RINA BARBER
Department of Statistics
University of Chicago

Inference in High Dimensions with the Knockoff Filter

SATURDAY, MAY 14, 2016, at 9:30 AM
Max Palevsky Cinema at Ida Noyes Hall, 1212 E. 59th Street

ABSTRACT

We develop a framework for testing for associations in a possibly high-dimensional linear model where the number of features $p$ may far exceed the sample size $n$. Our goal is to select a model, i.e. a set of features, while controlling the proportion of false discoveries and sign errors in our model. For our method, the observations are split into two groups, where the first group is used to screen for a set of potentially relevant variables, whereas the second is used for inference over this reduced set of variables; we also develop strategies for leveraging information from the first part of the data at the inference step for greater accuracy. The inference step is carried out by applying the knockoff filter, which creates a knockoff copy—a fake variable serving as a control—for each screened variable. We prove that this procedure controls the directional false discovery rate (FDR) in the reduced model that controls for the effects of all screened variables; this says that our high-dimensional knockoff procedure ‘discovers’ important variables as well as the directions (signs) of their effects, in such a way that the expected proportion of wrongly chosen signs is low. This result is non-asymptotic, and holds regardless of sparsity level, effect size, or correlation structure among the features. We apply our method to a genome-wide association study to find locations on the genome that are possibly associated with cholesterol levels.

This work is joint with Emmanuel Candes.