Two-Stage Adaptive Testing of Large-Scale Mediation Hypotheses

Yueqi Xu, **Kwun Chuen Gary Chan**

Department of Biostatistics, University of Washington

ABSTRACT

In testing many mediation hypotheses, we propose to use a pair of asymptotically independent p-values constructed from the exposure-mediator and mediator-outcome path-specific p-values from each hypothesis, to test a more stringent joint null hypothesis and a composite mediation null hypothesis respectively. A first-stage screening procedure based on the p-values testing a joint hypothesis can effectively reduce the number of mediation hypotheses to be tested in the second stage. The procedure controls false discovery rate while being consistently well powered across a wide range of scenarios.

Keywords: Composite null hypothesis, False-discovery rate control, Joint significance, Ordered statistics.

Leveraging Multi-Study, Multi-Outcome Data to Improve External Validity and Efficiency of Clinical Trials for Medications for Opioid Use Disorder

Amy Pitts[1], Oliver Hines[2], Rachael Ross[3], Kara Rudolph[2], **Caleb Miles[2]**

Regeneron Pharmaceuticals

Columbia University

University of Utah

ABSTRACT

As data sources have become more plentiful and readily accessible, the practice of data fusion has become increasingly ubiquitous. However, when the focus is on a causal effect on a particular outcome, a major limitation is that this outcome may not be available in all data sources. In fact, different randomized experiments or observational studies of a common exposure will often focus on potentially related yet distinct outcomes. One such example is the medication for opioid use disorder (MOUD) clinical trials network (CTN), which consists of several randomized trials of the comparative effectiveness of different MOUDs with inconsistent quality of life measures across studies. The causally principled methodology is developed for fusing data sets when multiple outcomes are observed across studies, which leverages mediators and outcomes of secondary interest as informative proxies for the missing outcome of primary interest, thereby maximizing power and efficiency by making full use of the available data. As this methodology relies on a key transportability assumption, methods are also developed to assess the degree of sensitivity to violations of this assumption. This methodology is applied to data from the CTN trials to make improved causal inferences about the comparative effectiveness of medications for opioid use disorder.

Keywords: Causal inference, data fusion, external validity, generalizability, missing data, transportability

Mediation Analysis with Graph Mediator

Yixi Xu[1], **Yi Zhao[1,*]**

[1]*Department of Biostatistics and Health Data Science, Indiana University School of Medicine*

ABSTRACT

This study introduces a mediation analysis framework when the mediator is a graph. A Gaussian covariance graph model is assumed for graph presentation. Causal estimands and assumptions are discussed under this presentation. With a covariance matrix as the mediator, a low-rank representation is introduced and parametric mediation models are considered under the structural equation modeling framework. Assuming Gaussian random errors, likelihood-based estimators are introduced to simultaneously identify the low-rank representation and causal parameters. An efficient computational algorithm is proposed and asymptotic properties of the estimators are investigated. Via simulation studies, the performance of the proposed approach is evaluated. Applying to a resting-state fMRI study, a brain network is identified within which functional connectivity mediates the sex difference in the performance of a motor task.

Keywords: Common diagonalization; Covariance regression; Decomposition method; Gaussian covariance graph model; Mediation analysis

Toward Flexible and Efficient Counterfactual Density Estimation

Kwangho Kim[1], Jisu Kim[2], Edward Kennedy[3]

[1]*Department of Statistics, Korea University*

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

[3]*Department of Statistics and Data Science, Carnegie Mellon University*

ABSTRACT

Comparing full counterfactual distributions provides richer measures of causal effects than conventional summaries such as means or quantiles. We study the problem of estimating the entire counterfactual density in a fully nonparametric setting and develop three complementary approaches. First, we introduce a doubly robust–style estimator that directly targets a kernel–smoothed counterfactual density. We establish its large-sample properties, derive finite-sample risk bounds, and construct uniform confidence bands via a bootstrap procedure. Second, we propose a diffusion-informed bump that adapts to the intrinsic geometry of the outcome manifold. By replacing the standard kernel with a diffusion-informed smoother, this estimator reduces bias near complex supports and attains faster convergence rates for high-dimensional outcomes when intrinsic dimension is low. Third, we develop a score-based method that targets the smoothed counterfactual score rather than the density itself. Focusing on the score enables even faster rates under common structural assumptions, with smaller constant factors, and can be paired with efficient density recovery when desired. We compare the three estimators and clarify when each is preferable, with particular emphasis on the advantages of diffusion-based smoothing for learning counterfactual distributions in high-dimensional settings. Together, these results provide a unified toolkit for flexible and statistically efficient counterfactual density estimation.

Keywords: causal inference; density estimation; influence function; semiparametric theory; diffusion model