Robust Powerful Methods for Understanding Gene-Gene and Gene-Environment Interactions

Raymond J. Carroll
Distinguished Professor of Statistics, Nutrition and Toxicology
Department of Statistics
Texas A&M University

Abstract

We consider population-based case-control studies of gene-environment and gene-gene interactions using prospective logistic regression models. Data sets like this arise when studying pathways based on haplotypes as well as in multistage genome wide association studies (GWAS). In a typical case-control study, logistic regression is used and there is little power for detecting interactions. However, in many cases it is reasonable to assume that, for example, genotype and environment are independent in the population, possibly conditional on factors to account for population stratification. In such as case, we have developed an extremely statistically powerful semiparametric approach for this problem, showing that it leads to much more efficient estimates of gene-environment interaction parameters and the gene main effect than the standard approach: decreases of standard errors for the former are often by factors of 50% and more. The issue of course that arises is the very assumption of conditional independence, because if that assumption is violated, biases result so that one can announce gene-environment interactions or gene effects even though they do not exist. We will describe a simple, computationally fast approach for gaining robustness without losing statistical power, one based on the idea of Empirical Bayes methodology. Examples to colorectal adenoma studies of the NAT2 gene and prostate cancer in the VDR pathway are described to illustrate the approaches.