STATISTICS SEMINAR

Jin Liu, Ph.D.
University of Chicago

Thursday, October 16, 2014
Dickens Hall, Room 207, 4:00-5:00 pm
Refreshments: Dickens 108, 3:30-4:00 pm

An Empirical Bayesian Approach to Integrate GWAS with Gene Expressions from Multiple "Tissues"

Abstract: To date, a large number of genome-wide association studies (GWAS) have been conducted. Most of them used multiple testing corrections to adjust for single-marker p-values while penalized regressions have been implemented in GWAS to conduct variable selection and parameter estimation simultaneously. With advancement of array techniques, there are a large number of genomic data from multiple sources: e.g., gene expression data from multiple tissues. The integration of gene expression from multiple tissues with GWAS can increase the statistical power of a single GWAS. Traditional frequentist methods lack the ability to consider the hierarchical structures among different data sources while the Bayesian methods are too computational intensive in genomic studies. We propose to use empirical Bayes as a framework to model the hidden variables, which indicate whether a gene-level p-value is from null or non-null (associated with trait), together with gene expression data from limited number of tissues. Expectation-Maximization (EM) algorithm is implemented to optimize the corresponding complete log-likelihood function. These methods can jointly analyze two or more GWAS at the same time to test for the "pleiotropic" effects. To integrate multiple tissues, we propose a three-stage strategy to transform gene expression from multiple tissues to the new predictor with much lower dimension and the proposed model with this new predictor. An efficient algorithm based on EM is developed. Meanwhile, we estimate the corresponding local false discovery rate (FDR) and formulate the hypothesis testing for "pleiotropy" and identification of associated tissues. Simulation studies are used to evaluate finite sample performance. We make comparison under different level of "pleiotropy" using generative model. Rheumatoid arthritis and type-1 diabetes from the Wellcome Trust Case Control Consortium (WTCCC) together with gene expression data from multiple tissues are analyzed using the proposed approach.