RUTGERS
School of Arts and Sciences

Department of Statistics

Armeen Taeb
Department of Statistics
University of Washington

Causal discovery: optimal algorithms and false positive error control
Wednesday, November 6th, 2024
12:00 PM
110 Frelinghuysen Road, Hill Center, Room 116

Zoom Meeting: Meeting ID: 969 0606 4706
Password: 745339

https://rutgers.zoom.us/j/96906064706?pwd=ZklvbExpRVBJQ3c5dUhhYTFuR2ZrZz09

Light refreshments will be served in Hill 452, 11:20am

Abstract:

The goal of causal discovery is to learn causal structure from data. We consider two open questions in causal discovery: i) can we design fast optimization algorithms that yield provably optimal models? and ii) how do we measure and control for false discoveries?

In the first part of the talk, I will describe our work on designing fast and optimal algorithms for causal discovery. Existing algorithms either do not provide optimality guarantees or are computationally challenging to solve, especially for moderately sized models. We consider an L0 penalized maximum likelihood estimator for this problem, which is known to have favorable statistical properties but is intractable to solve. We propose a new coordinate descent algorithm to approximate this estimator. Our algorithm is fast to compute. Furthermore, we prove that despite the non-convexity of the loss function, as the sample size tends to infinity, the objective value of the coordinate descent solution converges to the optimal objective value of the L0 penalized maximum likelihood estimator. Finite-sample statistical consistency guarantees are also established.

In the second part of the talk, I will describe our work on quantifying and controlling for false positive error in complex model spaces, with causal discovery being a prominent example. In model selection problems such as variable selection and graph estimation, models are characterized by an underlying Boolean structure such as the presence or absence of a variable or an edge. Therefore, false positive error or false negative error can be conveniently specified as the number of variables/edges that are incorrectly included or excluded in an estimated model. However, in many modern data analysis tasks—such as causal discovery, clustering, ranking, and others—the underlying decision space is not naturally characterized by a Boolean logical structure, which leads to difficulties in defining and controlling false positive error. We present a generic approach to endow a collection of models with a partial order structure, which facilitates systematic approaches for defining natural generalizations of false positive error and methodologies for controlling this error.

Bio: Armeen is an assistant professor in the Department of Statistics at the University of Washington. Before UW, he was a postdoc at ETH Zurich and received his PhD at Caltech. His research interests lie at the interface of optimization and statistics. His work currently focuses on developing efficient methods for graphical modeling and latent-variable modeling, learning provably optimal causal models from data, domain adaptation, and false positive error control in non-traditional settings.