Statistical methods for testing genetic effects in the presence of possible gene-gene and gene-environment interactions

Arnab Maity
Department of Biostatistics
Harvard School of Public Health

Abstract

In genome-wide association studies and epigenetic studies, it is of significant interest to develop powerful and robust tests for the genetic effect on an outcome in the presence of possible gene-gene and gene-environment interactions. We develop two such testing procedures. We first consider the case with a gene that possibly interacts with a single gene or a single environmental factor in a semiparametric model, where the effect of the latter is modeled nonparametrically. We consider the Tukey-type interaction formulation in a semiparametric regression model to increase the power. We show that the score-test parallel to that derived under the parametric setting is biased and requires undersmoothing of the nonparametric component for the test to be valid. Moreover, in the presence of repeated outcomes, the asymptotic distribution of the score test statistic depends on the estimation of functions that are defined as solutions of integral equations making implementation difficult and computationally taxing. We develop profiled score statistics which are unbiased, asymptotically efficient and can be easily implemented. We next consider the case with a gene that possibly interacts with a set of genes (e.g., genes in a pathway) or multiple correlated environmental exposures (e.g., different metal exposures or pollutants). We develop a general powerful testing procedure within the kernel machine framework. Specifically, our test is based on a garrote kernel method and is constructed as a score test. The key features of the proposed test is that it is flexible and developed for both parametric and nonparametric models within a unified framework, it accounts for the correlation among genes and environmental exposures, and often uses much smaller degrees of freedom resulting in a more powerful test than the standard test. We investigate the theoretical properties of the proposed testing procedures and present simulation studies to evaluate type-I error and power of the proposed methods. We demonstrate these procedures by applying them to the human DNA methylation study data on 140 subjects of the Normative Aging Study (NAS) cohort.