Examining Directional Association between Depression and Anxiety in US Medical Interns

Soumik Purkayastha[1], Peter X.-K. Song[2,*]

[1]*Department of Biostatistics and Health Data Science, University of Pittsburgh, Pittsburgh, USA.*

[2]*Department of Biostatistics, University of Michigan, Ann Arbor, USA.*

ABSTRACT

Utilizing a novel entropy loss (EL) metric, a causal discovery method is proposed to understand directional effects in the causal relationship between depression and anxiety among medical interns. This method advances existing methods of bivariate causal discovery with theoretical guarantees of causal effect identifiability and statistical inference and enjoys good computational performance. Using data from the intern health study (n=6,858), the proposed method reveals with high statistical confidence that depression scores (PHQ-9) consistently predispose anxiety scores (GAD-7) across four longitudinal visits of the study, controlling for demographic confounders. This finding provides crucial insights into the directional effect useful for mental health intervention strategies for medical interns. Simulation studies demonstrate that EL achieves nearly superior accuracy compared to existing approaches across various conditions with reduced computation time. The EL framework's ability to handle discrete clinical scores while adjusting for confounders makes it particularly valuable for psychiatric epidemiology and broader applications in causal discovery with discrete data.

Keywords: Information theory; Asymmetry; Causal discovery; Mental Health

Statistical Methods for Chemical Mixtures: A Roadmap for Practitioners Using Simulation Studies and a Sample Data Analysis in the PROTECT Cohort

Wei Hao[1], Amber L. Cathey[2], Max M. Aung[3], Jonathan Boss[1], John D. Meeker[2], Bhramar Mukherjee[4]

[1]*Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA*

[2]*Department of Environmental Health Sciences, University of Michigan, Ann Arbor, MI, USA*

[3]*Division of Environmental Health, University of South California, Los Angeles, CA, USA*

[4]*Yale School of Public Health, Yale University, New Haven, CT, USA*

ABTRACT

Quantitative characterization of the health impacts associated with exposure to chemical mixtures has received considerable attention in current environmental and epidemiological studies. With many existing statistical methods and emerging approaches, it is important for practitioners to understand which method is best suited for their inferential goals. The goal of this paper is to provide empirical simulation-based evidence regarding performance of mixture methods to help guide researchers on selecting the best available methods to address three scientific questions in mixtures analysis: identifying important components of a mixture, identifying interactions among mixture components and creating a summary score for risk stratification and prediction. We conduct a review and comparison of 11 analytical methods available for use in mixtures research, through extensive simulation studies for continuous and binary outcomes. In addition, we carry out an illustrative data analysis using the PROTECT birth cohort from Puerto Rico, to examine the associations between exposure to chemical mixtures—metals, polycyclic aromatic hydrocarbons (PAHs), phthalates and phenols—and birth outcomes. Our simulation results suggest that the choice of methods depends on the goal of analysis and there is no clear winner across the board. For selection of important toxicants in the mixtures and for identifying interactions, Elastic net by Zou et al. (Enet), Lasso for Hierarchical Interactions by Bien et al. (HierNet), Selection of nonlinear interactions by a forward stepwise algorithm by Narisetty et al. (SNIF) have the most stable performance across simulation settings. For overall summary or a cumulative measure, we find that using the Super Learner to combine multiple Environmental Risk Scores can lead to improved risk stratification and prediction properties. We develop an integrated R package "CompMix" that provides a platform for mixtures analysis where the practitioners can implement a pipeline that includes

several approaches for mixtures analysis. Our study offers guidelines for selecting appropriate statistical methods for addressing specific scientific questions related to mixtures research. We identify critical gaps where new and better methods are needed.

Keywords: Chemical mixtures, environmental risk score

Back to Sessions List

Supervised Fusion Learning of Physical Activity Features: Functional Frameworks and Longitudinal Analysis with L_0 Regularization

Dr. Margaret Banker

Northwestern University Feinberg School of Medicine

ABTRACT

Wearable devices are crucial in physical activity research because they provide continuous, real-time monitoring of various health metrics such as heart rate, physical activity, sleep patterns, and vital signs. These devices enable the collection of extensive, longitudinal data, offering insights into the daily lives and health trajectories of older adults. This information is invaluable for identifying early signs of health decline, assessing the effectiveness of interventions, and personalizing care plans. I consider wearable device data in a functional framework with an L_0 regularization approach, handling highly correlated micro-activity windows that serve as predictors in a scalar-on-function regression model. I develop a longitudinal functional framework with repeated wearable data to understand the influence of serially measured functional accelerometer data on longitudinal health outcomes. This method leverages Quadratic Inference Function (QIF) via mixed integer optimization for longitudinal data analysis to detect critical physical activity windows and assess their population-average effects on health outcomes.

Keywords: Actigraphy; Change-point detection; Mixed Integer Optimization (MIO); Scalar-on-function regression; Quadratic Inference Function (QIF)

Estimation and Inference of Quantile Spatially Varying Coefficient Models Over Complicated Domains

Myungjin Kim[1], Lily Wang[2,*], Huixia Judy Wang[3]

[1]*Department of Statistics, Kyungpook National University*

[2]*Department of Statistics, George Mason University*

[3]*Department of Statistics, Rice University*

ABSTRACT

This work presents a flexible quantile spatially varying coefficient model (QSVCM) for the regression analysis of spatial data. The proposed model enables researchers to assess the dependence of conditional quantiles of the response variable on covariates while accounting for spatial non-stationarity. Our approach facilitates learning and interpreting heterogeneity in spatial data distributed over complex or irregular domains. We introduce a quantile regression method that utilizes bivariate penalized splines in triangulation to estimate unknown functional coefficients. We establish the L2 convergence of the proposed estimators, demonstrating their optimal convergence rate under certain regularity conditions. An efficient optimization algorithm is developed using the alternating direction method of multipliers (ADMM). We develop wild bootstrap-based pointwise confidence intervals for the QSVCM quantile coefficients. Furthermore, we construct reliable conformal prediction intervals for the response variable using the proposed QSVCM. Numerical studies show the remarkable performance of the proposed methods.

Keywords: Alternating direction method of multiplier; Bivariate penalized spline; Nonparametric quantile regression; Triangulation