



香港中文大學統計學系

Department of Statistics

THE CHINESE UNIVERSITY OF HONG KONG

SEMINAR

DEPARTMENT OF STATISTICS
THE CHINESE UNIVERSITY OF HONG KONG

Root Cause Discovery

INVITED SPEAKER

Dr. Li Jinzhou
Postdoctoral Researcher
Department of Statistics
Stanford University

TIME

January 8th, 2025 (Wed) · 2:30 pm - 3:30 pm

Zoom meeting

Zoom ID: 914 3479 0590 · Password: cuhkstat · [Zoom link](#)

ABSTRACT

Although the statistical literature on causality is more focused on forward causal questions about the effects of causes, reverse causal questions about the causes of effects are equally important. In this talk, we address one such reverse causal problem: identifying the disease-causing gene in a patient affected by a monogenic disorder. This problem is an instance of root cause discovery. In particular, we aim to identify the intervened variable in one interventional sample using a set of observational samples as reference. We consider a linear structural equation model where the causal ordering is unknown. We begin by examining a simple method that uses squared z-scores and characterize the conditions under which this method succeeds and fails, showing that it generally cannot identify the root cause. We then prove, without additional assumptions, that the root cause is identifiable even if the causal ordering is not. Two key ingredients of this identifiability result are the use of permutations and the Cholesky decomposition, which allow us to exploit an invariant property across different permutations to discover the root cause. Furthermore, we characterize permutations that yield the correct root cause and, based on this, propose a valid method for root cause discovery. We also adapt this approach to high-dimensional settings. Finally, we evaluate the performance of our methods through simulations and apply the high-dimensional method to discover disease-causing genes in the gene expression dataset.