Whole Data Approaches to Large-Scale Multiple Hypothesis Testing

John Storey

Department of molecular biology
University of Princeton
jstorey@princeton.edu

A number of modern scientific problems involve simultaneously testing thousands or millions of hypotheses. Methodologies for these problems have typically been developed in the context of p-values or test-statistics, which are one-dimensional summaries of each hypothesis test's data. I will present some of our recent developments where the approaches are instead formulated at the level of the whole data set. This allows information shared across hypothesis tests to be directly captured and utilized. Specifically, solutions to multiple testing dependence and optimal frequentist multiple testing will be presented, both of which require careful computational considerations.