Bayesian High-Dimensional Biological Pathway-Guided Mediation Analysis with Application to Metabolomics

Zhang, Yuzi

Abstract:
With advances in mass spectrometry technologies, metabolomics data are increasingly used to investigate biological mechanisms underlying associations between exposures and health outcomes in clinical and epidemiological studies. Mediation analysis is a powerful framework and when applied to metabolomics data, a large number of correlated metabolites arising from metabolic pathways consisting of biochemical reactions need to be considered as mediators. To identify metabolic pathways as active mediators, existing approaches typically focus on first identifying individual metabolites as active mediators, followed by post-hoc metabolic pathway determination. These multi-stage procedures make statistical inference challenging. We propose a Bayesian biological pathway-guided mediation analysis that aims to jointly analyze all metabolites together, identify metabolic pathways directly, and estimate metabolic pathway-specific indirect effects. This is accomplished by incorporating existing biological knowledge of metabolic pathways to account for correlations among mediators, along with variable selection and dimension reduction techniques. Advantages of the proposed method is demonstrated in extensive simulation studies with real-word metabolic pathway structure. We apply the proposed method to two studies examining the role of metabolism in mediating (1) the effect of Roux-en-Y gastric bypass on glycemic control, and (2) the effect of prenatal exposure to per- and polyfluoroalkyl substances (PFAS) on gestational age at birth. Our analyses confirm metabolic pathways previously identified and provide additional uncertainty quantification for the mediation effects.