Yale SCHOOL OF PUBLIC HEALTH

Biostatistics

Presentation

"BAYESIAN SEMIPARAMETRIC ANALYSIS FOR TWO-PHASE STUDIES OF GENE-ENVIRONMENT INTERACTION"

Jaeil Ahn, PhD

Department of Bioinformatics and Computational Biology

MD Anderson Cancer Center

ABSTRACT

The two-phase sampling design is a cost-efficient way of collecting expensive covariate information on a judiciously selected sub-sample. It is natural to apply such a strategy for collecting genetic data in a sub-sample enriched for exposure to environmental factors for gene-environment interaction (G x E) analysis. In this talk, I will describe Bayesian semiparametric two-phase studies of G x E interaction where phase I data is available on exposure, covariates and disease status and stratified sampling is done to prioritize individuals for genotyping at phase II. I will address several important statistical issues: (i) a model with multiple genes and environmental factors and their pairwise interactions; (ii) 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) a flexible model for the joint distribution of the phase I categorical variables using the non-parametric Bayesian construction. I will present the results from simulation study to compare the proposed Bayesian method with other standard choices for analyzing two-phase data. I will also show the application of the method to an ongoing case-

control study of colorectal cancer, where the goal is to explore the interaction between the use of statins and genetic markers in the lipid metabolism/cholesterol synthesis pathway.

Thursday, February 7, 2013
60 College Street LEPH RM 101 11:30am-1:00pm