High-Dimensional Causal Mediation Analysis by Partial Sum Statistic and Split-Sample Strategy in Radiogenomics Study

Hung-Ching Chang[1], Yusi Fang[1], Michael Gorczyca[1], Kayhan Batmanghelich[2], George C. Tseng[1]

[1]*Department of Biostatistics, University of Pittsburgh, U.S.A.*
[2]*Department of Electrical and Computer Engineering, Boston University, U.S.A.*

Abstract

Causal mediation analysis with high-dimensional mediator brings new statistical challenges due to curse of dimensionality. To reduce noises, many existing methods perform variable selection by penalized regression or screening test, which violate cross-world assumption in theoretical causality and produce biased estimation. To circumvent the issue, recent methods conduct orthogonal transformation while the strategy fails in variable selection and interpretability. The highly correlated nature of mediators commonly seen in imaging and omics data also bring difficulty to prioritize mediator findings. To this end, we develop a framework with partial sum statistic and split-sample strategy, namely PS5, for high dimensional causal mediation and focus on a radiogenomics study in lung disease. The method provides a test for global indirect effect satisfying causality assumptions, followed by algorithms to select and prioritize detected mediators. Extensive simulations demonstrate proper control of type I error, increased statistical power, improved mediator selection accuracy at various mediation, sparsity, and correlation settings. Finally, we demonstrate the method in COPDGene data set showing causal mediation of lung images in the association between SNPs and lung function outcome. The result detects regions with mediation effects, estimates the global indirect effect, and concludes that the pleural cavity contributes majority of the global indirect effect. The method is also applied to an omics mediation analysis in METABRIC breast cancer data.

Keywords: High-dimensional inference; Causal mediation analysis; Partial sum statistic; sample splitting; Radiogenomics.