Poster

No.	Name	Topic
1	Wei-Chun Kang	Cyclic SOAs and Moving Window Criteria for Space-Filling Designs
2	Li-Sheng Zhuang	Nonparametric Mediation Analysis of Non-Markov Illness-Death Model
3	Aman Prakash	Order Restricted and Unrestricted Inference for the Recursive Stage Life Testing Model
4	Anil Maurya	Analysing Load-Sharing System under Progressive Censoring Using Statistical and Machine Learning Approach
5	Vaibhav N. Dhameliya	Compound Optimal Design Strategy for Life-Testing Experiment under Progressive Hybrid Censoring Scheme
6	Chien-Tai Lin	Optimal Progressively Censored Reliability Sampling Plans for the Log-Location-Scale Distribution
7	Erjia Cui	Quantifying Physical Activity Intervention Effects via Functional Regression
8	Yun-Ting Wei	Comparison and Ensemble Strategies of Measurement Error Correction Methods in Multiple Regression with Validation Data
9	Jihoon Kim	A Deep Learning Random-Effect Modelling Approach for clustered Count Data
10	Chin-Sheng Teng	A Statistical Framework Leveraging Single-Cell Data for Cell Type Deconvolution in Bulk Transcriptomics
12	Sung-Hyuk Choi	Prediction of Deterioration of Patients with Dyspnea in Emergency Department
13	En-Yu Lai	Reframing Cross-World Independence for Identifying Path-Specific Effects
14	Shu-Hsien Cho	VACANT: An Adaptive Framework for Rare-Variant Association Testing that Leverages Continuous Functional Annotations
15	Chi Chun Yeh	Causal Mediation Analysis with Survival Data and a Recurrent Mediator
16	Kai-Yuan Wu	Authoritarian Regime Types and Their Pathways to Democratization via Coups
17	Mei Dong	Marginal Causal Effect Estimation with Continuous Instrumental Variables
18	Wei-Yang Yu	Automated Analysis of Experiments using Hierarchical Garrote
19	Jiuqian Shang	Uncertainty Quantification for Noisy Low-tubal-rank Tensor Completion
20	Elvin Tseng	Linearly Constrained Symmetric Rank-One Approximation for Pre-Image Recovery in Nonlinear Matrix Completion
21	Jia-Ying Su	Investigating the Pleiotropic Effects of Splice-Affecting Variants on Cancer and Metabolic Diseases
23	Mengqi Lin	Controlling the False Discovery Proportion in Observational Studies with Hidden Bias
24	Lars Skaaret-Lund	LBBNN: An R Package for Sparse Bayesian Neural Networks
25	Tianyu Liu	Cramér-Type Moderate Deviation and Berry–Esseen Bounds in the p-Spin Curie–Weiss Model
26	Kong Xin	Concentration-QTC Analysis to Support ICH E14 with a Real Case

Back to Sessions List

Cyclic SOAs and Moving Window Criteria for Space-Filling Designs

Wei-Chun, Kang, Cheng-Yu, Sun

National Tsing Hua University, Taiwan

ABSTRACT

Space-filling designs are essential in computer experiments to ensure that design points are uniformly distributed across the input space. Among them, strong orthogonal arrays (SOAs) are widely used for their stratification properties on fixed grids formed by collapsing adjacent factor levels. However, these grids lack flexibility and cannot adapt to local structures. A moving window approach is introduced that evaluates uniformity over sliding regions across the space. By specifying a window size and selecting a subset of positions, a more fine-grained space-filling criterion is defined that generalizes several existing methods. Within this framework, cyclic SOAs are further proposed: SOAs that preserve their stratification properties under cyclic shifts of factor levels. These designs exhibit a structural invariance that is particularly useful in settings involving periodicity or level relabeling. Their optimality properties are established, and construction methods are presented, positioning cyclic SOAs as a flexible and robust addition to the space-filling design toolkit.

Keywords: Computer experiments, Design of experiments, Orthogonal arrays, Strong orthogonal arrays, Centered L_2-discrepancy, Wrap-around L_2-discrepancy, Rectangular Uniformity criterion, Cyclic strong orthogonal arrays

Nonparametric Mediation Analysis of Non-Markov Illness-Death Model

Li-Sheng Zhuang[1], Jih-Chang Yu[2], Yen-Tsung Huang[3]

[1,3]*Institute of Statistical Science, Academia Sinica*

[2]*Department of Statistics, National Taipei University*

ABSTRACT

The illness-death model is widely used to characterize disease progression over time. Previous work focuses either on estimation without causal interpretation or on causal interpretation under a strong Markov assumption where the terminal event depends on the status but not the timing of the intermediate event. To bridge the research gap, we propose a new definition of counterfactual hazard that relaxes the Markov assumption by considering the entire history of the intermediate event. We derive an identification formula that involves an integral with respect to the probability density function of the intermediate event time. Direct and indirect effects refer to the influence of an exposure on the terminal event not mediated by, and mediated through, the intermediate event, respectively. We propose nonparametric kernel estimators for the two effects and study their asymptotic properties. We conduct numerical simulations to examine the proposed estimators' finite-sample performance. Applying the method to a hepatitis study where the Markov assumption is violated, we show that the effect of hepatitis C on mortality is not mediated through septicemia.

