PROJECT TITLE: Bayesian hidden Markov models for dependent large-scale hypotheses testing

Xia Wang  
Division of Statistics and Data Science  
Department of Mathematical Sciences  
College of Arts and Sciences  
4428E French Hall West  
Cincinnati, OH 45221  
Phone: 513-556-3295  
E-mail: xia.wang@uc.edu

Project Description

1. area of the research: Statistics  
2. research tasks the student will be performing: Simulation studies using the R package; summarize and report the simulation results by tables/figures  
3. training that the mentor will provide to the WISE student: (1) Introduction to Bayesian methods, Hidden Markov Model, Multiple Hypotheses testing; (2) Use of R and the R package; (3) data visualization and results reporting.  
4. specific requirements, if any, that the mentor expects the student to meet: strongly motivated and interested in the topics

Project Description:

Large-scale hypotheses testing with dependence is common in many application areas, such as engineering, disease monitoring, genomics and brain imaging. Early research in multiple hypotheses testing and false discovery rate (FDR) control largely ignored the dependence structure among hypotheses. This ignorance has led to the problem of degrading statistical accuracy and high variability of testing results and hence irreproducibility of scientific findings. Current attention has shifted to developing multiple comparison procedures for dependent hypotheses.

We developed an optimal and flexible hypotheses testing procedure for large-scale, dependent data through the hidden Markov model (HMM) under the Bayesian framework. An R package has been developed for the normal time series data as well as the Poisson count data. We would like to further work on multiple extensions of the current model, which includes (1) development of multiple hypotheses testing methods with a Bayesian Markov dynamic
switching regression model, which is useful when dynamic features present in the hypothesis testing problem; (2) development of hypotheses testing methods for spatiotemporal data; (3) theoretical investigation on the optimality and posterior consistency of the Bayesian methods. Large scale simulation studies using the R package developed would be the tool to investigate the performance of the proposed model. The model will also be applied to real data examples to obtain statistical inference on key parameters.