



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS

Statistics Colloquium

ZHIMEI REN

Department of Statistics
University of Chicago

“Stable Variable Selection with Knockoffs”

MONDAY, JANUARY 23, 2023, at 2:00 PM
Jones 303, 5747 S. Ellis Avenue

ABSTRACT

A common problem in many modern statistical applications is to find a set of important variables—from a pool of many candidates—that explain the response of interest. For this task, model-X knockoffs offers a general framework that can leverage any feature importance measure to produce a variable selection algorithm: it discovers true effects while rigorously controlling the number or fraction of false positives, paving the way for reproducible scientific discoveries. The model-X knockoffs, however, is a randomized procedure that relies on the one-time construction of synthetic (random) variables. Different runs of model-X knockoffs on the same dataset often result in different sets of selected variables, which is not desirable for the reproducibility of the reported results.

In this talk, I will introduce derandomization schemes that aggregate the selection results across multiple runs of the knockoffs algorithm to yield stable selection. In the first part, I will present a derandomization scheme that controls the number of false positives, i.e., the per family error rate (PFER) and the k family-wise error rate (k -FWER). In the second part, I will talk about an alternative derandomization scheme with provable false discovery rate (FDR) control. Equipped with these derandomization steps, the knockoffs framework provides a powerful tool for making reproducible scientific discoveries. The proposed methods are evaluated on both simulated and real data, demonstrating comparable power and dramatically lower selection variability when compared with the original model-X knockoffs.