer for the selected parameters, tends to inflate their estimated effects. The result is an increased chance of finding a statistically significant result when the selected variable in fact has no effect. Several strategies have been developed recently to ensure valid inferences after learning a model: some of these are easy to compute, while others fare better in terms of inferential power. In this talk, we discuss a new method for selective inference through approximate maximum likelihood estimation. Our goal is to: (i) efficiently utilize hold-out information from selection with the aid of randomization, (ii) bypass expensive MCMC sampling from exact conditional distributions that are hard to evaluate in closed forms. At the core of our new method is the solution to a simple convex optimization problem in a few dimensions. .

268. Balancing on the Leaves: Exploring Causal Roots of Decision Tree Algorithms

[29.A2.I73, (page 47)]

Bhuvanesh PAREEK, *IIM Bangalore*
Bikram KARMAKAR, *University of Florida*

This paper proposes a decision tree-based nonparametric method, which is agnostic to the outcome model, to estimate the average treatment effect in observational settings. We show that by leveraging the building idea of popular decision trees, we can achieve conditional independence between a self-selected multi-level treatment and the potential outcomes. We prove that the leaf nodes of a decision tree built using the joint distribution of the covariates and the observed treatments act as a balancing score. Conditioning on this discrete score creates a stratification where each stratum resembles a randomized experiment or captures a non-overlapping region. Additionally, this balancing score based causal inference method achieves exact balance asymptotically. For a finite sample, we provide an error bound on the estimated average potential outcome. Through simulation studies, we show that our method provides better covariate balance and more efficient treatment effect estimates than popular state-of-the-art methods. Finally, we apply the proposed method in two empirical studies of practical interest.

269 . Likelihood-based inference for skewed responses in a crossover trial setup.

[27.A2.C8, (page 23)]

Savita PAREEK, *IIT Bombay*
Siuli MUKHOPADHYAY, *IIT Bombay*
Kalyan DAS, *IIT Bombay*

This work proposes a statistical model for crossover trials with multiple skewed responses measured in each period. A 3*3 crossover trial data where different doses of a drug were administered to subjects with a history of seasonal asthma rhinitis to grass pollen is used for motivation. In each period, gene expression values for ten genes were measured from each subject. It considers a linear mixed effect

model with skew normally distributed random effect or random error term to model the asymmetric responses in the crossover trials. The paper examines cases (i) when a random effect follows a skew-normal distribution, as well as (ii) when a random error follows a skew-normal distribution. The EM algorithm is used in both cases to calculate maximum likelihood estimates of parameters. Simulations and crossover data from the gene expression study illustrate the approach.

270. A Double Machine Learning Approach to Combining Experimental and Observational Data

[29.A2.I73, (page 47)]

Harsh PARIKH, *Duke University*
Marco MORUCCI, *NYU*
Vittorio ORLANDI, *Duke University*
Alexander VOLFOVSKY, *Duke University*

Experimental and observational studies are both essential approaches for investigating causal questions. Experiments are prized for random assigned treatment being independent of potential outcomes, allowing for unbiased estimation of in-sample treatment effects. However, an experimental sample may not be representative of the population, violating external validity. On the other hand, while observational studies generally satisfy external validity, conditional ignorability may be violated due to the presence of unobserved confounders. In this work, we consider a setting where we have access to both an experiment and an observational study, and aim to leverage both to infer the precision and generalizability of our causal analysis. We propose formal tests for violations of (i) conditional ignorability in the observational data and (ii) external validity of the experimental data. Further, we introduce an estimator for the population average treatment effect (PATE) that draws strength from both samples and is *resilient* to violations of either of these two assumptions. Our estimators are asymptotically normal and unbiased and build on double machine learning approaches, efficiently using ML methods that can be selected depending on the data at hand. We apply our methodology on synthetically generated data as well as on two real-world case studies: (i) Project STAR data to estimate the impact of small class sizes in early schooling on future test scores and (ii) the Coronary Artery Surgery Study to study the effect of coronary bypass surgery on mortality in patients with ischemic heart disease. Overall, our framework allows practitioners to take advantage of both experimental and observational data at their disposal. Furthermore, our methods resilience to violations of certain assumptions makes it more applicable in real-world settings.

271. Robust probabilistic inference via a constrained transport metric

[28.A1.I48, (page 31)]

Debdeep PATI, *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 article, we offer a novel getaround 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. Importantly, our novel formulation of the modified Wasserstein metric enjoys great computational simplicity, exploiting the Sinkhorn regularization of discrete optimal transport problem, and being inherently parallelizable. We demonstrate superior performance of our methodology when compared against the state-of-the-art robust Bayesian inference methods. Moreover, 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 fields of entropy-based portfolio allocation and group fairness in machine learning problems.

272. Estimation of spectra of linear processes in high dimensions

[27.M1.I15, (page 11)]

Debashis PAUL, *University of California, Davis*
Jamshid NAMDARI, *University of California, Davis*
Alexander AUE, *University of California, Davis*

We propose a method for estimating the joint spectra of the coefficient matrices of a class of linear processes in a high-dimensional setting, where the dimension and the sample sizes are proportional. The class of processes under consideration is characterized by the fact that the coefficients, and the covariance matrix of the innovations, are simultaneously diagonalizable in an unknown orthogonal basis. The

estimation procedure leverages the existence of limiting spectral distributions of weighted sample periodograms of the process through minimization of an appropriate loss function. Consistency of the estimation procedure is established under the assumption that the support of the joint spectral distribution of the coefficients is known. We also provide an estimate of the coefficient matrices themselves by borrowing information from the sample eigenprojections. The method illustrated through numerical simulations and an application to data on daily stock prices.

(This is a joint work with Jamshid Namdari and Alexander Aue)

273. Wavelets and Machine Learning Techniques for Forecasting Time Series

[30.M1.I76, (page 51)]

Ranjit Kumar PAUL, *ICAR-Indian Agricultural Statistics Research Institute*

It is generally assumed that the time series is made of two things: actual signal and noise. It is difficult to extract actual signal from noisy data. Moreover, observed time series may contain both linear and nonlinear pattern. Therefore, a single model may not be suitable to satisfactorily express the underlying relationship. Wavelet decomposition is one of the powerful nonparametric techniques which helps in de-noise the series and decomposing it at different frequency levels. The presence of high chaotic nature and nonlinear structure in the noisy data sometimes distorts the model specification. Wavelet decomposition is one of the proven efficient tool in extracting actual signal from noisy data. Each decomposed sub series contains different frequency information. After decomposing the series, different models/techniques may be applied to individual sub-series based on the pattern existed in it. In the present investigation, an attempt has been made to apply wavelet based machine leering techniques as well as stochastic models for forecasting prices of different agricultural commodities in India. Different stochastic models viz. Autoregressive integrated moving average (ARIMA) model, Generalized autoregressive conditional heteroscedasitc (GARCH) model and machine learning techniques viz. artificial neural network (ANN), support vector regression (SVR) and random forest (RF) have been applied to each of wavelet decomposed series. Finally prediction is obtained by means of inverse wavelet transform. Different wavelet filters have been used for implementing the above algorithm. The prediction accuracy of different models has been empirically compared in terms of four criteria namely Mean error (ME), Mean absolute error (MAE), Root mean square error (RMSE) and Mean absolute percentage error (MAPE). The improvement over conventional techniques is obtained to a great extent by using wavelet-based combination approach with ML technique as exhibited through empirical evidence.

274. Modeling Temporal Networks of Relational Events Data

[28.A1.I51, (page 32)]

Subhadeep PAUL, *The Ohio State University*

Continuous-time temporal networks or relational events data are commonly encountered in several application problems, including online social media communications, human mobility, financial transactions, and international relations. Such datasets consist of directed instantaneous interaction events among entities at specific time points. For example, in online social media, users interact with each other through events that occur at specific time instances such as liking, mentioning, commenting, or sharing another user's content. In international relations and conflicts, nations commit acts of hostility or disputes through discrete time-stamped events. The relational events data often exhibit community structure and strong dependence patterns through mutual excitations among node pairs. We will introduce statistical models and methods for analyzing such datasets combining network models and multivariate point processes. We will also describe scalable estimation methods and study the asymptotic properties of the estimators. Finally, we will demonstrate that the models can fit several real datasets well and predict temporal structures in those datasets.

275. DocLens: Digitizing hard-to-read handwritten healthcare documents

[28.M2.I44, (page 29)]

Sujoy PAUL, *Google Research*

”What was the last time I took an antibiotic? What was my SGPT level a year back?” - these are questions which become important for diagnosis, early prediction and so on. People in developing countries often do not have proper records to answer such questions about their own personal health, because clinical documents in these countries are mostly handwritten or printed without any digital record. After its use, these documents are not stored properly, often get lost, thus rendering them useless. The ability to answer previous questions about a persons health is of paramount importance for a wide range of health related analysis from diagnosis, early prediction, local and global health analysis and so on. Our project, DocLens, aims at digitizing medical documents. In this talk, we are primarily going to talk about medicine name extraction from hard-to-read handwritten prescriptions of Indian doctors. The challenge here is unstructured documents, poor handwriting, and scarcity of unlabeled or labeled data to train machine learning models.

276. Semiparametric regression analysis of doubly censored recurrent event data

[26.A1.I1, (page 3)]

P. G. SANKARAN, *Cochin University of Science and Technology*
Hari S., *Cochin University of Science and Technology*
Sreedevi E. P., *Maharaja's College, Ernakulam*

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.

277. Combinatorial games with “jump” moves on rooted Galton-Watson trees

[26.A1.I2, (page 4)]

Moumanti PODDER, *Indian Institute of Science Education and Research (IISER) Pune*
Dhruv BHASIN, *Indian Institute of Science Education and Research (IISER) Pune*

A *normal* game on a rooted Galton-Watson (GW) tree involves two players and a token. The tree itself is visualized as a directed graph where an edge $\{u, v\}$ between a parent vertex u and a child v is directed from u to v. The players take turns to move the token along these directed edges, and the first player to fail to make a move (i.e. the first player to arrive at a leaf vertex) loses the game. The probabilities of win and loss (for the player who plays the first round of the game) and that of the game ending in a draw have been investigated by Holroyd and Martin (2021). We study a significantly more complicated generalized version of this game: fixing a $k \in \mathbb{N}$ *a priori*, we let each player, when it is her turn, move the token from its current vertex u to a vertex v such that v is a descendant of u and v is at distance at most k away from u. We study the probabilities of the game's outcomes in this new set-up. The probability that the first player losing the game is expressed as the minimum positive fixed point of a function that depends on k and the offspring distribution of the GW tree. For each fixed k, when the offspring distribution is Poisson(λ), we show that the probability of the game ending in a draw is strictly positive for all λ sufficiently large. We also show that for $k = 2$, there is a unique critical $\lambda\_c \approx 2.41$ such that the draw probability is 0 for all $\lambda < \lambda\_c$ and it is strictly positive for all $\lambda > \lambda\_c$. We strongly believe that in fact, this phase transition is continuous. We also discuss some results pertaining to the average duration of the game for small values of k.

278. An Integrated Bayesian Framework for Multi-omics Prediction and Classification

[Poster Session, (page 35)]

Anupreet PORWAL, *University of Washington*
Himel Mallick, Erina Paul, Satabdi Saha, Vladimir Svetnik ,

With the growing commonality of multi-omics datasets, there is now increasing evidence that integrated omics profiles lead to more efficient discovery of clinically actionable biomarkers that enable better disease outcome prediction and patient stratification. Several methods exist to perform host phenotype prediction from cross-sectional, single-omics data modalities but decentralized frameworks that jointly analyze multiple time-dependent omics data to highlight the integrative and dynamic impact of repeatedly measured biomarkers is currently limited. In this article, we propose a novel Bayesian

ensemble method to consolidate prediction by borrowing information across several longitudinal and cross-sectional omics data layers. Unlike existing frequentist paradigms, our approach enables uncertainty quantification in prediction as well as interval estimation for a variety of quantities of interest based on posterior summaries. We apply our method to four published multi-omics datasets and demonstrate that it recapitulates known biology in addition to providing novel insights while also outperforming existing methods in estimation, prediction, and uncertainty quantification. Our open-source software is publicly available at: https://github.com/himelmallick/IntegratedLearner.

279. Prediction rules, random partitions and borrowing of information: an invitation to Bayesian Nonparametrics

[Plenary Lecture 3, (page 30)]

Igor PRUENSTER, *Bocconi University*

The availability of complex-structured data has sparked new research directions in Statistics and Machine Learning. Bayesian Nonparametrics is at the forefront of this trend thanks to two crucial features: its coherent probabilistic framework, which naturally leads to principled uncertainty quantification and prediction, and its infinite-dimensionality, which exempts from parametric restrictions and ensures full modeling flexibility. In this talk we provide an overview of Bayesian Nonparametrics starting from its foundations and the Dirichlet process, the most popular nonparametric prior. First we frame the Dirichlet process, as well as most of its generalizations, as transformations of completely random measures. Then, we tackle more complex data structures by building dependent versions of these models via additive, hierarchical and nested constructions. We investigate the dependence structure, the borrowing of information and, more generally, the variety of discrete random structures induced by the underlying completely random measures. Illustrations on multilayer networks, species sampling and survival analysis are provided.

280 . Care2Vec: A Deep learning approach for the classification of self-care problems in physically disabled children

[26.A1.I4, (page 4)]

Sayan PUTATUNDA, *sayanp@iima.ac.in*

Accurate classification of self-care problems in children who suffer from physical and motor affliction is an important problem in the healthcare industry. This is a difficult and a time consuming process and it needs the expertise of occupational therapists. In recent years, healthcare professionals have opened up to the idea of using expert systems and artificial intelligence in the diagnosis and classification of self-care problems. In this study, we propose a new deep learning based approach named Care2Vec for solving the self-care classification problem. We use a real world self care activities dataset that is based on a conceptual framework designed by the World Health Organization (WHO). The conceptual framework is known as the International Classification of Functioning, Disability, and Health for Children and Youth (ICF-CY), which is a widely used standard framework for analyzing self-care activity records. Deep learning is a form of representation learning and in recent years, it has been very successful in various fields such as computer vision, speech processing and more. Care2Vec is a mix of unsupervised and supervised learning where we use Autoencoders and Deep neural networks as a two step modeling process. We compare Care2Vec with traditional methods reported in the literature for solving the self-care classification problem (such as, Artificial Neural Networks and Decision trees) in both multi-class classification and binary classification settings. We use k-fold cross validation while applying the different methodologies. The evaluation metrics used were the Mean Cross Validation (CV) score and the Mean Area under the curve (AUC). We found that the Care2Vec method has better prediction accuracy than the prevalent methods and so a recommended approach for the self care classification problem. The adoption of Care2Vec can help expert therapists in making better diagnostic decisions and will thus lead to better treatment.

281. Estimating and improving the performance of machine learning under natural distribution shifts

[28.M2.I43, (page 29)]

Aditi RAGHUNATHAN, *Carnegie Mellon University*
Ananya KUMAR, *Stanford*
Christina BAEK, *Carnegie Mellon University*

Machine learning systems often fail catastrophically under the presence of distribution shiftwhen the test distribution differs in some systematic way from the training distribution. If we can mathematically

characterize a distribution shift, we could devise appropriate robust training algorithms that promote robustness to that specific class of shifts. However, the resulting robust models show limited gains on shifts that do not admit the structure they were specifically trained against. Naturally occurring shifts are both hard to predict a priori and intractable to mathematically characterize necessitating different approaches to addressing distribution shifts in the wild.

In this talk, we first discuss how to estimate the performance of models under natural distribution shiftsthe shift could cause a small degradation or a catastrophic drop. Obtaining ground truth labels is expensive and requires the a priori knowledge of when and what kind of distribution shifts are likely to occur. We present a phenomenon that we call agreement-on-the-line that allows us to effectively predict performance under distribution shift from just unlabeled data. Next, we investigate a promising avenue for improving robustness to natural shiftsleveraging representations pre-trained on diverse data. Via theory and experiments, we find that the de facto fine-tuning of pre-trained representations does not maximally preserve robustness. Using insights from our analysis, we provide two simple alternate fine-tuning approaches that substantially boost robustness to natural shifts.

282 . APPLICATIONS OF PROBABILISTIC MODELS IN HYDROMETEOROLOGICAL DROUGHT MONITORING & FREQUENCY ANALYSIS

[29.M2.I65, (page 44)]

Meenu RAMADAS, *Assistant Professor, School of Infrastructure, IIT Bhubaneswar*

Alok SAMANTARAY, *Student, School of Infrastructure, IIT Bhubaneswar*

There are numerous ongoing studies on changes in the regional drought characteristics: severity, duration and frequency that are likely in future periods due to potential climate change, for instance, increase in drought severity and duration, and increase in frequency and spatial extent of drought impacts. In this regard, there are potential applications of probabilistic and statistical approaches to characterize droughts, to model their spatiotemporal evolution under uncertainty, and to perform risk assessment for future periods. We present a study on changes in regional severity-duration-frequency (S-D-F) relationships under a warming future climate scenario over the Eastern India region (comprising of states of Odisha, Chattisgarh, Jharkhand and Bihar). The Standardized Precipitation Index (SPI) was used to compute the meteorological drought characteristics, and the Coordinated Regional Downscaling Experiment (CORDEX) South Asia Regional Climate Model (RCM)-based future rainfall projections under Representative Concentration Pathway (RCP) 4.5 scenario were utilized for future drought frequency analysis. In this study, regionalization of meteorological drought characteristics was performed using the simple k-means clustering technique and bivariate copula was utilized to model the dependence among the drought characteristics to develop the S-D-F curves for both historical (1971-2003) and future (2042-2068) time periods. The changes in regional S-D-F curves under future potential climate change were then investigated. In the second study, we demonstrate development of a new meteorological drought index: a bivariate copula-based multiscalar Rainfall Pattern Drought Index (RPDI) that incorporates the number of rainy days information in addition to magnitude of rainfall to characterize drought severity. Future RPDI drought characteristics based on Coupled Model Intercomparison Project phase 6 (CMIP6) multi-model ensemble under two emission scenarios or Shared Socioeconomic Pathways (SSPs): SSP126 and SSP245 were then computed. When compared with their baseline period (19852014) values, drought characteristics such as maximum drought duration and maximum drought severity are projected to increase in the near future (20202049) and far future (2070-2099) periods under potential climate change.

283. Doob's theorem revisited

[Special Invited Session 2, (page 28)]

R V RAMAMOORTHI, *ramamoor@msu.edu*

Doob's theorem states that when the observations are i.i.d P then for any prior for P, the posterior is consistent on a set of prior measure 1. We extend the result to more general case and examine its connection with sufficient sigma algebras. This talk is based on discussions with Professors B.V, Rao and Karthik Sriram

284 . Beyond mean-field limits for stochastic processes on networks

[Special Invited Session 3, (page 42)]

Kavita RAMANAN, *Brown University*

Large-scale networks capture various models in engineering and operations research, ranging from opinion dynamics, the spread of diseases and load balancing in computer networks, and their dynamics can often be modelled in terms of coupled stochastic processes whose interactions are local with respect to an underlying graph. The dynamics of these networks are complex, and not amenable to exact analysis. When the interaction graph is the complete graph, so-called mean-field approximations are well known to serve as tractable approximations for these processes. We will describe novel approximations that are valid in the case when the interaction graph is truly sparse, and which can be rigorously shown to be accurate in a suitable asymptotic regime. We will also demonstrate their efficacy of these approximations and describe other potential applications of these results. This is based on joint works with Ankan Ganguly and Juniper Cocomello.

285. Focused Model Selection for Time Series Data

[28.M2.I45, (page 29)]

Thekke Variyam RAMANATHAN, *Savitribai Phule Pune University*

Information criteria such as AIC, BIC, etc. have been extensively used for model selection problems for independent as well as dependent data. In this talk, we consider the Focused Information Criteria (FIC) for model selection problems associated with dependent data. Some of the recent developments in this area will be presented in this talk.

286. A Predictive Analytics Approach for the Impact Assessment of Patent Characteristics on the Renewal Life of a Patent

[26.A1.I4, (page 4)]

Pritam RANJAN, *pritamr@iimidr.ac.in*
Ashit KUMAR, *IIM Indore*
Shadab DANISH, *BASE University*

Over the period of time, patent filing in India has gone up significantly. The patent value refers to the economic reward the patentee is able to extract from the patent by excluding others from making, using, or selling an innovation. The life of a patent is at most 20 years, and in India, there is an annual renewal cost for keeping a patent active. Once the patent is filed and granted, if the assessed value of a patent exceeds the renewal cost, then the patentee chooses to renew it. There are several patent characteristics that may influence the value of a patent, for example, number of claims, inventor size, grant lag, ownership, etc. Our objective is to identify the significant patent characteristics that influence the patent value. Since the patents are not bought or sold in the usual sense, often the renewal life of a patent is used as a proxy for the patent value.

In this paper, we conduct a comparative study of a host of statistical and machine learning techniques and Gaussian Process regression model to fit the renewal model and subsequently assess the significance of patent characteristics that influences the renewal life of a patent.

287. Bayesian Analysis of Weibull Quantile Regression model using Hamiltonian Monte Carlo Algorithm when the Observations are Interval Censored

[27.M1.I16, (page 12)]

Rakesh RANJAN, *Benares Hindu University*

Koenker and Bassett (1978) established the idea of estimation of regression quantiles, and many other authors have since studied it. There are relatively few studies that address the estimation of the regression quantile for lifetime data. This paper presents the Bayesian analysis of the interval-censored Weibull quantile regression model. We have considered interval censoring because it is a general type of censoring, and both left and right censoring are its specific instances. We performed Bayesian analysis using sample-based approaches, particularly the Hamiltonian Monte Carlo technique, to deal with analytically intractable posterior. A numerical example for the real data set is provided, and the considered quantile regression model is compared with the null model.

288 . Data Arising from Destructive Samples: Some Informative Models: The Case of Spina Bifida

[28.M2.I42, (page 28)]

Marepalli RAO, *University of Cincinnati*
Rigwed TATU, *Cincinnati Children's Hospital, Cincinnati*
Anand SETH, *SK Patent Associates, Columbus*
Shesh RAI, *Cancer Data Science Center, University of Cincinnati*

A fetus is diagnosed with spina bifida. It is a birth defect that skin has not formed in some sections of the spinal column. The spinal fluid is seeping out of the spinal column. Amniotic fluid is getting into the spinal column reaching the developing brain. This will have serious health consequences for the baby. A surgery (fetoscopy) on the fetus can be performed to cover the gap by a patch, hoping that skin gets generated under the patch. Once the skin forms under the skin, another surgery has to be performed to remove the patch. In our lab at the University of Cincinnati, one of our collaborators has created a biodegradable patch by fusing two polymers in certain proportions to strike a balance on the number of days the patch takes to degrade. The inserted patch faces amniotic fluid on one side and spinal fluid on the other side.

In the course of the experiments, we need to measure how rough the patch is at 0, 4, 8, 12, and 16 weeks. The rougher the patch is, the better the nutrients latch onto the patch. The better the nutrients are, the speedier the skin will grow under the patch. We have a choice: dip the patch either in amniotic fluid or phosphate buffered saline. The question is which liquid makes the patch rougher on average over time (mimicking the spinal fluid). The issue can be settled by experimentation and data collection over time. In the course of measuring roughness, the patch becomes unusable after that. The patch is destroyed. We have introduced a novel model of the joint distribution of roughness at 0, 4, 8, 12, and 16 weeks so that the marginal distributions have information on the joint distribution. This is the focus of the presentation.

289 . A novel meta-analysis approach in the presence of disparate confounder information from studies

[27.A1.I30, (page 19)]

Debashree RAY, *Johns Hopkins University*
Alvaro MUOZ, *Johns Hopkins University*
Lisa JACOBSON, *Johns Hopkins University*
Bryan LAU, *Johns Hopkins University*

A practical challenge in meta-analyzing studies is that important confounders are likely not measured across all cohorts since each cohort may have been independently funded with independent study protocols. Some may report adjusted estimates of exposure-outcome association while others report unadjusted estimates, and there is no consensus on how to synthesize these estimates. There exist nave solutions such as meta-analyzing only unadjusted estimates, or only adjusted estimates, or do both and qualitatively assess conclusion from each. In this talk, I will present CIMBAL, a practical yet valid statistical method for meta-analyzing independently sampled cohorts/studies with imbalance in measurement of confounders. CIMBAL leverages theoretical asymptotic relations between adjusted and unadjusted estimates to impute the adjusted estimates for studies with missing confounders and finally provides a meta-analyzed adjusted estimate that appropriately accounts for the dependence between estimates arising due to borrowing of information across studies. I will discuss assumptions behind CIMBAL, show special cases, and illustrate its performance in both simulated and real data.

290. Clustering consistency with Dirichlet process mixtures

[28.M1.I40, (page 25)]

Giovanni REBAUDO, *The University of Texas at Austin*
Filippo ASCOLANI, *Bocconi University*
Antonio LIJOI, *Bocconi University*
Giacomo ZANELLA, *Bocconi University*

Dirichlet process mixtures are flexible nonparametric models, particularly suited to density estimation and probabilistic clustering. In this work we study the posterior distribution induced by Dirichlet process mixtures as the sample size increases, and more specifically focus on consistency for the unknown number of clusters when the observed data are generated from a finite mixture. Crucially, we consider the situation where a prior is placed on the concentration parameter of the underlying Dirichlet process. Previous findings in the literature suggest that Dirichlet process mixtures are typically not consistent for the number of clusters if the concentration parameter is held fixed and data come from a finite mixture. Here we show that consistency for the number of clusters can be achieved if the concentration parameter is adapted in a fully Bayesian way, as commonly done in practice. Our results are derived for data coming from a class of finite mixtures, with mild assumptions on the prior for the concentration parameter and for a variety of choices of likelihood kernels for the mixture.

291. Bayesian nonparametric taxonomic Gibbs priors

[27.M2.I22, (page 16)]

Tommaso RIGON, *University of Milano-Bicocca*
Ching-Lung HSU, *Duke University*

David DUNSON,

Gibbs-type priors are a popular Bayesian nonparametric tool for modeling exchangeable discrete data. However, few of these priors are tailored for the nested clustering mechanism that occurs in the biological taxonomy. We propose a novel framework, taxonomic Gibbs-type priors, that accommodate data coming from such taxonomic structure and provide a finer control in modeling uncertainty. We describe its link to enriched processes, present a simple predictive scheme, and show that the resulting marginal of each layer still admits a Gibbs-type representation.

