High-dimensional Multivariate Mediation with Application to Neuroimaging Data

Oliver Y. Chén[1], Elizabeth L. Ogburn[1], Ciprian M. Crainiceanu[1],
Brian S. Caffo[1], Tor D. Wager[2], Martin A. Lindquist[1]

[1] Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

[2] Department of Psychology and Neuroscience
University of Colorado Boulder

1

Abstract

Mediation analysis has become an important tool in the behavioral sciences for investigating the role of intermediate variables that lie in the path between a randomized treatment and an outcome variable. The influence of the intermediate variable on the outcome is often determined using structural equation models (SEMs), with model coefficients interpreted as effects. While there has been significant research on the topic in recent years, little work has been done on mediation analysis when the intermediate variable (mediator) is a high-dimensional vector. In this work we present a new method for performing mediation in this setting called the directions of mediation (DMs). The first DM is defined as the linear combination of the elements of a high-dimensional vector of potential mediators that maximizes the likelihood of the SEM. The subsequent DMs are defined as linear combinations of the elements of the high-dimensional vector that are orthonormal to the previous DMs and maximize the likelihood of the SEM. We provide an estimation algorithm and establish the asymptotic properties of the obtained estimators. This method is well suited for cases when many potential mediators are measured. Examples of high-dimensional potential mediators are brain images composed of hundreds of thousands of voxels, genetic variation measured at millions of SNPs, or vectors of thousands of variables in large-scale epidemiological studies. We demonstrate the method using a functional magnetic resonance imaging (fMRI) study of thermal pain where we are interested in determining which brain locations mediate the relationship between the application of a thermal stimulus and self-reported pain.

Keywords directions of mediation, principal components analysis, fMRI, mediation analysis, structural equation models, high-dimensional data