



Cornell CIS  
**Statistical Science**

## **STATISTICS SEMINAR**



**Wednesday, October 11, 2017**

**4:15 pm**

**G01 Biotechnology**

**NATHAN KALLUS**

Operations Research and Information  
Engineering, and Cornell Tech,  
Cornell University

### **Generalized Optimal Matching for Causal Inference and Policy Learning**

I will present recent advances in using modern optimization and representation learning for learning causal effects and causal-effect-maximizing policies from observational data. Central to these is a new, encompassing framework for matching, covariate balancing, and doubly-robust methods for causal inference called generalized optimal matching (GOM). The framework is given by generalizing a new functional-analytical formulation of classic optimal matching, giving rise to the class of GOM methods, for which I provide a single unified theory to analyze tractability, consistency, and efficiency. Many commonly used existing methods are included in GOM and, using their GOM interpretation, can be extended to optimally and automatically trade off balance for variance and outperform their standard counterparts. As a subclass, GOM gives rise to kernel optimal matching (KOM), which, as supported by new theory, I will show is notable for combining the interpretability of matching methods, the non-parametric model-free consistency of optimal matching, the efficiency of well-specified regression, the efficiency and robustness of augmented inverse propensity weight estimators, the judicious sample size selection of monotonic imbalance bounding methods, and the model-selection flexibility of Gaussian-process regression. New

inference methods for KOM enable partial (interval) identification of causal effects even in the absence of necessary overlap conditions.

The GOM approach is particularly apt for the problems of evaluating and learning personalized decision policies from observational data of past contexts, decisions, and outcomes, where only the outcome of the enacted decision is available and the historical policy is unknown. These problems arise in personalized medicine using electronic health records and in internet advertising, where existing approaches rely on fragile and inefficient plug-in approaches leading to high-variance evaluation and ineffective learning. I will present a new GOM-based policy learner that proceeds as a bilevel optimization problem and demonstrably outperforms existing approaches both in evaluation and learning, as supported by new regret bounds for the method. I will conclude by discussing exciting new directions for balancing as a new paradigm for causal inference methods in other settings as well as new applications and challenges for policy learning from observational data.

**Nathan Kallus** is Assistant Professor in the School of Operations Research and Information Engineering and Cornell Tech at Cornell University. Kallus's research revolves around data-driven decision making, the interplay of optimization and statistics in decision making and in inference, and the analytical capacities and challenges of observational, large-scale, and web-driven data. He holds a PhD in Operations Research from MIT as well as a BA in Mathematics and a BS in Computer Science both from UC Berkeley. Before coming to Cornell, Kallus was a Visiting Scholar at USC's Department of Data Sciences and Operations and a Postdoctoral Associate at MIT's Operations Research and Statistics group.

*Refreshments will be served following the seminar in 1181 Comstock Hall.*