Cornell University Library

We gratefully acknowledge support from the Simons Foundation and member institutions

arXiv.org > stat > arXiv:1305.5695

Search or Article-id

(Help | Advanced search)

All papers ▼ | Go!

Statistics > Applications

Bayesian semiparametric analysis for two-phase studies of gene-environment interaction

Jaeil Ahn, Bhramar Mukherjee, Stephen B. Gruber, Malay Ghosh

(Submitted on 24 May 2013)

The two-phase sampling design is a cost-efficient way of collecting expensive covariate information on a judiciously selected subsample. It is natural to apply such a strategy for collecting genetic data in a subsample enriched for exposure to environmental factors for gene-environment interaction (G x E) analysis. In this paper, we consider two-phase studies of G x E interaction where phase I data are available on exposure, covariates and disease status. Stratified sampling is done to prioritize individuals for genotyping at phase II conditional on disease and exposure. We consider a Bayesian analysis based on the joint retrospective likelihood of phases I and II data. We address several important statistical issues: (i) we consider a model with multiple genes, environmental factors and their pairwise interactions. We employ a Bayesian variable selection algorithm to reduce the dimensionality of this potentially high-dimensional model; (ii) we use the assumption of gene-gene and gene-environment independence to trade off between bias and efficiency for estimating the interaction parameters through use of hierarchical priors reflecting this assumption; (iii) we posit a flexible model for the joint distribution of the phase I categorical variables using the nonparametric Bayes construction of Dunson and Xing [J. Amer. Statist. Assoc. 104 (2009) 1042-1051].

Submission history

From: Jaeil Ahn [view email]

Which authors of this paper are endorsers?