Surrogate Variable Analysis for Multivariate Functional Responses in Imaging Data

Chao Huang
Department of Biostatistics
University of North Carolina at Chapel Hill

With the rapid growth of modern technology, many large-scale biomedical studies, e.g., Alzheimer's disease neuroimaging initiative (ADNI) study, have been conducted to collect massive datasets with large volumes of complex information from increasingly large cohorts. Despite the numerous successes of biomedical studies, the imaging heterogeneity has posed many challenges in both data integration and disease etiology. Specifically, imaging heterogeneity often represents at three different levels: subject level, group level, and study level. This talk mainly focuses on the heterogeneity at study level. The study-level heterogeneity can result from the difference in study environment, population, design, and protocols, which are mostly unknown. Surrogate variable analysis (SVA), which is a powerful tool in tackling this heterogeneity, has been widely used in genomic studies. However, the imaging data is usually represented as functional phenotype while no existing SVA procedures work for functional responses. To address these challenges, a functional latent factor regression model (FLFRM) is proposed to handle the unknown factors. Several inference procedures are established for estimating the unknown parameters and detecting the latent factors. The consistency of estimate of latent variables and the weak convergence of estimate of parameters are systematically investigated. The finite-sample performance of proposed procedures is assessed by Monte Carlo simulations and a real data example on hippocampal surface data from ADNI study.