High-Dimensional Knockoffs Inference for Time Series Data

Jinchi Lv

Department of Data Sciences and Operations, Marshall School of Business, University of Southern California

Abstract

The recently introduced framework of model-X knockoffs provides a flexible tool for exact finite-sample false discovery rate (FDR) control in variable selection in arbitrary dimensions without assuming any dependence structure of the response on covariates. It also completely bypasses the use of conventional p-values, making it especially appealing in high-dimensional nonlinear models. Existing works have focused on the setting of independent and identically distributed observations. Yet time series data is prevalent in practical applications in various fields such as economics and social sciences. This motivates the study of model-X knockoffs inference for time series data. In this paper, we make some initial attempt to establish the theoretical and methodological foundation for the model-X knockoffs inference for time series data. We suggest the method of time series knockoffs inference (TSKI) by exploiting the idea of subsampling to alleviate the difficulty caused by the serial dependence. We establish sufficient conditions under which the original model-X knockoffs inference combined with subsampling still achieves the asymptotic FDR control. Our technical analysis reveals the exact effect of serial dependence on the FDR control. To alleviate the practical concern on the power loss because of reduced sample size cause by subsampling, we exploit the idea of knockoffs with copies and multiple knockoffs. Under fairly general time series model settings, we show that the FDR remains to be controlled asymptotically. To theoretically justify the power of TSKI, we further suggest the new knockoff statistic, the backward elimination ranking (BE) statistic, and show that it enjoys both the sure screening property and controlled FDR in the linear time series model setting. The theoretical results and appealing finite-sample performance of the suggested TSKI method coupled with the BE are illustrated with several simulation examples and an economic inflation forecasting application. This is a joint work with Chien-Ming Chi, Yingying Fan and Ching-Kang Ing.