Keywords: illness-death model, non-Markov assumption, causal inference, mediation analysis, kernel density estimation

Order Restricted and Unrestricted Inference for the Recursive Stage Life Testing Model

Aman Prakash[1*], Raj Kamal Maurya[1], Debashis Samanta[2], Debasis Kundu[3]

[1]*Department of Mathematics, Sardar Vallabhbhai National Institute of Technology, Surat, India*

[2]*Department of Mathematics and Statistics, Aliah University, Kolkata, West Bengal, India*

[3]*Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, India*

ABSTRACT

In this study, we propose a recursive stage life testing (RSLT) framework that advances stage life testing experiments by allowing each surviving unit to move sequentially through multiple stages. In this setup, any unit that survives one stage moves on to the next, facing higher stress levels step by step. We develop classical inference procedures for the RSLT model under order-restricted and unrestricted scenarios. To obtain the estimate of the model parameters, maximum likelihood estimators have been used for both cases. Under unrestricted inference, parameters are estimated freely, and in order-restricted inference, we incorporate the natural assumption that higher stress levels lead to early failure. By comparing these two approaches, we see how using known patterns in the data can improve the accuracy of the estimates. We have discussed the proposed methodology using three widely used lifetime distributions: exponential, Weibull, and Chen. An extensive Monte Carlo simulation study evaluates estimator performance under unrestricted and order-restricted settings for each distribution. To illustrate its practical relevance, the proposed RSLT framework is applied to an aerospace electrical connector failure dataset, validating the flexibility and accuracy of the model.

Keywords: Stage life testing; Progressive censoring; Proportional hazard model; Maximum likelihood estimates; Order restriction; Confidence interval

Analysing Load-Sharing System under Progressive Censoring Using Statistical and Machine Learning Approach

Anil Maurya[1*], Raj Kamal Maurya[1]

[1]*Department of Mathematics, Sardar Vallabhbhai National Institute of Technology, Surat, India*

ABSTRACT

This study focuses on the reliability analysis of a multi-component load-sharing system under a progressive censoring scheme. In such a system, where a component fails, its load is redistributed among the surviving components, increasing their failure rates. Such systems are known as load-sharing systems. The lifetimes of the components follow the Weibull distribution; we estimate system parameters and reliability measures using both classical and Bayesian frameworks. In the classical approach, the maximum likelihood estimation method is employed, with asymptotic confidence intervals derived for system parameters. For the Bayesian framework, we use informative priors to obtain Bayesian estimates and construct the highest posterior density intervals. A Monte Carlo simulation study is presented to compare the performance of these estimation methods. Furthermore, real data is used to validate the proposed methodologies, demonstrating the effectiveness of the approaches. Furthermore, we introduce machine learning techniques to analyze the reliability of the model. We have employed three different machine learning methods, such as support vector regressor, random forest regressor, and gradient boosting regressor, to estimate reliability. A comparative analysis of all three methods has been discussed using performance metrics. Finally, the influence of model parameters on reliability estimation has been explored.

Keywords: Load-sharing system, Progressive censoring scheme, Maximum likelihood estimator, Bayes estimator, Machine learning prediction, Performance evaluation

Compound Optimal Design Strategy for Life-Testing Experiment under Progressive Hybrid Censoring Scheme

Vaibhav N. Dhameliya[1*], Raj Kamal Maurya[1], Ritwik Bhattacharya[2]

[1]*Department of Mathematics, Sardar Vallabhbhai National Institute of Technology, Surat, India*

[2]*Department of Mathematical Sciences, University of Texas at El Paso, El Paso, Texas, USA*

ABSTRACT

In life-testing experiments, optimal designs based on single objectives or constraints have been extensively studied, whereas research on multi-objective approaches is still in its early stages of development. In this study, we have introduced a compound design strategy for optimizing life-testing plans under a progressive hybrid censoring scheme. Our approach utilizes a graphical method to develop and evaluate efficient test designs, balancing cost with factors such as trace, determinant, and variance of the inverse Fisher information matrix. With the help of illustrative examples, we demonstrate the advantages of compound optimal designs compared with traditional constraint-based methods. Furthermore, we have discussed the advantages of compound optimal designs over traditional constraint-based methods and used sensitivity analysis to assess design robustness and limitations. Additionally, we analyze real-life data to validate the practical applicability of our proposed methodology, showcasing its effectiveness in optimizing life-testing experiments.

