The Principal Direction of Mediation

Oliver Chén, Elizabeth Ogburn, Ciprian Crainiceanu,
Brian Caffo, Martin A. Lindquist

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

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. As a motivating example, consider a functional magnetic resonance imaging (fMRI) study of thermal pain where we are interested in determining which brain measurements (over hundreds of thousands of voxels) mediate the relationship between the application of a thermal stimulus and self-reported pain. To address the problem of high-dimensional mediators in the context of linear SEMs, we propose a framework called the principal direction of mediation (PDM). It is philosophically similar to principal component analysis (PCA), but addresses a fundamentally different problem. The first PDM is the linear combination of the elements of a high-dimensional vector of potential mediators that maximizes the likelihood of the SEM. Like PCA, subsequent directions can thereafter be found that maximizes the likelihood of the SEM conditional on being orthogonal to previous directions. We provide an estimation algorithm and prove some asymptotic properties of the obtained estimates. The efficacy of the approach is illustrated through simulations and an application to data from an fMRI study of thermal pain.

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