Division of Biostatistics

Seminar Talk: *Bayesian Knockoff Generators for Robust Inference Under Complex Data Structure*

JOIN US! *Tuesday, December 7, 2021 3:30 – 4:30 PM*

Presented virtually by:

Mike Martens, PhD

The recent proliferation of medical data, such as genetics and electronic health records (EHR), offers new opportunities to find novel predictors of health outcomes. Presented with a large set of candidate features, interest often lies in selecting the ones most likely to be predictive of an outcome for further study, so that the goal is to control the false discovery rate (FDR) at a specified level. Knockoff filtering is an innovative strategy for FDR-controlled feature selection; but, existing knockoff methods often make strong distributional assumptions that hinder their applicability to real world data.

We propose Bayesian models for generating high quality knockoff copies that utilize available knowledge about the data structure, thus improving the resolution of prognostic features. Applications to two feature sets are considered: those with categorical and/or continuous variables possibly having a population substructure, such as in EHR; and those with microbiome features having a compositional constraint and phylogenetic relatedness. Through simulations and real data applications, these methods are shown to identify important features with good FDR control and high power.