Asymptotic FDR Control with Model-X Knockoffs: Is Moments Matching Sufficient?

Yingying Fan[1], Lan Gao, Jinchi Lv, Xiaocong Xu[2,*]

Data Sciences and Operations, University of Southern California

ABSTRACT

We propose a unified theoretical framework for studying the robustness of the model-X knockoffs framework by investigating the asymptotic false discovery rate (FDR) control of the practically implemented approximate knockoffs procedure. This procedure deviates from the model-X knockoffs framework by substituting the true covariate distribution with a user-specified distribution that can be learned using in-sample observations. By replacing the distributional exchangeability condition of the model-X knockoff variables with three conditions on the approximate knockoff statistics, we establish that the approximate knockoffs procedure achieves the asymptotic FDR control. Using our unified framework, we further prove that an arguably most popularly used knockoff variable generation method--the Gaussian knockoffs generator based on the first two moments matching--achieves the asymptotic FDR control when the two-moment-based knockoff statistics are employed in the knockoffs inference procedure. For the first time in the literature, our theoretical results justify formally the effectiveness and robustness of the Gaussian knockoffs generator. Simulation and real data examples are conducted to validate the theoretical findings.

Keywords: Model-X knockoffs; Gaussian knockoffs generator; moments matching; asymptotic FDR control; robustness

Detection of Dynamic Instability by Dispersion Ratios in Local Block Lyapunov Exponent Diagrams

Rintaro Ichiya[1], Rinka Sagawa[1], **Yan Liu[2,*]**

[1]*Department of Applied Mathematics, Waseda University*

[2]*Faculty of Science and Engineering, Waseda University*

ABSTRACT

This paper presents statistical methods for quantifying local dynamic instability arising from chaotic behaviors in stochastic processes. We introduce the local block Lyapunov exponent and the diagonal Lyapunov dispersion ratio as fundamental statistical tools to distinguish chaotic behaviors in stochastic processes. The diagonal Lyapunov dispersion ratio is used as a macroscopic measure to investigate the distributional distortion in each block of stochastic processes. We develop the asymptotic theory for these statistical tools under a general setting. Numerical simulations under different parameter settings illustrate the satisfactory performance of our statistical approach. We also apply this method to the financial market data, providing evidence for the possible local dynamic instability in the data.

Keywords: Dynamic instability; Diagonal Lyapunov dispersion ratio; Local block Lyapunov exponent; Nonlinear time series

Variable Selection for High-Dimensional Heteroscedastic Regression and Its Applications

Po-Hsiang Peng[1], Hai-Tang Chiou[2], **Hsueh-Han Huang[3]**, Ching-Kang Ing[1]

[1]*Institute of Statistics and Data Science, National Tsing Hua University, Hsinchu, Taiwan*

[2]*Department of Mathematics, National Chung Cheng University,Chiayi, Taiwan*

[3]*Institute of Statistical Science, Academia Sinica, Taipei, Taiwan*

Abstract

We are examining variable selection in high-dimensional linear heteroscedastic models. Drawing inspiration from the connection between the linear heteroscedastic function and the interaction model, we develop a two-stage algorithm to identify the relevant variables in the model mentioned above. We demonstrate the selection consistency of our proposed two-stage method and highlight its efficacy through numerical simulations. Furthermore, we leverage our method to pinpoint defective tools during the semiconductor manufacturing process.

Keywords: High-dimensional; interaction model; Linear heteroscedasticity; Model selection; Multiplicative heteroscedasticity

Adaptive High-Dimensional Model Selection via Chebyshev's Greedy Algorithm

Chien-Tong Lin[1], **Chi-Shian Dai[2]**, You-Lin Chen[3], Ching-Kang Ing[4]

Department of Applied Mathematics, National Sun Yat-sen University

Department of Statistics and Data Science, National Cheng Kung University Amazon.com, Inc.

Insititute of Statistics and Data Science, National Tsing Hua University

ABSTRACT

Sparsity assumptions on regression coefficients play a central role in high-dimensional model selection. However, the sparsity level is typically unknown in practice, motivating the development of procedures that perform robustly across a range of sparsity regimes. In this paper, we investigate the convergence behavior of Chebyshev's Greedy Algorithm (CGA) under varying sparsity levels and propose selecting the number of CGA iterations using a high-dimensional information criterion (HDIC). We show that the resulting procedure, CGA combined with HDIC (CGA+HDIC), is adaptive in the sense that it automatically achieves the optimal trade-off between variance and squared bias without prior knowledge of the sparsity level. As a key application, we demonstrate that CGA+HDIC attains the optimal convergence rate (up to a logarithmic factor in the sample size) in high-dimensional generalized linear models. Theoretical results are supported by extensive simulation studies and real data analyses.

Keywords: Chebyshev greedy algorithm; generalized linear models; high-dimensional information criterion; sparsity levels