Keywords: Life testing, Compound optimal design, Weibull distribution, Cost function, Progressive hybrid censoring scheme, Sensitivity analysis

Optimal Progressively Censored Reliability Sampling Plans for the Log-Location-Scale Distribution

Chien-Tai Lin[1,*], Shao-Lun Jane[1] and N. Balakrishnan[2]

Department of Mathematics, Tamkang University, Taiwan

Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

ABSTRACT

We introduce here a variable neighborhood search (VNS) algorithm-based approach to determine the minimum sample sizes required for progressively censored reliability sampling plans within the flexible family of log-location-scale family of distributions, which includes Weibull and log-logistic distributions. The proposed method significantly reduces sample sizes compared to previously developed approaches, demonstrating its feasibility, especially for small sample sizes, in contrast to complete search methods (CSM). Optimal censoring plans are identified using A- and D-optimality criteria, and a variance-measure criterion. The proposed approach consistently outperforms the method discussed in Ng et al. (2004) for the case of Weibull distribution.

Keywords: Acceptance sampling; Log-logistic distribution; Optimal criteria; Variable neighborhood search algorithm; Weibull distribution

Back to Poster List

Quantifying Physical Activity Intervention Effects via Functional Regression

Nidhi Pai, Yu Lu, Kristin A. Linn, **Erjia Cui**

Division of Biostatistics and Health Data Science, University of Minnesota, Minneapolis, Minnesota, U.S.A.

ABSTRACT

Physical activity (PA) intervention studies often collect repeated intensity measurements over long observation periods. Quantifying the variation in intervention effects over the study period is critical to evaluating and improving intervention strategies, yet many analyses reduce PA data into scalar summary measures, resulting in limited insights. We propose a functional regression framework, which captures time-varying intervention effects by modeling the entire PA trajectory as a functional observation. From both methodological and practical perspectives, we demonstrate the advantages of function-on-scalar regression (FoSR) over the traditional two-step approach of applying functional principal components analysis (FPCA) followed by regressing scores on covariates. The FoSR is further extended to a function-on-function regression (FoFR) for studying the association of PA across time periods. Methods are applied to daily step counts from the Social incentives to Encourage Physical Activity and Understand Predictors (STEP UP) study, revealing distinct and highly interpretable time-varying effects of three intervention strategies on PA and differences in their sustainability. Our case study highlights the feasibility of functional data analysis techniques for uncovering novel insights in intervention studies with high-dimensional endpoints.

Keywords: Functional data analysis; Functional regression; Physical activity intervention; Wearable devices

Comparison and Ensemble Strategies of Measurement Error Correction Methods in Multiple Regression with Validation Data

Yun-Ting Wei, Jia-Ren Tsai*

Department of Statistics and Information Science, Fu Jen Catholic University, New Taipei City, Taiwan (ROC)

ABSTRACT

Validation data provide researchers with access to replicate or gold-standard measurements on a subset of observations, which in turn facilitate the direct estimation of error variances within measurement error models. In the absence of validation data, measurement error is typically non-identifiable, leading to systematic attenuation bias in the estimation of regression coefficients. However, by incorporating validation data, the observed variance can be decomposed into true variance and error variance, allowing for the estimation of reliability ratios or calibration factors. These quantities are subsequently incorporated into correction procedures, yielding regression coefficients and variable selection outcomes that more closely approximate those based on the unobserved true values. This study integrates several methods with measurement error correction capabilities and further proposes an ensemble estimation strategy to leverage the strengths of different models, aiming to achieve optimal predictive performance. The simulation study encompasses both structural and functional models, settings with independent and correlated covariates, varying sample sizes, and different levels of measurement error. We examine how changes in the proportion of error-free observations within the validation sample affect model performance, thereby providing a comprehensive assessment of the applicability and predictive accuracy of each method and ensemble strategy across diverse scenarios. The results indicate that the presence of validation samples substantially reduces estimation bias and improves predictive accuracy. As the proportion of error-free observations increases, differences across methods gradually diminish, and their overall performance converges. Further comparisons under varying conditions reveal that the proposed ensemble strategy consistently achieves the lowest mean squared prediction error across all simulation scenarios as well as in empirical analyses of Visceral Adipose Tissue data, demonstrating superior robustness and accuracy in diverse settings. These findings underscore the critical role of validation data design and confirm the practical value of ensemble strategies for empirical research involving measurement error.