292. Sampling using Adaptive Regenerative Processes

[27.M1.I16, (page 12)]

Christian ROBERT, *Université Paris Dauphine PSL and University of Warwick*
Hector MCKIMM, *University of Warwick*
Andi WANG, *University of Bristol*
Murray POLLOCK, *University of Newcastle*

Enriching Brownian Motion with regenerations from a fixed regeneration distribution at a particular regeneration rate results in a Markov process that has a target distribution as its invariant distribution. We introduce a method for adapting the regeneration distribution, by adding point masses to it. This allows the process to be simulated with as few regenerations as possible, which can drastically reduce computational cost. We establish convergence of this self-reinforcing process and explore its effectiveness at sampling from a number of target distributions. The examples show that our adaptive method allows regeneration-enriched Brownian Motion to be used to sample from target distributions for which simulation under a fixed regeneration distribution is computationally intractable.

293. A Novel Image Monitoring Procedure Using Jump Regression

[27.A2.C8, (page 23)]

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

Image monitoring is a comparatively new research area in the field of statistics and machine learning that has wide applications in different areas including medical diagnostics and disease monitoring, satellite imaging, security systems, and so on. Traditionally, quality control techniques have been used widely in manufacturing industries to inspect a process over time to detect any non-random or special cause of variation in the observed data. Due to rapid progress in image acquisition techniques, sequences of images are now a popular data format. Therefore, detection of changes in images over time has become an important research area. In the literature, existing intensity based image surveillance methods to detect out of control images are often unreasonable in many real-life applications where a change in contrast between the background and foreground of an image should not indicate an out-of-control image as long as the boundaries of the image objects remain unchanged. In this regard, we propose a control chart to monitor grayscale images using detected edges of the images. Specifically, we focus on monitoring the Hausdorff distance between the point-set of detected edge pixels in each image from the corresponding point-set of the estimated true in-control image. The proposed control chart should be easy to execute for the practitioners in many real-life applications. Numerical studies show that it outperforms several competing methods in various examples.

294. Soft-thresholded scalar on vector-valued image regression with application to DTI data

[27.M1.I13, (page 11)]

Arkaprava ROY, *University of Florida*
Zhou LAN,

In this paper, we develop a novel spatial variable selection method for scalar on vector-valued image regression in a multi-group setting. Here, vector-valued image refers to the imaging datasets that contain vector-valued information at each pixel/voxel location, such as in RGB color images, multimodal medical images, DTI imaging, etc. The focus of this work is to identify the spatial locations in the image having an important effect on the scalar outcome measure. Specifically, the overall effect of each voxel is of interest. We thus develop a novel shrinkage prior by soft-thresholding the 2 norm of a latent multivariate Gaussian process. It allows us to estimate sparse and piecewise-smooth spatially varying vector-valued regression coefficient function. Motivated by the real data, we further develop a double soft-thresholding based framework when there are multiple known discrete subgroups. For posterior inference, an efficient MCMC algorithm is developed. We compute the posterior contraction rate for parameter estimation and also establish consistency for variable selection of the proposed Bayesian model, assuming that the true regression coefficients are H older smooth. Finally, we demonstrate the advantages of the proposed method in simulation studies and further illustrate in an ADNI dataset for modeling MMSE scores based on DTI-based vector-valued imaging markers.

295. A General Framework for Powerful Confounder Adjustment in Omics Association Studies

[Student Paper Competition 1, (page 13)]

Asmita ROY, *Texas A&M University*
Jun Chen, Xianyang Zhang ,

Genomic data are subject to various sources of confounding, such as demographic variables, biological heterogeneity, and batch effects. To identify genomic features associated with a variable of interest in the presence of confounders, the traditional approach involves fitting a confounder-adjusted regression model to each genomic feature, followed by multiplicity correction. This study shows that the traditional approach was sub-optimal and proposes a new two-dimensional false discovery rate control framework (2dFDR+) that provides significant power improvement over the conventional method and applies to a wide range of settings. 2dFDR+ uses marginal independence test statistics as auxiliary information to filter out less promising features, and FDR control is performed based on conditional independence test statistics in the remaining features. 2dFDR+ provides (asymptotically) valid inference from samples in settings where the conditional distribution of the genomic variables given the covariate of interest and the confounders is arbitrary and completely unknown. To achieve this goal, our method requires the conditional distribution of the covariate given the confounders to be known or can be estimated from the data. We develop a new procedure to simultaneously select the two cutoff values for the marginal and conditional independence test statistics. 2dFDR+ is proved to offer asymptotic FDR control and dominate the power of the traditional procedure. Promising finite sample performance is demonstrated via extensive simulations and real data applications.

296 . Stable random fields indexed by discrete amenable groups

[29.A1.I69, (page 46)]

Parthanil ROY, *Indian Statistical Institute*
Mahan MJ, *Tata Institute of Fundamental Research, Mumbai*
Sourav SARKAR, *University of Cambridge*

Random fields indexed by amenable groups arise naturally in machine learning algorithms for structured and dependent data. On the other hand, mixing properties of such fields are extremely important tools for investigating asymptotic properties of any method/algorithm in the context of space-time statistical inference. In this work, we find a necessary and sufficient condition for weak mixing of a left-stationary symmetric stable random field indexed by an amenable group in terms of its Rosinski representation. The main challenge is ergodic theoretic - more precisely, the unavailability of an ergodic theorem for nonsingular (but not necessarily measure preserving) actions of amenable groups even along a tempered Flner sequence. We remove this obstacle with the help of a truncation argument along with the seminal work of Lindenstrauss (2001) and Tempelman (2015), and finally applying the Maharam skew-product. This work extends the domain of application of the speakers previous paper connecting stable random fields with von Neumann algebras via the group measure space construction of Murray and von Neumann (1936). In particular, weak mixing has now become W*- rigid properties for stable random fields indexed by any amenable group, not just Z^d.

This talk is based on an ongoing joint work with Mahan Mj (TIFR Mumbai) and Sourav Sarkar (University of Cambridge).

297. Asymptotic Bayes optimality under sparsity for equicorrelated multivariate normal test statistics

[28.A2.C12, (page 36)]

Rahul ROY, *Indian Statistical Institute, Kolkata*
Subir Kumar BHANDARI, *Indian Statistical Institute, Kolkata*

Here we address dependence among the test statistics in connection with asymptotically Bayes' optimal tests in presence of sparse alternatives. Extending the setup in Bogdan et. al. (2011) we consider an equicorrelated (with equal correlation rho) multivariate normal assumption on the joint distribution of the test statistics, while conditioned on the mean vector mu. Rest of the set up is identical to Bogdan et. al. (2011) with a slight modification in the asymptotic framework. We exploit an well known result on equicorrelated multivariate normal variables with equal marginal variances to decompose the test statistics into independent random variables. We then identify a set of independent yet unobservable gaussian random varibales sufficient for the multiple testing problem and chalk out the necessary and sufficient conditions for single cutoff tests to be ABOS based on those dummy variables following Bogdan et. al. (2011). Further we replaced the dummy variables with deviations of the statistics from their arithmatic means which were easily calculable from the observations due to the decomposition used earlier. Additional conditions are then derived so that the necessary and sufficient conditions for single cutoff tests to be ABOS using the independent dummy variables plays the same role with the replacement variable as well (with a deviation of order o(1)). Next with the same additional assumption, necessary and sufficient conditions for single cutoff tests to control the Bayesian FDRs are derived and as a consequence under various sparsity assumptions we proved that the classical Bonferroni, Sidak and Benjamini-Hochberg methods of multiple testing are ABOS if the same conditions are satisfied.

298. Co-existence in probability models

[29.M1.I59, (page 39)]

Rishideep ROY, *IIM Bangalore*
Kumarjit SAHA,

We study coexistence in discrete time multi-type frog models. We first show that for two types of particles on Zd, for d 2, for any jumping parameters p1,p2 (0,1], coexistence occurs with positive probability for sufficiently rich deterministic initial configuration. We extend this to the case of random distribution of initial particles. We study the question of coexistence for multiple types and show positive probability coexistence of 2d types on Zd for rich enough initial configuration. We also show an instance of infinite coexistence on Zd for d 3 provided we have sufficiently rich initial configuration.

299 . Regularized high dimension low tubal-rank tensor regression

[28.M1.I36, (page 24)]

Samrat ROY, *University of Pennsylvania*
George MICHAILIDIS, *University of Florida*

Tensor regression models are of emerging interest in diverse fields of social and behavioral sciences, including neuroimaging analysis, neural networks, image processing and so on. Recent theoretical advancements of tensor decomposition have facilitated significant development of various tensor regression models. The focus of most of the available literature has been on the Canonical Polyadic (CP) decomposition and its variants for the regression coefficient tensor. A CP decomposed coefficient tensor enables estimation with relatively small sample size, but it may not always capture the underlying complex structure in the data. In this work, we leverage the recently developed concept of tubal rank and develop a tensor regression model, wherein the coefficient tensor is decomposed into two components: a low tubal rank tensor and a structured sparse one. We first address the issue of identifiability of the two components comprising the coefficient tensor and subsequently develop a fast and scalable Alternating Minimization algorithm to solve the convex regularized program. Further, we provide finite sample error bounds under high dimensional scaling for the model parameters. The performance of the model is assessed on synthetic data and is also used in an application involving data from an intelligent tutoring platform.

300 . On Progressive Type-I Interval Censoring Schemes under Competing Risks

[30.M2.I82, (page 54)]

Soumya ROY, *Indian Institute of Management Kozhikode*

In reliability engineering literature, Type-I and Type-II censoring schemes have drawn a lot of interest. However, a major drawback with these censoring schemes is that they require continuous monitoring of the life test experiment, which may not always be feasible due to resource constraints. Furthermore, these traditional censoring schemes do not permit intermediate withdrawals of the test units. In view of this, Progressive Type-I Interval Censoring (PIC-I) schemes are proposed in the literature. A PIC-I scheme, which is essentially an extension of the wellknown Interval Censoring (IC) scheme, allows intermediate withdrawals of the test units at the prefixed inspection times. Moreover, as in the case of IC schemes, the failures of test units are only observed in groups at each inspection time. This article deals with such PIC-I datasets for statistical inference. Furthermore, we assume that the test units have multiple failure modes. Moreover, the test units fail as soon as one of the modes of failures strikes. This article presents both classical and Bayesian inference in detail, assuming that the lifetime distributions corresponding to the failure modes belong to a log-location-scale family of distributions. We further discuss optimal design of such PIC-I schemes.

301 . ON BAYESIAN BEST RESPONSE LEARNING IN GAMES

[26.A1.I2, (page 4)]

Souvik ROY, *ISI Kolkata*
Sayan MUKHERJEE, *ISI Kolkata*

We introduce the notion of regularized Bayesian best response (RBBR) learning dynamic in heterogeneous population games. We obtain such a dynamic via perturbation by an arbitrary lower semicontinuous, strongly convex regularizer in Bayesian population games. We provide a sufficient condition for the existence of rest points of the RBBR learning dynamic, and hence, for the existence of regularized Bayesian equilibrium in Bayesian population games. We also explore the fundamental properties of the RBBR learning dynamic, which includes the existence of unique continuous solutions from arbitrary initial conditions, as well as the continuity of the solution trajectories thus obtained with respect to the initial conditions. Finally, as an application, we introduce the notions of Bayesian potential and Bayesian negative semidefinite games and provide convergence results for such games.

302. Statistical methods for finding cytometric markers associated with Systemic Lupus Erythematosus

[27.M1.I18, (page 13)]

Pratyaydipta RUDRA, *Oklahoma State University*

Systemic lupus erythematosus (SLE) is a severe multi-organ rheumatologic disease characterized by immune dysregulation with a heterogeneous and unpredictable disease course. Owing to its clinical and immunological heterogeneity, SLE continues to be a major diagnostic and therapeutic challenge. We develop different statistical approaches to analyze mass cytometry data collected from a study on pediatric SLE patients and age and sex-matched healthy controls.

Mass cytometry data are often clustered into cell sub-populations first, which can then be used to answer scientific questions regarding the abundance of cell types and expressions of specific parameters (e.g. surface markers, signaling proteins, cytokines) across groups, such as disease and control groups, or stimulation regimes. The statistical questions about the tree-structured cell population data can be visualized in two layers. First, it is clinically interesting to know if the abundance of the cell subpopulations is different across two or more groups and/or conditions. Given the proportion of cell types for each sample, the next question is whether there is any differential expression of signaling proteins or cytokines (functional measurements of the cell populations studied).

Modeling data with multiple layers of correlation using a classical parametric model often becomes a challenging task. In order to tackle this, we developed two new kernel-based statistical learning methodology frameworks to test differential cell type abundance and differential expressions across disease groups and stimulation conditions. We used both simulations and real data analysis to study the performance of these methods.

We also propose a network analysis approach to integrate mass cytometry and gene expression data to study biomarkers that may help monitor disease activity. We combine the single-cell mass cytometry data with (bulk) gene expression microarray data and clinical disease activity measures (i.e. SLEDAI score, Low Disease Activity endpoints).

The different approaches of analyzing the data revealed clinically relevant patterns such as differential cell type abundance between the disease and the control group, and also differential expression of several cytokines.

303. On the Robustness to Misspecification of -Posteriors and Their Variational Approximations

[29.A1.I67, (page 45)]

Cynthia RUSH, *Columbia University*
Marco AVELLA MEDINA, *Columbia University*
Jos MONTIEL OLEA, *Cornell University*
Amilcar VALEZ, *Northeastern University*

Variational inference (VI) is a machine learning technique that approximates difficult-to-compute probability densities by using optimization. While VI has been used in numerous applications, it is particularly useful in Bayesian statistics where one wishes to perform statistical inference about unknown parameters through calculations on a posterior density. In this talk, I will review the core concepts of VI and introduce some new ideas about VI and robustness to model misspecification. In particular, we will study -posteriors, which distort standard posterior inference by downweighting the likelihood, and their variational approximations. We will see that such distortions, if tuned appropriately, can outperform standard posterior inference when there is potential parametric model misspecification.

304. Preservation of Log-concavity under Multi-state Series and Multi-state Parallel Systems

[27.M2.I24, (page 17)]

Sanjeev SABNIS, *IIT Bombay*
Priyanka MAJUMDER, *IIT Bombay*
Shyamal GHOSH, *IIIT Guwahati*

Log-concavity of multivariate distributions is an important concept in general and has a very special place in the field of Reliability Theory. The preservation of univariate unimodality under binary coherent systems of n independent binary components has been studied by Sabnis and Nair [11] as an extension ofAlams [1] result for koutof n systems. Here an attempt has been made to study the preservation of continuous version of multivariate log-concavity under multi-state series and multi-state parallel systems made up of n independent components and states of both, systems and components, taking values in set S = 0, 1, 2, ... , M and under the assumption that random variables representing times spent by these systems are available in specific forms. Similar preservation results for discrete version of multivariate logconcavity for multi-state series and multi-state parallel systems consisting of n independent components have been established for a subset 0, 1, 2 of S. These results for continuous and discrete versions of log-concavity have also been extended to systems that are formed using both multi-state series and multistate-parallel systems. The log-concavity and these preservation results have, in turn, enabled obtaining bounds on relevant joint probabilities for the systems under consideration.

305. Limit Theorems for Entropic Maps

[29.M1.C16, (page 41)]

Ritwik SADHU, *Cornell University*
Ziv GOLDFELD, *Cornell University*
Kengo KATO, *Cornell University*
Gabriel RIOUX, *Cornell University*

We study limit theorems for entropic optimal transport (EOT) maps, dual potentials, and the Sinkhorn divergence. The key technical tool we use is a first and second order Hadamard differentiability analysis of EOT potentials with respect to the underlying distributions. Given the differentiability results, the functional delta method is used to obtain central limit theorems for empirical EOT potentials and maps. The second order functional delta method is leveraged to establish the limit distribution of the empirical Sinkhorn divergence under the null. Building on the latter result, we further derive the null limit distribution of the Sinkhorn independence test statistic and characterize the correct order. Since our limit theorems follow from Hadamard differentiability of the relevant maps, as a byproduct, we also obtain bootstrap consistency and asymptotic efficiency of the empirical EOT map, potentials, and Sinkhorn divergence.

306 . A Game-Theoretic approach for disease propagation

[28.A1.I50, (page 32)]

Soumyarup SADHUKHAN, *Indian Institute of Technologuy Kanpur*
Sayar KARMAKAR , *Department of Statistics, University of Florida*
Moumanti PODDER , *Department of Mathematics, Indian Institute of Science Education and Research, Pune*
Souvik ROY, *Applied Statistics Unit, Indian Statistical Institute, Kolkata*

We study a game theoretic model of disease propagation where there are finitely many individuals who are bound to interact with each other every day for their day-to-day life, and the disease spreads through such interactions. Every day, each individual needs to decide an action representing the extent to which she will go out. While going out gives her some utility, it also has some cost due to the possibility of getting infected. Each individual has some immunity power, which determines the probability of getting infected as a function of her exposure to the virus. Some individuals are infected from the beginning. At every subsequent time point, a randomly selected individual updates her action to maximize her ”payoff”. The payoff depends on her utility from going out and the cost of being infected. Also, the cost of being infected, in turn, depends on the actions chosen by the other individuals, which makes the model game theoretic. The (random) infected set gets updated at the next time point based on her action, and the process reaches a new state.

We discuss the convergence of the limiting distribution of the number of infected agents after starting from negligible initial contagion/infection. For a wide range of scenarios comprising different values of initial action profile, immunity we find this limiting distribution. Starting from simple static action-based dynamics, we also extend our result when the graph can potentially evolve based on the current status of the agents.

307 . Efficient Model Compression for Sparse Bayesian Neural Networks

[Student Paper Competition 2, (page 26)]

Diptarka SAHA, *University of Illinois, Urbana - Champaign*
Zihe Liu, Feng Liang ,

Model Compression has drawn much attention within the deep learning community recently. Compressing a dense neural network offers many advantages including lower computation cost, deployability to devices of limited storage and memories, and importantly resistance to adversarial attacks. This may be achieved via node pruning or fully discarding certain input features. Here we demonstrate a novel strategy to emulate principles of Bayesian model selection in a deep learning setup. Given a fully connected Bayesian neural network with spike-and-slab priors trained via a variational algorithm we obtain the posterior inclusion probability for every node that typically gets lost. We employ these probabilities for pruning and feature selection on a host of simulated and real world benchmark data, and find evidence of better generalisability of the pruned model in all our experiments.

308. Convergence to the Brownian web for a perturbed Howard model

[27.A1.I28, (page 19)]

Kumarjit SAHA, *Ashoka University*

Understanding the structure of random directed forests constructed on random sets of points and studying their scaling limits has been extensively studied in the literature. For most models, the corresponding vertex sets satisfy independence assumption over disjoint regions. A natural question is what happens for point processes if such an assumption does not hold. In a joint work with Subhroshekhar Ghosh, we explored such questions where lattice points are perturbed over compact regions. In this talk I will describe a recent work with Rahul Roy and Anish Sarkar where we are dealing with perturbed lattice points and perturbations are no longer restricted to compact regions. The dependency of the generated point process is quite challenging to deal with and require non-trivial modification of existing techniques.

309. A smooth alternative to check loss in Deep inference tasks

[27.M2.I20, (page 15)]

Snehanshu SAHA, *snehanshus@goa.bits-pilani.ac.in*

Quantile Regression allows the estimation of full conditional quantiles, as opposed to estimating conditional means in a regression setting. It is known that minimizers of Check loss are the conditional quantiles. We propose log cosh as an alternative to the check loss and apply it in the context of learning regimens driven by neural networks. We show that log cosh, an approximation to check loss, possesses many desirable properties such as convexity, smoothness, robustness, and its ability to approach mean absolute error (MAE/L1) and mean squared error (MSE/L2) in the limit. Further, I'll talk about how log cosh can be extended to binary classification problems, and demonstrate the interpretability of quantiles in that setting. In addition, I will also show how the second-order differentiability can be leveraged to facilitate faster convergence.

310 . Interacting Urns with Multiple drawings

[27.M2.I23, (page 16)]

Neeraja SAHASRABUDHE, *IISER Mohali*

We consider N interacting two-colour urns with multiple drawings on a finite graph and study the dependence of the asymptotic properties of the fraction of balls of either colour on the type of reinforcement and the structure of the underlying graph.

311 . Predictive role of neutrophil-lymphocyte ratio (NLR) in patients with metastatic melanoma: A post hoc exploratory analysis from phase 3 trial

[27.A1.I25, (page 18)]

Braja Gopal SAHOO, *Novartis Healthcare Pvt. Ltd.*
Hiya BANERJEE,

Background: An elevated NLR is associated with poor survival in patients with a variety of solid cancers, including those treated with immunotherapies.

This post hoc exploratory analysis, based on machine learning (ML) techniques, from a phase 3 trial aimed at evaluating the potential predictive role of NLR in patients with BRAFV600mutant advanced melanoma

Methods: All patients enrolled in the trial were included in this analysis. NLR values at baseline and subsequent treatment (Tx) cycles (up to cycle 7) were selected based on exploratory analysis. Final predictive variable selection step was performed using knockoff framework. NLR measurements along with other baseline variables were found to be correlated with progression-free survival (PFS) using Survival tree and KaplanMeier curve

Results: Baseline NLR values ranged from 2 to 4 for most of the patients in either of the Tx arms that reduced to 1-3 upon Tx. Survival tree analysis indicated that in patients receiving placebo, the threshold NLR value obtained through mathematical model was associated with poor PFS outcome in patients with baseline NLR 1.693 (n=114) vs ¡1.693 (n=19). In patients with NLR 1.693, those with sum of the longest diameter (mm) 61 (n=43) had worst PFS outcome. Baseline NLR 1.693 was associated with poor PFS outcome in placebo arm vs target treatment arm. Literature-based cutoff for baseline NLR 5 was associated with shorter mPFS in placebo arm (n=37) vs target treatment arm (n=40; hazard ratio, 0.642 [95% confidence interval, 0.377-1.1]).

Conclusion: The findings suggest that knockoff based variable selection outcome is congruent with classical approach, with known level of false discovery chance. Baseline NLR comes out as an independent factor predicting poor PFS outcome in patients with metastatic melanoma treated with the combination of immune checkpoint inhibitors and targeted therapy. The predictive value of NLR and the selected threshold will have to be validated in future studies.

312. A Graphical Lasso model for Hermitian matrices to detect global time-lagged Teleconnections

[26.A1.I6, (page 5)]

Indranil SAHOO, *Virginia Commonwealth University*
Joe GUINNESS, *Cornell University*
Brian REICH, *North Carolina State University*

Teleconnections refer to spatially and temporally connected large-scale anomalies that influence the variability of atmospheric phenomena. Since teleconnections influence the global climate system, it is important to understand the abnormal behavior and interactions of these phenomena and identify them accurately. In this paper, we provide a mathematical definition of teleconnections based on a spatio-temporal model using spherical needlet functions. Spherical needlets are exactly localized at several overlapping intervals corresponding to different frequencies in the frequency domain and form a tight frame. This ensures the perfect reconstruction property of an orthonormal basis. We also extend the famous graphical Lasso algorithm to incorporate Hermitian matrices and use it to estimate the inverse covariance matrix of needlet coefficients after projecting them onto the Fourier domain. The proposed method is demonstrated by simulation studies and detection of possible global teleconnections in the HadCM3 model output air temperature data.

313 . On estimation of the covariate's scale effect on failure time using martingale residuals

[29.M1.I58, (page 39)]

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

In the analysis of censored survival data, it is common to look for a suitable alternative to the proportional hazards (PH) model if the proportional hazards (PH) assumption is violated. However, it may then still be possible to continue using a modified version of the PH model with some adjustments. In this talk, we discuss the problem of scale-disparity in the PH model, specifically assuming that one covariate has a non-proportional effect in the form of a scale change of the failure time, while the other covariates have a proportional effect on the hazard. The focus is on estimating the corrective scale factor if this alternative model appears to hold. We propose a method for estimating the scale factor based on a suitable transformation of martingale residuals. The performance of the estimator is also examined via Monte Carlo simulations. The procedure is then illustrated by analyzing the survival data of the patients with an apparently benign monoclonal gammopathy.

314. Ordering and ageing properties of sequential order statistics governed by the Archimedean copula

[26.A2.C3, (page 9)]

Tanmay SAHOO, *Indian Institute of Technology Jodhpur*

Nil Kamal HAZRA, *Indian Institute of Technology Jodhpur*

In reliability theory, k-out-of-n systems play an important role because k-out-of-n structures can be found in a variety of technical systems and subsystems. But, in the usual modeling of this structure, the failure of one component is assumed to have no effect on others. Sequential order statistics is a useful tool for describing the lifetimes of components of a system, where the failure of one component of the system with independent components may affect the performance of the remaining components running at this age. However, the independence between components of engineering systems is sometimes overestimated in real-world situations. Usually, components are non-identically distributed and dependent on each other due to the sophisticated nature of the problem. The sequential order statistics model may not be suitable for expressing the lifetime of such a system with dependent components. The copula is a frequently used technique to describe the dependency structure between the components of a system. The Archimedean copula family is the most popular of all the existing copulas because it captures a wide range of dependencies. In this paper, we investigate the characteristics of the lifetimes of the coherent system obtained by the sequential order statistics, whose dependence structure has been described by the Archimedean copula. In particular, by using the signature of the system, we construct a mixture representation. Using this representation, we study some stochastic comparison results for the coherent system with dependent components obtained by the sequential order statistics. In order to obtain these comparisons, we obtain some ordering properties for the sequential order statistics, which can be used to represent the lifetimes of k-out-of-n systems. In addition, some ageing properties of sequential order statistics are investigated, and some adequate conditions are provided under which the distribution of a sequential order statistic is ILR, IFR, IFRA, or NBU.

315 . A Fiducial Confidence Interval for Agreement Measurement Among Raters in a Generalized Linear Mixed Effects Model

[26.A2.C1, (page 8)]

Soumya SAHU, *University of Illinois Chicago*
Dulal BHAUMIK, *University of Illinois Chicago*
Thomas MATHEW, *University of Maryland Baltimore County*

Measurements of agreement among different data-generating sources, referred to as raters, are needed to assess the acceptability of a new or generic process. A rater can be a chemist, a psychologist, a technique, or even a formula. Additionally, recently in many instances, computer measurement has replaced human measurement, for example, image-based measures in medical research. Measurements can be dichotomous, polychotomous, ordinal, count, or continuous. Furthermore, a subject can be observed repeatedly over time by each of several raters. We plan to develop statistical methods for assessing the agreement or reproducibility of these longitudinal measurements within a rater or among multiple raters. We will use the fiducial approach to construct a confidence interval for concordance correlation coefficient (CCC), a measure of agreement. We model our longitudinal data by a mixed effect generalized linear model and generalize the concept of CCC for our model. We show via simulation that our proposed method performs better than the large sample-based method, especially when used for small samples. In addition, we explore the robustness of our approach using a mixer distribution and compare our method with parametric bootstrapping to show it has better performance in terms of coverage probability. Results are illustrated with a real-life data set. Data consists of longitudinal measurements of radiographs of knee joints in individuals participating in a clinical trial. The measurements were taken by two professionals and one computer algorithm. We numerically computed three pairwise CCCs and one combined CCC and their corresponding confidence intervals by the proposed fiducial approach and evaluated the reliability of the computer algorithm based on the confidence intervals.

