THE UNIVERSITY OF CHICAGO

Department of Statistics
DISSERTATION PRESENTATION AND DEFENSE

Joelle Mbatchou

Department of Statistics
The University of Chicago

Assessments of significance for genetic association analysis in structured samples.

Monday, April 01, 2019, at 10:30 AM
Room 304, 5747 S. Ellis Avenue

ABSTRACT

In this dissertation, we develop methods to address several problems that arise in the assessment of significance for genetic association analysis of complex traits in samples with population structure and/or relatedness. In Chapter 2, we focus on phenotype resampling methods for binary trait analysis in structured samples. We develop BRASS, a permutation-based approach to testing association between a binary trait and an arbitrary predictor in samples with population structure and/or related individuals. This method is applicable in various contexts, including (1) correction for multiple comparisons when testing for region-wide or genome-wide significance and (2) assessment of significance for tests that combine test statistics that perform well in different scenarios. Previous methods are applicable only to analysis of a quantitative trait and do not perform well for a binary trait. BRASS allows for covariates, ascertainment and simultaneous testing of multiple markers, and it does not place strong restrictions on the test statistic used. We use an estimating equation approach that can be viewed as a hybrid of logistic regression and linear mixed-effects model methods, and we use a combination of principal components and a genetic relatedness matrix to account for sample structure. In simulation studies, we demonstrate that BRASS maintains correct control of type 1 error. We illustrate the proposed approach in two genome-wide analyses of binary traits in domestic dog.

In Chapter 3, we focus on assessment of significance in genetic association analysis of single or multi-dimensional phenotypes, including high-dimensional phenotypes, where we consider test statistics of a certain form, allow association to be tested with either a single genetic marker or with multiple genetic markers simultaneously, and where there is population structure and/or relatedness in the sample. Existing approaches that can be used to assess significance in this context are either computationally burdensome (permutation-based approaches), or do not perform well in settings such as small samples, high-dimensional traits, or misspecified phenotype model (asymptotic approximations based on prospective models), or require unrealistically accurate estimation of the correlation structure among a set of rare variants in order to be applied with rare variants (retrospective methods), or require an assumption of second-order exchangeability of individuals' genotypes, possibly after correction for a few ancestry-informative covariates (existing moment-matching methods for detecting association of two matrices). We develop JASPER, which can be viewed as an extension of existing moment-matching methods for detecting association of two matrices, to allow very general population structure and relatedness in the sample. JASPER can be used for a reasonably broad class of test statistics currently used in genetic association analysis, including most linear mixed model-based score tests and kernel-based test statistics. Notable features of JASPER are that it (1) is insensitive to misspecification of the phenotype model, (2) does not require knowledge of the distribution of the test statistic under the null hypothesis, (3) allows population structure, related individuals, covariates, ascertainment, rare variants, and multiple traits, and (4) with rare variant mapping, it does not require knowledge of the correlation structure among the rare variants. Through simulation studies for a range of settings and test statistics, including for high-dimensional traits, we demonstrate that JASPER properly controls type 1 error in the presence of sample structure and can in some cases provide substantial power gains compared to large-sample-based assessments of significance. JASPER is applied in a study of the genetic regulation of gene expression levels within biological pathways in data from the Framingham Heart Study.