STATISTICS SEMINAR

Dunfu Yang

Preliminary Exam

Wednesday, November 3, 2021
Via Zoom, 2:00 pm (Central Time)

Bayesian approaches to sparse reduced-rank multivariate regression

In multivariate regression analysis, reduced-rank regression (RRR) has received considerable attention. In particular, with the advent of high-dimensional data, variable selection becomes a crucial step to break the curse of high-dimensionality. In this dissertation, we aim to develop novel methods to handle both rank reduction and variable selection simultaneously under various reduced-rank multivariate regression settings with high-dimensional data.

In the first project, we develop a fully Bayesian approach to sparse reduced-rank regression. A major difficulty that occurs in a fully Bayesian framework is that the dimension of parameter space varies with the selected variables and the reduced-rank. Due to the varying-dimensional problems, traditional Markov Chain Monte Carlo (MCMC) methods such as Gibbs sampler and Metropolis-Hastings algorithm are inapplicable in our Bayesian framework. To address this issue, we propose a new posterior computation procedure based on the Laplace approximation within the collapsed Gibbs sampler. A key feature of our fully Bayesian method is that the model uncertainty is automatically integrated out by the proposed MCMC computation. The proposed method is examined via simulation study and real data analysis.

In the second project, we extend our Bayesian method in project 1 to a general framework that can be applied to various multivariate outcomes such as binary, count, or mixed-types. Our motivation is based on the fact that the multiple responses are often correlated and the sparse reduced-rank regression improves predictive accuracy. However, due to the use of the nonlinear link function, the computation complexity becomes extremely cumbersome. To cope with the computational intensity, we propose to use the Markov chain Monte Carlo model composition within the partially collapsed Gibbs sampling scheme. In simulation study, we examine the performance of the proposed method for the problem of multivariate binary responses.