316 . Consistent Bayesian Variable Selection in High-Dimensional Hierarchical Regression

[28.M1.I37, (page 24)]

Srijata SAMANTA, *M D Anderson Cancer Center*
Kshitij KHARE,
George MICHAILIDIS,

In several high-dimensional regression problems there is often an inherent partial ordering which stipulates that a "higher priority" predictor should be included before a "lower priority" predictor. The most common examples are time series applications where many predictors are time lags of the same variable. Penalized methods which use some nested group lasso penalties are commonly used but these can get unwieldy in high-dimensional settings. A Bayesian approach with a more compact penalty/shrinkage structure was recently developed. This method can only handle restricted chain-based hierarchy and corresponding posterior consistency results are not provided. In this work we develop novel Bayesian methodology which significantly generalizes the hierarchy structure among the predictors. We also establish high dimensional posterior model selection and estimation consistency under regularity assumptions. Furthermore, we demonstrate the statistical efficacy of the proposed approach through simulation studies and apply the same on real datasets.

317. Bayesian Semiparametric Hidden Markov Tensor Partition Models for Longitudinal Data with Local Variable Selection

[27.A1.I27, (page 18)]

Abhra SARKAR, *The University of Texas at Austin*
Giorgio PAULON, *Berry Consultants*
Peter MUELLER, *The University of Texas at Austin*

We present a flexible Bayesian semiparametric mixed model for longitudinal data analysis in the presence of potentially high-dimensional categorical covariates. Building on a novel hidden Markov tensor decomposition technique, our proposed method allows the fixed effects components to vary between dependent random partitions of the covariate space at different time points. The mechanism not only allows different sets of covariates to be included in the model at different time points but also allows the selected predictors influences to vary flexibly over time. Smooth time-varying additive random effects are used to capture subject specific heterogeneity. We establish posterior convergence guarantees for both function estimation and variable selection. We design a Markov chain Monte Carlo algorithm for posterior computation. We evaluate the methods empirical performances through synthetic experiments and demonstrate its practical utility through real world applications.

318. Some Inadmissibility Results on Estimating Ordered Scale Parameters of Two Exponential Populations with Different Locations using Progressive Type-II Censored Data

[Poster Session, (page 35)]

Mojammel Haque SARKAR, *National Institute of Technology Rourkela*
Manas Ranjan Tripathy ,

Due to the vast application, researchers are interested in making inferences about the unknown parameter associated with a statistical model using censored or incomplete data. In view of this, we consider the point estimation of ordered scale parameters of two exponential populations when the location parameters are unknown and unequal using progressive type-II censoring data. The progressive type-II censoring scheme is a generalization of conventional type-II and hence the complete data. In particular, we derive improved estimators for the best equivariant estimators (BEE), maximum likelihood estimators (MLEs) and uniformly minimum variance unbiased estimators (UMVUEs) of the two scale parameters under order restriction using a general class of bowl-shaped loss functions. The percentage of risk improvement for each of the improved estimators has been shown numerically using the Monte Carlo simulation method. The risk comparison of all the estimators using entropy loss function, a member of the underlying class of loss function, has been presented for illustration purposes.

319. Posterior Consistency of Covariance matrices Under Some Classes of Noninformative Priors

[27.M1.C4, (page 14)]

Partha SARKAR, *University of Florida*
Kshitij KHARE, *University of Florida*
Malay GHOSH, *University of Florida*

Covariance matrix estimation arises in multivariate problems including multivariate normal regression models. A Bayesian analysis of these problems requires a prior on the covariance matrix that is generally taken as Inverse-Wishart to maintain conditional conjugacy properties of the model. However, there are a least two problems with the IW prior. First, the uncertainty for all variance parameters is controlled by the single degree of freedom parameter and thus provides no flexibility to incorporate different amounts of prior knowledge to different variance components (see Sec 19.2 of Gelman et al. (2003)[1]). Second, in many cases, the implied scaled inv-2 distribution on each individual variance has extremely low density in a region near zero and thus causes bias in the result posteriors for these variances (Gelman 2006)[2]. This has led to the consideration of alternative priors for the covariance matrices (see e.g. Huang

& Wand (2013)[3], Mulder & Pericchi(2018)[4]). All of them are mostly computationally sound but none has proved any results so far related posterior consistency of that methods. We have taken a classical multivariate normal linear regression model and applied that priors on the variance matrix of error distributions, say . Also, number of variables(say p) can grow with sample size(say n) but p should be o(n). Under this setup we quantified a rate at which the posterior distribution of contracts to the true value of (say _0) at an exponential rate. Also, our method is robust if the original data generating method is sub-Gaussian instead of simple Gaussian. Additionally, posterior consistency of the regression coefficients is also taken care of.

References ——————— [1] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003), Bayesian data analysis, Chapman and Hall.

[2] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. Bayesian Analysis, 1: 515533.

[3] Huang, Alan and Matt P. Wand. Simple Marginally Noninformative Prior Distributions for Covariance Matrices. Bayesian Analysis 8 (2013): 439-452.

[4] Joris Mulder, Luis Raul Pericchi The Matrix-F Prior for Estimating and Testing Covariance Matrices, Bayesian Analysis, Bayesian Anal. 13(4), 1193-1214, (December 2018)

320. Detection of Anomalous Behaviour of Aircraft's Engines for Predictive Maintenance

[Poster Session, (page 35)]

Shibajee SARKAR, *Ramakrishna Mission Vivekananda Educational and Research Institute*
Srirup Mitra, Dr. Sudipta Das, Prof. Debasis Sengupta ,

Aviation industry is essential in present times and maintenance of air-crafts is important for the industry to provide a safe travelling experience. Anomaly detection in the aircraft's engines can be useful to predict the failure of an engine as well as to optimise the maintenance period. The variables of an engine under use provide a rich set of information for the above mentioned problem. Here we approach the problem of predicting the failure of aircraft's engines through two different procedures. Anomalous behaviour of an engine(failure) is detected using: Machine learning algorithms,Survival analysis techniques. In both the cases, the variables are used as features. A classification approach(in supervised way) is taken using machine learning algorithms, where we classified engines into failed(anomaly) and not failed(non-anomaly).In the unsupervised way, different machine learning algorithms for anomaly detection are executed. The engines which did not fail are considered as censored. Using survival analysis techniques we tried to predict the failure of the engines.

321. A novel GEE type approach for analyzing zero-inflated clustered ordinal longitudinal data with an application to dental studies

[Poster Session, (page 35)]

Shoumi SARKAR, *University of Florida*
Somnath Datta, Peihua Qiu, Jeremy Gaskins, Steven Levy ,

We consider ordinal responses that are clustered at several levels and have excessive zeros (or some other category) compared to what is predicted by a fitted model otherwise. This data structure was motivated by multi-year fluorosis data obtained from a study on Iowa school children. Methods for inflated clustered longitudinal ordinal data are very limited in existing literature. We propose a novel GEE type approach, implementing latent variable modelling for the inflated category. A pseudo-likelihood is used model the mixture distribution. We develop a regression model connecting the latent continuous fluorosis levels with the observed ordinal responses. An expectation-solution (ES) approach is employed to solve the estimating equations. Extensive simulation studies are conducted to assess the performance of the model. The statistical findings demonstrate that this methodology is useful for analyzing such data and beyond. The real data analysis findings have tremendous clinical applications; factors like water sources, dietary factors, and age groups vulnerable to fluorosis can be identified and preventive interventions can be prescribed early on.

322. Covariance Networks for Functional Data on Multidimensional Domains

[30.M2.I82, (page 54)]

Soham SARKAR, *Indian Statistical Institute, Delhi Centre*

Covariance estimation is ubiquitous in functional data analysis. Yet, the case of functional observations over multidimensional domains introduces computational and statistical challenges, rendering the standard methods effectively inapplicable. We introduce

Covariance Networks (CovNet) as a modeling and estimation tool to address this problem. The CovNet model is universal; it can be used to approximate any covariance up to desired precision. Moreover, the model can be fitted efficiently and its neural network architecture allows us to employ modern computational tools in the implementation. The CovNet model also admits a closed-form eigendecomposition, which can be computed efficiently, without constructing the covariance itself. This facilitates easy storage and subsequent manipulation of the estimator. Moreover, the proposed estimator comes with theoretical guarantees in the form of consistency and rate of convergence. We demonstrate the usefulness of the proposed method on resting-state fMRI data.

Based on joint works with Victor M. Panaretos.

323. Statistical methods for evaluating gene-exposure interactions

[29.A2.I75, (page 48)]

Jaya SATAGOPAN, *Rutgers School of Public Health*

Clinical and epidemiological studies of anticancer therapies increasingly seek to identify predictive biomarkers to obtain insights into variation in treatment benefit. For time to event endpoints, a predictive biomarker is typically assessed using the interaction between the biomarker (or gene) and treatment (or exposure) in a proportional hazards model. Interactions are contrasts of summaries of outcomes and depend upon the choice of the outcome scale. This work investigates gene-exposure interactions under three scales of the outcome the natural logarithm of hazard ratio, the natural logarithm of survival probability, and survival probability at a prespecified time. The investigations show that a nonzero gene-exposure interaction can occur on the survival or the logarithm of survival probability scales even when there is no interaction (i.e., even if the model is additive) on the logarithm of hazard ratio scale. Since survival probabilities have clinically useful interpretation and are easier to convey to patients than hazard ratios, we recommend evaluating the role of a biomarker or gene on survival outcome using survival probabilities. Empirical illustration of the properties of three scales of gene-exposure interaction is provided using data from a published melanoma study.

324 . INFERENCE WITH JOINT MODELS UNDER MISSPECIFIED RANDOM EFFECTS DISTRIBUTIONS

[28.M2.I46, (page 30)]

Abdus SATTAR, *Case Western Reserve University*
Sanjoy SINHA, *Carleton University*

Abstract: Joint models are often used to analyze survival data with longitudinal covariates or biomarkers. Latent random effects that are used to describe the relationship between longitudinal and survival outcomes are typically assumed to follow a multivariate Gaussian distribution. A joint likelihood analysis of the data provides valid inferences under a correctly specified random effects distribution. However, the maximum likelihood method may produce biased estimators under a misspecified random effects distribution, and hence may provide invalid inferences. In this paper, we explore the empirical properties of the maximum likelihood estimators under various types of random effects, and propose a skewnormal distribution to address uncertainties in random effects. An extensive Monte Carlo study shows that our proposed method provides robust and efficient estimators under various types of model misspecifications. We also present an application of the proposed method using a large clinical dataset obtained from the genetic and inflammatory markers of sepsis (GenIMS) study.

325. Model-based imputation methods for small area estimation

[Student Paper Competition 2, (page 26)]

Aditi SEN, *University of Maryland, College Park*
Dr. Partha Lahiri ,

There is a growing demand to produce reliable estimates of different characteristics of interest for small geographical areas (e.g., states) or domains obtained by a cross classification of different demographic factors such as age, sex, race/ethnicity. The information on the outcome variable(s) of interest often comes from a sample survey that targets reliable estimation for large areas (e.g., national level). In this paper, we will discuss how model-based imputation methods can be used to improve inferences about different small area or domain parameters. The proposed method essentially uses suitable statistical models that can be used to extract information from multiple data sources. We illustrate the proposed methodology in the context of election projection for small areas.

326 . Copula-likelihood Maximization under Bivariate Random Censoring

[30.M1.I77, (page 51)]

Arusharka SEN, *Concordia University*
Magloire LOUDEGUI DJIMDOU, *Concordia University*
Yogendra CHAUBEY, *Concordia University*

In this work we extend the copula-based pseudo-likelihood estimation method of Genest et al. (1995) to bi-variate data subject to bi-variate random censoring. The extension is based on a recently developed multi-variate version (Sen and Stute (2013)) of the well-known Kaplan-Meier estimator. We also propose a new estimator for Kendalls tau based on the same estimator. The methods are illustrated using simulated data as well as a Canadian life-insurance data-set first used by Frees et al. (1995).

327 . Multivariate Distribution-free Nonparametric Testing using Optimal Transport

[Special Invited Session 3, (page 42)]

Bodhisattva SEN, *Columbia University*
Nabarun DEB, *University of British Columbia, Vancouver*
Bhaswar BHATTACHARYA, *University of Pennsylvania*

We propose a general framework for distribution-free nonparametric testing in multi-dimensions, based on a notion of multivariate ranks defined using the theory of optimal transport (see e.g., Villani (2003)). We demonstrate the applicability of this approach by constructing exactly distribution-free tests for two classical nonparametric problems: (i) testing for the equality of two multivariate distributions, and (ii) testing for mutual independence between two random vectors. In particular, we propose (multivariate) rank versions of Hotelling T^2 and kernel two-sample tests (e.g., Gretton et al. (2012), Szekely and Rizzo (2013)), and kernel tests for independence (e.g., Gretton et al. (2007), Szekely et al. (2007)) for scenarios (i) and (ii) respectively. We investigate the consistency and asymptotic distributions of these tests, both under the null and local contiguous alternatives. We also study the local power and asymptotic (Pitman) efficiency of these multivariate tests (based on optimal transport), and show that a subclass of these tests achieve attractive efficiency lower bounds that mimic the remarkable efficiency results of Hodges and Lehmann (1956) and Chernoff and Savage (1958) (for the Wilcoxon-rank sum test). To the best of our knowledge, these are the first collection of multivariate, nonparametric, exactly distribution-free tests that provably achieve such attractive efficiency lower bounds. We also study the rates of convergence of the rank maps (aka optimal transport maps).

328. Network Modeling of Contagion in High Dimensional Financial Time Series

[26.A2.I12, (page 7)]

Rituparna SEN, *Indian Statistical Institute, Bangalore*
Sumanta BASU, *Cornell University*
Georgia SMITS, *Cornell University*

Financial contagion is defined as a significant increase in cross-linkages after a shock to an individual financial entity. Networks, broadly understood as a collection of nodes and links between nodes, can be a useful representation of the financial system. In most of the related works in literature, network representation of the stock market is based on linear methods of association such as the cross-correlations of stock price fluctuations. We model a tail-based association network of financial linkages with contagion detected by using the method of Residual and Recurrence Times (RRT). Empirically, the resulting tail-based networks accurately capture the increase in market linkages at turbulent times, and discover new connections that linear methods of associations fail to detect.

329 . Scaling limits of random graph models at criticality: edge-weighted graphs converging to an L^3 graphon

[30.M2.I81, (page 53)]

Sanchayan SEN, *Indian Institute of Science*

Consider critical bond percolation on a sequence of edge-weighted graphs converging to an L^3 graphon. We use a general universality principle to establish the scaling limit of these random graphs viewed as metric measure spaces. This gives the first such result where no assumptions on the particular functional form of the graphon is required. As a corollary, we obtain the critical metric scaling limit of the RGIV model (random graphs with immigrating vertices) studied by Aldous and Pittel. Based on joint work with Jnaneshwar Baslingker, Shankar Bhamidi, Nicolas Broutin, and Xuan Wang.

330 . A semi-parametric model of ice accummulation rate based on temperature and age of ice for Antarctic ice core data

[29.M1.I58, (page 39)]

Debasis SENGUPTA, *Indian Statistical Institute*
Radhendushka SRIVASTAVA, *IIT Bombay*
Sreelekshmy S, *IIT Bombay*

In this paper we propose a semiparametric model for the apparent accumulation rate of ice at two Antarctic locations over the past 420 thousand years. The model postulates that the said rate is a simple product of a decreasing function of the age of the ice and a linear function of the average sea surface temperature prevailing at that age. The decreasing function represents thinning of the layers of ice with passing of age. The large sample properties of the model parameters are worked out, and the extent of their applicability at finite sample size are explored through simulations. Ice core data from Vostok and Epica Dome C for the relevant period fits the model very well. The close fit of the thinning adjusted accumulation rate to temperature, possibly due to absence of other influencing factors in the desert climate, gives one the unique opportunity to calibrate the exclusive effect of temperature, which works out to be an increase by about 7% with every degree Celsius rise in temperature. This rate is found to match the known rate of increase in the moisture absorbing capacity of the atmosphere with rising temperature.

331. Sequential Estimation for the Multiple Linear Regression Models with Balanced Loss Functions

[29.A2.I72, (page 47)]

Raghu Nandan SENGUPTA, *Indian Institute if Technology Kanpur*
Ren SCHILLING, *Institutfr Math. Stochastik, TU Dresden, GERMANY*
Sundeep R. BAPAT, *Indian Institute of Management Indore*
Neeraj JOSHI, *Indian Institute if Technology Kanpur*

Sequential analysis (SA) as a sampling technique has notable advantages like smaller average sample size and reduced value of risk, over similarly comparable fixed sampling techniques. In this study we first propose few models for the estimation of regression parameters or functions of the parameters, for the multiple linear regression (MLR) set-up under a balanced loss function (BLF). The respective proofs of the closed form solutions of the risk functions are general in nature while the proposed lemmas corroborate the admissibility of the estimators. Depending on the proposed stopping criteria (based on the concept of bounded risk) we propose different multi-stage sampling methodologies [viz. (i) Two-stage sampling, (ii) Three-stage sampling, (iii) Purely sequential sampling and (iv) Batch sequential sampling procedures], and corroborate the same with some detailed simulation runs.

332. HODOR: A two-stage hold-out design for online controlled experiments on networks

[28.M1.I36, (page 24)]

Srijan SENGUPTA, *North Carolina State University*
Nicholas LARSEN, *North Carolina State University*
Jonathan STALLRICH, *North Carolina State University*

A/B tests are standard tools for estimating the average treatment effect in online controlled experiments (OCEs). The majority of OCE theory relies on the Stable Unit Treatment Value Assumption, which presumes the response of individual users depends only on the assigned treatment, not the treatments of others. Violations of this assumption occur when users are subjected to network interference, a common phenomenon in social media platforms. Standard methods for estimating the average treatment effect typically ignore network effects and produce heavily biased results. Additionally, unobserved user covariates, such as offline information or variables hidden due to privacy restrictions, that influence user response and network structure also bias current estimators of the average treatment effect. In this paper, we demonstrate that network-influential lurking variables can heavily bias popular network clustering-based methods, making them unreliable. We propose a two-stage design and estimation technique called HODOR (Hold-Out Design for Online Randomized experiments) to address this problem. We show that HODOR is unbiased for the average treatment effect, has a minimizable variance, and provides reliable estimation even when the underlying network is partially-unknown or uncertain.

333. PREDICTING SUCCESS FOR A NEW SHOW ON A KIDS CHANNEL

[29.A2.C19, (page 50)]

Richa SHARMA, *Great Learning Mumbai India (Stu-*

dent)
MARVIN JOHN CARDOZA, *Great Learning Mumbai India*
RATNAKAR VIJAY ACHREKAR , *Great Learning Mumbai India*
UDAYAKUMAR D, *Great Learning Mumbai India*

Abstract Post pandemic all the television channels with kids content are going through a tough phase, online content with cheaper internet tariffs threatening the channels profitability. The space is highly cluttered with 11 channels. To keep the children engaged, new offerings need to be planned continuously. To maintain profitability, creation of new content and accurate risk assessment is the need of the hour.

Currently, success of a new show is measured by cost intensive interview-based research, the experience of the programming-head and the success rate of the production house. This study will solve 3 business questions on whether the show will be success or failure, important factors (like time-band of Launch etc) to be taken care during the launch and how current content can work harder if bucketed under different clusters and treated accordingly.

To address the problem, viewership for top kids channel Nickelodeon, for the fiscal year 2021 was considered, sourced from the Broadcast Audience Research Council of India. Data has 24896 rows and 18 feature variables. Each row represents minute wise viewership and columns represents time, program etc when the viewership was captured.

Data has gone through feature engineering 12 filters were created derived from 18 categorical variables deriving sharper insights.

Being a decision-making problem with categorical variable and binary outcome, Classification Algorithm became algorithm of choice

Logistic regression, random forest (RF), nave bayes, KNN, Cat-boost Model, Combination Model (logistic regression, RF and Cat boost Model) were utilized selecting feature variables which are important in defining Success or Failure of the show and model utilizing these variables for predicting success or failure of a new show. Combination model is best with 97% accuracy and 87% Precision. The model can give 6 most important features cancelling out lot of noise as well. This model will help to take data-based decisions on investments unlike qualitative analysis so far. To utilize existing content better, K-means clustering was utilized giving 4 cluster.

Keywords Television Channels, Kids Content, Profitability

334 . Modeling Hydrologic Extremes with Multivariate Extreme Value Models

[29.M2.I65, (page 44)]

SHAILZA SHARMA, *IISc*

Hydrologic extremes occur due to unusual combinations of multiple causative factors interacting in complex ways. The complex nature of these multidimensional extremes requires a multivariate viewpoint for robust risk assessment, especially in a warming climate. Investigating the extremal dependence structure in multiple dimensions is also crucial for understanding the interactions of plausible causative factors. Traditionally, univariate risk assessment methods have been used due to their simplicity; however, ignoring the dependence could lead to underestimation of the associated risk. This work aims at investigating the extremal behaviour of hydrologic extremes with multivariate extreme value models. Marginal distributions are specified and transformed into unit Frchet margins to characterize the complex dependence structure. Standardized distributions are represented by a Poisson point process and the coordinates of data points are further transformed to pseudo-polar coordinates to make the dependence form more explicit. Extremal dependence structure is described through angular densities on the unit simplex to gain the first insight and summarized using the extremal coefficient and coefficient of tail dependence. Estimated extremal dependence is further used to compute probabilities of critical combinations when all the components are simultaneously extremes and when the subset of components is extreme.

335 . Branching Brownian motion seen as a Markov process and its invariant measures

[27.A1.I28, (page 19)]

Atul SHEKHAR, *TIFR-CAM*
Xinxin CHEN, *Beijing Normal University*
Christophe GARBAN, *University Claude Bernard Lyon 1,*

We give a construction of branching Brownian motion as a Feller process on the space of point measures. We then answer some natural questions about it, e.g. invariant measures, domain of attraction, etc. This is based on joint works with Xinxin Chen and Christophe Garban.

336 . Game-theoretic formulation of sequential nonparametric two-sample tests

[29.M2.I60, (page 42)]

Shubhanshu SHEKHAR, *Department of Statistics and Data Science, Carnegie Mellon University*
Aaditya RAMDAS,

In this talk, I will describe a general framework for constructing sequential nonparametric two-sample tests. Our strategy relies on the principle of testing-by-betting, that reframes the task of sequential testing into that of selecting payoff functions that maximize the wealth of a fictitious bettor, betting against the null in a repeated game. To select the payoff functions, I will describe a simple strategy that proceeds by constructing predictable estimates of the witness function associated with a class of integral probability metrics (IPMs). The statistical properties of the resulting test can then be characterized in terms of the regret of this prediction strategy. I will then instantiate the general testing strategy for two popular IPMs: the Kolmogorov-Smirnov metric, and the kernel-MMD metric. I will conclude the talk by discussing some extensions to other testing problems.

This is based on joint work with Aaditya Ramdas (https://arxiv.org/abs/2112.09162).

337. Bayesian Estimation in Correlated Gamma Frailty Models

[28.A2.C12, (page 36)]

CHAKRDHAR SHELAKE, *Savitribai Phule Pune University*
CHAKRDHAR SHELAKE, *Savitribai Phule Pune University*
RAOSAHEB LLATPATE, *Savitribai Phule Pune University*
DAVID HANAGAL , *Savitribai Phule Pune University*

Frailty models are used in the survival data to account for the unobserved heterogeneity in individual risk to disease and death. The shared frailty models allow for the unbiased heterogeneity or statistical dependence between the observed survival data. To analysis the bivariate data on related survival time (e.g. matched pairs experiments, twins or family data), the shared frailty models were suggested. Shared frailty models are used despite their limitations. To overcome their disadvantages correlated frailty models may be used. In this paper, we consider correlated gamma frailty model with four different baseline distributions namely, the Weighted weibull distribution, Extended Weibull distribution, Pareto distribution and Linear failure rate. With these four baseline distributions we propose four different correlated frailty models. We also compare these models with the models where the above mentioned distributions are considered without frailty. We develop the Bayesian estimation procedure using Markov Chain Monte Carlo (MCMC) technique to estimate the parameters involved in these models. We present a simulation study to compare the true values of the parameters with the estimated values. A search of the literature suggests that currently no work has been done for these four baseline distributions with a correlated gamma frailty so far. We also apply these four models by using a real life bivariate survival data set of McGilchrist and Aisbett [15] related to the kidney infection data and a better model is suggested for the data.

338. Spatial Stochastic Frontier Model for Count Data

[29.A2.C17, (page 48)]

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

This paper introduces Spatial Stochastic Frontier Model (SSFM) for count type output data which incorporates the effect of spatial dependence as well as spatial heterogeneity in estimation of technical efficiency of decision making units. The model proposed builds upon the stochastic frontier models developed for count data but extends these models by taking into consideration the effects of spatial dependency as well as spatial heterogeneity. The proposed generalized form of SSFM for count data nests the Spatial Lag Model, Spatial Durbin Model as well as the Spatial Error Model as special cases in the Stochastic Frontier Model framework. Though flexible, the generalized form of SSFM faces identifiability problems and hence feasible specifications are persued and studied in detail. The effects on the estimation of spatial dependence parameter under various alternate specifications of the Spatial Weight Matrix are also studied. Bayesian approach has been used for estimation. The proposed Spatial Stochastic Frontier Model for count type data has been applied to district-wise data of severe crime and estimation of technical efficiency of states in India in handling of severe crimes has been attempted. Results suggest that the presence of spatial dependency as well as heterogeneity if ignored, leads to biased estimates of efficiency.