(* corresponding author: 141628@mail.fju.edu.tw)

Keywords: Measurement error model, Validation data, Ensemble strategy

A Deep Learning Random-Effect Modelling Approach for clustered Count Data

Jihoon Kim[1], VU TUAN ANH[2], Il Do Ha[1]

[1]*Department of Statistics and Data Science, Pukyong National University, Busan, South Korea*

[2]*Department of Artificial Intelligence Convergence, Pukyong National University, Busan, South Korea*

ABSTRACT

The deep neural network (DNN) model, a core model of deep learning, provides high predictive power to predict and classify output variables involved in various fields by modeling the nonlinear functional relationship between input and output variables through hidden layers. However, since the DNN has been mainly developed when output variables are independent, applying it immediately when there is a correlation between output variables has a disadvantage in that predictive performance is greatly reduced. This correlation is usually modeled via random effects, but the random effect model has been mainly developed under the assumption of linearity in the functional relationship between input and output variables.

In this talk, we propose a new Poisson DNN random-effect model. The output variables are correlated count outcomes which are obtained from the repeated measures over time. For estimation (learning) of the proposed model, we develop an optimization algorithm based on negative marginal likelihood as a loss function. Simulation and real data analysis demonstrate the validity of the proposed method. In particular, the simulation results confirmed that the proposed DNN model provides higher predictive performance than the existing prediction models.

Keywords: DNN, random effect, count data, marginal likelihood

A Statistical Framework Leveraging Single-Cell Data for Cell Type Deconvolution in Bulk Transcriptomics

Chin-Sheng Teng[1], Pei-Ling Lin[1], Yijia Xue[1], Ying Lin[1], Ryan Kitagawa[1], Makena Grigsby[1], RJ Honicky[2], Eugene Bolotin[2], Xiaoqian Liu[1], Weixin Yao[1], Xinping Cui[1,*]

[1]*Department of Statistics, University of California, Riverside, USA*

[2]*Industry Partner, Miraomics, USA*

ABSTRACT

Bulk transcriptomics measures averaged gene expression across all cells in a sample, masking the contributions of individual cell types. Abnormal cell type proportions are closely linked to disease initiation and progression, making accurate deconvolution both biologically and clinically important. Single-cell RNA sequencing (scRNA-seq) provides detailed cellular resolution, but its use is limited by cost and sample availability. A framework is therefore needed to leverage existing scRNA-seq data to develop and evaluate deconvolution methods for bulk datasets.

We developed a framework using real scRNA-seq data to generate pseudo-bulk mixtures with known ground-truth proportions, ensuring that natural biological heterogeneity is preserved. Within this framework, we introduced an alternative statistical approach and compared it with conventional deconvolution methods. Conventional approaches rely on predefined reference profiles for each cell type and use regression to estimate cell type proportions. In contrast, our approach directly learns gene-level weights that map bulk expression to cell type composition, reframing deconvolution as a supervised learning problem. This design enables flexible statistical modeling, including regression-based methods such as ridge and lasso regression, and tree-based methods such as Random Forest and XGBoost.

We applied our framework to the CZ CellxGene hippocampus scRNA-seq dataset to generate pseudo-bulk mixtures for evaluation. Once trained, the same models can be applied to bulk RNA-seq data such as GTEx, where true cell type proportions are not available, to estimate cellular composition in real tissue samples. While the hippocampus dataset is relatively small and limits generalization, our framework is flexible and scalable, making it well suited for application to larger single-cell and bulk datasets. This provides a practical tool for advancing statistical methods for cell type deconvolution in bulk transcriptomics.

Keywords: Cell type deconvolution, Bulk transcriptomics, Single-cell RNA-seq, Statistical framework, Supervised learning

Prediction of Deterioration of Patients with Dyspnea in Emergency Department

Sung-Hyuk Choi[1], Byeong-Cheul Ko[2]

[1]*Korea University Guro Hospital*

[2]*Korea Medical Institute*

ABSTRACT

The initial severity triage of patients with dyspnea is essential to provide adequate care. The appropriate value of the factor to 154,383 patients visiting the emergency department from January 1, 2015 to December 31, 2018. Actual hospitalization and discharge in patients with dyspnea who are visiting the emergency department from January 1, 2019 to December 31, 2020. The rapid predicition of hospitalization and discharge using Rasch analysis in ED was highly accurate when combined with more efficient factors, similar to the analysis of artificial intelligence.

Keywords: triage, severity, dyspnea

Reframing Cross-World Independence for Identifying Path-Specific Effects

En-Yu Lai[1], Jih-Chang Yu[2], Yen-Tsung Huang[1]

[1]*Institute of Statistical Science, Academia Sinica*

[2]*Department of Statistics, National Taipei University*

ABSTRACT

The challenge of identifying causal mechanisms in real-world problems often involves multiple factors and requires the evaluation of path-specific effects within a multi-mediator setting. Identification in this context depends not only on standard causal assumptions but also on the demanding cross-world independence assumption. To address this issue, Lin et al. (2017) introduced an alternative causal framework using an interventional approach, which fulfills the cross-world independence by redefining path-specific effects. Later, Stensrud et al. (2021) proposed the dismissible component conditions assumption to identify separable effects in the presence of competing events. In this study, we employed SWIGs to systematically investigate the underlying causal concepts of the three causal semantics for identifying path-specific effects. Specifically, we extended the notion of separable effects and formulated the corresponding assumptions required for identifying path-specific effects. We emphasized that violations of cross-world independence arise from mediators excluded from the model. By analogy with how exchangeability between actual and counterfactual outcomes is achieved through sufficient control of confounders, we argue that cross-world independence can be approximated in practice by incorporating a sufficient set of mediators.

