THE UNIVERSITY OF CHICAGO

Department of Statistics

DISSERTATION PRESENTATION AND DEFENSE

MIAOYAN WANG

Department of Statistics
The University of Chicago

Mixed-Model and Quasi-Likelihood Methods for Genetic Association Studies in Samples with Related Individuals and Population Structure

WEDNESDAY, July 1, 2015, at 2:00 PM
Eckhart 110, 5734 S. University Avenue

ABSTRACT

We develop statistical methods for two problems in genetic association analysis. At the heart of our methods are approaches to addressing relatedness and hidden sample structure. The first part of the thesis is concerned with the selection of an optimal subset of individuals, chosen from a set of pedigrees, to genotype for a genetic association study. We consider samples that include arbitrarily related individuals, with the kinship matrix assumed known. Dependence is an important feature of this type of family data --- dependence not only between genotypes and phenotypes, but also among genotypes and among phenotypes. The retrospective version of a linear mixed model provides a natural way to incorporate partial information by making use of this dependence when both phenotype and pedigree information are available for a larger set of individuals from the same pedigree. We propose G-STRATEGY, which uses simulated annealing to maximize the non-centrality parameter of the quasi-score test for association in the retrospective model. We demonstrate that G-STRATEGY compares very favorably to existing methods and achieves robustness of power against a wide range of alternative models.

The second part of the thesis focuses on joint genetic mapping studies involving two interactive organisms, e.g.,\ a highly coupled host-pathogen pair. Existing approaches to association mapping typically consider a single organism type and aim to identify genes on the genome that are statistically associated with the phenotype of interest. However, in a host-pathogen interactive system, the response (e.g.,\ infection) often depends on the specific pairing of host and pathogen. Our solution to this problem is to develop a two-way mixed model to account for the background genetic heterogeneity due to the polygenic contributions from both organisms and their inter-genome interaction. We use empirical genetic relatedness matrices for both organisms to account for hidden sample structure, such as population stratification, admixture and/or cryptic relatedness. We propose various score tests, depending on the specific goal, for assessing genetic association effects of individual SNPs and/or SNP pairs.

For information about building access for persons with disabilities, please contact Laura Rigazzi at 773.702-0541 or send an email to lrigazzi@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: https://lists.uchicago.edu/web/arc/statseminars.