339. Sparse estimation in the finite mixture of accelerated failure time regression models

[27.A2.I33, (page 21)]

Farhad SHOKOOHI, *University of Nevada Las Vegas*
Masoud ASGHARIAN, *McGill University, Canada*
Abbas KHALILI, *McGill University, Canada*
Shili LIN, *Ohio State University, USA*

There has been a growing demand for analysis of high-dimensional data subject to censoring and intangible heterogeneity of population, specifically in biology and health sciences. For instance, we may be interested in relationship between disease free time after surgery among ovarian cancer patients and their DNA methylation profiles of genomic features. Such studies pose additional challenges beyond the typical big data problem due to population substructure and censoring. In this talk, we first lay down the challenges arise due to the complex structure of high-dimensional data including sparse estimation, and then we present our method to address some of these challenges. We specifically present the properties of our proposed method both theoretically and numerically. Finally, we analyze a dataset on high-grade serous ovarian cancer in an attempt to identify risky genes among over 9,000 genes.

340. On computationally efficient algorithm to estimate the two-dimensional chirp model parameters with the interaction term in phase

[26.A2.C2, (page 8)]

Abhinek SHUKLA, *IIT Kanpur*
Rhythm GROVER, *Mehta Family School of Data Science and Artificial Intelligence, Indian Institute of Technology Guwahati, Assam-781039, India*
Debasis KUNDU, *Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur - 208016, India*
Amit MITRA, *Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur - 208016, India*

We consider the two-dimensional (2D) chirp model (see Shukla et al. 2022) with interaction coefficient term in the phase. The deterministic part of the model is contaminated with linear stationary errors. The interaction term in such chirp models is an essential characteristic of numerous measurement interferometric signals or radar signal returns. The parameter estimation problem for these models is encountered in many real-life applications such as 2D-homomorphic signal processing, magnetic resonance imaging (MRI), optical imaging, interferometric synthetic aperture radar (INSAR), modeling non-homogeneous texture patterns, etc.

Due to the presence of the interaction term coefficient, the estimation of non-linear parameters becomes more difficult as the estimators of frequency parameters of both dimensions are no longer independent, making their computation and study of theoretical analysis more challenging. Several methods have been proposed for parameter estimation of these models. These methods, however are either statistically sub-optimal with a high signal-to-noise ratio (SNR) threshold, or they are computationally burdensome.

We contribute in the following ways: providing a computationally efficient algorithm to estimate the parameters of the 2D chirp model and further establishing theoretical asymptotic properties of the proposed estimators. The key attributes of the proposed method are that it is computationally faster and attains the same rates of convergence as the optimal LSEs. The proposed estimators of the chirp rate parameters have the same asymptotic variance as that of the traditional LSEs and are statistically optimal. We also present some simulation results validating the theoretical findings, and compared our estimators with one of the most classical multi-lag high order function based estimators (see Barbarossa et al. 2014).

References: Barbarossa, S., Di Lorenzo, P., & Vecchiarelli, P. (2014). Parameter estimation of 2D multi-component polynomial phase signals: an application to SAR imaging of moving targets. IEEE Transactions on Signal Processing, 62(17), 4375-4389.

Shukla, A., Grover, R., Kundu, D., & Mitra, A. (2022). Approximate least squares estimators of a two-dimensional chirp model. Journal of Multivariate Analysis, 105045.

341 . Bayes Analysis of Generalized Gamma and its component models using gradient based Markov Chain Monte Carlo Methods

[26.A2.C3, (page 9)]

Asmita SHUKLA, *Banaras Hindu University*
Rakesh RANJAN, *Banaras Hindu University*
Richa SRIVASTAVA, *Jaipuria Institute of Management, Lucknow*

The article performs Bayesian analysis of the Generalized gamma model and its important component models (i.e., Weibull and Gamma) using few important gradient-based Markov Chain Monte Carlo methods. These methods include Hamiltonian Monte Carlo, Metropolis adjusted Langevin algorithm and Metropolis-Hastings Algorithm with Barker proposal. The objective is to find an efficient gradient based MCMC method. Besides, the paper also compares the considered model and its component models via Bayes factor, where marginal likelihoods are calculated using Bridge sampling. Finally, a numerical illustration is provided based on a survival data set with right censored observations.

342. Option pricing using Hawkes Process

[28.A1.I53, (page 33)]

Shubhangi SIKARIA, *Philips India Limited*
Rituparna SEN, *ISI Bangalore, India*

We propose a methodology for European options pricing in which the Hawkes process drives variations in asset prices. We associate two point processes corresponding to the sum of the asset's positive and negative jumps. The point processes have a self and mutually exciting stochastic intensities with an exponential kernel. We employ a mean signature plot to estimate the parameters. We also demonstrate the hedging strategy for constant and stochastic intensities. We use Bayesian filtering in the Markov process to obtain the optimal hedge ratios for the stochastic intensities case. We examine the model's implementation in real data applications and compare it with the Black-Scholes formula.

343. Modeling and forecasting of crop yield using AI/MLT incorporating with weather variables

[30.M1.I76, (page 51)]

Kamlesh SINGH, *ICAR-IASRI, New Delhi*
Bishal GURUNG, *ICAR-IASRI, New Delhi*
Achal LAMA, *ICAR-IASRI, New Delhi*
Md Ashraful HAQUE, *ICAR-IASRI, New Delhi*

Modeling and forecasting of crop yield using AI/MLT incorporating with weather variables

K N Singh1, Bishal Gurung1, Achal Lama1, Md Ashraful Haque1, Manoj Varma1 and K K Singh2 1ICAR-Indian Agricultural Statistics Research Institute, New Delhi 2India Meteorological Department, Ministry of Earth Science, Govt. of India, New Delhi

Abstract: Crop yield is determined by numerous input parameters, it is vital to find important variables and omit the other redundant ones which may decrease the accuracy of predictive models. The machine learning driven feature selection algorithms assist in selecting only those features that are relevant in the predictive algorithms. Instead of a complete set of features, feature subsets give better results for the same algorithm with less computational time. Feature selection has the potential to play an important role in the agriculture domain, with the crop yield depending on multiple factors. Crop yield prediction is a complex phenomenon and has many underlining nonlinear patterns. Such, datasets are difficult to deal with stringent assumptions of the statistical models. Hence, machine learning (ML) techniques which has very few prior assumptions and are data driven provides great deal of flexibility for modelling and forecasting the crop yield. Various researchers have applied different ML techniques for forecasting crop yield and have obtained satisfactory results. We have attempted to build crop yield forecasting model based on features (weather indices) selected by using two very popular machine learning algorithms, i.e., Least Absolute Shrinkage and Selection Operator (LASSO) and Random Forest Variable Importance (RFVarImp). Further, we have also applied Random Forest (RF) and Support Vector Regression (SVR) directly to the data sets and have made a comparative analysis among them using appropriate statistical measures. Further, for ease of implementation and reaching out to more researchers and users, a webtool named WIAYFS (Weather Indices based Automated Yield Forecasting System, http://wiayfs.icar.gov.in/wiayfs) has been developed. In this webtool both statistical and machine learning models has been implemented and it also compares the performances of the models. At the end, user is delivered with the best fitted model along with forecast value for the used dataset. Keywords: Feature selection, Machine learning, crop yield, LASSO, Random forest, SVR.

344 . Inference of adaptive type-II progressively censored competing risks data for generalized inverted exponential distribution

[Poster Session, (page 35)]

Kundan SINGH, *Department of Mathematics, IIT Patna*
Yogesh Mani Tripathi ,

This paper consider the estimation of the unknown parameters, survival and hazard rate function of the generalized inverted exponential distribution using the adaptive type-II progressive censored competing risks data. We assumed that the lifetime random variables of the individual failure cause is independently distributed with a common scale and different shape parameters. For the estimation purpose maximum likelihood estimates are obtained and the existence and uniqueness are also proved. Subsequently, asymptotic confidence intervals of parameters, survival and hazard rate functions are constructed. Further MCMC techniques obtain Bayes estimates and highest posterior density credible intervals. The performance of the Bayes estimators is investigated and compared with maximum likelihood estimates using simulation. A real data set is also analysed in support of a competing risks model. Finally optimal censoring plans are established to improve the efficiency of experiments under three optimal criteria. We present a simulation study to check the performance of the optimized criteria.

345. Bayesian Estimation from Pareto Type II Distribution with Record Values using Monte-Carlo Techniques

[29.A2.C18, (page 49)]

SABHIYA SINGH, *University of Lucknow*
Rajeev PANDEY, *University of Lucknow*

In this paper, Pareto type II Distribution for obtaining Bayes estimators of unknown parameters considering the gamma prior for the parameters. The present study proposes Markov Chain Monte-Carlo (MCMC) techniques in finding sample from posterior distribution and computing bayes estimators. Maximum Likelihood is also proposed for point estimation and comparison of estimates obtained by using non-informative priors has been shown using Monte-Carlo Simulation by giving illustration of one real data set.

346. Cultural and Socio-demographic determinants affecting Female sterilization over other methods of Family planning in India: Evidence from Recent round of NFHS surveys.

[28.A2.C11, (page 36)]

Sanjiv SINGH , *Department of Statistics, Institute of Science , Banaras Hindu University, Varanasi(U.P) India*
Dr. Kaushlendra Kumar SINGH, *Department of Statistics, Institute of Science Banaras Hindu University, Varanasi (221005)*

Abstract Background- Fertility levels have dropped substantially in India in recent decades, fuelled by increased contraceptive use and notably a method mix skewed towards female sterilization. According to latest report of National family health survey TFR of India is 2.0 which takes it below the replacement level. InSite of this India is all set to surpass China as most populous nation in 2023 according to recently released World population prospectus because of momentum of population growth it had in the past. The recent decline in India has been unprecedented especially in Southern India, where fertility almost exclusively controlled by means of permanent contraceptive methods, mainly female sterilization, which more than half of overall contraceptive use. Many Indian women undergo sterilization at relatively young ages as a consequence of early marriage and childbearing in short birth intervals. Objective- Aim of our study is to investigate the socioeconomic factors determining dominating trends and patterns in female sterilization over reversible and traditional method among young married women in India. Methods- For the purpose of our study we have used data from three recent rounds of the National Family Health Surveys (NFHS). A multilevel analytical approach is used to estimate the impact of socioeconomic factors on contraceptive choices, differentiating temporary modern or traditional methods versus sterilization. The analysis is done with the help of software STATA. Results- The results of our study show significant decreases in the female sterilization and increase in other methods of family planning in fifth round of survey when compared to third round of the survey. Over the period, trends and pattern in uses of various methods of contraceptives is highly influenced by the heterogenous characteristics of population like age-group, maternal education, religion, caste, wealth index, family planning awareness, son preference, age at marriage, zonal classification etc. Although female sterilization has been decreasing over the time by a significant amount but still it is dominating among all available choices of contraceptives specially in the women with no education. Parity of women, number of living sons are other factors affecting trends in female sterilization. Conclusion- The study was successful in identifying important factors that revealed why the likelihood of a womans having undergone contracep-

tive sterilization was dominant over time and has declined significantly for the first time. Our study has proven importance of family planning awareness, level of education, parity of women, age at marriage, religion and size of the family. Keyword- DHS, NFHS, Multinomial Logistic regression,

347. Bayesian Optimal Stepped Wedge Design

[30.M2.I80, (page 53)]

Satya Prakash SINGH, *Indian Institute of Technology Kanpur*

Recently, there has been a growing interest in designing the cluster trials using stepped wedge design (SWD). A SWD is a type of cluster–crossover design in which clusters of individuals are randomized unidirectional from a control to an intervention at certain time points. Intra–class correlation coefficient (ICC) that measures the dependency of subject within a cluster plays an important role in design and analysis of a SWD. In this work we discussed a Bayesian approach to address the dependency of SWD on ICC and proposed robust SWDs.

348. Sculpting Human Fertility Using Variance-Adjusted Logistic Family of Distributions

[29.M1.C16, (page 41)]

SHAMBHAVI SINGH, *BANARAS HINDU UNIVERSITY*
AKANKSHA GUPTA, *BANARAS HINDU UNIVERSITY*
SATYANSHU K. UPADHYAY, *BANARAS HINDU UNIVERSITY*

With changes in the fertility patterns, the earlier developed demographic models fall short in imitating the changes that occur with respect to both time and geographic locations. Models providing a good fit for classical fertility patterns prove to be inadequate in case of distorted patterns, whereas those useful for distorted data are inefficient and can have poor predictive performance for traditional curves. In this paper, the Logistic distribution is taken as a base and new models based on logistic distribution are proposed for modelling the gradual changes in the age-specific fertility rates. The work consists of differentiating between the pre-modal and the post-modal variability and explores Bayes paradigm to deal with such problems. To show the relevance of the models in current scenario, the real-life age-specific fertility rate data of three countries, namely Denmark, India, and Ireland, having different age-specific fertility shapes for different years are considered and the posterior samples are generated for further analysis using the Metropolis algorithm. The proposed models are found compatible and satisfactory results are obtained for their respective usages. Finally, the proposed models are compared through some model comparison tools and the best among the proposed models is suggested. Keywords: Age-specific fertility rate modelling, Bayes paradigm, Bayesian information criterion (BIC), bimodality, predictive loss criteria (PLC), Logistic distribution, prior, skewness.

349 . Examining the factors influencing enrollment and grade appropriate learning outcomes in Indian rural schools

[28.A2.C13, (page 37)]

VIJAY SINGH, *Great Learning*
Srabashi BASU, *Great Learning, India*
Saathveenaa MOHAN, *Great Learning, India*
Yukta JOSHI, *Great Learning, India*

Annual Status of Education Report (Rural) 2018 (ASER 2018) has found mixed results among school children regarding enrollment status and learning outcome. Whereas out of school proportion in India is at 2.8%, only 50% of children in Standard V can read Standard II level text and less than 45% in Standard VIII can perform division. It is hypothesized that infrastructure and teaching quality impact school enrollment and learning outcome. ASER 2018 has published state level data on school infrastructure parameters and on a number of facets of enrollment, reading ability and performance on basic arithmetic operations.

In this study state level data collected by ASER 2018 has been investigated to discover the impact of infrastructure and teaching quality on school enrollment and grade-level learning. The data presents unique challenges due to relatively smaller number of records (26 states) compared to the number of possible predictors. To eliminate non-orthogonality among the predictors, principal component analysis has been performed. The top 5 principal components are considered as they are able to explain over 85% of variability. Principal component regression has been performed to predict out of school proportion and grade appropriate reading and arithmetic performance.

It is conjectured that learning at grade level progressively depends on a combination of school infrastructure and how much has been learnt at the previous grade. A state-level analysis is performed based on the available data. Initial results show that reading and arithmetic ability at the lower grade(s) is a significant predictor of performance at the current grade along with a combination of school infrastructures. Ability to perform subtraction by grade 3 is a significant predictor of ability to perform division by grade 4. The same effect is observed in reading performance, ability to read grade 2 paragraphs significantly depends on ability to read letters in grade 1. The school infrastructure plays a significant role on learning outcomes, some prominent results from our analysis show that learning outcomes are significantly dependent on pupil teacher ratio compliance, availability of usable toilets, electricity, and mid-day meals.

Keywords: ASER 2018, school infrastructure, school enrollment, grade-level learning, principal component regression

350 . Bayesian regression analysis of skewed tensor responses

[26.A1.I6, (page 5)]

DEBAJYOTI SINHA, *FLORIDA STATE UNIVERSITY*
Inkoo LEE, *Rice University, USA*
Dipankar BANDYOPADHYAY, *Virginia Commonwealth University*
Qing MAI, *Florida State University*

Tensor regression analysis is finding vast emerging applications in a variety of clinical settings, including neuroimaging, genomics, and dental medicine. The motivation for this paper is a study of periodontal disease (PD) with an order-3 tensor response: multiple biomarkers measured at prespecified toothsites within each tooth, for each participant. A careful investigation would reveal considerable skewness in the responses, in addition to missing responses. To mitigate the shortcomings of existing analysis tools, we propose a new Bayesian tensor response regression method that facilitates interpretation of covariate effects on both marginal and joint distributions of highly skewed tensor responses, and accommodates missing-at-random responses under a closure property of our tensor model. Furthermore, we present a prudent evaluation of the overall covariate effects while identifying their possible variations on only a sparse subset of the tensor components. Our method promises Markov chain Monte Carlo (MCMC) tools that are readily implementable. We illustrate substantial advantages of our proposal over existing methods via simulation studies and application to a real data set derived from a clinical study of PD. The R package BSTN available in GitHub implements our model.

351 . Efficient Estimation of the Additive Risks Model for Interval-Censored Data

[27.M1.I17, (page 12)]

Samiran SINHA, *Texas A&M University*
Tong WANG, *School of Statistics and Data Science, Nankai University, Tianjin, China*
Dipankar BANDYOPADHYAY, *Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA*

The semiparametric additive risks model (ARM) offers an attractive additive specification, allowing for direct assessment of the changes or the differences in the hazard function for changing values of the covariates. The ARM is a flexible model, allowing the estimation of both time-independent and time-varying covariates. It has a nonparametric component and a regression component identified by a finite-dimensional parameter. An efficient approach for maximum-likelihood (ML) estimation of the nonparametric and finite-dimensional components will be presented. Specifically, a novel minorize-maximize (MM) algorithm will be presented that has facilitated the efficient computation of the estimator for case-II interval-censored data. The operating characteristics of the proposed MM approach will be illustrated via simulation studies and the analysis of a breast cancer dataset. It is expected that the proposed computational approach will not only provide scalability to the ML estimation scenario but may also simplify the computational burden of other complex likelihoods or models.

352 . Constrained inference in mixed models for clustered data

[28.M2.I46, (page 30)]

Sanjoy SINHA, *Carleton University, Ottawa, Canada*

Mixed models are widely used for analyzing clustered data, including longitudinal data and repeated measurements. Unrestricted maximum likelihood (ML) methods were extensively studied in the literature for analyzing generalized, linear, and mixed

models. However, constraints or parameter orderings may occur in practice, and in such cases, we can improve the efficiency of a statistical method by incorporating parameter constraints into the ML estimation and hypothesis testing. In this talk, I will discuss constrained inference in generalized linear mixed models (GLMMs) under linear inequality constraints. Methods will be assessed using both Monte Carlo simulations and actual survey data from a health study.

353. Spectral regularized kernel two-sample test

[27.A1.I26, (page 18)]

Bharath SRIPERUMBUDUR, *Pennsylvania State University*

Over the last decade, an approach that has gained a lot of popularity to tackle non-parametric testing problems on general (i.e., non-Euclidean) domains is based on the notion of reproducing kernel Hilbert space (RKHS) embedding of probability distributions. The main goal of our work is to understand the optimality of two-sample tests constructed based on this approach. First, we show that the popular MMD (maximum mean discrepancy) two-sample test is not optimal in terms of the separation boundary measured in Hellinger distance. Second, we propose a modification to the MMD test based on spectral regularization by taking into account the covariance information (which is not captured by the MMD test) and prove the proposed test to be minimax optimal with a smaller separation boundary than that achieved by the MMD test. Third, we propose an adaptive version of the above test which involves a data driven strategy to choose the regularization parameter and show the adaptive test to be almost minimax optimal up to a logarithmic factor. Moreover, our results hold for the permutation variant of the test where the test threshold is chosen elegantly through the permutation of the samples. Through numerical experiments on synthetic and real-world data, we demonstrate superior performance of the proposed test in comparison to the MMD test.

[Joint work with Omar Hagrass (PSU) and Prof. Bing Li (PSU)]

354. Gaussian Process Regression and Classification Using International Classification of Disease Codes as Covariates

[28.M1.I36, (page 24)]

Sanvesh SRIVASTAVA, *The University of Iowa*
Sanvesh SRIVASTAVA, *University of Iowa*
Yunyi LI, *University of Iowa*
Stephanie GILBERTSON-WHITE, *University of Iowa*

International Classification of Disease (ICD) codes are widely used for encoding diagnoses in electronic health records (EHR). Automated methods have been developed to predict biomedical responses using EHR that borrow information from similar patients. Relatively less attention has been paid to developing patient similarity measures that model the structure of ICD codes and the presence of multiple chronic conditions, where a chronic condition is defined as a set of ICD codes. Motivated by this problem, we first develop a type of string kernel function for measuring similarity between a pair of subsets of ICD codes, which uses the definition of chronic conditions. Second, we extend this similarity measure to define a family of covariance functions on subsets of ICD codes. Using this family, we develop Gaussian process (GP) priors for Bayesian nonparametric regression and classification using diagnoses and other demographic information as covariates. Markov chain Monte Carlo (MCMC) algorithms are used for posterior inference and predictions. The proposed methods are free of any tuning parameters and are well-suited for automated prediction of continuous and categorical biomedical responses that depend on chronic conditions. We evaluate the practical performance of our method on EHR data collected from 1660 patients at the University of Iowa Hospitals and Clinics (UIHC) with six different primary cancer sites. Our method has better sensitivity and specificity than its competitors in classifying different primary cancer sites and estimates the marginal associations between chronic conditions and primary cancer sites.

355. Optimal pooling and distributed inference for the tail index and extreme quantiles

[27.A2.I32, (page 20)]

Gilles STUPFLER, *ENSAI & University of Angers*
Abdelaati DAOUIA, *Toulouse School of Economics*
Simone PADOAN, *Bocconi University*

We investigate pooling strategies for tail index and extreme quantile estimation from heavy-tailed data. To fully exploit the information contained in several samples, we present general weighted pooled Hill estimators of the tail index and weighted pooled

Weissman estimators of extreme quantiles calculated through a nonstandard geometric averaging scheme. We develop their large-sample asymptotic theory across a fixed number of samples, covering the general framework of heterogeneous sample sizes with different and asymptotically dependent distributions. Our results include optimal choices of pooling weights based on asymptotic variance and MSE minimization. In the important application of distributed inference, we prove that the variance-optimal distributed estimators are asymptotically equivalent to the benchmark Hill and Weissman estimators based on the unfeasible combination of subsamples, while the AMSE-optimal distributed estimators enjoy a smaller AMSE than the benchmarks in the case of large bias. We consider additional scenarios where the number of subsamples grows with the total sample size and effective subsample sizes can be low. An application to insurance data across several US states is presented. Joint work with Abdelaati Daouia (Toulouse School of Economics) and Simone Padoan (Bocconi University).

356. Tampered Random Variable Modeling for Multiple Step-stress Life Test

[26.A1.I5, (page 5)]

Farha SULTANA, *Assistant Professor*
Anup DEWANJI, *Professor*

In this paper, we introduce the Tampered Random Variable (TRV) modeling in multiple step-stress life testing experiments. We define the tampered random variable in multiple step-stress scenario and calculate the PDF, CDF, and Hazard rate for the proposed tampered variable. We derive a general expression for the expectation of TRV under different number k of stress levels and also obtain some results on stochastic ordering for different k. All these results are obtained under arbitrary baseline (under normal stress condition with normal stress level) life distribution. In particular, we consider exponential distribution and Weibull distribution for specific expressions. We also prove some results on equivalence of the TRV modeling with the two other existing models for step-stress life testing, namely, Cumulative Exposure (CE) and Tampered Failure rate (TFR). Finally, we consider some variations of the modeling approach for TRV to include incorporation of the stress levels, discrete life time, bivariate or multivariate life times.

357. A Modern CLT for the classical AIPW: variance inflation, cross-fit covariance and beyond

[26.A2.I8, (page 6)]

Pragya SUR, *Harvard University*
Kuanhao JIANG, *Harvard University*
Subhabrata SEN, *Harvard University*
Rajarshi MUKHERJEE , *Harvard University*

Estimating the average treatment effect (ATE) is a central problem in causal inference. Modern advances in the field studied estimation and inference for the ATE in high dimensions through varied approaches. Doubly robust estimators such as the augmented inverse probability weighting (AIPW) form a popular approach in this context. However, the high-dimensional literature surrounding these estimators relies on sparsity conditions, either on the outcome regression (OR) or the propensity score (PS) model. This talk will introduce a new central limit theorem for the classical AIPW estimator, that applies agnostic to such sparsity-type assumptions. Specifically, we will study properties of the cross-fit version of the estimator under well-specified OR and PS models, and the common modern regime where the number of features and samples are both large and comparable. Under assumptions on the covariate distribution, our CLT will uncover two crucial phenomena among others: (i) the cross-fit AIPW exhibits a substantial variance inflation that can be precisely quantified in terms of the signal-to-noise ratio and other problem parameters, (ii) the asymptotic covariance between the estimators used while cross-fitting is non-negligible even on the root-n scale. These findings are strikingly different from their classical counterparts, and open a vista of possibilities for studying similar other high-dimensional effects. On the technical front, our work utilizes a novel interplay between three distinct toolsapproximate message passing theory, the theory of deterministic equivalents, and the leave-one-out approach. Time permitting, I will outline some of these techniques.

358. Reliability modeling of a two component parallel load sharing system with non-identical components

[27.M1.C5, (page 14)]

Santosh SUTAR, *Shivaji University, Kolahpur*

The systems in which the failure of a component affects the residual life of the surviving components are known as load sharing systems. We consider a two component parallel load system with non-identical components. We model the load sharing phenomenon observed in such system through the exponentiated conditional distribution function of ordered failure times. The proposed model leads to a different bivariate family of distributions for ordered random variables. Estimation procedure for the proposed model with baseline distribution as an exponential is discussed. The simulation study is carried out to check the performance of the proposed estimation procedure. We also carry out the data analysis to demonstrate the applicability of the proposed model.

359. Quantile regression with a mixture of function-valued covariate prone to complex heteroscedastic measurement errors and a scalar-valued covariate prone to classical measurement error

[29.M2.I61, (page 42)]

Carmen TEKWE, *Indiana University at Bloomington*
Xiwei CHEN, *Indiana University at Bloomington*
Yuanyuan LUAN, *Indiana University at Bloomington*
Roger ZOH, *Indiana University at Bloomington*

Many current recommendations for dietary intake (DI) and physical activity (PA) to maintain optimal health and minimize risks for chronic health conditions, such as obesity, are based on statistical analyses of data prone to measurement error, including those collected from self-reported questionnaires and wearable devices. Self-reported measures based on food frequency questionnaires are often used in DI assessments, however, they are prone to recall bias. Wearable devices enable the continuous monitoring of PA but generate complex functional data with poorly characterized systematic errors. In this work, we propose the sparse conditional quantile regression model with function- and scalar- valued covariates prone to measurement errors. We develop semiparametric and parametric approaches to correct for measurement errors associated with the mixture of functional and scalar covariates prone to errors in quantile regression settings. Extensive simulations are performed to assess the finite sample properties of the proposed methods. The developed methods are applied to investigate the influence of wearable-device-based PA and self-reported measures of total caloric intake on quantile functions of fat mass index (FMI), an indicator for obesity. The device-based measures of PA are assumed to be prone functional covariates prone to complex arbitrary heteroscedastic errors, while DI is assumed to be a scalar-valued covariate prone to error. The developed methods are applied to National Health and Examination Survey data to assess the relationship between PA and DI with quantile functions of FMI among community dwelling adults living in the United States.