Keywords: cross-world independence, generalized separable effects, interventional approach, path-specific effect, causal assumptions, multimediator model, SWIG

VACANT: An Adaptive Framework for Rare-Variant Association Testing that Leverages Continuous Functional Annotations

Shu-Hsien Cho[1,2], Yao Yu[2], Ryan Bohlender[2], Chad Huff[2]

[1]*UTHealth Graduate School of Biomedical Sciences*

[2]*Department of Epidemiology, University of Texas MD Anderson Cancer Center*

ABSTRACT

To address this, we developed the Variant Annotation Clustering Association Test (VACANT), a novel and flexible statistical framework. VACANT first partitions rare variants (MAF < 0.5%) into ordered, disjoint sets based on a continuous annotation score, a strategy designed to isolate association signals by stratifying variants according to their functional impact. The framework offers two testing modalities: 1) a univariate approach, where each variant set is independently tested and the resulting p-values are combined using the Aggregated Cauchy Association Test (ACAT), effectively preserving the signal from high-impact variant set; and 2) a multivariate approach, where all sets are simultaneously included as predictors in a single model, with significance assessed via a comprehensive goodness-of-fit test. For parameter estimation in both modalities, VACANT employs Firth's penalized logistic regression to ensure robust and well-calibrated results, even with sparse variant counts. To account for population structure and genetic relatedness, the framework can optionally incorporate a genetic relatedness matrix (GRM), shifting the underlying statistical mechanism to a generalized linear mixed model (GLMM).

Through extensive simulations and analysis of whole-exome sequencing data for key breast cancer susceptibility genes (ATM, BRCA1/2, CHEK2, PALB2) from the UK Biobank, we demonstrate the utility of our approach. VACANT consistently achieved similar or higher empirical power than established methods, including SKAT-O and STAAR-O. For instance, our method surpassed 95% power for BRCA1 and BRCA2 with approximately 2,000 cases, a threshold at which competing methods showed negligible power. Critically, VACANT maintained rigorous control of the Type I error rate across all tested scenarios. Its computational efficiency and modular architecture make VACANT a powerful and statistically robust framework for enhancing rare-variant association discovery in modern biobank-scale analyses.

Keywords: Rare-Variant Association Test, Functional Annotation, Case-Control Imbalances, Biobank Data

Causal Mediation Analysis with Survival Data and a Composite Time-to-Event Mediator

Chi-Chun Yeh[1] and Yen-Tsung Huang[2]

[1,2]*Institute of Statistical Science, Academia Sinica*

ABSTRACT

Mediation analysis aims to investigate how an exposure affects an outcome through mediators. It decomposes the effect from the exposure to the outcome into the effect through mediator(s) (indirect effect) and the effect direct to the outcome (direct effect). Recently, mediation analysis has been proposed for the setting where both mediator (e.g., diseases) and outcome (e.g., death) are time-to-event variables. In medicine, many disease categories or syndromes encompass simpler component diseases, as seen in Cardiovascular Kidney Metabolic (CKM) Syndrome. In this study, we build on this concept by considering a composite mediator, in which each component is a time-to-event variable. We propose a counterfactual hazard difference as our effect estimand and derive a counting process-based estimator, which can be simplified as a Nelson-Aalen estimator with time-varying weights. Asymptotic properties, including consistency and asymptotic normality, of the estimator are established using the martingale theory. Extensive numerical experiments are conducted to evaluate the performance of our proposed estimator. We demonstrate the method by applying it to REVEAL data to assess how HCV causes death through multiple chronic diseases (i.e., the CKM Syndrome).

Keywords: Causal inference, Mediation analysis, Mediation group

Authoritarian Regime Types and Their Pathways to Democratization via Coups

Kai-Yuan Wu[1], Jih-Chang Yu[2], Wen-Chin Wu[3], Yen-Tsung Huang[4,*]

