

STATISTICS SEMINAR

Molly Roberts, Ph.D.
University of California, San Diego

Thursday, December 3, 2015
Dickens Hall, Room 207, 4:00-5:00 pm
Refreshments: Dickens 108, 3:30-4:00 pm



Matching Methods for High-Dimensional Data with Applications to Text

Matching is a popular technique for preprocessing observational data to facilitate causal inference and reduce model dependence by ensuring that treated and control units are balanced along pre-treatment covariates. While most applications of matching balance on a small number of covariates, we identify situations where matching with thousands of covariates may be desirable, such as causal inference where confounders are measured with text. With high-dimensional covariates, traditional matching methods are less effective and may be difficult or impossible to implement. We characterize the problem of matching in a high-dimensional context as a tradeoff between dimension reduction and imbalance bounding. We develop a new method called Topical Inverse Regression Matching (TIRM) that optimizes this tradeoff by including both a low-dimensional projection of covariates and information about the probability of treatment. We illustrate our approach by estimating the effect of censorship on the writing of Chinese bloggers, the effects of gender on citation counts in international relations, and the effects of targeted killings and capture by counterterrorists on the popularity of jihadist writings.