360. Mixed Membership Models for Functional Data

[27.M2.I22, (page 16)]

Donatello TELESCA, *UCLA*
Nicholas MARCO, *UCLA*
Damla SENTURK, *UCLA*

Partial membership models, or mixed membership models, are a flexible unsupervised clustering method that allows observations to belong to multiple clusters at the same time. In this paper, we propose a Bayesian partial membership model for functional data. By using the multivariate Karhunen-Loeeve theorem, we are able to derive a scalable representation that maintains data-driven covariance structures and establish conditional posterior consistency. Compared to previous work on partial membership models, our proposal allows for increased flexibility, with the benefit of direct interpretation of the mean and covariance functions. Our work is motivated by studies in functional brain imaging of children with Autism Spectrum Disorder (ASD).

361. A spatial copula interpolation in a random field with application in air pollution data

[Student Paper Competition 2, (page 26)]

Debjoy THAKUR, *Indian Institute of Technology, Tirupati*
Dr. Ishapathik Das, Ms. Shubhashree Chakravarty ,

Interpolating a skewed conditional spatial random field with missing data is cumbersome in the absence of Gaussianity assumptions. Copulas can capture different types of joint tail characteristics beyond the Gaussian paradigm. Maintaining spatial homogeneity and continuity around the observed random spatial point is also challenging. Especially when interpolating along a spatial surface, the boundary points also demand focus in forming a neighborhood. As a result, importing the concept of hierarchical clustering on the spatial random field is necessary for developing the copula model with the interface of the Expectation-Maximization algorithm and concurrently utilizing the idea of the Bayesian framework. This article introduces a spatial cluster-based C-vine copula and a modified Gaussian distance kernel to derive a novel spatial probability distribution. To make spatial copula interpolation compatible and efficient, we estimate the parameter by employing different techniques. We apply the proposed spatial interpolation approach to the air pollution of Delhi as a crucial circumstantial study to demonstrate this newly developed novel spatial estimation technique.

362. Joint Calibration Approach-Based Estimator for Dual Frame Surveys using Two Auxiliary Variables

[27.A2.C8, (page 23)]

Shiwani TIWARI, *Banaras Hindu University*

Piyush RAI, *Department of Statistics, Banaras Hindu University, Varanasi, Uttar Pradesh, India*

The present article deals with the estimation of finite population total from dual-frame surveys utilizing two auxiliary variables using calibration approach. An empirical study has been carried out using real data for numerical illustration to show the efficiency of the proposed joint calibrated estimator. In addition, a simulation study is also performed where the proposed estimator is found to be more efficient than the existing estimator in terms of average absolute relative bias (ARB), simulated relative standard error (SRSE), and percent relative efficiency (PRE) for both the cases, i.e., equal sample allocation as well as in proportional sample allocation.

363. Guided unsupervised outlier detection using domain knowledge

[Poster Session, (page 35)]

Vinay TIWARI, *Savitribai Phule Pune University*

Dr. Akanksha Kashikar ,

Outlier detection has always been an area of interest for many domains including fraud detection, intrusion detection, medical analysis, etc. This problem can be solved in both supervised as well as unsupervised way. Several state of the art (SOTA) unsupervised methods are available for outlier detection. However, these methods lack in two departments - first, they do not utilize domain knowledge of an analyst and second, these methods lack interpretability. These SOTA unsupervised methods perform outlier detection only based the structure of the explanatory/X variables without any information about the relationship with target/Y variable. In this paper, we propose two algorithms - first, Guided Empirical Cumulative Outlier Detection (GECOD) algorithm which allows the analyst to define feature relationship with target variable and second, Guided Logistic Regression Outlier Detection (GLROD) algorithm which is a supervised approximation of GECOD and it provides interpretability at a global level. Extensive experiments have been performed using both simulations and real datasets to show that these algorithms not only provide improved interpretability but also provide higher accuracy compared to the baseline unsupervised SOTA methods.

Keywords: Domain knowledge, Global interpretability, Logistic regression, Unsupervised outlier detection

364. Bayesian Weibull Quantile Regression and Variable Selection with Application Uterine Serous Carcinoma patient Survival

[26.A2.C3, (page 9)]

Tripti TRIPATHI, *Banaras Hindu University*

Rakesh RANJAN, *DST-CIMS, BHU, Varanasi*

Weibull distribution is one well known model when it comes to study probabilistic nature of time-to-event phenomena. This article provides the full Bayesian analysis of the Weibull Quantile regression model. In order to deal with analytically intractable posterior, Bayesian analysis is performed using sample-based approaches, in particular the Metropolis algorithm. A symmetric Multivariate Gaussian kernel is used to execute the algorithm. To perform variable selection, all possible subsets of regressor variables are fitted with response variable, and calculation of Bayes information criterion (BIC), deviance information criterion (DIC) and log pseudo-marginal likelihood (LPML) have been done. Finally, numerical illustration has been provided for real data set of Uterine Serous Carcinoma patient survival.

365. An Application of Elliptical Spatial Linear Models

[Poster Session, (page 35)]

ASHIS RANJAN UDGATA, *ICAR-Indian Agricultural Statistics Research Institute*

ANIL RAI, *ICAR-Indian Agricultural Statistics Research Institute*

ANKUR BISWAS, *ICAR-Indian Agricultural Statistics Research Institute*

NOBIN CHANDRA PAUL, *ICAR-Indian Agricultural Statistics Research Institute*

Spatial analysis is the technique applied to analyze of geographic data which includes attribute data as well as topological data. In recent years there is high demand for modeling spatial dependence structure and estimating parameters in various disciplines such as agriculture, geology, soil science, hydrology, ecology, oceanography, forestry, meteorology, and climatology. As per the distributional form of the errors there are several procedures for estimation of the parameters. Elliptical distributions are broad family of probability distributions that generalize the multivariate normal distribution. In this study the family of elliptical distributions is used to estimate the spatial dependence in georeferenced data. The methodology is applied to a real data set and effect of sample data points has been studied.

Keywords: Elliptical distributions, Oceanography, Spatial analysis, Spatial dependence.

366. Adaptive schemes for sampling in infinite dimensions

[27.M2.I23, (page 17)]

Sreekar VADLAMANI, *TIFR CAM*
Jonas WALLIN,
Sibsankar SINGHA,

Latent Gaussian processes are widely applied in many fields like statistics, inverse problems, and machine learning. A popular method for inference is through the posterior distribution, which is typically carried out by Markov Chain Monte Carlo (MCMC) algorithms. However, the infinite dimensional framework creates certain technical hurdles which need to be addressed with care. Taking cue from recent developments in adaptive Metropolis adjusted MCMC algorithms, we propose a family of proposals to sample from "good" infinite dimensional measures. We discuss the relevant issues concerned with convergence of our proposed schemes, and also demonstrate their efficiency via standard computational examples.

367. Assessing the feasibility of feature selection algorithms for wheat yield forecasting in regression framework

[29.A2.C18, (page 49)]

Manoj VARMA, *PhD Scholar, ICAR-IASRI, New Delhi*
Kamalesh Narain SINGH, *Principal Scientist and Head (F and ASM), ICAR-IASRI, NEW DELHI*
Achal LAMA, *Scientist, ICAR_IASRI, New Delhi*

Assessing the feasibility of feature selection algorithms for wheat yield forecasting in regression framework *M. VARMA, K. N. SINGH AND A. LAMA ICAR Indian Agricultural Statistics Research Institute, New Delhi, 110012 *Presenting Authors Email: manojvarma57282@gmail.com Abstract

Wheat is arguably the most important food crop in the world. Approximately 750 million metric tons (MMT) of wheat are produced worldwide today. India produces around 107 MMT (2021-22). Therefore, the wheat yield forecasting is crucial for ensuring the safety of the nation's food supply. It is crucial to deliver timely information for the best management of growing wheat, including early determination of the import/export plan and price. The dimensionality issue prevents the regression framework from being utilized for wheat yield forecasting because wheat yield is determined by many input parameters. It is necessary to separate the most crucial variables and get rid of those that can compromise the predictive model's accuracy. The feature selection algorithms help us to choose only relevant features for the predictive algorithms. In current work, the implementation of feature selection algorithms, including forward selection, backward selection, correlation-based feature selection (CBFS), and other machine learning algorithm based approaches, was done on three different wheat datasets of the Amritsar, Jalandhar and Gurdaspur district of Punjab. All the three datasets contain observations of past 37-year yield (1982-2018) and the weekly weather parameters includes maximum temperature, minimum temperature, relative humidity, surface pressure, wind speed and surface moisture. Yield forecasting models under regression framework were developed for each district individually based on the features selected. Statistical indicators such as Root Mean Square Error (RMSE), Mean Absolute Prediction Error (MAPE) and Mean Absolute Deviation (MAD) were used to compare the forecasting performance of the proposed models. The developed models were compared with the weather indices based stepwise regression model and encouraging results were obtained. The developed models can be implemented for forecasting yield of other crops as well.

Keywords: Feature selection, Wheat yield, Weather indices, Regression model, Forecasting.

368. A unified approach to a class of distributions

[Student Paper Competition 1, (page 13)]

Durga VASUDEVAN, *Department of Statistics, Cochin University of Science and Technology*
Dr. Asha Gopalakrishnan ,

In this paper we generalize bivariate distributions via a addition equation which reduces to the Cauchy functional equation under certain conditions. We use a * operator which is reducible and associative binary operator corresponding to a continuous and strictly increasing function g(.). Necessary and sufficient conditions for distributions to belong to this general class is studied. Finally we illustrate how these results help to construct distributions belonging to the general class of distributions proposed.

369 . Output Analysis for Parallel MCMC

[28.M1.I41, (page 26)]

Dootika VATS, *Indian Institute of Technology Kanpur*
Agarwal MEDHA, *University of Washington, Seattle*
Kushagra GUPTA, *Stanford University*

The ever-increasing power of the personal computer has led to easy parallel implementations of Markov chain Monte Carlo (MCMC). However, almost all work in estimating the variance of Monte Carlo averages, focuses on averaging estimates from single-chain MCMC runs. We demonstrate that simply averaging covariance matrix estimators from multiple chains (average BM) can yield critical underestimates in small sample sizes, especially for slow mixing Markov chains. Using global-centering, we propose a simple alteration to the estimator, that utilizes information from parallel chains, thereby correcting for the underestimation. Consequences of this improved estimation are discussed in the context of autocorrelation plots and effective sample size.

370. Hybrid Statistical and Deep Learning Technique Over Different Viable Approaches for Improved Predictive Performance of Rainfall

[28.A2.C12, (page 36)]

VEERSHETTY, *PhD Scholar at ICAR-IASRI*
G AVINASH, *PhD Scholar at ICAR-IASRI*
Harish NAYAK. G.H., *PhD Scholar at ICAR-IASRI*
Moumita BAISHYA, *PhD Scholar at ICAR-IASRI*

Rainfall estimation is one of the most significant and difficult environmental issues in a present situation. Predicting rainfall requires complex and effective models and simulations because it is a very nonlinear and dynamic process. Hence, robust and accurate rainfall forecasting models need to be developed by applying Machine Learning (ML), Deep Learning (DL) approaches and hybrid models. Several automatic systems were created to predict weather, but it depends on the type of weather pattern, season and location, which leads in maximizing the processing time. Therefore, in this study, different neural network techniques are used and a simple classical time series Auto-Regressive Integrated Moving Average (ARIMA) model is hybridized with Long Short-Term Memory (LSTM) deep learning network. In this study, monthly rainfall data of Coastal Karnataka over the period of 115 years (i.e., from 1901-2015) is used as more rain occurs in the region. The gathered information is analysed using ARIMA, LSTM, Generalized Regression Neural Network (GRNN), Recurrent Neural Network (RNN) and ARIMA-LSTM models, which are capable of processing the time series data and predict the dependency between the data effectively. The most challenging phase of ARIMA model is failing to capture the non-linearity. For obtaining the non-linearity, deep learning neural network i.e., LSTM was hybridized. Proposed hybrid model was compared with ARIMA and other nonlinear volatility capturing techniques like LSTM, GRNN and RNN. Results indicated that the prediction accuracy of the proposed hybrid model is superior compared to other four competitive models for predicting the rainfall in Coastal Karnataka.

371 . EFFICIENT SEQUENTIAL THIRD ORDER RESPONSE SURFACE DESIGNS

[27.A2.C7, (page 22)]

ANKITA VERMA, *STUDENT*
SEEMA JAGGI, *ADG (HRD) ,ICAR , New Delhi-110012*
ELDHO VARGHESE, *Senior Scientist, ICAR-CMFRI, Kochi-682018*
ANINDITA DATTA, *Scientist, ICAR-IASRI, New Delhi-110012*

EFFICIENT SEQUENTIAL THIRD ORDER RESPONSE SURFACE DESIGNS Ankita Verma1, Seema Jaggi2, Eldho Varghese3, Arpan Bhowmik4, Cini Varghese1 and Anindita Datta1

1ICAR-Indian Agricultural Statistics Research Institute,New Delhi - 110 012 2Agricultural Education Division, ICAR, New Delhi - 110 012 3ICAR-

Central Marine Fisheries Research Institute, Kochi - 682 018 4ICAR-Indian Agricultural Research Institute, Assam

Abstract Sequential third-order response surface designs are advantageous when there is significance of lack of fit of a fitted second-order model while establishing the relationship between the input and response variables. The sequential experimentation approach comprises of conducting the trials one at a time, in contrast to the non-sequential experimentation strategy, which executes the complete runs at once. The practical use of sequential experimentation is that it is more economical and uses less experimental resources. For example, in agriculture, ongoing crop cultivation may deplete previously available mineral elements, necessitating a sequential addition of the mineral elements that over time become deficient in the soil. A method of construction of sequential third order design is suggested for symmetric as well as mixed level factors that satisfy the required condition of moment matrix and ensure rotatability. Factorial points and axial points are used in the first stage of the construction method, which is then supplemented with a balanced incomplete block design to create a complete third order design. The first stage design may be utilized to fit the second order model, and with a few more runs, the third order model can be fitted without discarding the first stage design. Additionally, the proposed designs have smaller sizes, making it more cost-effective to attain the best response using the proposed design. A list of proposed class of designs has been presented along with their G-efficiency.

Key words: Moment matrix; Lack of fit; Response surface methodology; Rotatability; Sequential design; Third order model.

372 . PARTIALLY BALANCED BIPARTITE BLOCK DESIGNS

[27.A2.C6, (page 22)]

VINAYAKA, *PhD Scholar (Agricultural Statistics), ICAR-IASRI, New Delhi-110012*

Rajender PARSAD, *Principal Scientist, ICAR-IASRI, Pusa, New Delhi-110012*

B. N. MANDAL, *Senior Scientist, ICAR-IASRI, Pusa, New Delhi-110012*

Sukanta DASH, *Scientist, ICAR-IASRI, Pusa, New Delhi-110012*

This article provides some new construction methods of partially balanced bipartite block (PBBPB) designs for comparing test treatments with more than one control. Partially balanced incomplete block (PBIB) designs based on some association schemes such as triangular association, Latin-square association, group divisible association, and cyclic association are used for developing these methods of construction. A catalogue of PBBPB designs is included for parameter values v1 (number of test treatments) 10, v2 (number of control treatments) =2, r1 (replications of test treatments) 10 and r2 (replications of control treatments) 15 along with computed variances using SAS software.

373 . AIM-AHEAD: Artificial Intelligence and Machine Learning to address health disparities and achieve health equity

[29.A2.I74, (page 47)]

Jamboor VISHWANATHA, *Univ. of North Texas Health Science Center*

The rapid increase in the volume of data generated through electronic health records (EHR) and other biomedical research presents exciting opportunities for developing data science approaches (e.g., AI/ML methods) for biomedical research and improving healthcare. Many challenges hinder more widespread use of AI/ML technologies, such as the cost, capability for widespread application, and access to appropriate infrastructure, resources, and training. Additionally, lack of diversity of both data and researchers in the AI/ML field runs the risk of creating and perpetuating harmful biases in its practice, algorithms, and outcomes, thus fostering continued health disparities and inequities. Many underrepresented and underserved communities, which are often disproportionately affected by diseases and health conditions, have the potential to contribute expertise, data, diverse recruitment strategies, and cutting-edge science, and to inform the field on the most urgent research questions, but may lack financial, infrastructural, and data science training capacity to apply AI/ML approaches to research questions of interest to them.

The AIM-AHEAD Coordinating Center (A-CC) is a consortium of institutions and organizations that have a core mission to serve under-represented or underserved groups (minority populations, low socioeconomic, rural, sexual gender minorities) impacted by health disparities (e.g., Historically Black Colleges and Universities, Tribally Controlled Colleges and Universities, etc.). The A-CC focused initially on coordination, assessment, planning, and capacity

building to enhance the use of artificial intelligence (AI) and machine learning (ML) in research among the consortium institutions and organizations; and to build and sustain trusted relationships between the consortium and groups impacted by health disparities. Through the AIM-AHEAD CONNECT, the networking platform of the A-CC, training, mentoring and networking resources available to the stakeholders will be demonstrated.

This research was, in part, funded by the National Institutes of Health (NIH) Agreement No. 1OT2OD032581-01. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NIH.

374. Building the Infrastructure and Co-design of Data Sets for AI/ML applications to address health equity

[29.A2.I74, (page 47)]

Lavanya VISHWANATHA , *University of North Texas Health Science Center*

Lavanya Vishwanatha AIM-AHEAD Coordinating Center, University of North Texas Health Science Center, Fort Worth, Texas, USA

The AIM-AHEAD Coordinating Center (A-CC) is a consortium of institutions and organizations that have a core mission to serve under-represented or underserved groups (minority populations, low socioeconomic, rural, sexual gender minorities) impacted by health disparities (e.g., Historically Black Colleges and Universities, Tribally Controlled Colleges and Universities, etc.). The A-CC focused initially on coordination, assessment, planning, and capacity building to enhance the use of artificial intelligence (AI) and machine learning (ML) in research among the consortium institutions and organizations; and to build and sustain trusted relationships between the consortium and groups impacted by health disparities.

Through the Research, Infrastructure and Training cores, the A-CC has developed resources for a national audience of stakeholders. In this presentation, various models of infrastructure that meet the needs of different stakeholders will be presented. These models accommodate needs of stakeholders that need concierge services and existing datasets, stakeholders that have datasets that never leaves their institution/organization, to centralized hosting and computing in the cloud. Co-design of data sets to integrate social determinants along with EHR and providing access to these resources to various stakeholders will be discussed. A data landscape has been developed that includes publicly available data and privately held data, and these data will be made accessible through the infrastructure models. Data curation and data governance will be discussed.

This research was, in part, funded by the National Institutes of Health (NIH) Agreement No. 1OT2OD032581-01. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NIH.

375. Incorporating mechanistic knowledge in causal inference

[29.A1.I67, (page 45)]

Alexander VOLFOVSKY, *Duke University*

At their core, the assumptions needed for causal inference are concerned with removing the effects of potentially unobserved quantities. We may know that a drug is given, but maybe not when, or we may observe where a disease is transmitted but maybe not exactly from whom, yet in both settings we might be interested in causal questions: Does the drug have an effect? Does a mitigation strategy work to prevent future transmission? Because these processes are governed by established biological mechanisms, mechanistic models can provide invaluable insights into the interactions between biological objects (drug diffusion in the body, transmission probabilities between individuals). Conditioning on these models can provide more credibility to the necessary assumptions for causal inference. We present two case studies of leveraging these types of models for causal inference: (1) we analyze observational data of critically ill patients and identify the effect of seizures if they were not treated, and (2) we employ a mechanistic model of disease transmission to help design a trial for evaluating a non-pharmaceutical intervention.

376. Interim Monitoring in Sequential Multiple Assignment Randomized Trials

[26.A2.I9, (page 6)]

Abdus WAHED, *University of Pittsburgh*
Liwen WU, *University of Pittsburgh*
Junyao WANG, *University of Pittsburgh*

A sequential multiple assignment randomized trial (SMART) facilitates the comparison of multiple adaptive treatment strategies (ATSs) simultaneously. Previous studies have established a framework to test the homogeneity of multiple ATSs by a global Wald test through inverse probability weighting. SMARTs are generally lengthier than classical clinical trials due to the sequential nature of treatment randomization in multiple stages. Thus, it would be beneficial to add interim analyses allowing for an early stop if overwhelming efficacy is observed. We introduce group sequential methods to SMARTs to facilitate interim monitoring based on the multivariate chi-square distribution. Simulation studies demonstrate that the proposed interim monitoring in SMART (IM-SMART) maintains the desired type I error and power with reduced expected sample size compared to the classical SMART. Finally, we illustrate our method by reanalyzing a SMART assessing the effects of cognitive behavioral and physical therapies in patients with knee osteoarthritis and comorbid sub-syndromal depressive symptoms.

377. Online Nonparametric Monitoring of Heterogeneous Data Streams with Partial Observations based on Thompson Sampling

[26.A2.I11, (page 7)]

Xiaochen XIAN, *University of Florida*

With the rapid advancement of sensor technology driven by various Internet-of-Things (IoT) enabled applications, nowadays tremendous amounts of measurements of heterogeneous data streams are frequently acquired and utilized for online process monitoring and quick anomaly detection. Such massive data, involving a large number of data streams with high sampling frequency, incur high costs on data collection, transmission, and analysis in practice. As a result, the resource constraint often restricts the data observability to only a subset of data streams at each data acquisition time, posing significant challenges for practitioners in many online monitoring applications. Unfortunately, existing methods do not provide a general framework for monitoring heterogeneous data streams with partial observations, as heterogeneous data streams contain distinct physical characteristics and follow different distributions. In this article, we propose a nonparametric monitoring and sampling algorithm integrated with Thompson sampling to quickly detect abnormalities occurring to heterogeneous data streams. In particular, an approximation framework is incorporated with an antirank-based cumulative sum (CUSUM) procedure to collectively estimate the underlying status of all data streams based on partially observed data. Furthermore, an intelligent sampling strategy based on Thompson sampling is proposed to dynamically observe the informative data streams and balance between exploration and exploitation to facilitate quick anomaly detection. Theoretical justification of the proposed algorithm is also investigated. Both simulations and case studies are comprehensively conducted to evaluate the performance and demonstrate the superiority of the proposed method.

378 . Likelihood-based Inference for Stochastic Epidemic Models via Data Augmentation

[29.A1.I67, (page 45)]

Jason XU, *Duke University*

Stochastic epidemic models such as the Susceptible-Infectious-Removed (SIR) model are widely used to model the spread of disease at the population level, but fitting these models to observational data present significant challenges. In particular, the marginal likelihood of such stochastic processes conditioned on observed endpoints a notoriously difficult task. As a result, likelihood-based inference is typically considered intractable in the presence of missing data, and practitioners often resort to simulation methods or approximations. We discuss some recent contributions that enable ”exact” inference using the likelihood of observed data, focusing on a perspective that makes use of latent variables to explore configurations of the missing data within a Markov chain Monte Carlo framework. Motivated both by count data from large outbreaks and high-resolution contact data from mobile health studies, we show how our data-augmented approach successfully learns the interpretable epidemic parameters and scales to handle large realistic data settings efficiently

379. A Flexible Micro-Randomized Trial Design and Sample Size Considerations

[28.A2.C13, (page 37)]

Jing XU, *Duke National University of Singapore*

Jing XU, *Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore*

Xiaoxi YAN, *Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore*

Bibhas CHAKRABORTY, *Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore*

Technological advancements have made it possible to deliver mobile health interventions to individuals. A novel framework that has emerged from such advancements is the just-in-time adaptive intervention (JITAI), which aims to suggest the right support to the individuals when their needs arise. The micro-randomized trial (MRT) design has been proposed recently to test the proximal effects of these JITAIs. However, the extant MRT framework only considers components with a fixed number of categories added at the beginning of the study. We propose a flexible MRT (FlexiMRT) design, which allows addition of more categories to the components during the study. The proposed design is motivated by collaboration on the DIAMANTE study, which learns to deliver text messages to encourage physical activity among the patients with diabetes and depression. We developed a new test statistic and the corresponding sample size calculator for the FlexiMRT using an approach similar to the generalized estimating equation (GEE) for longitudinal data. Simulation studies was conducted to evaluate the sample size calculators and an R shiny application for the calculators was developed.

380. Statistical Inference for Generalized half-normal distribution for simple step stress case in the presence of hybrid censoring

[27.A2.C7, (page 22)]

Priya YADAV, *Department of Statistics, Central University of Haryana, Mahendergarh, India*
Devendra KUMAR, *Department of Statistics, Central University of Haryana, India*
Sukhdev SINGH, *Department of Statistics, Amity University, Mohali, India*

Accelerated life testing experiments have become more popular in recent years. Our key objective in accelerated life testing experiments is to produce lifetime data in a short period of time for statistical analysis, and these types of experiments are particularly more effective when the units under test are extremely reliable. The problem of a simple step stress accelerated life test for a generalized half-normal distribution in the presence of hybrid censoring is considered in this study. To relate the lifetime of products at different stress levels, we use the Cumulative Exposure Model (CEM). Both classical and Bayesian inferential approaches for estimation of parameters are discussed. The stochastic expectation maximization (SEM) algorithm is used to obtain maximum likelihood estimates. Further Bayes estimates are computed using the squared error loss function, and a simulation study is conducted to evaluate the efficiency of the suggested approaches in various situations. Finally, for an illustrative purpose, a real data set is analyzed.

381. Modelling of sparse conditional spatial extremes processes subject to left-censoring

[Poster Session, (page 35)]

Rishikesh YADAV, *HEC Montreal*
Leo R. Belzile, Nicholas R. Beck ,

The conditional spatial extremes model of Wadsworth and Tawn, which focuses on extreme events given threshold exceedance at a site, has garnered a lot of attention as a flexible way to model large-scale spatio-temporal events. We consider extensions that combine Gaussian Markov random field residual processes along with data augmentation schemes for dealing with left-censored realizations, exploiting the sparsity of the precision matrix obtained through the basis function approximation of the Gaussian process. Models are fitted using Markov chain Monte Carlo methods through a combination of Metropolis-within-Gibbs, Langevin steps, and Laplace approximations. We showcase the scalability of the approach using precipitation data from British Columbia.

