Zhou, Ruixuan Rachel; Wang, Liewei; Zhao, Sihai Dave
Estimation and inference for the indirect effect in high-dimensional linear mediation models.
(English) [Zbl 1451.62157]
Biometrika 107, No. 3, 573-589 (2020).

Summary: Mediation analysis is difficult when the number of potential mediators is larger than the sample size. In this paper we propose new inference procedures for the indirect effect in the presence of high-dimensional mediators for linear mediation models. We develop methods for both incomplete mediation, where a direct effect may exist, and complete mediation, where the direct effect is known to be absent. We prove consistency and asymptotic normality of our indirect effect estimators. Under complete mediation, where the indirect effect is equivalent to the total effect, we further prove that our approach gives a more powerful test compared to directly testing for the total effect. We confirm our theoretical results in simulations, as well as in an integrative analysis of gene expression and genotype data from a pharmacogenomic study of drug response. We present a novel analysis of gene sets to understand the molecular mechanisms of drug response, and also identify a genome-wide significant noncoding genetic variant that cannot be detected using standard analysis methods.

MSC:
62P10 Applications of statistics to biology and medical sciences; meta analysis
62R07 Statistical aspects of big data and data science
62M10 Time series, auto-correlation, regression, etc. in statistics (GARCH)

Keywords:
high-dimensional inference; integrative genomics; mediation analysis

Full Text: DOI arXiv