[1]*Institute of Statistical Science, Academia Sinica, Taipei*

[2]*Department of Statistics, National Taipei University*

[3]*Institute of Political Science, Academia Sinica, Taipei*

[4]*Institute of Statistical Science, Academia Sinica, Taipei*

ABSTRACT

The relationship between coups and democracy has long been debated. However, the mechanisms through which authoritarian regimes transition to democracy remain unclear. This study investigates how different types of authoritarian regimes transition to democracy, using a causal mediation framework within a semi-competing risks setting. It focuses particularly on the role of coups. We compiled data from 116 countries that experienced authoritarian rule between 1950 and 2020. In the proposed framework, regime type is treated as the exposure, time to coup as the mediator, and time to democratization as the outcome. The exposure and confounders are time- varying, allowing past country-specific information to improve estimates of the causal relationships between authoritarian regimes, coups, and democratic transitions. We employ a semiparametric estimator to identify two effects: an indirect effect, the effect of regime type on democratization via coups, and a direct effect, the effect not mediated through coups. Our proposed semiparametric mediation analyses show that the original type of authoritarian regime affects the chance of later democratization, which may be mediated through the occurrence of coups. Our findings provide new insights into the pathways through which authoritarian regimes either maintain power or transition to democracy, underscoring that the effect of coups depends on the type of authoritarian regime.

Keywords: causal mediation model; semi-competing risk; coups; democracy

Back to Poster List

Marginal Causal Effect Estimation with Continuous Instrumental Variables

Mei Dong[1], Lin Liu[2], Dingke Tang[3], Geoffrey Liu[4], Wei Xu[1,5], Linbo Wang[6]

[1]*Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto*

[2]*Institute of Natural Sciences, MOE–LSC; School of Mathematical Sciences, CMA–Shanghai; SJTU–Yale Joint Center for Biostatistics and Data Science, Shanghai Jiao Tong University*

[3]*Department of Mathematics and Statistics, University of Ottawa*

[4]*Princess Margaret Cancer Centre, University Health Network*

[5]*Department of Biostatistics, University Health Network*

[6]*Department of Statistical Sciences, University of Toronto*

ABSTRACT

Instrumental variables (IVs) are often continuous, arising in diverse fields including economics, epidemiology, and the social sciences. Existing approaches for continuous IVs typically impose strong parametric models or homogeneous treatment effects, while fully nonparametric methods may perform poorly in moderate to high dimensions. We propose a new framework for identifying and estimating the average treatment effect (ATE) with continuous IVs via the conditional weighted average derivative effect (CWADE). Using a conditional Riesz representation, our framework also unifies continuous and categorical IVs. In this framework, the ATE is typically overidentified, leading to a semiparametric observed data model with nonlinear constraints on the tangent space. Addressing these nonlinear constraints requires delicate tools: we develop semiparametric efficiency theory using a second-order parametric submodel, which, to the best of our knowledge, has not been standard practice in this literature. For estimation, we characterize a class of conditional reverse Riesz representers for the CWADE, yielding an easy-to-implement, triply robust estimator that is also locally efficient. We apply our methods to a novel dataset from the Princess Margaret Cancer Centre to examine the so- called obesity paradox in oncology, assessing the causal effect of excess body weight on two-year mortality among patients with non–small cell lung cancer.

Keywords: Mendelian randomization; Riesz representers; Semiparametric methods; Unmeasured confounding

Automated Analysis of Experiments using Hierarchical Garrote

Wei-Yang Yu and V. Roshan Joseph*

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332

ABSTRACT

In this work, we propose an automatic method for the analysis of experiments that incorporates hierarchical relationships between the experimental variables. We use a modified version of the nonnegative garrote method for variable selection which can incorporate hierarchical relationships. The nonnegative garrote method requires a good initial estimate of the regression parameters for it to work well. To obtain the initial estimate, we use generalized ridge regression with the ridge parameters estimated from a Gaussian process prior placed on the underlying input-output relationship. The proposed method, called HiGarrote, is fast, easy to use, and requires no manual tuning. Analysis of several real experiments are presented to demonstrate its benefits over the existing methods.

Keywords: Gaussian process; Generalized ridge regression; Nonnegative garrote; Variable selection

Uncertainty Quantification for Noisy Low-Tubal-Rank Tensor Completion

Jiuqian Shang, Jingyang Li, Yang Chen

Department of Statistics, University of Michigan

ABSTRACT

The low-tubal-rank tensor model has been used for multidimensional data to capture signals in the frequency domain. Algorithms have been developed to estimate low-rank third-order tensors from partial and corrupted entries. However, uncertainty quantification and statistical inference for these estimates remain largely unclear. We introduce a flexible framework for inference on general linear forms of a large tensor whenever an entry-wise consistent estimator is available. Under mild regularity conditions, we construct asymptotically normal estimators of these linear forms through double-sample debiasing and low-rank projection. These estimators allow us to construct confidence intervals and perform hypothesis testing. Simulation studies support our theoretical results. We apply the method to the total electron content (TEC) reconstruction problem, demonstrating that it delivers more robust reconstructions and informative entry-wise confidence intervals.