382 . Baxter-type convergence results for a block Toeplitz system under long memory with an application to linear prediction problem

[28.A1.I51, (page 32)]

Junho YANG, *Academia Sinica*
Akihiko INOUE,

The Wiener-Hopf (WH) equation is a semi-infinite Toeplitz system of equations that has diverse applications. In practice, truncation of the system is inevitable to calculate the solution of the WH equation in a finite time. Baxter's inequality provides an L_1-bound for the approximation error between the WH solution and its finite-section approximation. However, this inequality is only valid for a symbol with short memory. In this talk, we derive the Baxter-type convergence results for a block Toeplitz system when the corresponding matrix-valued symbol has a long memory. A key ingredient is using a series expansion of the inverse of a finite-order Toeplitz matrix. Based on these results, we show the Baxter-type convergence for linear prediction problems for multivariate long memory stationary processes. This is a joint work with Akihiko Inoue.

383. Effective spatio-temporal disease surveillance by using covariate information

[26.A2.I11, (page 7)]

Kai YANG, *Medical College of Wisconsin*
Peihua QIU, *University of Florida*
Kai YANG, *Medical College of Wisconsin*

Effective surveillance of infectious diseases, cancers and other deadly diseases is critically important for public health and safety of our society. Incidence data of such diseases are often collected spatially from different clinics and hospitals through a regional, national or global disease reporting system. In such a system, new batches of data keep being collected over time, and a decision needs to be made immediately after new data are collected regarding whether there is a disease outbreak at the current time point. This is the spatio-temporal disease surveillance problem that will be focused in the talk. There are some existing methods for solving this problem, most of which use the disease incidence data only. In practice, however, disease incidence is often associated with some covariates, including the air temperature, humidity, and other weather or environmental conditions. In this talk, we will introduce a new methodology for spatio-temporal disease surveillance which can make use of the helpful covariate information to improve its effectiveness. A novelty of this new method is behind its property that only the covariate information which is associated with a true disease outbreak can help trigger a signal from the new disease surveillance method. The new method can accommodate seasonality, spatio-temporal data correlation, and nonparametric data distribution. These features make it feasible to use in many real applications.

384. Poisson process approximation under stabilization

[26.A2.I7, (page 6)]

D YOGESHWARAN, *Indian Statistical Institute*

We present some Poisson process approximation results for stabilizing functionals of Poisson (or Binomial) processes that arise in stochastic geometry. Our bounds are derived for the Kantorovich-Rubinstein distance between a point process and an appropriate Poisson point process. We will discuss application to largest k-nearest neighbour distances. This is based on a joint project with Omer Bobrowski (Technion) and Matthias Schulte (Hamburg Institute of Technology) [see arXiv:2104.13261].

385. Flexible and Interpretable Learning for High-Dimensional Complex Data

[Plenary Lecture 2, (page 17)]

Helen ZHANG, *University of Arizona*

Modern data are often massive, high-dimensional, and intrinsically complex. When learning with such data, it is desired to build a decision rule that is both flexible and interpretable. The final model should be able to capture nonlinear patterns and also provide interpretable insights. In this talk, I will present a broad class of regularization frameworks for high-dimensional nonparametric regression and classification, including their theory, computation, and real applications.

386. Bayesian Semi-Parametric Scalar-On-Function Quantile Regression with Measurement Error using GAL

[29.M2.I61, (page 42)]

Roger ZOH, *Indiana University*

Quantile regression provides a consistent approach to investigating the association between covariates and various aspects of the distribution of the response. Most methods assume that the covariates in the regression are precisely measured with no potential for errors. Here, we propose extending the Bayesian measurement error and Bayesian quantile regression literature to allow for functional covariates prone to measurement error. Our approach uses the Generalized Asymmetric Laplace (GAL) distribution. The family of Gal distribution has recently emerged as a more flexible distribution family in Bayesian quantile regression modeling. We compare the proposed approach's performance to that of a naive approach that ignores measurement error. We finally apply our approach to the analysis of an NHANES dataset.

Directory

A, Praveenkumar
praveenkumarupm@gmail.com
`Ph.D. scholar, ICAR-IASRI, New Delhi`
Speaker: 29.A2.C18, p. 49, §1, p. 59

ABBOTT, Sam

Panelist: Panel Discussion 4, p. 44,
Speaker: 29.M1.I56, p. 38, §2, p. 59

ACHARYYA, Satwik
University of Michigan
`satwika@umich.edu`
Speaker: 27.A2.I34, p. 21, §3, p. 59

ADHIKARI, Kartick
IISER Bhopal
`kartick@iiserb.ac.in`
Speaker: 30.M2.I81, p. 53, §4, p. 59

AGASHE, Nehatai
PGS IARI, New Delhi
`nehaagashe29@gmail.com`
Speaker: 26.A2.C2, p. 9, §5, p. 60

APTE, Amit
Indian Institute of Science Education and Research (IISER) Pune
`apte@iiserpune.ac.in`
Speaker: 28.M1.I41, p. 26, §6, p. 60

ARGIENTO, Raffaele
Università degli studi di Bergamo
`raffaele.argiento@unibg.it`
Chair: 27.M1.I16, p. 12,
Chair: 27.A2.I35, p. 21,
Speaker: 27.M1.I16, p. 12, §7, p. 60

ARYA, Sakshi
Penn State University
`arya.sakshi44@gmail.com`
Organizer: 27.A1.I26, p. 18,
Speaker: 27.A1.I26, p. 18, §8, p. 60

ASHKAMINI,
Banaras Hindu University
`ashkamini99@gmail.com`
Chair: 29.M1.C15, p. 40,
Speaker: 29.M1.C15, p. 41, §9, p. 60

AVINASH, G
PhD Scholar at ICAR-Indian Agricultural Statistics Research Institute
`avinash143stat@gmail.com`
Speaker: 28.A2.C11, p. 35, §10, p. 61

AWALE, Manik
Savitribai Phule Pune University
`manik.stats@gmail.com`
Organizer: 27.A2.I31, p. 20,
Speaker: 27.A2.I31, p. 20, §11, p. 61

AYYER, Arvind
Indian Institute of Science
`Arvind@iisc.ac.in`
Speaker: 30.M1.I78, p. 52, §12, p. 61

BABEL, Sneha
H.V.Desai College, Pune
`snehashingvi23@gmail.com`
Speaker: 29.A2.C19, p. 49, §13, p. 62

BAISHYA, MOUMITA
Ph.D. Scholar
`moumitabaishya1194@gmail.com`
Speaker: 29.A2.C17, p. 48, §14, p. 62

BALAJI, Srinivasan
Department of Statistics, George Washington University
`balaji@gwu.edu`
Chair: 26.A1.I5, p. 5,
Speaker: 26.A1.I5, p. 5, §15, p. 62

BALAKRISHNAN, Sivaraman
Carnegie Mellon University
`the.seeing.stone@gmail.com`
Organizer: 30.M2.I79, p. 53,
Speaker: 30.M2.I79, p. 53, §16, p. 62

BALASUBRAMANIAN, Krishnakumar
University of California, Davis
`kbala@ucdavis.edu`
Speaker: 29.M2.I60, p. 42, §17, p. 63

BANDYOPADHYAY, Antar
Indian Statistical Institute, Delhi Centre
`antar@isid.ac.in`
Speaker: 26.A1.I2, p. 4, §18, p. 63

BANDYOPADHYAY, Dipankar
Virginia Commonwealth University
`bandyopd@gmail.com`
Organizer: 27.M1.I14, p. 11

BANDYOPADHYAY, Soutir
Colorado School of Mines
`bsoutir@gmail.com`
Speaker: 26.A2.I8, p. 6, §19, p. 63

BANDYOPADHYAY, Subhadip
Principal Data Scientist, GAIA, Ericsson, Bangalore
`subhadip.bandyopadhyay@ericsson.com`

Chair and organizer: 29.A1.I71, p. 46,
Speaker: 29.A1.I71, p. 46, §20, p. 63

BANERJEE, Anjishnu
Medical College of Wisconsin
`abanerjee@mcw.edu`
Chair: 26.A2.I11, p. 7,
Speaker: 26.A2.I11, p. 7, §21, p. 64

BANERJEE, Arka
arkabee96@gmail.com
`Indian Institute of Technology, Kanpur`
Speaker: 28.M1.C9, p. 27, §22, p. 64

BANERJEE, Hiya
Novartis Pharmaceuticals
`hiyabanerjee@gmail.com`
Panelist: Panel Discussion 2, p. 17,
Organizer: 27.A1.I25, p. 18,
Speaker: 28.M2.I46, p. 30, §23, p. 65

BANERJEE, Imon
Purdue University
`ibanerj@purdue.edu`
Speaker: Student Paper Competition 1, p. 13, §24, p. 65

BANERJEE, Jyotirmoy
Principal Data Scientist, GAIA, Ericsson, Bangalore
`jyotirmoy.a.banerjee@ericsson.com`

Speaker: 29.A1.I71, p. 46, §25, p. 65

BANERJEE, Moulinath
University of Michigan
moulib@umich.edu
Chair and organizer: In Memorium:, p. 21,
Speaker: 26.A1.I5, p. 5, §26, p. 65

BANERJEE, Rudrani
GSK
rudranibanerjee@gmail.com
Speaker: 29.A1.I68, p. 45, §27, p. 66

BAPAT, Sudeep
Indian Institute of Management Indore
sudeepb@iimidr.ac.in
Chair and organizer: 29.A2.I72, p. 47

BASAK, Anirban
ICTS-TIFR
anirban.basak@icts.res.in
Speaker: 30.M1.I78, p. 52, §28, p. 66

BASU, Pallavi
Indian School of Business
pallavi_basu@isb.edu
Speaker: 26.A1.I3, p. 4, §29, p. 66

BASU, Riddhipratim
International Centre for Theoretical Sciences
rbasu@icts.res.in
Chair: 28.A1.I50, p. 32,
Speaker: 28.A1.I50, p. 32, §30, p. 66

BASU, Sanjib
University of Illinois Chicago
sbasu@uic.edu
Speaker: 29.A2.I73, p. 47, §31, p. 67

BASU, Saonli
University of Minnesota
saonli@umn.edu
Speaker: In Memorium:, p. 21,
Speaker: Special Invited Session 4, p. 52, §32, p. 67

BHADRA, Anindya
Purdue University
bhadra@purdue.edu
Speaker: 28.M1.I37, p. 24, §33, p. 67

BHAKTA, RAJU
National Institute of Technology Rourkela
bhakta.r93@gmail.com
Speaker: Poster Session, p. 33, §34, p. 67

BHAT K, Shrikrishna
Pondicherry University
skbhat.in@pondiuni.ac.in
Speaker: 28.M1.C10, p. 27, §35, p. 68

BHATTACHARJEE, Abhishek
Pfizer

Chair: 28.M1.C10, p. 27,
Speaker: 28.M1.C10, p. 28, §36, p. 68

BHATTACHARJEE, Madhuchhanda
University of Hyderabad, Hyderabad, India
mbsm@uohyd.ac.in
Organizer: 27.M1.I15, p. 11,
Speaker: 28.A1.I49, p. 31, §37, p. 68

BHATTACHARJEE, Monika
Department of Mathematics, IIT Bombay
monaiidexp.gamma@gmail.com
Speaker: 26.A1.I3, p. 4, §38, p. 69

BHATTACHARJEE, Satarupa
UC Davis
sbhattacharjee@ucdavis.edu
Speaker: 28.A1.I49, p. 31, §39, p. 69

BHATTACHARYA, Arkajyoti
Post doctoral researcher, University of Groningen, The Netherlands
a.bhattacharya@rug.nl
Chair: 28.M1.C9, p. 27,
Speaker: 28.M1.C9, p. 27, §40, p. 69

BHATTACHARYA, Ayan
Assistant Professor, Department of Mathematics, IIT Bombay
ayanbhattacharya.isi@gmail.com
Speaker: 29.A1.I69, p. 45, §41, p. 70

BHATTACHARYA, Bhaswar
University of Pennsylvania
bhaswar.bhattacharya@gmail.com
Speaker: 28.A1.I52, p. 32, §42, p. 70

BHATTACHARYA, Rahul
Professor
rlbstat@caluniv.ac.in
Speaker: 29.M2.I62, p. 43, §43, p. 70

BHATTACHARYA, SUMANGAL
Department of Mathematics and Statistics, IIT Tirupati
sumangal.82@gmail.com
Speaker: 27.M1.C5, p. 14, §44, p. 70

BHATTACHARYYA, Arinjita
Merck
arinjita.bhattacharyya@merck.com

Speaker: 29.M1.I55, p. 38, §45, p. 71

BHATTACHARYYA, Sharmodeep
Oregon State University
sharmodeep.bhattacharyya@oregonstate.edu

Organizer: 29.M2.I63, p. 43

BHOGLE, Srinivas

Panelist: Panel Discussion 1, p. 3

BHOWMICK, Amiya Ranjan
Institute of Chemical Technology, Mumbai
amiyaiitb@gmail.com
Speaker: 28.M2.I45, p. 29, §46, p. 71

BHUYAN, Prajamitra
Indian Institute of Management, Calcutta
prajamitra.bhuyan@gmail.com
Speaker: 29.A1.I70, p. 46, §47, p. 72

BISWAS, Anushua
IISER Pune
anushua.biswas@students.iiserpune.ac.in

Speaker: 27.A1.I30, p. 19, §48, p. 72

BISWAS, Arunangshu
GSK
aruangshu.x.biswas@gsk.com
Speaker: 29.A1.I68, p. 45, §49, p. 73

BISWAS, Debjit

Chair: 26.A2.I10, p. 7

BISWAS, Shilpi
Research Scholar, IIT Guwahati
shilpi.biswas@iitg.ac.in
Speaker: 27.M1.C4, p. 14, §50, p. 73

BISWAS, Suparna
Indian Statistical Institute, Bangalore
suparnabsws4@gmail.com
Speaker: 26.A2.I12, p. 7, §51, p. 73

BISWAS, Swati
University of Texas at Dallas
swati.biswas@utdallas.edu
Speaker: 29.A2.I75, p. 48, §52, p. 73

BOSE, Arup
Indian Statistical Institute
bosearu@gmail.com
Speaker: 27.M1.I15, p. 12, §53, p. 74

BOUEZMARNI, Taoufik
Universit de Sherbrooke
Taoufik.Bouezmarni@USherbrooke.ca

Speaker: 30.M1.I77, p. 51, §54, p. 74

BS, Yashavanth
yashavanthbs@gmail.com
ICAR-National Academy of Agricultural Research Management
Speaker: 28.M1.C9, p. 27, §55, p. 74

C. S., Soorya
Department of Statistics, Cochin University of Science and Technology
sooryasurendran93@gmail.com
Speaker: Poster Session, p. 33, §86, p. 84

CARONE, Marco
University of Washington
mcarone@uw.edu
Speaker: 27.A2.I33, p. 20, §56, p. 75

CATALANO, Marta
University of Warwick
marta.catalano@warwick.ac.uk
Speaker: 27.A2.I35, p. 21, §57, p. 75

CHAKRABARTY, Arijit
Indian Statistical Institute, Kolkata
arijit.isi@gmail.com
Speaker: 26.A2.I7, p. 5, §58, p. 75

CHAKRABORTY, Anirvan
IISER Kolkata
anirvan.c@iiserkol.ac.in
Speaker: 30.M2.I82, p. 54, §59, p. 76

CHAKRABORTY, Antik
Purdue University
antik015@purdue.edu
Chair: 29.M1.C14, p. 40,
Speaker: 29.M1.C14, p. 40, §60, p. 76

CHAKRABORTY, Bibhas
National University of Singapore & Duke University
bibhas.chakraborty@duke-nus.edu.sg

Speaker: Short Course 1, p. 3,
Chair and organizer: 28.M1.I39, p. 25,
Speaker: 28.M1.I39, p. 25, §61, p. 76

CHAKRABORTY, Kaustav
University of Illinois at Urbana-Champaign
kaustav2@illinois.edu
Speaker: Poster Session, p. 33, §62, p. 77

CHAKRABORTY, Nilanjana
nchakraborty@ufl.edu
University of Pennsylvania
Speaker: 27.A2.C7, p. 22, §63, p. 77

CHAKRABORTY, Sounak
University of Missouri
chakrabortys@missouri.edu
Speaker: 27.M1.I14, p. 11, §64, p. 77

CHAKRAVARTTY, Arunava
Novartis
archnova@gmail.com
Speaker: 27.A1.I25, p. 18, §65, p. 78

CHANDNA, Swati
Birkbeck, University of London
s.chandna@bbk.ac.uk
Speaker: 29.M2.I63, p. 43, §66, p. 78

CHANDRA, Noirrit Kiran
The University of Texas at Dallas
noirritchandar@gmail.com
Speaker: 28.M1.I40, p. 25, §67, p. 78

CHANDRA, PRAKASH
Indian Institute of Technology, Patna
prakash_2021ma16@iitp.ac.in
Speaker: 29.M1.C14, p. 40, §68, p. 78

CHATTERJEE, Kiranmoy
Department of Statistics, Bidhannagar College, Kolkata
kiranmoy07@gmail.com
Speaker: 29.M1.C15, p. 40, §69, p. 79

CHATTERJEE, Nilanjan
Johns Hopkins University
nilanjan@jhu.edu
Speaker: 28.M1.I38, p. 25, §70, p. 79

CHATTERJEE, Sabyasachi
Assistant Professor
sc1706@illinois.edu
Speaker: 29.M2.I60, p. 42, §71, p. 79

CHATTERJEE, Sabyasachi
University of Illinois at Urbana Champaign
sc1706@illinois.edu
Chair and organizer: 29.M2.I60, p. 42

CHATTERJEE, Saptarshi
Eli Lilly and Co.
chatterjee_saptarshi@lilly.com
Speaker: 29.M1.I57, p. 39, §72, p. 80

CHATTERJEE, Saptarshi
Eli Lilly and Co
chatsap@gmail.com
Organizer: 29.M1.I57, p. 39

CHATTERJEE, Shirshendu
City University of New York, Graduate Center & City College
shirchat1@gmail.com
Speaker: 29.M2.I63, p. 43, §73, p. 80

CHATTERJEE, Snigdhansu
University of Minnesota
`chatt019@umn.edu`
Chair: 26.A2.I8, p. 6,
Chair and organizer: 27.M2.I21, p. 16,
Chair: Plenary Lecture 2, p. 17,
Panelist: Panel Discussion 2, p. 17,
Organizer: 29.M1.I54, p. 38,
Speaker: 27.M1.I17, p. 12, §74, p. 81

CHATTERJEE, Suvo
Indiana University, Bloomington
`suvo2chatterjee@gmail.com`
Speaker: 29.M1.I55, p. 38, §75, p. 81

CHATTOPADHYAY, Bhargab
Indian Institute of Technology Jodhpur
`bhargab@iitj.ac.in`
Speaker: 29.A2.I72, p. 47, §76, p. 81

CHAUBEY, Yogendra
Concordia University
`yogen.chaubey@concordia.ca`
Speaker: 30.M1.I77, p. 51, §77, p. 81

CHAUDHURI, Sanjay
National University of Singapore
`stasc@nus.edu.sg`
Chair: 29.M1.I59, p. 39,
Chair: Special Invited Session 3, p. 42,
Organizer: 30.M1.I76, p. 51,
Speaker: 27.M2.I21, p. 16, §78, p. 82

CHENG, Yu
University of Pittsburgh
`yucheng@pitt.edu`
Speaker: 26.A2.I9, p. 6, §79, p. 82

CHENJI, Suresh
Eli Lilly
`chenji_suresh@lilly.com`
Speaker: 27.M1.C5, p. 14, §80, p. 82

CHETIYA, Anuradha Rajkonwar
Ramjas College, University of Delhi
`anuradha_rc@hotmail.com`
Speaker: 28.A2.C12, p. 36, §81, p. 83

CHITTURI, Pallavi
Temple University

Panelist: Panel Discussion 3, p. 31,
Organizer: Panel Discussion 3, p. 31

CHOUDHARY, Kapil
ICAR-IASRI, New Delhi
`kapiliasri@gmail.com`
Speaker: 29.M2.I64, p. 43, §82, p. 83

CHOWDHURY, Joydeep
King Abdullah University of Science and Technology
`joydeepchowdhury01@gmail.com`
Speaker: 28.M2.I47, p. 30, §83, p. 83

CHOWDHURY, Shovan
Indian Institute of Management Kozhikode
`shovanc@iimk.ac.in`
Chair and organizer: 27.A1.I29, p. 19,
Speaker: 27.A1.I29, p. 19, §84, p. 84

CHOWDHURY, Shrabanti
Icahn School of Medicine at Mount Sinai
`shrabanti.chowdhury@mssm.edu`
Speaker: 30.M2.I80, p. 53, §85, p. 84

CUI, Yifan
Zhejiang University
`yifanc095@gmail.com`
Speaker: 28.M1.I39, p. 25, §87, p. 84

DALAL, ASHUTOSH
ICAR-INDIAN AGRICULTURAL STATISTICS RESEARCH INSTITUTE
`ashutosh.dalal97@gmail.com`
Speaker: 27.A2.C8, p. 23, §88, p. 85

DAN, Biltu
Indian Institute of Science
`biltudanmath@gmail.com`
Speaker: 30.M2.I81, p. 53, §89, p. 85

DAS, Kalyan
kalyanstat@gmail.com
`Former Professor IIT Bombay`
Speaker: Plenary Lecture 1, p. 3, §90, p. 85

DAS, Kiranmoy
Indian Statistical Institute, Kolkata
`kiranmoy.das@gmail.com`
Speaker: 29.M2.I62, p. 43, §91, p. 86

DAS, Priyam
Department of Biostatistics, Virginia Commonwealth University
`pdasvcu@gmail.com`
Chair: 27.M1.I14, p. 11,
Speaker: 27.M1.I14, p. 11, §92, p. 86

DAS, Sourish
CMI
`sourish@cmi.ac.in`
Chair: 26.A2.I12, p. 7,
Speaker: 26.A2.I12, p. 7, §93, p. 86

DAS, Sudipta
Ramakrishna Mission Vivekananda Educational and Research Institute
`jusudipta@gmail.com`
Chair: 29.A2.I74, p. 47,
Speaker: 29.A2.I74, p. 48, §94, p. 87

DAS, Sudipta
Ramakrishna Mission Vivekananda Educational and Research Institute
`jusudipta@gmail.com`
Organizer: 29.M1.I58, p. 39

DASGUPTA, Kohinoor
Novartis Healthcare Pvt Ltd
`kohinoor.dasgupta@novartis.com`
Speaker: 27.A1.I25, p. 18, §95, p. 87

DASGUPTA, Sayan
Fred Hutchinson Cancer Center
`sdasgup2@fredhutch.org`
Speaker: 27.M1.I18, p. 13, §96, p. 87

DASGUPTA, Shibasish
Pfizer and Chennai Mathematical Institute (CMI)
`shibasish.dasgupta@pfizer.com`
Chair and organizer: 29.M2.I62, p. 43,
Speaker: 26.A2.I10, p. 7, §97, p. 88

DASH, Sukanta
ICAR-Indian Agricultural Statistics Research Institute

`sukanta.iasri@gmail.com`
Speaker: 28.A2.C11, p. 36, §98, p. 88

DATTA, Abhirup
Johns Hopkins University
`abhidatta@jhu.edu`
Chair: 27.M1.I18, p. 13,
Chair: 29.M1.I54, p. 38,
Speaker: 29.M1.I54, p. 38, §99, p. 88

DATTA, Anindita
Scientist
`anindita.datta@icar.gov.in`
Speaker: 29.M1.C16, p. 41, §100, p. 88

DATTA, Jyotishka
Virginia Tech
`jyotishka@vt.edu`
Organizer: 26.A1.I3, p. 4,
Organizer: 28.M1.I37, p. 24,
Speaker: 28.M1.I37, p. 24, §101, p. 89

DATTA, Somnath
University of Florida
`somnath.datta@ufl.edu`
Speaker: 27.M1.I13, p. 11, §102, p. 89

DATTA, Susmita
University of Florida
`susmdatta@gmail.com`
Organizer: Panel Discussion 2, p. 17,
Speaker: 29.M2.I62, p. 43, §103, p. 89

DE BLASI, Pierpaolo
University of Torino and Collegio Carlo Alberto, Italy
`pierpaolo.deblasi@unito.it`
Organizer: 27.M1.I16, p. 12,
Organizer: 27.M2.I22, p. 16,
Organizer: 28.M1.I40, p. 25

DE, Shyamal Krishna
Indian Statistical Institute Kolkata
`shyamalkd@gmail.com`
Chair and organizer: 30.M2.I82, p. 54,
Speaker: 29.M2.I63, p. 43, §104, p. 90

DEB, Navonil
Cornell University
`navonil1997@gmail.com`
Speaker: 29.M1.C15, p. 40, §105, p. 90

DEB, Soudeep
Indian Institute of Management Bangalore
`soudeep@iimb.ac.in`
Chair and organizer: 29.M2.I64, p. 43,
Chair: 29.A1.I66, p. 44,
Speaker: 29.A1.I66, p. 44, §106, p. 90

DENNIS, DIVYA
Division of Cancer Epidemiology & Biostatistics, Regional Cancer Centre, Thiruvananthapuram, Kerala, India
`divyadennis8@gmail.com`
Speaker: Poster Session, p. 33, §107, p. 91

DEO, Vishal
Ramjas College, University of Delhi
`vishaaldeo@gmail.com`
Speaker: 29.A2.C19, p. 49, §108, p. 91

DEY, Monitirtha
Indian Statistical Institute, Kolkata
`monitirthadey3@gmail.com`
Speaker: 27.A2.C6, p. 22, §109, p. 92

DEY, Tanujit
Harvard Medical School
`tdey@bwh.harvard.edu`
Speaker: 27.M1.I14, p. 11, §110, p. 92

DHAMALE, Prashant
SVKM'S NMIMS Deemed to be University, Mumbai/ Savitribai Phule Pune University, Pune
`prashantdhamale4@gmail.com`
Chair: 29.M1.C16, p. 41,
Speaker: 29.M1.C16, p. 41, §111, p. 92

DHUM, Shantaram
Savitribai Phule Pune University, Pune
`shantaramdhum13@gmail.com`
Speaker: 29.A2.C18, p. 49, §112, p. 93

DIYALI, Bishal
Department of Statistics, Central University of Haryana
`BISHALDIYALI09@GMAIL.COM`
Speaker: 27.A2.C7, p. 22, §113, p. 93

DUBEY, Paromita
University of Southern California
`paromita@marshall.usc.edu`
Speaker: 28.M2.I45, p. 29, §114, p. 93

DUTTA, Diptavo
National Cancer Institute
`dutta.diptavo@gmail.com`
Speaker: 29.A2.I75, p. 48, §115, p. 94

DUTTA, Subhajit
IIT Kanpur, India
`duttas@iitk.ac.in`
Organizer: 28.M2.I47, p. 30

