THE UNIVERSITY OF GEORGIA
DEPARTMENT OF STATISTICS

Colloquium Series

JOINT WITH AI SERIES

Jun S Liu

Department of Statistics, Harvard University

3:50 PM, Thursday, Oct. 21, 2021
Zoom link:
https://zoom.us/j/99986325350?pwd=QUVqdldrMm1OMVNaNzJEai9jZkVTUT09

Data Splitting for Graphical Model Selection With FDR Control

Simultaneously finding multiple influential variables and controlling the false discovery rate (FDR) for statistical and machine learning models is a problem of renewed interest recently. A classical statistical idea is to introduce perturbations and examine their impacts on a statistical procedure. We here explore the use of data splitting (DS) for controlling FDR in learning linear, generalized linear, and graphical models. Our proposed DS procedure simply splits the data into two halves at random, and computes a statistic reflecting the consistency of the two sets of parameter estimates (e.g., regression coefficients). The FDR control can be achieved by taking advantage of such a statistic, which possesses the property that, for any null feature its sampling distribution is symmetric about 0. Furthermore, by repeated sample splitting, we propose *Multiple Data Splitting* (MDS) to stabilize the selection result and boost the power. Interestingly, MDS not only helps overcome the power loss caused by DS with the FDR still under control, but also results in a lower variance for the estimated FDR compared with all other considered methods. DS and MDS are straightforward conceptually, easy to implement algorithmically, and efficient computationally. Simulation results as well as a real data application show that both DS and MDS control the FDR well and MDS is often the most powerful method among all in consideration, especially when the signals are weak and correlations or partial correlations are high among the features. Our preliminary tests on nonlinear models such as generalized linear models and neural networks also show promises. The presentation is based on joint work with Chenguang Dai, Buyu Lin, and Xin Xing.