Phase Transition and Regularized Bootstrap in Large-scale t-tests with False Discovery Rate Control

Weidong Liu[*]and Qi-Man Shao[†]

April 15, 2014

Abstract

Applying the Benjamini and Hochberg (B-H) method to multiple Student's t tests is a popular technique for gene selection in microarray data analysis. Because of the non-normality of the population, the true p-values of the hypothesis tests are typically unknown. Hence, it is common to use the standard normal distribution $N(0,1)$, Student's t distribution t_{n-1} or the bootstrap method to estimate the p-values. In this paper, we prove that, when the population has the finite 4-th moment and the dimension m and the sample size n satisfy $\log m = o(n^{1/3})$, the B-H method controls the false discovery rate (FDR) and the false discovery proportion (FDP) at a given level α asymptotically with p-values estimated from $N(0,1)$ or t_{n-1} distribution. However, a phase transition phenomenon occurs when $\log m \geq c_0 n^{1/3}$. In this case, the FDR as well as the FDP of B-H method may be larger than α or even tend to one. In contrast, the bootstrap calibration is accurate for $\log m = o(n^{1/2})$ as long as the underlying distribution has the sub-Gaussian tails. However, such light tailed condition can not be weakened in general. The simulation study shows that for the heavy tailed distributions, the bootstrap calibration is very conservative. In order to solve this problem, a regularized bootstrap correction is proposed and is shown to be robust to the tails of the distributions. The simulation study shows that the regularized bootstrap method performs better than the usual bootstrap method.

[*]Department of Mathematics and Institute of Natural Sciences, Shanghai Jiao Tong University Email: weidongl@sjtu.edu.cn

[†]Department of Statistics, The Chinese University of Hong Kong.