Principal Stratification with U-Statistics under Principal Ignorability

Fan Li[1], Xinyuan Chen[2]
[1]*Department of Biostatistics, Yale School of Public Health, CT, USA*

[2]*Department of Mathematics and Statistics, Mississippi State University, MS, USA*

ABSTRACT

Principal stratification is a popular framework for causal inference in the presence of an intermediate outcome. While the principal average treatment effects have traditionally been the default target of inference, it may not be sufficient when the interest lies in the relative favorability of one potential outcome over the other within the principal stratum. We introduce the principal generalized causal effect estimands, which extend the principal average causal effects to accommodate arbitrary nonlinear contrast functions. Under principal ignorability, we expand the existing theoretical results to a much wider class of causal estimands in the presence of a binary intermediate variable. We develop identification formulas and derive the efficient influence functions of the generalized causal estimands for principal stratification analyses. These efficient influence functions motivate a set of multiply robust estimators and lay the ground for obtaining efficient debiased machine learning estimators via cross-fitting based on U-statistics. The proposed methods are illustrated through simulations and the analysis of a data example.

Keywords: Causal inference, efficient influence function, principal stratification, multiply robust estimation, win ratio, probabilistic index

Back to Sessions List

Using a Two-Parameter Sensitivity Analysis Framework to Efficiently Combine Randomized and Non-randomized Studies

Ruoqi Yu[1], Bikram Karmakar[2], Jessica Vandeleest[3], Eleanor Bimla Schwarz[4]

[1]*Department of Statistics, University of Illinois Urbana-Champaign*

[2]*Department of Statistics, University of Florida*

[3]*California National Primate Research Center, University of California Davis*

[4]*Department of Medicine, University of California San Francisco*

ABSTRACT

Causal inference is vital for informed decision-making across fields such as biomedical research and social sciences. Randomized controlled trials (RCTs) are considered the gold standard for the internal validity of inferences, whereas observational studies (OSs) often provide the opportunity for greater external validity. However, both data sources have inherent limitations preventing their use for broadly valid statistical inferences: RCTs may lack generalizability due to their selective eligibility criterion, and OSs are vulnerable to unobserved confounding. This paper proposes an innovative approach to integrate RCT and OS that borrows the other study's strengths to remedy each study's limitations. The method uses a novel triplet matching algorithm to align RCT and OS samples and a new two-parameter sensitivity analysis framework to quantify internal and external validity biases. This combined approach yields causal estimates that are more robust to hidden biases than OSs alone and provides reliable inferences about the treatment effect in the general population. We apply this method to investigate effects of lactation on maternal health using a small RCT and a long-term observational health records dataset from the California National Primate Research Center. This application demonstrates the practical utility of our approach in generating scientifically sound and actionable causal estimates.

Keywords: Causal inference; Generalizability bias; Matching, Sensitivity analysis, Unmeasured confounding

Robust Sensitivity Analysis via Augmented Percentile Bootstrap under Simultaneous Violations of Unconfoundedness and Overlap

Cui Han[1], **Xinran Li[2]**

[1]*Department of Statistics, University of Illinois at Urbana-Champaign*

[2]*Department of Statistics, University of Chicago*

ABSTRACT

The identification of causal effects in observational studies typically relies on two standard assumptions: unconfoundedness and overlap. However, both assumptions are often questionable in practice: unconfoundedness is inherently untestable, and overlap may fail in the presence of extreme unmeasured confounding. While various approaches have been developed to address unmeasured confounding and extreme propensity scores separately, few methods accommodate simultaneous violations of both assumptions. In this paper, we propose a sensitivity analysis framework that relaxes both unconfoundedness and overlap, building upon the marginal sensitivity model. Specifically, we allow the bound on unmeasured confounding to hold for only a subset of the population, thereby accommodating heterogeneity in confounding and allowing treatment probabilities to be zero or one. Moreover, unlike prior work, our approach does not require bounded outcomes and focuses on overlap-weighted average treatment effects, which are both practically meaningful and robust to non-overlap. We develop computationally efficient methods to obtain worst-case bounds via linear programming, and introduce a novel augmented percentile bootstrap procedure for statistical inference. This bootstrap method handles parameters defined through over-identified estimating equations involving unobserved variables and may be of independent interest. Our work provides a unified and flexible framework for sensitivity analysis under violations of both unconfoundedness and overlap.

Keywords: unconfoundedness; overlap; extreme confounding; overlap weights; parameter augmentation