

Empirical Likelihood Based Longitudinal Studies

Tharshanna Nadarajah
PhD Candidate

Supervisors: Dr. Asokan Variyath and Dr. J.C. LoredO-Osti

Department of Mathematics and Statistics, Memorial University

October 19, 2015

HH 3017

1:00-2:00 pm

Abstract

In longitudinal data analysis, our primary interest is on the regression parameters for the marginal expectations of the longitudinal responses and the longitudinal correlation parameters being of secondary interest. The joint likelihood function for longitudinal data is challenging, particularly for correlated discrete outcome data. In such situations, the marginal models like Generalized Estimating Equations (GEEs) has received much attention in the longitudinal regression setup. These methods assume the estimates of the first two moments of the data and working correlation structure. In this approach, the confidence regions and test of hypothesis are constructed based on asymptotic normality. In addition, these methods are at the risk of model mis-specification such as variance function and working correlation structure. Such mis-specification can cause these estimates to be inefficient, inconsistent, and inferences based on the wrong assumption can be completely inappropriate. To overcome this problem, we proposed an empirical likelihood procedure based on a set of estimating equations of parameter of interest. Then we extended to penalized empirical likelihood variable selection for high-dimensional longitudinal data with large number of covariates. Identification of a sub-model that adequately represents the data is necessary for easy interpretation. Also inclusion of redundant variables in the model may leads to hinder accuracy and efficiency for estimation and inference. In such situation, we propose penalized empirical likelihood variable selection based on generalized estimating equations (GEEs). In this approach, the variable selection and the estimation of the coefficients are carried out simultaneously. Our performance analysis show that when model assumptions are true, its performance is comparable to that of the existing methods and when the model is mis-specified, our proposed method has clear advantages over the existing methods. We have implemented the proposed method in two real case examples.