DUTTA, Subhankar
National Institute of Technology, Rourkela
`subhankar.dta@gmail.com`
Speaker: Poster Session, p. 33, §116, p. 94

E P., Sreedevi
Maharaja's College , Ernakulam
`sreedeviep@gmail.com`
Speaker: 26.A1.I1, p. 3, §117, p. 94

ESLAVA, Laura
IIMAS-UNAM
`laura@sigma.iimas.unam.mx`
Chair: 26.A2.I7, p. 5,
Speaker: 26.A2.I7, p. 6, §118, p. 95

FONSECA, Rodney
Weizmann Institute of Science
`rodney.fonseca@weizmann.ac.il`
Speaker: 29.A1.I66, p. 45, §119, p. 95

FRANZOLINI, Beatrice
franzolini@pm.me
`Agency for Science, Technology and Research`
Speaker: 28.M1.I40, p. 26, §120, p. 95

GADEKAR, Shital
Department of Statistics Savitribai Phule Pune University Pune
`skgadekar871@gmail.com`
Speaker: 28.A2.C12, p. 36, §121, p. 95

GAFFI, FRANCESCO
Bocconi University
francesco.gaffi@phd.unibocconi.it

Speaker: 27.A2.I35, p. 21, §122, p. 96

GANGULY, Ayon
Indian Institute of Technology Guwahati
aganguly@iitg.ac.in
Speaker: 29.M1.I58, p. 39, §123, p. 96

GARAI, Sandip
Phd Scholar (Ag. Statistics)
sandipnicksandy@gmail.com
Speaker: Poster Session, p. 34, §124, p. 96

GARG, Naresh
IIT Kanpur
nareshng@iitk.ac.in
Speaker: 27.M1.C4, p. 14, §125, p. 97

GHANDWANI, Disha
Stanford University
disha123@stanford.edu
Speaker: 26.A2.C1, p. 8, §126, p. 97

GHARDE, Yogita
ICAR-Directorate of Weed Research
yogitagharde@gmail.com
Speaker: 30.M1.I76, p. 51, §127, p. 97

GHODAKE, Aishwarya
Department of Statistics, Savitribai Phule Pune University, Pune
ghodakeaish@gmail.com
Speaker: 29.A2.C19, p. 49, §128, p. 98

GHOSAL, Nairita
Merck & Co., Inc., Rahway, NJ, USA
nairita.ghosal89@gmail.com
Speaker: 27.M1.I17, p. 12, §129, p. 98

GHOSH, Abhik
Indian Statistical Institute
abhik.ghosh@isical.ac.in
Speaker: 29.M2.I64, p. 43, §130, p. 99

GHOSH, ARKA
apghosh@iastate.edu
Iowa State University
Speaker: 29.M1.I59, p. 39, §131, p. 99

GHOSH, Arka
Iowa State University
apghosh@iastate.edu
Organizer: 29.M1.I59, p. 39

GHOSH, Aurobindo
Singapore Management University
aurogh@gmail.com
Speaker: 26.A2.I10, p. 7, §132, p. 99

GHOSH, Dhrubajyoti
Duke University
d.ghosh@wustl.edu
Speaker: 29.A1.I66, p. 45, §133, p. 99

GHOSH, Joyee
The University of Iowa
joyee123in@gmail.com
Chair: Student Paper Competition 2, p. 26,
Speaker: 26.A1.I6, p. 5, §134, p. 100

GHOSH, Malay
University of Florida

Speaker: 27.M2.I21, p. 16, §135, p. 100

GHOSH, Palash
Department of Mathematics, Indian Institute of Technology Guwahati, India
palash.ghosh@iitg.ac.in
Speaker: 28.M1.I39, p. 25, §136, p. 100

GHOSH, Rochishnu
GSK India Global Services Pvt. Ltd.
rochishnu.k.ghosh@gsk.com
Speaker: 27.M1.C5, p. 14, §137, p. 100

GHOSH, Samiran
University of Texas School of Public Health
Samiran.Ghosh@uth.tmc.edu
Chair: 26.A2.I9, p. 6,
Speaker: 26.A2.I9, p. 6, §138, p. 101

GHOSH, Shanta
The University of Illinois at Chicago
sghosh32@uic.edu
Speaker: Poster Session, p. 34, §139, p. 101

GHOSH, Shyamal
Assistant Professor, Dept. of Mathematics, IIIT Guwahati
shyamalmath2012@gmail.com
Speaker: 27.M2.I24, p. 17, §140, p. 101

GHOSHAL, Subhashis
North Carolina State University
sghosal@ncsu.edu
Panelist: Panel Discussion 2, p. 17,
Organizer: 28.M1.I38, p. 24,
Chair: Special Invited Session 2, p. 28,
Organizer: 29.A2.I74, p. 47

GOPALAKRISHNAN, Asha
asha@cusat.ac.in
Cochin University of Science & Technology
Speaker: Special Invited Session 4, p. 52, §141, p. 102

GOPALAKRISHNAN, Harithasree
Great Learning
shivharitha@gmail.com
Chair: 29.A2.C18, p. 49,
Speaker: 29.A2.C18, p. 49, §142, p. 102

GOSWAMI, Subhajit
Tata Institute of Fundamental Research
goswami@math.tifr.res.in
Speaker: 28.A1.I50, p. 32, §143, p. 102

GOYAL, Dheeraj
Department of Mathematics, Indian Institute of Technology Jodhpur, 342037, Karwar, Rajasthan, India
goyal.13@iitj.ac.in
Speaker: 26.A2.C3, p. 9, §144, p. 102

GUHA, Apratim
XLRI, Xavier School of Management
apratim@xlri.ac.in
Speaker: 29.M1.I59, p. 39, §145, p. 103

GUHA, Nilabja
UMASS Lowell
`nilabja_guha@uml.edu`
Speaker: 28.A1.I52, p. 32, §146, p. 103

GUHA, SAURAV
ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India
`saurav.guha1990@gmail.com`
Speaker: 29.M1.C14, p. 40, §147, p. 103

GUHA, Sharmistha
Texas A&M University
`sharmistha@tamu.edu`
Chair and organizer: 27.M1.I17, p. 12,
Speaker: 27.A1.I27, p. 18, §148, p. 104

GUHA, Subharup
University of Florida
`s.guha@ufl.edu`
Chair: 27.M1.I13, p. 11,
Speaker: 27.M1.I13, p. 11, §149, p. 104

GUHANIYOGI, Rajarshi
Texas A & M University
`rajguhaniyogi@tamu.edu`
Chair and organizer: 26.A1.I6, p. 5,
Speaker: 28.M2.I44, p. 29, §150, p. 104

GUINDANI, Michele
University of California, Los Angeles
`mguindani@ucla.edu`
Chair: 28.M2.I44, p. 29,
Speaker: 27.A1.I27, p. 18, §151, p. 105

GULANIKAR, Neela
Department of Statistics, Savitribai Phule Pune University
`ngulanikar@gmail.com`
Speaker: 28.M1.C10, p. 28, §152, p. 105

GUNTUBOYINA, Adityanand
University of California Berkeley
`aditya@stat.berkeley.edu`
Chair: Student Paper Competition 1, p. 13,
Speaker: 29.M1.I54, p. 38, §153, p. 106

GUPTA, Jayanti
NishKash Consulting
`jayantigupta@live.com`
Panelist: Panel Discussion 3, p. 31,
Speaker: 29.A1.I68, p. 45, §154, p. 106

GUPTA, Pritam
Pfizer Healthcare India Pvt Ltd
`pritam.gupta@pfizer.com`
Speaker: Special Invited Session 1, p. 15, §155, p. 106

H. G., Ranjani
Principal Data Scientist, GAIA, Ericsson, Bangalore
`ranjani.h.g@ericsson.com`
Speaker: 29.A1.I71, p. 46, §160, p. 107

HABIBA, Ummey
Integral University, Lucknow
`say2habiba@gmail.com`
Chair: 26.A2.C2, p. 8,
Speaker: 26.A2.C2, p. 9, §156, p. 106

HARUN, Mohd
Scientist, ICAR-Indian Agricultural Statistics Research Instiyute, New Delhi-12
`Mohd.Harun@icar.gov.in`
Speaker: 28.A2.C11, p. 36, §157, p. 107

HAZRA, Arnab
Indian Institute of Technology Kanpur
`ahazra@iitk.ac.in`
Speaker: 27.A2.I32, p. 20, §158, p. 107

HAZRA, Nil Kamal
IIT Jodhpur
`nilkamal@iitj.ac.in`
Speaker: 27.A1.I29, p. 19, §159, p. 107

HIROSE, Masayo
Institute of Mathematics for Industry, Kyushu University
`masayo@imi.kyushu-u.ac.jp`
Speaker: 26.A2.I8, p. 6, §161, p. 108

IISA,

Organizer: 26.A1.I5, p. 5,
Organizer: 26.A2.I8, p. 6,
Organizer: 28.M2.I42, p. 28

IYER, Srikanth
Indian Institute of Science Bangalore
`skiyer@iisc.ac.in`
Chair: 27.A1.I28, p. 19,
Speaker: 27.A1.I28, p. 19, §162, p. 108

JADHAV, Sneha
Wake Forest University
`jadhavs@wfu.edu`
Speaker: 29.M2.I61, p. 42, §163, p. 108

JAIN, Prateek
Google Research India
`prajain@google.com`
Speaker: Special Invited Session 1, p. 15, §164, p. 109

JAIN, Riya
Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University Jalgaon, India
`riyarjain3110@gmail.com`
Speaker: 29.M1.C14, p. 40, §165, p. 109

JAISWAL, Ronit
PhD Scholar, ICAR-INDIAN AGRICULTURAL RESEARCH INSTITUTE, New Delhi
`ronitjaiswal2912@gmail.com`
Chair: 27.A2.C6, p. 21,
Speaker: 27.A2.C6, p. 22, §166, p. 109

JANA, Indrajit
IIT Bhubaneswar
`ijana@iitbbs.ac.in`
Chair: 27.M1.I15, p. 11,
Speaker: 27.M1.I15, p. 12, §167, p. 109

JANA, Kaushik
Ahmedabad University
`kaushik.jana@ahduni.edu.in`
Chair and organizer: 27.A2.I32, p. 20,
Speaker: 27.A2.I32, p. 20, §168, p. 110

JANA, Sayantee
IIT Hyderabad
`sayantee.jana@math.iith.ac.in`
Panelist: Panel Discussion 3, p. 31,
Speaker: 29.M1.C16, p. 41, §169, p. 110

JHA, Jayant
Indian Statistical Institute
jayantjha@gmail.com
Speaker: 29.A1.I70, p. 46, §170, p. 111

JOHNSON, Wesley
UC Irvine
WJOHNSON@UCI.EDU
Speaker: 27.M2.I22, p. 16, §171, p. 111

JONES, Galin
University of Minnesota
galin@umn.edu
Speaker: 28.M1.I41, p. 26, §172, p. 111

JOSEPH, Mathew
Indian Statistical Institute
m.joseph@isibang.ac.in
Chair: 30.M1.I78, p. 52,
Speaker: 30.M1.I78, p. 52, §173, p. 112

JOSHI, Neeraj
joshin@iitk.ac.in
Indian Institute of Technology Kanpur
Speaker: 29.A2.I72, p. 47, §174, p. 112

K D, Mariyamma
Christ College (Autonomous) Irinjalakuda, Calicut University
srjosephinedavid@gmail.com
Speaker: 27.A2.I31, p. 20, §175, p. 112

K, Shiva Kumar
Pondicherry University
shivakumar.sk137@pondiuni.ac.in
Speaker: 28.M1.C10, p. 27, §176, p. 112

KAMALJA, Kirtee
Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, Maharashtra
kirteekamalja@gmail.com
Speaker: 29.M1.C14, p. 40, §177, p. 113

KAMAT, Gauri
Brown University
kgauri1231@gmail.com
Speaker: 27.A2.C7, p. 23, §178, p. 113

KARANDIKAR, Rajeeva

Panelist: Panel Discussion 1, p. 3

KARIM, Md Aktar Ul
Institute of Chemical Technology, Mumbai
mdaktarulkarim829@gmail.com
Speaker: Poster Session, p. 34, §179, p. 113

KARMAKAR, Bikram
University of Florida
bkarmakar@ufl.edu
Chair: 26.A1.I3, p. 4,
Speaker: 26.A1.I3, p. 4, §180, p. 114

KARMAKAR, Sayantani
ICAR-Indian Agricultural Research Institute
sayantanikarmakar20@gmail.com
Speaker: 26.A2.C2, p. 8, §181, p. 114

KARMAKAR, Sayar
University of Florida
sayarkarmakar@ufl.edu
Speaker: 29.M2.I64, p. 43, §182, p. 114

KASHIKAR, Akanksha
Department of Statistics, Savitribai Phule Pune University
akanksha.kashikar@gmail.com
Speaker: 27.A2.I31, p. 20, §183, p. 115

KATTI, Rohit
S & P Global
rohit.katti@ihsmarkit.com
Speaker: 27.A2.C8, p. 23, §184, p. 115

KATTUMANNIL, Sudheesh
Indian Statistical Institute
skkattu@isichennai.res.in
Speaker: 26.A1.I1, p. 3, §185, p. 115

KATTUMANNIL, Sudheesh
Indian Statistical Institute
skkattu@isichennai.res.in
Organizer: 26.A1.I1, p. 3

KAUR, Amarjot
Merck

Moderator: Panel Discussion 3, p. 31,
Organizer: Panel Discussion 3, p. 31

KERALAPURA NAGARAJU, Ksheera Sagar
Purdue University
kkeralap@purdue.edu
Speaker: Poster Session, p. 34, §186, p. 115

KHAMARU, Koulik
Rutgers University
koulik@berkeley.edu
Speaker: 27.M2.I19, p. 15, §187, p. 116

KHATUN, Habiba
National Institute of Technology Rourkela
habibakhatun7860@gmail.com
Speaker: Poster Session, p. 34, §188, p. 116

KLEIJN, Bas
b.j.k.kleijn@uva.nl
Korteweg-de Vries Institute for Mathematics
Speaker: Special Invited Session 2, p. 28, §189, p. 116

KOLEY, Arnab
IIM Indore
arnabk@iimidr.ac.in
Speaker: 30.M2.I80, p. 53, §190, p. 117

KOLEY, Tamalika
Birla Institute of Technology Mesra

Speaker: 28.A1.I53, p. 33, §191, p. 117

KORE, Rutuja
Department of Statistics, Savitribai Phule Pune University
rutujav98@gmail.com
Speaker: 29.A2.C19, p. 49, §192, p. 117

KOUL, Hira
Michigan State University
koul@msu.edu
Chair: Bahadur memorial lecture, p. 44

KRISHNA, Hare
Department of Statistics, Chaudhary Charan Singh University, Meerut (INDIA)
hkrishnastats@yahoo.com
Speaker: 28.A2.C13, p. 37, §193, p. 117

KUCHIBHOTLA , Arun Kumar

Carnegie Mellon University

karun3kumar@gmail.com
Speaker: 30.M2.I79, p. 53, §194, p. 118

KULKARNI, Leena

Department of Statistics, School of Mathematics, Applied Statistics & Analytics, NMIMS Deemed-to-be University, Mumbai, India

leena@dictation.co.in
Speaker: 27.M2.I24, p. 17, §195, p. 118

KUMAR, Pushkal

National Institute of Technolgy, Rourkela

pushkalkumar0812@gmail.com
Speaker: Poster Session, p. 34, §196, p. 118

KUMAR, Rajat

S&P Global

rajat.kumar@ihsmarkit.com
Speaker: 27.M1.C4, p. 14, §197, p. 118

KUMAR, Sanjay

PhD Student

sanjaymk@iitk.ac.in
Speaker: Poster Session, p. 34, §198, p. 119

KUMAR, Shubham

shubh.stat@bhu.ac.in

Banaras Hindu University, Varanasi
Speaker: 28.A2.C13, p. 37, §199, p. 120

KUMARI, Anita

Department of Statistics, Central University of Haryana, Mahendragarh India

meera97bhalothia@gmail.com
Speaker: 27.A2.C6, p. 22, §200, p. 120

KUMARI, Rani

National Institute of Technology Patna

ranik.phd19.ma@nitp.ac.in
Speaker: 28.M1.C10, p. 27, §201, p. 120

KUNDU, Amarjit

Professor, Department of Mathematics, Raiganj University, Raiganj

bapai_k@yahoo.com
Speaker: 27.A1.I29, p. 19, §202, p. 120

KUNDU, Debasis

Indian Institute of Technology Kanpur

kundu@iitk.ac.in
Chair: 30.M2.I80, p. 53,
Speaker: 27.M2.I20, p. 15, §203, p. 121

KUNDU, Subrata

George Washington University

kundu@gwu.edu
Chair: Special Invited Session 4, p. 52

L.N., Vinaykumar

ICAR - Indian Agricultural Research Institute

vinaymandya123@gmail.com
Speaker: 26.A2.C2, p. 9, §212, p. 123

L, INIYA

Madurai Kamaraj University

iniya254@gmail.com
Speaker: 29.A2.C17, p. 48, §204, p. 121

LAHA, Arnab

Indian Institute of Management Ahmedabad

arnab@iima.ac.in
Chair: 29.A1.I70, p. 46,
Speaker: 28.M2.I42, p. 28, §205, p. 121

LAHA, Nilanjana

Texas A&M University

nlaha@tamu.edu
Chair and organizer: 27.M2.I19, p. 15,
Speaker: 27.M2.I19, p. 15, §206, p. 122

LAHIRI, Ananya

IIT Tirupati

ananya@iittp.ac.in
Speaker: 28.A1.I53, p. 33, §207, p. 122

LAHIRI, Parthasarathi

University of Maryland, College Park

plahiri@umd.edu
Speaker: 27.M2.I21, p. 16, §208, p. 122

LAHIRI, Soumendra

Washington University in St. Louis

s.lahiri@wustl.edu
Speaker: Plenary Lecture 3, p. 44, §209, p. 123

LINERO, Antonio

University of Texas at Austin

antonio.linero@austin.utexas.edu

Speaker: 28.A1.I48, p. 31, §210, p. 123

LIPTON, Zachary

Carnegie Mellon University

zlipton@cmu.edu
Speaker: 30.M2.I79, p. 53, §211, p. 123

LU, Qiongshi

University of Wisconsin-Madison

qlu@biostat.wisc.edu
Speaker: 27.A1.I30, p. 19, §213, p. 124

M, Anas

Research Scholar, D/o Statistics & Operations Research, Aligarh Muslim University, Aligarh

anasmelethil@gmail.com
Speaker: 29.M1.C15, p. 41, §214, p. 124

MAITRA, Samopriyo

Chair: 29.A1.I68, p. 45

MAJUMDAR, Arunabha

Assistant Professor, Department of Mathematics, Indian Institute of Technology Hyderabad

arun.majum@math.iith.ac.in
Speaker: 28.M2.I42, p. 28, §215, p. 125

MAJUMDAR, Sourav

Indian Institute of Management Ahmedabad

phd18souravm@iima.ac.in
Speaker: Student Paper Competition 1, p. 13, §216, p. 125

MAJUMDER, Priyanka

Postdoctoral Fellow

priyankamjmdr@gmail.com
Speaker: 26.A2.C3, p. 9, §217, p. 125

MAJUMDER, Tuhin

Speaker: 28.A1.I51, p. 32, §218, p. 126

MALLICK, Himel
Merck Research Labs
`himel.stat.iitk@gmail.com`
Organizer: 29.M1.I55, p. 38

MANATUNGA, Amita
Emory University
`amanatu@emory.edu`
Panelist: Panel Discussion 3, p. 31,
Speaker: 28.M1.I38, p. 25, §219, p. 126

MANDAL, Abhyuday
University of Georgia
`amandal@stat.uga.edu`
Organizer: 27.M2.I20, p. 15,
Organizer: 30.M2.I80, p. 53,
Speaker: 27.M2.I20, p. 15, §220, p. 126

MANDAL, Siddhartha
Center for Chronic Disease Control, New Delhi, India
`sid.mandal.85@gmail.com`
Speaker: 27.M1.I18, p. 13, §221, p. 127

MANDREKAR, Jay
MAYO CLINIC
`MANDREKAR.JAY@MAYO.EDU`
Chair: 29.M1.I57, p. 39,
Speaker: 29.M1.I57, p. 39, §222, p. 127

MANDREKAR, Sumithra
Department of Quantitative Health Sciences, Mayo Clinic
`mandrekar.sumithra@mayo.edu`
Speaker: 29.M1.I57, p. 39, §223, p. 127

MANJUNATH, BG
University of Hyderabad
`bgmanjunath@gmail.com`
Chair: 29.A2.C17, p. 48,
Speaker: 29.A2.C17, p. 48, §224, p. 128

MANJUNATHA, B
Indian Agricultural Statistics Research Institute-New Delhi
`bmanjunathabpt@gmail.com`
Speaker: 28.M1.C9, p. 27, §225, p. 128

MASIHUDDIN,
INDIAN INSTITUTE OF TECHNOLOGY KANPUR
`masih.iitk@gmail.com`
Speaker: 26.A2.C1, p. 8, §226, p. 128

MATHEWS, Irene
S&P Global
`irene.mathews@spglobal.com`
Speaker: 28.M1.C9, p. 27, §227, p. 129

MAULIK, Krishanu
Indian Statistical Institute
`krishanu@isical.ac.in`
Speaker: In Memorium:, p. 21,
Speaker: 27.M2.I23, p. 16, §228, p. 129

MAZUMDAR, Madhu
Icahn School of Medicine at Mount Sinai
`mam2073@gmail.com`
Speaker: 29.M1.I55, p. 38, §229, p. 129

MCVITTIE, James
University of Regina
`james.mcvittie@uregina.ca`
Speaker: 27.A2.I33, p. 20, §230, p. 130

MESTRY, Dipali V.
Institute of Chemical Technology, Mumbai
`dipalimestry96@gmail.com`
Speaker: Poster Session, p. 34, §231, p. 130

MIRA, Antonietta
Universit della Svizzera italiana and University of Insubria
`antonietta.mira@usi.ch`
Speaker: 28.M2.I44, p. 29, §232, p. 130

MISHRA, Swapnil
University of Copenhagen
`swapnmilmishra.anu@gmail.com`
Panelist: Panel Discussion 4, p. 44,
Speaker: 29.M1.I56, p. 38, §233, p. 131

MITRA, Adway
Indian Institute of Technology Kharagpur
`adway.cse@gmail.com`
Speaker: 29.M2.I65, p. 44, §234, p. 131

MITTAL, Anjali
Indian Institute of Technology Kanpur
`anjalimittal107@gmail.com`
Speaker: Poster Session, p. 34, §235, p. 132

MODOK, BIDHAN
RAIGANJ UNIVERSITY
`bidhanmodok95@gmail.com`
Speaker: 26.A2.C1, p. 8, §236, p. 132

MOHAMMED, Shariq
Boston University School of Public Health
`shariqm@bu.edu`
Organizer: 27.A2.I34, p. 21

MOHAMMED, Shariq
Boston University
`shariqm@bu.edu`
Speaker: 27.A2.I34, p. 21, §237, p. 132

MOLSTAD, Aaron
University of Florida
`amolstad@ufl.edu`
Organizer: 28.M1.I36, p. 24

MONDAL, Anirban
Case Western Reserve University
`axm912@case.edu`
Speaker: 29.M1.I54, p. 38, §238, p. 133

MONDAL, Arpita
Indian Institute of Technology (IIT) Bombay
`arpita567@gmail.com`
Chair and organizer: 29.M2.I65, p. 44

MONDAL, Debashis
Washington University in St Louis
`mondal@wustl.edu`
Organizer: 29.A1.I66, p. 44

MONDAL, Sagnik
King Abdullah University of Science and Technology
`sagnik.mondal@kaust.edu.sa`
Chair: 28.A2.C11, p. 35,
Speaker: 28.A2.C11, p. 36, §239, p. 133

MUELLER, Peter
UT Austin
`pmueller@math.utexas.edu`
Chair and organizer: 27.A1.I27, p. 18,
Organizer: 27.A2.I35, p. 21,
Organizer: 28.M2.I44, p. 29,
Chair: Plenary Lecture 3, p. 30,
Speaker: 28.A1.I48, p. 31, §240, p. 134

MUKHERJEE, Abhishek

Panelist: Panel Discussion 1, p. 3

MUKHERJEE, Anirbit
The University of Manchester
`anirbit.mukherjee@manchester.ac.uk`

Speaker: 27.A1.I26, p. 18, §241, p. 134

MUKHERJEE, Arghya
Indian Institute of Technology, Kanpur
`arghyam21@iitk.ac.in`
Speaker: Poster Session, p. 34, §242, p. 134

MUKHERJEE, Bhramar
University of Michigan
`bhramar@umich.edu`
Chair and organizer: 29.M1.I56, p. 38,
Organizer: Panel Discussion 4, p. 44,
Speaker: 29.M1.I56, p. 38, §243, p. 135

MUKHERJEE, Gourab
University of Southern California
`gourab@usc.edu`
Organizer: 28.A1.I52, p. 32,
Speaker: 28.A1.I52, p. 32, §244, p. 135

MUKHERJEE, Partha Sarathi
Indian Statistical Institute, Kolkata
`psmukherjee.statistics@gmail.com`

Organizer: 29.A1.I70, p. 46,
Speaker: 29.A1.I70, p. 46, §245, p. 136

MUKHERJEE, Rajarshi
Harvard University
`ram521@mail.harvard.edu`
Speaker: 27.M2.I19, p. 15, §246, p. 136

MUKHERJEE, Sayan
Indian Statistical Institute, Kolkata
`sayanmukherjee012@gmail.com`
Speaker: 27.A2.C6, p. 22, §247, p. 136

MUKHERJEE, Somabha
National University of Singapore
`somabha@nus.edu.sg`
Speaker: 28.M2.I47, p. 30, §248, p. 136

MUKHERJEE, Ujjal Kumar
College of Business, and Carle Illinois College of Medicine, UIUC
`ukm@illinois.edu`
Speaker: 28.A1.I49, p. 31, §249, p. 137

NAG, Pratik
King Abdullah University of Science and Technology
`pratik.nag@kaust.edu.sa`
Speaker: Student Paper Competition 2, p. 26, §250, p. 137

NAGARAJAN, Vaishnavh
Google
`vaishnavh@google.com`
Speaker: 28.M2.I43, p. 29, §251, p. 138

NANDI, Debarghya
Adjunct Professor
`dnandi4@uic.edu`
Speaker: Poster Session, p. 34, §252, p. 138

NANDI, Shinjini
Montana State University
`shinjini.nandi2@gmail.com`
Speaker: 29.A2.C17, p. 48, §253, p. 138

