DOCTORAL THESIS

The estimation and inference of complex models

Zhou, Min

Date of Award:
2017

Abstract

In this thesis, we investigate the estimation problem and inference problem for the complex models. Two major categories of complex models are emphasized by us, one is generalized linear models, the other is time series models. For the generalized linear models, we consider one fundamental problem about sure screening for interaction terms in ultra-high dimensional feature space; for time series models, an important model assumption about Markov property is considered by us.

The first part of this thesis illustrates the significant interaction pursuit problem for ultra-high dimensional models with two-way interaction effects. We propose a simple sure screening procedure (SSI) to detect significant interactions between the explanatory variables and the response variable in the high or ultra-high dimensional generalized linear regression models. Sure screening method is a simple, but powerful tool for the first step of feature selection or variable selection for ultra-high dimensional data. We investigate the sure screening properties of the proposal method from theoretical insight. Furthermore, we indicate that our proposed method can control the false discovery rate at a reasonable size, so the regularized variable selection methods can be easily applied to get more accurate feature selection in the following model selection procedures. Moreover, from the viewpoint of computational efficiency, we suggest a much more efficient algorithm-discretized SSI (DSSI) to realize our proposed sure screening method in practice. And we also investigate the properties of these two algorithms SSI and DSSI in simulation studies and apply them to some real data analyses for illustration.

For the second part, our concern is the testing of the Markov property in time series processes. Markovian assumption plays an extremely important role in time series analysis and is also a fundamental assumption in economic and financial models. However, few existing research mainly focused on how to test the Markov properties for the time series processes. Therefore, for the Markovian assumption, we propose a new test procedure to check if the time series with beta-mixing possesses the Markov property. Our test is based on the Conditional Distance Covariance (CDCov). We investigate the theoretical properties of the proposed method. The asymptotic dis-

tribution of the proposed test statistic under the null hypothesis is obtained, and the power of the test procedure under local alternative hypothesizes have been studied. Simulation studies are conducted to demonstrate the finite sample performance of our test.

Keywords: Variable selection; Interaction screening; Log-likelihood functions; Boolean representation; Data discretization; Markov property; Conditional independence; Conditional distance covariance; β-mixing; Complex models; Generalized linear models; Time series models.

Table of Contents