Public Abstract
First Name:Zhenyu
Middle Name:
Last Name:Wang
Adviser's First Name:Sounak
Adviser's Last Name:Chakraborty
Co-Adviser's First Name:(Tony) Jianguo
Co-Adviser's Last Name:Sun
Graduation Term:SS 2014
Department:Statistics
Degree:PhD
Title:Bayesian Non-Linear Methods for Survival Analysis and Structural Equation Models

High dimensional data are more common nowadays, because the collection of such data becomes larger and more complex due to the technology advance of the computer science, biology, etc. The analysis of high dimensional data is different from traditional data analysis, and variable selection for high dimensional data becomes very challenging. Structural equation modeling (SEM) analyzes the relationship between manifest variables and latent variables. The structural equation focuses on analyzing the relationship between latent variables. New proposed methods of these topics are discussed in the dissertation.

In the first chapter, we review the basic concept of survival analysis, SEM, and current method of variable selection in those two scenarios. We also introduce the available software package for current methods and relevant data set.

In the second chapter, we develop a Bayesian kernel machine model with incorporating existing information on pathways and gene networks in the analysis of DNA microarray data. Each pathway is modeled nonparametrically using reproducing kernel Hilbert space. The pathways and the genes are selected via assigning mixture priors on the pathway indicator variable and the gene indicator variable. This approach helped us in flexible modeling of the pathway effects, which can capture both linear and non-linear effect. Moreover, the model can also pinpoint the important pathways and the important active genes within each pathway. We have also developed an efficient Markov Chain Monte Carlo (MCMC) algorithm to fit our model. We used simulations and a real data analysis, [van 't Veer et al., 2002] breast cancer microarray data, to illustrate the proposed method.

In the third chapter, we extend the idea of semiparametric structural equation model where the nonlinear functional relationships are approximated using basis expansions [Guo et al., 2012]. Many basis expansion methods, including cubic splines, are known to induce correlations. In this chapter we compare standard Lasso, Fused Lasso and Elastic Net to account for correlations in both the covariate and basis expansions. To illustrate the usefulness of the proposed methods, a simulation study and a real data study have been performed. The semiparametric structural equation models based on Bayesian fused Lasso and Bayesian elastic-net outperform the Bayesian Lasso model.

In the fourth chapter, we apply Bayesian Graph Laplacian Model, developed by [Liu et al., 2014] and generalized the graph Laplacian allowing both positively and negatively correlated variable, to analyze gene expression data from Michigan prostate cancer study [Dhanasekaran et al., 2001]. We find out the underlie gene network and interaction related to prostate cancer and discuss the possible extensions for Bayesian Graph Laplacian Model, including analyzing multiple pathways simultaneously and pathways selection, right censored data as response variable and binomial or multinomial data as response variable.