NANDY, Karabi
University of Texas Southwestern Medical Center
`Karabi.Nandy@UTSouthwestern.edu`
Chair: 28.M1.I38, p. 24,
Speaker: 28.M1.I38, p. 25, §254, p. 138

NANDY, Rajesh
UNT Health Science Center
`Rajesh.Nandy@unthsc.edu`
Speaker: 28.M2.I47, p. 30, §255, p. 139

NAVAS T, Muhammed
Research scholar
`navasmku@gmail.com`
Speaker: 29.M1.C15, p. 40, §256, p. 139

NAYAK.G.H., Harish
Ph.D. Scholar at ICAR-IASRI, New Delhi
`harishnayak626@gmail.com`
Speaker: 28.A2.C13, p. 37, §257, p. 139

NETRAPALLI, Praneeth
Google Research India
`praneethn@gmail.com`
Speaker: 28.M2.I43, p. 28, §258, p. 140

NI, Yang

Chair: 27.M2.I22, p. 16

NIPHADKAR, Shubham
niphadkarshubham@gmail.com
`Indian Institute of Technology Bombay`
Speaker: 27.M1.C5, p. 14, §259, p. 140

OJHA, Abhishek
University of Illinois, Urbana-Champaign
`ojha4@illinois.edu`
Speaker: Student Paper Competition 1, p. 13, §260, p. 140

PAHLAJANI, Chetan
Indian Institute of Technology Gandhinagar
`cdpahlajani@iitgn.ac.in`
Speaker: 29.A1.I69, p. 46, §261, p. 141

PAL, RUPAM
GSK Pharmaceutical
`rupam.r.pal@gsk.com`
Speaker: 26.A2.I10, p. 7, §262, p. 141

PAL, Sandip
S&P Global
`ef20sandipk@iimidr.ac.in`
Speaker: 26.A1.I4, p. 4, §263, p. 141

PAL, Subrata
Department of Statistics, Iowa State University
subrata@iastate.edu
Speaker: Student Paper Competition 2, p. 26, §264, p. 142

PANCHGAUR, Richa
Research Scholar
panchgaurricha@gmail.com
Chair: 26.A2.C1, p. 8,
Speaker: 26.A2.C1, p. 8, §265, p. 142

PANDA, Mahesh Kumar
Associate Professor of Statistics, Department of Statistics, Ravenshaw University, Cuttack, Odisha
mahesh2123ster@gmail.com
Speaker: 27.M1.C4, p. 13, §266, p. 142

PANIGRAHI, Snigdha
University of Michigan
psnigdha@umich.edu
Speaker: 27.A2.I34, p. 21, §267, p. 143

PAREEK, Bhuvanesh
IIM Bangalore
bhuvanesh.pareek@iimb.ac.in
Chair and organizer: 29.A2.I73, p. 47,
Speaker: 29.A2.I73, p. 47, §268, p. 143

PAREEK, Savita
IIT Bombay
savy.pareek@gmail.com
Speaker: 27.A2.C8, p. 23, §269, p. 143

PARIKH, Harsh
Duke University
harsh.parikh@duke.edu
Speaker: 29.A2.I73, p. 47, §270, p. 144

PATEL, Ayush
University of Oxford

Speaker: Short Course 2, p. 51,
Speaker: Short Course 4, p. 52,
Speaker: Short Course 6, p. 54

PATI, Debdeep
Texas A&M University
debdeep@stat.tamu.edu
Speaker: 28.A1.I48, p. 31, §271, p. 144

PAUL, Debashis
University of California, Davis
debpaul@ucdavis.edu
Chair and organizer: 28.A1.I49, p. 31,
Speaker: 27.M1.I15, p. 11, §272, p. 144

PAUL, Ranjit Kumar
ICAR-Indian Agricultural Statistics Research Institute
ranjit.paul@icar.gov.in
Chair: 30.M1.I76, p. 51,
Speaker: 30.M1.I76, p. 51, §273, p. 145

PAUL, Subhadeep
The Ohio State University
paul.963@osu.edu
Speaker: 28.A1.I51, p. 32, §274, p. 145

PAUL, Sujoy
Google Research
sujoyp@google.com
Speaker: 28.M2.I44, p. 29, §275, p. 145

PODDER, Moumanti
Indian Institute of Science Education and Research (IISER) Pune
moumantip3@gmail.com
Chair and organizer: 26.A1.I2, p. 4,
Organizer: 26.A2.I7, p. 5,
Organizer: 27.M2.I23, p. 16,
Organizer: 27.A1.I28, p. 19,
Organizer: 28.A1.I50, p. 32,
Organizer: 29.A1.I69, p. 45,
Organizer: 30.M1.I78, p. 52,
Organizer: 30.M2.I81, p. 53,
Speaker: 26.A1.I2, p. 4, §277, p. 146

PORWAL, Anupreet
University of Washington
porwaa@uw.edu
Chair: 29.M1.I55, p. 38,
Speaker: Poster Session, p. 35, §278, p. 146

PRUENSTER, Igor
Bocconi University
igor.pruenster@unibocconi.it
Chair: 28.M1.I40, p. 25,
Speaker: Plenary Lecture 3, p. 30, §279, p. 147

PUTATUNDA, Sayan
sayanp@iima.ac.in
Walmart
Speaker: 26.A1.I4, p. 4, §280, p. 147

QIU, Peihua
University of Florida
sghosal@ncsu.edu
Organizer: 27.M1.I13, p. 11

RAGHUNATHAN, Aditi
Carnegie Mellon University
raditi@cmu.edu
Chair and organizer: 28.M2.I43, p. 28,
Speaker: 28.M2.I43, p. 29, §281, p. 147

RAILKAR, Radha
Merck

Panelist: Panel Discussion 3, p. 31

RAMADAS, Meenu
Assistant Professor, School of Infrastructure, IIT Bhubaneswar
meenu@iitbbs.ac.in
Speaker: 29.M2.I65, p. 44, §282, p. 148

RAMAMOORTHI, R V
ramamoor@msu.edu
Prof (Emeritus) Michigan State University; Adjunct Prof., Chennai Math Inst.
Speaker: Special Invited Session 2, p. 28, §283, p. 148

RAMANAN, Kavita
Brown University
kavita_ramanan@brown.edu
Speaker: Special Invited Session 3, p. 42, §284, p. 148

RAMANATHAN, Thekke Variyam
Savitribai Phule Pune University
ram@unipune.ac.in
Chair: 27.A2.I31, p. 20,
Chair: 28.M2.I45, p. 29,
Speaker: 28.M2.I45, p. 29, §285, p. 149

RANJAN, Pritam
pritamr@iimidr.ac.in
IIM Indore
Chair: 27.M2.I20, p. 15,
Speaker: 26.A1.I4, p. 4, §286, p. 149

RANJAN, Rakesh
Benares Hindu University
rakesh.ranjan@bhu.ac.in
Speaker: 27.M1.I16, p. 12, §287, p. 149

RAO, Arvind
Department of Computational Medicine and Bioinformatics, University of Michigan-Ann Arbor

Chair: 27.A2.I34, p. 21,
Organizer: 27.A2.I34, p. 21

RAO, Marepalli
University of Cincinnati
marepalli.rao@uc.edu
Chair: 28.M2.I42, p. 28,
Speaker: 28.M2.I42, p. 28, §288, p. 149

RAY, Debashree
Johns Hopkins University
dray@jhu.edu
Chair and organizer: 27.A1.I30, p. 19,
Speaker: In Memorium:, p. 21,
Speaker: 27.A1.I30, p. 19, §289, p. 150

REBAUDO, Giovanni
The University of Texas at Austin
rebaudo.giovanni@gmail.com
Chair: 28.A1.I48, p. 31,
Speaker: 28.M1.I40, p. 25, §290, p. 150

RIGON, Tommaso
University of Milano-Bicocca
tommaso.rigon@unimib.it
Speaker: 27.M2.I22, p. 16, §291, p. 150

ROBERT, Christian
Université Paris Dauphine PSL and University of Warwick
xian@ceremade.dauphine.fr
Speaker: 27.M1.I16, p. 12, §292, p. 151

ROY, Anik
Indian Statistical Institute, Kolkata
anikroy7713@gmail.com
Speaker: 27.A2.C8, p. 23, §293, p. 151

ROY, Arkaprava
University of Florida
ark007@ufl.edu
Speaker: 27.M1.I13, p. 11, §294, p. 151

ROY, Asmita
Texas A&M University
asmita112358@tamu.edu
Speaker: Student Paper Competition 1, p. 13, §295, p. 152

ROY, Parthanil
Indian Statistical Institute
parthanil.roy@gmail.com
Organizer: 26.A2.I7, p. 5,
Organizer: 27.M2.I23, p. 16,
Organizer: 27.A1.I28, p. 19,
Organizer: 28.A1.I50, p. 32,
Chair: 29.A1.I69, p. 45,
Organizer: 29.A1.I69, p. 45,
Organizer: 30.M1.I78, p. 52,
Organizer: 30.M2.I81, p. 53,
Speaker: 29.A1.I69, p. 46, §296, p. 152

ROY, Rahul
Indian Statistical Institute, Kolkata
rahul_r@isical.ac.in
Speaker: 28.A2.C12, p. 36, §297, p. 152

ROY, Rishideep
IIM Bangalore
rishideeproy@gmail.com
Speaker: 29.M1.I59, p. 39, §298, p. 153

ROY, Samrat
University of Pennsylvania
roysa@wharton.upenn.edu
Chair: 28.M1.I36, p. 24,
Speaker: 28.M1.I36, p. 24, §299, p. 153

ROY, Sandipan

Chair: 29.M2.I63, p. 43

ROY, Soumya
Indian Institute of Management Kozhikode
soumya@iimk.ac.in
Speaker: 30.M2.I82, p. 54, §300, p. 153

ROY, Souvik
ISI Kolkata
souvik.2004@gmail.com
Speaker: 26.A1.I2, p. 4, §301, p. 154

RUDRA, Pratyaydipta
Oklahoma State University
prudra@okstate.edu
Speaker: 27.M1.I18, p. 13, §302, p. 154

RUSH, Cynthia
Columbia University
cynthia.rush@columbia.edu
Speaker: 29.A1.I67, p. 45, §303, p. 154

S, Rukmini
Independent Data Journalist
rukminiwrites@gmail.com
Panelist: Panel Discussion 4, p. 44

SABNIS, Sanjeev
IIT Bombay
svs@math.iitb.ac.in
Chair and organizer: 27.M2.I24, p. 17,
Speaker: 27.M2.I24, p. 17, §304, p. 155

SADHU, Ritwik
Cornell University
rs2526@cornell.edu
Speaker: 29.M1.C16, p. 41, §305, p. 155

SADHUKHAN, Soumyarup
Indian Institute of Technologuy Kanpur
soumyarups@iitk.ac.in
Speaker: 28.A1.I50, p. 32, §306, p. 155

SAHA, Diptarka
University of Illinois, Urbana - Champaign
saha12@illinois.edu
Speaker: Student Paper Competition 2, p. 26, §307, p. 156

SAHA, Kumarjit
Ashoka University
kumarjit.saha@ashoka.edu.in
Speaker: 27.A1.I28, p. 19, §308, p. 156

SAHA, Snehanshu
snehanshus@goa.bits-pilani.ac.in
BITS PIlani Goa Campus
Speaker: 27.M2.I20, p. 15, §309, p. 156

SAHASRABUDHE, Neeraja
IISER Mohali
neeraja@iisermohali.ac.in
Speaker: 27.M2.I23, p. 16, §310, p. 156

SAHOO, Braja Gopal
Novartis Healthcare Pvt. Ltd.
braja_gopal.sahoo@novartis.com
Chair: 27.A1.I25, p. 18,
Speaker: 27.A1.I25, p. 18, §311, p. 156

SAHOO, Indranil
Virginia Commonwealth University
sahooi@vcu.edu
Speaker: 26.A1.I6, p. 5, §312, p. 157

SAHOO, Shyamsundar
Associate Professor, Department of Statistics, Haldia Government College
sssahoo.stats@gmail.com
Chair: 29.M1.I58, p. 39,
Speaker: 29.M1.I58, p. 39, §313, p. 157

SAHOO, Tanmay
Indian Institute of Technology Jodhpur
sahoo.4@iitj.ac.in
Speaker: 26.A2.C3, p. 9, §314, p. 157

SAHU, Soumya
University of Illinois Chicago
ssahu6@uic.edu
Speaker: 26.A2.C1, p. 8, §315, p. 158

SAMANTA, Srijata
M D Anderson Cancer Center
srijata.s06@gmail.com
Chair: 28.M1.I37, p. 24,
Speaker: 28.M1.I37, p. 24, §316, p. 158

SANKARAN, P. G.
Cochin University of Science and Technology
sankaran.pg@gmail.com
Chair: 26.A1.I1, p. 3,
Speaker: 26.A1.I1, p. 3, §276, p. 146

SARKAR, Abhinanda
Great Learning
sarkar.abhinanda@gmail.com
Chair: Special Invited Session 1, p. 15,
Speaker: Short Course 3, p. 51,
Speaker: Short Course 5, p. 52,
Speaker: Short Course 7, p. 54

SARKAR, Abhra
The University of Texas at Austin
abhra.stat@gmail.com
Organizer: 28.A1.I48, p. 31,
Speaker: 27.A1.I27, p. 18, §317, p. 159

SARKAR, Angshuman
GSK India Global Service Private Ltd
angshuman.x.sarkar@gsk.com
Organizer: 26.A2.I10, p. 7,
Organizer: 29.A1.I68, p. 45

SARKAR, Mojammel Haque
National Institute of Technology Rourkela
mojammelmhs3@gmail.com
Speaker: Poster Session, p. 35, §318, p. 159

SARKAR, Partha
University of Florida
sarkarpartha@ufl.edu
Chair: 27.M1.C4, p. 13,
Speaker: 27.M1.C4, p. 14, §319, p. 159

SARKAR, Shibajee
Ramakrishna Mission Vivekananda Educational and Research Institute
shibajee.prime001@gmail.com
Speaker: Poster Session, p. 35, §320, p. 160

SARKAR, Shoumi
University of Florida
shoumisarkar@ufl.edu
Speaker: Poster Session, p. 35, §321, p. 160

SARKAR, Soham
Indian Statistical Institute, Delhi Centre
sohamsarkar1991@gmail.com
Chair: 28.M2.I47, p. 30,
Speaker: 30.M2.I82, p. 54, §322, p. 160

SATAGOPAN, Jaya
Rutgers School of Public Health
satagopj@sph.rutgers.edu
Speaker: 29.A2.I75, p. 48, §323, p. 161

SATTAR, Abdus
Case Western Reserve University
sattar@case.edu
Chair and organizer: 28.M2.I46, p. 30,
Speaker: 28.M2.I46, p. 30, §324, p. 161

SEN, Aditi
University of Maryland, College Park
asen123@umd.edu
Speaker: Student Paper Competition 2, p. 26, §325, p. 161

SEN, Arusharka
Concordia University
arusharka.sen@concordia.ca
Chair and organizer: 27.A2.I33, p. 20,
Chair and organizer: 30.M1.I77, p. 51,
Speaker: 30.M1.I77, p. 51, §326, p. 161

SEN, Bodhisattva
Columbia University
bodhi@stat.columbia.edu
Speaker: Special Invited Session 3, p. 42, §327, p. 162

SEN, Rituparna
Indian Statistical Institute, Bangalore
ritupar.sen@gmail.com
Chair: Plenary Lecture 1, p. 3,
Chair and organizer: 26.A1.I4, p. 4,
Organizer: 26.A2.I12, p. 7,
Chair and organizer: 28.A1.I53, p. 33,
Speaker: 26.A2.I12, p. 7, §328, p. 162

SEN, Sanchayan
Indian Institute of Science
sanchayan.sen1@gmail.com
Chair: 30.M2.I81, p. 53,
Speaker: 30.M2.I81, p. 53, §329, p. 162

SENGUPTA, Arunabha

Organizer: Panel Discussion 1, p. 3

SENGUPTA, Debasis
Indian Statistical Institute
shairiksengupta@gmail.com
Speaker: 29.M1.I58, p. 39, §330, p. 162

SENGUPTA, Raghu Nandan
Indian Institute if Technology Kanpur
raghus@iitk.ac.in
Speaker: 29.A2.I72, p. 47, §331, p. 163

SENGUPTA, Srijan
North Carolina State University
ssengup2@ncsu.edu
Organizer: 27.M1.I18, p. 13,
Speaker: 28.M1.I36, p. 24, §332, p. 163

SHARMA, Richa
Great Learning Mumbai India (Student)
richas155@gmail.com
Chair: 29.A2.C19, p. 49,
Speaker: 29.A2.C19, p. 50, §333, p. 163

SHARMA, SHAILZA
IISc
shailzakashyap139@gmail.com
Speaker: 29.M2.I65, p. 44, §334, p. 164

SHEKHAR, Atul
TIFR-CAM
atul@tifrbng.res.in
Speaker: 27.A1.I28, p. 19, §335, p. 164

SHEKHAR, Shubhanshu
Department of Statistics and Data Science, Carnegie Mellon University
shubhan2@andrew.cmu.edu
Speaker: 29.M2.I60, p. 42, §336, p. 164

SHELAKE, CHAKRDHAR
Savitribai Phule Pune University
chakradharshelake@gmail.com
Speaker: 28.A2.C12, p. 36, §337, p. 165

SHETE, Sanjay
UT MD Anderson Cancer Center
sshete@mdanderson.org
Chair and organizer: 29.A2.I75, p. 48

SHIRKE, Shrinivas
Department of Statistics, Savitribai Phule Pune University
shrinivasshirke@gmail.com
Speaker: 29.A2.C17, p. 48, §338, p. 165

SHOKOOHI, Farhad
University of Nevada Las Vegas
farhad.shokoohi@unlv.edu
Speaker: 27.A2.I33, p. 21, §339, p. 166

SHUKLA, Abhinek
IIT Kanpur
abhushukla@gmail.com
Speaker: 26.A2.C2, p. 8, §340, p. 166

SHUKLA, Asmita
Banaras Hindu University
asmishukla23@gmail.com
Chair: 26.A2.C3, p. 9,
Speaker: 26.A2.C3, p. 9, §341, p. 166

SIKARIA, Shubhangi
Philips India Limited
shubhangisikariya@gmail.com
Speaker: 28.A1.I53, p. 33, §342, p. 167

SINGH , Sanjiv
Department of Statistics, Institute of Science , Banaras Hindu University, Varanasi(U.P) India
sanjiv20stats@bhu.ac.in
Speaker: 28.A2.C11, p. 36, §346, p. 168

SINGH, Kamlesh
ICAR-IASRI, New Delhi
kn.singh@icar.gov.in
Speaker: 30.M1.I76, p. 51, §343, p. 167

SINGH, Kundan
Department of Mathematics, IIT Patna
kundan.shg@gmail.com
Speaker: Poster Session, p. 35, §344, p. 167

SINGH, SABHIYA
University of Lucknow
singhsabhiya30@gmail.com
Speaker: 29.A2.C18, p. 49, §345, p. 168

SINGH, Satya Prakash
Indian Institute of Technology Kanpur
singhsp@iitk.ac.in
Speaker: 30.M2.I80, p. 53, §347, p. 169

SINGH, SHAMBHAVI
BANARAS HINDU UNIVERSITY
shambhavisingh010@gmail.com
Speaker: 29.M1.C16, p. 41, §348, p. 169

SINGH, VIJAY
Great Learning
vijay.sg@gmail.com
Speaker: 28.A2.C13, p. 37, §349, p. 169

SINGHVI, Somya

Chair: 28.A1.I52, p. 32

SINHA, DEBAJYOTI
FLORIDA STATE UNIVERSITY
DSINHA@FSU.EDU
Speaker: 26.A1.I6, p. 5, §350, p. 170

SINHA, Samiran
Texas A&M University
sinha@stat.tamu.edu
Speaker: 27.M1.I17, p. 12, §351, p. 170

SINHA, Sanjoy
Carleton University, Ottawa, Canada
sinha@math.carleton.ca
Speaker: 28.M2.I46, p. 30, §352, p. 170

SRIPERUMBUDUR, Bharath
Pennsylvania State University
bks18@psu.edu
Speaker: 27.A1.I26, p. 18, §353, p. 171

SRIVASTAVA, Sanvesh
The University of Iowa
sanvesh-srivastava@uiowa.edu
Speaker: 28.M1.I36, p. 24, §354, p. 171

STUPFLER, Gilles
ENSAI & University of Angers
gilles.stupfler@ensai.fr
Speaker: 27.A2.I32, p. 20, §355, p. 171

SUBBARAO, Suhasini
suhasini@tamu.edu
stasc@nus.edu.sg
Organizer: 28.A1.I51, p. 32

SULTANA, Farha
Assistant Professor
farha.sultana@iiitg.ac.in
Speaker: 26.A1.I5, p. 5, §356, p. 172

SUR, Pragya
Harvard University
pragya@fas.harvard.edu
Speaker: 26.A2.I8, p. 6, §357, p. 172

SUTAR, Santosh
Shivaji University, Kolahpur
sutarssantosh@gmail.com
Chair: 27.M1.C5, p. 14,
Speaker: 27.M1.C5, p. 14, §358, p. 172

TEKWE, Carmen
Indiana University at Bloomington
ctekwe@iu.edu
Panelist: Panel Discussion 2, p. 17,
Chair and organizer: 29.M2.I61, p. 42,
Speaker: 29.M2.I61, p. 42, §359, p. 173

TELESCA, Donatello
UCLA
donatello.telesca@gmail.com
Speaker: 27.M2.I22, p. 16, §360, p. 173

THAKUR, Debjoy
Indian Institute of Technology, Tirupati
debjoythakur95@gmail.com
Speaker: Student Paper Competition 2, p. 26, §361, p. 173

TIWARI, Shiwani
Banaras Hindu University
shiwanitiwari15@gmail.com
Chair: 27.A2.C8, p. 23,
Speaker: 27.A2.C8, p. 23, §362, p. 174

TIWARI, Vinay
Savitribai Phule Pune University
vinaystiwari786@gmail.com
Speaker: Poster Session, p. 35, §363, p. 174

TOKDAR, Surya
Duke University
surya.tokdar@duke.edu
Organizer: 29.A1.I67, p. 45

TRIPATHI, Tripti
Banaras Hindu University
tripti.tripathi@bhu.ac.in
Speaker: 26.A2.C3, p. 9, §364, p. 174

TURNER, Heather
University of Warwick

Speaker: Short Course 2, p. 51,
Speaker: Short Course 4, p. 52,
Speaker: Short Course 6, p. 54

UDGATA, ASHIS RANJAN
ICAR-Indian Agricultural Statistics Research Institute
ashisu93@gmail.com
Speaker: Poster Session, p. 35, §365, p. 174

VADLAMANI, Sreekar
TIFR CAM
sreekar@tifrbng.res.in
Chair: 27.M2.I23, p. 16,
Speaker: 27.M2.I23, p. 17, §366, p. 175

VARMA, Manoj
PhD Scholar, ICAR-IASRI, New Delhi
manojvarma57282@gmail.com
Speaker: 29.A2.C18, p. 49, §367, p. 175

VASUDEVAN, Durga
Department of Statistics, Cochin University of Science and Technology
durgavvn10@gmail.com
Speaker: Student Paper Competition 1, p. 13, §368, p. 175

VATS, Dootika
Indian Institute of Technology Kanpur
dootika@iitk.ac.in
Chair: 27.A1.I26, p. 18,
Chair and organizer: 28.M1.I41, p. 26,
Speaker: 28.M1.I41, p. 26, §369, p. 176

VEERSHETTY,
PhD Scholar at ICAR-IASRI
biradarveershetty4@gmail.com
Chair: 28.A2.C12, p. 36,
Speaker: 28.A2.C12, p. 36, §370, p. 176

VERMA, ANKITA
STUDENT
ankiverma95@gmail.com
Speaker: 27.A2.C7, p. 22, §371, p. 176

VINAYAKA,
PhD Scholar (Agricultural Statistics), ICAR-IASRI, New Delhi-110012
vinayaka.b3vs@gmail.com
Speaker: 27.A2.C6, p. 22, §372, p. 177

VISHWANATHA , Lavanya
University of North Texas Health Science Center
lavanya.vishwanatha@unthsc.edu
Speaker: 29.A2.I74, p. 47, §374, p. 178

VISHWANATHA, Jamboor
Univ. of North Texas Health Science Center
jamboor.vishwanatha@unthsc.edu
Speaker: 29.A2.I74, p. 47, §373, p. 177

VOLFOVSKY, Alexander
Duke University
alexander.volfovsky@duke.edu
Speaker: 29.A1.I67, p. 45, §375, p. 178

WAHED, Abdus
University of Pittsburgh
wahed@pitt.edu
Organizer: 26.A2.I9, p. 6,
Speaker: 26.A2.I9, p. 6, §376, p. 178

XIAN, Xiaochen
University of Florida
xxian@ufl.edu
Speaker: 26.A2.I11, p. 7, §377, p. 179

XU, Jason
Duke University
jason.q.xu@duke.edu
Chair: 29.A1.I67, p. 45,
Speaker: 29.A1.I67, p. 45, §378, p. 179

XU, Jing
Duke National University of Singapore
kenny.xu@duke-nus.edu.sg
Chair: 28.A2.C13, p. 37,
Speaker: 28.A2.C13, p. 37, §379, p. 179

YADAV, Priya
Department of Statistics, Central University of Haryana, Mahendergarh, India
rao.priya228@gmail.com
Chair: 27.A2.C7, p. 22,
Speaker: 27.A2.C7, p. 22, §380, p. 180

YADAV, Rishikesh
HEC Montreal
`rishikesh.yadav@hec.ca`
Speaker: Poster Session, p. 35, §381, p. 180

YANG, Junho
Academia Sinica
`junhoyang@stat.sinica.edu.tw`
Chair: 28.A1.I51, p. 32,
Speaker: 28.A1.I51, p. 32, §382, p. 180

YANG, Kai
Medical College of Wisconsin
`kayang@mcw.edu`
Organizer: 26.A2.I11, p. 7,
Speaker: 26.A2.I11, p. 7, §383, p. 181

YOGESHWARAN, D
Indian Statistical Institute
`d.yogesh@isibang.ac.in`
Speaker: 26.A2.I7, p. 6, §384, p. 181

ZHANG, Helen
University of Arizona
`hzhang @ math.arizona.edu`
Speaker: Plenary Lecture 2, p. 17, §385, p. 181

ZOH, Roger
Indiana University
`rszoh@iu.edu`
Speaker: 29.M2.I61, p. 42, §386, p. 181