Keywords: Spectral method; Asymptotic normality; Total electron content (TEC)

Linearly Constrained Symmetric Rank-One Approximation for Pre-Image Recovery in Nonlinear Matrix Completion

Elvin Tseng[1], Laura Balzano[1,2], Gregory Ongie[3]

[1]*Department of Statistics, University of Michigan*

[2]*Department of Electrical Engineering & Computer Science, University of Michigan*

[3]*Department of Mathematical and Statistical Sciences, Marquette University*

ABSTRACT

We study the linearly constrained rank-one approximation of a symmetric matrix, motivated by pre-image recovery and structured matrix estimation. We show that the problem admits a simple geometric reduction to a one-dimensional equation whose largest root yields the global minimizer. Under mild assumptions, this solution is unique. We further give a verifiable condition that distinguishes true local minima from spurious stationary points. We also clarify connections to generalized eigenvalue problem and trust-region formulations, revealing strong but non-trivial structural links. For large-scale instances where the required eigendecomposition is computationally intensive, we develop a scalable projected-gradient method that avoids explicit factorization and remains effective in high dimensions.

Keywords: Constrained optimization; Generalized eigenvalue problem; Nonlinear matrix completion; Trust-region subproblem; Spurious local minima

Investigating the Pleiotropic Effects of Splice-Affecting Variants on Cancer and Metabolic Diseases

Jia-Ying Su[1], En-Yu Lai[1], Chien-Ling Lin[2], Yen-Tsung Huang[1,*]

[1]*Institute of Statistical Science, Academia Sinica*

[2]*Institute of Molecular Biology, Academia Sinica*

ABSTRACT

Genome-wide association studies (GWAS) have identified numerous loci for complex diseases, but a key challenge is to understand their biological functions. A specific class of potent functional variants, known as Splice-Affecting Variants (SAVs), alter the RNA splicing process, which can lead to the production of aberrant proteins and disrupt cellular function. There is substantial evidence that metabolic dysregulation, such as that seen in type 2 diabetes and obesity, is a significant risk factor for various cancers. Therefore, we aim to investigate whether SAVs confer risk for both metabolic diseases and cancer through a shared genetic mechanism, a phenomenon known as pleiotropy.

We leveraged genotype and electronic health record data from over 560,000 participants in the Taiwan Precision Medicine Initiative (TPMI). Our analysis focused on 468 high-confidence SAVs predicted by our previously developed tool, spliceAPP. We then employed a novel statistical framework designed to integrate evidence from two different traits. This method substantially boosts the power to identify shared genetic architecture, even when one of the signals is modest.

Our analyses indicate that several SAVs are significantly associated with the risk of metabolic diseases while also demonstrating a concurrent association with cancer risk (FDR < 0.01). Specifically, 118 SAVs were identified between prostate cancer and type 2 diabetes, while 71 SAVs were found between rectal cancer and type 2 diabetes. Among these, a core set of 38 SAVs was significant in both relationships. Notably, several of these shared variants are located in genes with well-established roles in both cancer and metabolic diseases, including *CDKN1A*, *TNFRSF1B*, *C2*, *SERINC2*, and *ST3GAL6*. These findings lend support to our pleiotropy hypothesis, offering new insights into how SAVs may drive cancer pathophysiology through shared biological pathways and highlighting the potential of integrative analytical strategies to unravel the genetic basis of complex diseases.

Keywords: Pleiotropy; Splice-Affecting Variants (SAVs); Cancer Genetics; Metabolic Disease; Taiwan Precision Medicine Initiative (TPMI).

Controlling the False Discovery Proportion in Observational Studies with Hidden Bias

Menggi Lin, Colin Fogarty

University of Michigan

ABSTRACT

We propose an approach to exploratory data analysis in matched observational studies. We consider the setting where a single intervention is thought to potentially impact multiple outcome variables, and the researcher would like to investigate which of these causal hypotheses come to bear while accounting not only for the possibility of false discoveries, but also the possibility that the study is plagued by unmeasured confounding. For any candidate set of rejected hypotheses, our method provides sensitivity sets for the false discovery proportion (FDP), the proportion of rejected hypotheses that are actually true. For a set \cR containing |\cR| outcomes, the method describes how much unmeasured confounding would need to exist for us to believe that the proportion of true hypotheses is 0/|\cR|, 1/|\cR|, ..., all the way to |\cR|/|\cR|. Moreover, the resulting confidence statements intervals are valid simultaneously over all possible choices for the rejected set, allowing the researcher to look in an ad hoc manner for promising subsets of outcomes that maintain a large estimated fraction of true discoveries even if a large degree of unmeasured confounding is present. The approach is particularly well suited to sensitivity analysis, as conclusions that some fraction of outcomes were affected by the treatment exhibit larger robustness to unmeasured confounding than the conclusion that any particular outcome was affected. In principle, the method requires solving a series of quadratically constrained integer programs. That said, we show not only that a solution can be obtained in reasonable run time, but also that one can avoid running the integer program altogether with high probability in large samples. We illustrate the practical utility of the method through simulation studies and a data example.

