

Genetic Association Test Based on Nonparametric Stratification of Propensity Scores

Yaji Xu

Yale University, New Haven, CT, USA yaji.xu@yale.edu

Chi Song

Yale University, New Haven, CT, USA chi.Song@yale.edu

Yuan Jiang

Oregon State University, Oregon, USA yuan.jiang@stat.oregonstate.edu

Heping Zhang*

Yale University, New Haven, CT, USA heping.zhang@yale.edu

Adjusting for covariate effects is challenging for genetic association studies. In this presentation, we introduce a combination of tree-based methods and propensity scores. Propensity scores are usually estimated using parametric models such as logistic regression and then used as a matching or stratification criterion in an observational study. However, we propose tree models and use propensity scores to form stratifications for genetic association studies that may involve many confounding factors including population stratification and environmental variables, which can be misinterpreted as genetic effects and/or affect the genetic association signals. The terminal nodes of a classification tree serve naturally the strata of propensity scores. Mantel-Haenszel test is then employed to detect the disease associations after the stratification procedure. Simulation studies show that the tree-based stratification method is more robust than the parametric approach especially when numbers of confounding covariates are presented and the minor allele frequency at the disease locus is low. We applied our method to Genetic Association Information Network-Major Depressive Disorder (GAIN-MDD) data and revealed genes associated with MDD.

Key Words: covariate adjustment, complex diseases, genetic association analysis, case control study