Multiple Testing for Dependent Data

John Storey
Princeton University

We develop a framework for performing large-scale significance testing in the presence of arbitrarily strong dependence. We derive a low-dimensional set of random vectors, called a dependence kernel, that fully captures the dependence structure in an observed high-dimensional data set. This represents a surprising reversal of the curse of dimensionality in high-dimensional hypothesis testing. We then show theoretically that conditioning on a dependence kernel is sufficient to render statistical tests independent regardless of the level of dependence in the observed data. This solution to multiple testing dependence has implications in a variety of popular testing problems, such as in gene expression studies, brain imaging, and spatial epidemiology.