Keywords: Causal Inference; False Discovery Proportion; Matching; Sensitivity Analysis

LBBNN: An R package for Sparse Bayesian Neural Networks

Lars Skaaret-Lund, Eirik Høyheim, Aliaksandr Hubin

Bioinformatics and applied statistics, Norwegian University of Life Sciences

ABSTRACT

Artificial neural networks are highly popular and successful non-linear statistical models, used in a wide variety of applications. They do however have some drawbacks, mainly that they are heavily overparameterized, leading to overfitting. There are several ways of mitigating this, one is to incorporate parameter uncertainty through a Bayesian neural network (BNN). Model averaging also becomes straight-forward with BNNs, which typically improves performance. It is also of interest to reduce the amount of parameters of BNNs, to make them more interpretable, and reduce the memory footprint. This can be done in a principled way with latent binary Bayesian neural networks (LBBNN) [1], where a Bernoulli variable is placed in front of each weight in the network. While this incurs an extra parameter per weight compared to standard BNNs, the resulting sparse networks can make up for this. For example, both [1] and [2] demonstrate that networks with only 10 % of the original weights maintain similar accuracy and calibration metrics.

A further advance is to concatenate the input-layer to each hidden layer. Doing this allows for learning different functions based on the data. For example, in [3], we demonstrate that if we generate synthetic data that is linear, the network learns to remove all non-linear transformations and reduces to only a linear layer. This is in contrast to traditional statistical models where the assumptions on what function to model has to be made a priori. In [3], we also demonstrate that our method generates much sparser representations than standard LBBNNs, and also sparser than other related baselines, while maintaining high predictive power. Additionally, we introduce the concept of active paths, which is simply a path from an input node, either directly or via one or more hidden units to an output node. The motivation behind this is that pruning edges in a network may leave us with inactive nodes, meaning that no information going through these nodes will contribute to the output. This then can leave us with an even sparser representation than just considering the proportion of removed edges. Furthermore, the concept of active paths allows us to obtain both global and local explanations of predictions of the network. The goal of the R package LBBNN is to unify all the above mentioned approaches to make it usable for a wider audience. We allow the user to choose between standard LBBNNs, LBBNNs with normalizing flows (moving beyond mean-field Gaussian posteriors to allow for non-independent weights, as demonstrated in [2], and LBBNNs

with input-skip.

Keywords: Bayesian neural networks, variational inference, uncertainty, explainability

References:

[1] A. Hubin and G. Storvik. Sparse Bayesian neural networks: Bridging model and parameter uncertainty through scalable variational inference. Mathematics, 12(6):788, 2024.

[2] L. Skaaret-Lund, G. Storvik, and A. Hubin. Sparsifying Bayesian Neural Networks with Latent Binary Variables and Normalizing Flows. Transactions on Machine Learning Research, 10 2024. URL https://openreview.net/forum?id=d6kqUKzG3V.

[3] E. Høyheim, L. Skaaret-Lund, S. Sæbø, and A. Hubin. Explainable bayesian deep learning through input-skip latent binary bayesian neural networks. arXiv preprint arXiv:2503.10496, 2025

Back to Poster List

Cramér-Type Moderate Deviation and Berry–Esseen Bounds in the p-Spin Curie–Weiss Model

Tianyu Liu[1], Somabha Mukherjee[1], Bhaswar B. Bhattacharya[2]

[1]*Department of Statistics and Data Science, National University of Singapore, Singapore*

[2]*Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, USA*

ABSTRACT

Limit theorems for the magnetization in the p-spin Curie–Weiss model, for $p \geq 3$, have been derived recently. In this work, we strengthen these results by proving Cramér-type moderate deviation theorems and Berry–Esseen bounds for the magnetization (suitably centered and scaled). In particular, we show that the rate of convergence is on the order $O(N^{-1/2})$ when the magnetization has asymptotically Gaussian fluctuations, and $O(N^{-1/4})$ when the fluctuations are non-Gaussian. As an application, we derive a Berry–Esseen bound for the maximum pseudolikelihood estimate of the inverse temperature in the p-spin Curie–Weiss model with no external field, for all points in the parameter space where consistent estimation is possible.

Keywords: Central Limit Theorems; Stein's Method; Moderate Deviation; Spin Systems

Concentration-QTC Analysis to Support ICH E14 with a Real Case

Kong Xin[1], Xin Zhang[2]

Biostatistics, Sanofi

ABSTRACT

This presentation introduces ICH E14 and Concentration-QTC model. The PK and ECG data obtained from a Phase I study characterized the effects of the drug and had demonstrated there were no effects on QT/QTc intervals and other ECG parameters. The results were adequate to waive a formal TQT trial for regulatory submission.

Keywords: Concentration-QTC, Pharmacokinetics, ICH E14, Thorough QT/QTc, Early Phase