

Case-Control Studies of Genetic and Environmental Factors with Errors in Measurement,
Missing Genetic Information and Haplotype-phase Ambiguity

Iryna V. Lobach

PhD Candidate

Texas A&M University

Abstract:

It is widely believed that risks of many complex diseases are determined by genetic susceptibilities, including environmental exposures, and their interaction. Chatterjee and Carroll (2005) have recently developed an efficient retrospective maximum-likelihood method for analysis of case-control studies that exploits an assumption of gene-environment independence and leaves the distribution of the environmental covariates to be completely non-parametric. We generalize the semiparametric maximum-likelihood approach to situations when some of the environmental covariates are measured with error, and allow genetic information is missing on some subjects, e.g., unphased haplotypes.

Profile likelihood techniques and an EM algorithm are developed, resulting in a relatively simple procedure for parameter estimation. We prove consistency and derive the resulting asymptotic covariance matrix of parameter estimates when variance of measurement error is known and when it is estimated using replications. The performance of the proposed method is illustrated using simulation studies emphasizing the case when genetic information is in the form of a haplotype and missing data arises from haplotype-phase ambiguity and missing genetic data.

Inference is performed via a likelihood-ratio type procedure, one that we show has better small-sample performance than Wald-type inferences. An application of this method is illustrated using a case-control study of an association of calcium intake with early stages of colorectal tumor development.