

Colloquium
Michigan State University
Department of Statistics and Probability

Fred Wright
University of North Carolina at Chapel Hill
Department of Biostatistics

Useful resampling methods you've probably never used

Tuesday, October 30, 2012
10:20 a.m. - 11:10 a.m.
Refreshments 10:00am
C405 Wells Hall

Abstract

We describe several methods that involve resampling (primarily permutation), and are useful for high-dimensional multiple testing problems. First we introduce moment-corrected correlation (MCC), a simple approach to produce accurate parametric approximations to permutations of trend tests, in which two vectors are compared for evidence of association. We show that most popular trend tests, including those based on likelihood ratios, are one-to-one with unique values of the Pearson correlation coefficient, provided the outcomes are divided by the correlation sign. MCC provides values that are quickly computed, and comparable or superior in accuracy to those based on the saddlepoint.

Next, we describe safeExpress, an approach and software package for gene-set enrichment testing that provides fast approximations to permutation, for both so-called self-contained and competitive testing. Simulations and real data support the accuracy and power of the approach.

Finally, we describe the DiNAMIC approach to identify recurrent tumor copy number aberrations. The data structure involves a comparison of rows of a matrix, for which coincident occurrence of high or low value is indicative of an aberration. Previous work has been somewhat unclear on the nature of the hypothesis to be tested. Using simple stationarity arguments, we provide support for the use of cyclic shift permutation in this context, with high-dimension, low sample size convergence results. Time permitting, we will also describe a centered bootstrapping approach that enables testing when stationarity assumptions are unreasonable.

(Joint work with Yi-Hui Zhou, Vonn Walter, and Andrew Nobel)