False Discovery Control Under Arbitrary Dependence

Xu Han

Department of Statistics
Temple University

Multiple hypothesis testing is a fundamental problem in high dimensional inference, with wide applications in many scientific fields. In genome-wide association studies, tens of thousands of hypotheses are tested simultaneously to find if any genes are associated with some traits; in finance, thousands of tests are performed to see which fund managers have winning ability. In practice, these tests are correlated. False discovery control under arbitrary covariance dependence is a very challenging and important open problem in the modern research.

We propose a new methodology based on principal factor approximation, which successfully extracts the common dependence and weakens significantly the correlation structure, to deal with an arbitrary dependence structure. We derive the theoretical distribution for false discovery proportion (FDP) in large scale multiple testing when a common threshold is used for rejection, and provide a consistent estimate of FDP. Specifically, we decompose the test statistics into an approximate multifactor model with weakly dependent errors, derive the factor loadings and estimate the unobserved but realized factors which account for the dependence by L_1-regression. Asymptotic theory is derived to justify the consistency of our proposed method. This result has important applications in controlling FDR and FDP. The finite sample performance of our procedure is critically evaluated by various simulation studies. Our estimate of FDP compares favorably with Efron (2007)'s approach, as demonstrated by in the simulated examples. Our approach is further illustrated by some real data in genome-wide association studies.

This is joint work with Professor Jianqing Fan and Dr. Weijie Gu at Princeton University.