Statistical tests in Gene Set Analysis

Zhide Fang, Phd

Biostatistics Program, School of Public Health LSU Health Sciences Center

Abstract

Gene set analysis is widely used to facilitate biological interpretations in the analyses of differential expression from high throughput profiling data. Wilcoxon Rank-Sum (WRS) test is one of the commonly used methods in gene set enrichment analysis because it is easy to implement and it eliminates the dichotomization of genes into significant and non-significant in a competitive hypothesis testing. Due to the large number of genes being examined, it is impractical to calculate the exact null distribution for the WRS test. Therefore, the normal distribution is commonly used as an approximation. However, the normal approximation is problematic when a gene set with relative small number of genes is tested against the large number of genes in the complementary set. In this situation, a uniform approximation is substantially more powerful, more accurate, and less intensive in computation. We demonstrate the advantage of the uniform approximations in Gene Ontology (GO) term analysis using simulations and real data sets.