Joint Mixed Membership Modeling of Multivariate Longitudinal and Survival Data

Yuyang He[1], **Xinyuan Song**[2], Kai Kang[3]

Department of Statistics, The Chinese University of Hong Kong

ABSTRACT

This study develops a novel joint mixed membership model for multivariate longitudinal AD-related biomarkers and time of AD diagnosis. Unlike conventional finite mixture models that assign each subject a single subgroup membership, the proposed model assigns partial membership across subgroups, allowing subjects to lie between two or more subgroups. This flexible structure enables individualized disease progression and facilitates the identification of clinically meaningful neurological statuses that are often elusive in current mixed-effects models. We employ a spline-based trajectory model to characterize complex and possibly nonlinear patterns of multiple longitudinal clinical markers. A Cox model is then used to examine the effects of time-variant risk factors on the hazard of developing AD. We develop a Bayesian method coupled with efficient Markov chain Monte Carlo sampling schemes to perform statistical inference. The proposed approach is assessed through extensive simulation studies and an application to the Alzheimer's Disease Neuroimaging Initiative study, demonstrating better performance in AD diagnosis compared to existing joint models.

Keywords: Mixed membership model; longitudinal data; MCMC methods; survival data

Building a Dose Toxo-Equivalence Model from a Bayesian Meta-Analysis of Published Clinical Trials

Elizabeth A Sigworth[1], Samuel M Rubinstein[2], Jeremy L Warner[3], Yong Chen[4], **Qingxia Chen[1]**

[1]*Department of Biostatistics, Vanderbilt University.*

[2]*Division of Hematology, University of North Carolina School of Medicine.*

[3]*Department of Medicine, Vanderbilt University School of Medicine.*

[4]*Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania.*

ABSTRACT

In clinical practice, medications are often interchanged in treatment protocols when a patient negatively reacts to their first line of therapy. Although switching between medications is common, clinicians often lack structured guidance when choosing the initial dose and frequency of a new medication, given the former with respect to the risk of adverse events. In this paper, we propose to establish this dose toxic-equivalence relationship using published clinical trial results with one or both drugs of interest via a Bayesian meta-analysis model that accounts for both within- and between-study variances. With the posterior parameter samples from this model, we compute median and 95% credible intervals for equivalent dose pairs of the two drugs that are predicted to produce equal rates of an adverse outcome, relying solely on study-level information. Via extensive simulations, we show that this approach approximates well the true dose toxo-equivalence relationship, considering different study designs, levels of between-study variance, and the inclusion/exclusion of no confounder/nonmodifier subject-level covariates in addition to study-level covariates. We compare the performance of this study-level meta-analysis estimate to the equivalent individual patient data meta-analysis model and find comparable bias and minimal efficiency loss in the study-level coefficients used in the dose toxo-equivalence relationship. Finally, we present the findings of our dose toxo-equivalence model applied to two chemotherapy drugs, based on data from 169 published clinical trials.

Keywords: Bayesian methods; dose toxo-equivalence; individual patient data meta-analysis; study-level meta-analysis.

Quasi Instrumental Variable Methods for Stable Hidden Confounding and Binary Outcome

Zhonghua Liu[1], Baoluo Sun[2], Ting Ye[3], David Richardson[4], Eric Tchetgen Tchetgen[5]

[1]*Department of Biostatistics, Columbia University*

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

[3]Department of Biostatistics, University of Washington, Seattle

[4]Department of Environmental and Occupational Health, University of California, Irvine

[5]Department of Statistics and Data Science, University of Pennsylvania

ABSTRACT

Instrumental variable (IV) methods are central to causal inference from observational data, particularly when a randomized experiment is not feasible. However, of the three conventional core IV identification conditions, only one, IV relevance, is empirically verifiable; often one or both of the other conditions, exclusion restriction and IV independence from unmeasured confounders, are unmet in real-world applications. These challenges are compounded when the outcome is binary, a setting for which robust IV methods remain underdeveloped. A fundamental contribution of this paper is the development of a general identification strategy justified under a structural equilibrium dynamic generative model of so-called stable confounding and a quasi-instrumental variable (QIV), i.e. a variable that is only assumed to be predictive of the outcome. Such a model implies (a) stability of confounding on the multiplicative scale, and (b) stability of the additive average treatment effect among the treated (ATT), across levels of that QIV. The former is all that is necessary to ensure a valid test of the causal null hypothesis; together those two conditions establish nonparametric identification and estimation of the conditional and marginal ATT. To address the statistical challenges posed by the need for boundedness in binary outcomes, we introduce a generalized odds product re-parametrization of the observed data distribution, and we develop both a principled maximum likelihood estimator and a triply robust semiparametric locally efficient estimator, which we evaluate through simulations and an empirical application to the UK Biobank.

Keywords: Binary outcome; Invalid instrument; Mendelian randomization; Semiparametric theory; Unmeasured confounding

Clustering-Informed Shared-Structure Variational Autoencoder for Missing Data Imputation in Large-Scale Healthcare Data

Yuan Chen[1], Yasin Khadem Charvadeh, Kenneth Seier, Katherine S. Panageas, Danielle Vaithilingam, Mithat Gönen, Yuan Chen[2,*]

Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center

ABSTRACT

Despite advancements in healthcare data management, missing data in electronic health records (EHR) and patient-reported outcomes remain a persistent challenge, limiting their usability in healthcare analytics. Conventional imputation methods often struggle to capture complex non-linear relationships, require extensive computation time, and are limited in addressing various types of missing data mechanisms. To overcome these challenges, we propose the clustering-informed shared-structure variational autoencoder (CISS-VAE), which utilizes the strengths of Bayesian neural networks. This model can effectively capture complex associations and accommodate various missing data mechanisms, including missing not at random (MNAR). We also develop iterative learning algorithms that further enhance missing data imputation accuracy while preventing overfitting. Comprehensive simulations demonstrate the superior accuracy of our model compared to traditional and contemporary methods. We apply our method to EHR data from early-stage breast cancer patients at Memorial Sloan Kettering Cancer Center, aiming to mitigate the impact of missing data and enhance health monitoring and analyses.

Keywords: Missing Data Imputation; Variational Autoencoder; Missing Not at Random; Electronic Health Records