The Hong Kong University of Science and Technology

Seminar on Business Data Science

Department of ISOM

Testing composite null hypotheses with high-dimensional dependent data:

a computationally scalable FDR-controlling procedure

By

Professor Hongyuan CAO Florida State University Date: 14 April 2025 (Monday) Time: 10:30am – 11:30am Venue: Room 4047 (LSK Business Building)

Abstract

Testing composite null hypotheses can arise in various applications, such as mediation and replicability analyses. The problem becomes more challenging in high-throughput experiments where tens of thousands of features are examined simultaneously. Existing large-scale inference methods for composite null hypothesis testing do not explicitly incorporate the dependence structure, producing overly conservative or overly liberal results. In this work, we first develop a four-state hidden Markov model (HMM) to model a bivariate p-value sequence from replicability analysis with two studies, accounting for local feature dependence and study heterogeneity. Built upon the HMM, we propose a multipletesting procedure to control the false discovery rate (FDR). Extending the HMM to model the p-values from n studies requires a computational cost of exponential order of n. To address this challenge, we introduce a novel e-value framework, which ensures quadratic growth of computational cost with the number of studies while maintaining FDR control. We show that the proposed method asymptotically controls the FDR and exhibits higher power numerically than competing methods at the same FDR level. In a real data application to genome-wide association studies (GWAS), our method reveals new biological insights that existing methods overlook.

Bio

Hongyuan Cao is a professor of statistics at Florida State University. She got her Ph.D. from UNC-Chapel Hill. Her research interests include causal inference, multiple testing, survival analysis, and longitudinal data analysis. She is an elected fellow of ASA.

All interested are welcome!