RELAXING THE ASSUMPTIONS OF KNOCKOFFS BY CONDITIONING

Dongming Huang

Harvard University
E-mail: dhuang01@g.harvard.edu

Abstract: The recent paper Candes et al. (2018) introduced model-X knockoffs, a method for variable selection that provably and non-asymptotically controls the false discovery rate with no restrictions or assumptions on the dimensionality of the data or the conditional distribution of the response given the covariates. The one requirement for the procedure is that the covariate samples are drawn independently and identically from a precisely-known (but arbitrary) distribution. The present paper shows that the exact same guarantees can be made without knowing the covariate distribution fully, but instead knowing it only up to a parametric model. Although this idea is simple, even in Gaussian models conditioning on a sufficient statistic leads to a distribution supported on a set of zero Lebesgue measure, requiring techniques from topological measure theory to establish valid algorithms. We demonstrate how to do this for three models of interest, with simulations showing the new approach remains powerful under the weaker assumptions.