TIMOTHY THORNTON
Department of Statistics
The University of Chicago

"Statistical Inference for Genetic Analysis in Related Individuals"

Wednesday, November 9, 2005 at 4:00 PM
110 Eckhart Hall, 5734 S. University Avenue

ABSTRACT

Case-control studies have been extremely valuable in evaluating associations between candidate genes and complex diseases. Traditional case-control studies use unrelated subjects and compare allele or genotype frequencies of the cases and the controls at genetic markers. When affected related individuals are used in association studies, the power to detect an association is increased since affecteds with affected relatives have a higher expected frequency of the alleles that increase susceptibility for a genetic trait than do affected individuals that do not have affected relatives. When related individuals are used in a study, the correlations among the relatives must be taken into account to ensure validity of the test, and consideration of these correlations can also improve power. This thesis focuses on the problem of testing for association between a binary trait and a genetic marker when cases and/or controls are related. The emphasis is on methods that are computationally feasible even for hundreds of thousands of markers and/or extremely complex pedigrees.

Case-control association testing is essentially a comparison of the allele frequency distribution between cases and controls. We first consider the related problem of allele frequency estimation, which is important in its own right because many genetic mapping and population genetic analyses require allele frequency estimates. We propose a new estimator, which is an extension of the best linear unbiased estimator of McPeek et al. (2004), and which uses a random covariance matrix that depends on the observed genetic data at loci that are linked to the locus of interest to increase efficiency. Our estimator is unbiased if the linked loci are in linkage equilibrium with the locus of interest and it is not as computationally intensive as the MLE.

To test for allelic association with a binary trait, we propose a new test, the M_{QLS} test, that is an extension of the quasi-likelihood score test of Bourgain et al. (2003) and which takes advantage of the fact that affected individuals that have affected relatives are more likely to have the predisposing variants than individuals that do not have affected relatives. One of the motivations for using the M_{QLS} test is that for any arbitrary set of outbred individuals, the M_{QLS} statistic has maximal noncentrality parameter in a general class of linear statistics, for all 2 allele disease models, as the effect size tends to zero. We perform simulations to compare the type I error and the power of the M_{QLS} and competing methods. We apply the methods to analyze data on asthma-related phenotypes in a complex Hutterite pedigree (in collaboration with Dr. Carole Ober) and an alcoholism related phenotype in a sample of moderate size outbred Caucasian pedigrees from GAW 14 COGA data.

The M_{QLS}, the W_{QLS}, and the $W_{\chi^2_{corr}}$ are examples of allelic tests, which essentially treat the alleles, instead of the individual, as the sampling unit. These tests are valid under the assumption of Hardy-Weinberg equilibrium (HWE) in the population from which the pedigree founders are drawn. We introduce a new genotypic test for association which is an extension of the genotypic test of X. Wu (unpublished Master's thesis). Theses genotypic methods treat the individual as the sampling unit and are valid in the absence of HWE. Simulations are performed to compare the power and type I error of the new test with those of competitors. In collaboration with Dr. Xiaodong Wu, we apply the methods to test for association between hypertension and markers at a candidate gene in a sample of moderate-size African-American pedigrees.