The Chinese University of Hong Kong
Department of Statistics

Seminar

Statistical Methods for PheWAS using GWAS Summary Statistics and Large-scale Signal Detection for Mediation Effect in EWAS

By

Dr. LIU, Zhonghua
Department of Biostatistics
Harvard University

Abstract

Recently, detailed phenotype data from epidemiological studies, electronic health records (EHR), genome-wide omics profiling and real-time mobile health devices are becoming rapidly available, there is an increasing interest in identifying cross-phenotype associations, which holds great potentials for novel drug target discovery and precision medicine. The accumulation of massive GWAS results provides a cost-effective way for conducting cross-phenotype analysis via the phenome-wide association study (PheWAS) approach. We proposed a bundle of principal component association tests for discovering genetic variants associated with multiple phenotypes using GWAS summary statistics. The powers of those PC based tests can be well explained from a geometric perspective by the introduction of a novel concept called principal angle. All the p-values for those tests can be calculated analytically and have been implemented in a publicly available R package MPAT. An application of the PC based tests to the metabolic syndrome data sets collected from four international consortia identified promising novel genetic variants for future functional studies. It is often of scientific interest in assessing the mediation effects of a large number of mediators, e.g., DNA methylations in epigenome-wide association studies (EWAS), that lie in the causal pathway of an exposure on a clinical outcome. Testing for the mediation effect is statistically challenged by the fact that the null hypothesis is composite. We show that the standard mediation analysis tests using the maximum p-value method and the Wald test based on the product method fail and give too conservative tests in genome-wide mediation analysis when one needs to test for a large number of composite null hypotheses. We propose a divide-aggregate test (DAT) for assessing mediation for genome-wide epigenetic studies. We show that this composite testing procedure performs much better than existing methods for genome-wide epigenetic studies where the signals are usually very sparse. A fast Monte Carlo (MC) correction is also proposed when DAT indicates slight conservativeness. Simulation studies were conducted to evaluate the type I error rates and powers under a range of practical settings. An application to the Normative Aging Study (NAS) identified putative DNA methylation CpG sites as mediators in the causal pathway from smoking behavior to lung functions.

Date: February 6, 2018 (Tuesday)
Time: 2:30 p.m. - 3:30 p.m.
Venue: T.Y.Wong Hall (TYW) LT
The Chinese University of Hong Kong