Computational and Information-Theoretic Questions from Causal Inference

Leonard J. Schulman

California Institute of Technology, Pasadena, CA, USA

Abstract

Data, for the most part, is used in order to inform potential interventions: whether by individuals (decisions about education or employment), government (public health, environmental regulation, infrastructure investment) or business. The most common data analysis tools are those which identify correlations among variables – think of regression or of clustering. However, some famous paradoxes illustrate the futility of relying on correlations alone without a model for the causal relationships between variables.

Historically, causality has been teased apart from correlation through controlled experiments. But for a variety of reasons – cost, ethical constraints, or uniqueness of the system – we must often make do with passive observation alone. A theory based upon directed graphical models has been developed over the past three decades, which in some situations, enables statistically defensible causal inference even in the absence of controlled experiments.

Yet "some situations" is rather fewer than one would like. This limitation spurs a range of research questions. In this talk I will describe a couple of causality paradoxes along with how they are captured within the graphical model framework; this will lead naturally toward some of the computational and information-theoretic questions which arise in the theory.

2012 ACM Subject Classification Computing methodologies → Causal reasoning and diagnostics; Computing methodologies → Bayesian network models

Keywords and phrases Causal Inference, Bayesian Networks

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2023.3

Category Invited Talk

Funding *Leonard J. Schulman*: Research supported by NSF CCF-2321079.