Generalizability in Causal Inference (out-of-sample policy forecasting)

Elias Bareinboim
Computer Science Department, UCLA, USA, eb@cs.ucla.edu

Empirical scientists seek not just surface descriptions of the observed data, but deeper explanations of why things happened the way they did, and how the world would be like had things happened differently. With the unprecedented accumulation of data (or, "big data"), researchers are becoming increasingly aware of the fact that traditional statistical techniques, including those based on artificial intelligence and machine learning, must be enriched with two additional ingredients in order to construct such explanations:

1. the ability to integrate data from multiple, heterogeneous sources, and
2. the ability to distinguish causal from associational relationships.

In this talk, I will present a theory of causal generalization that provides a principled way for fusing pieces of empirical evidence coming from multiple, heterogeneous sources. I will first introduce a formal language capable of encoding the assumptions necessary to express each problem instance. I will then present conditions and algorithms for deciding whether a given problem instance admits a consistent estimate for the target effects and, if feasible, fuse information from various sources to synthesize such an estimate. These results subsume the analyses conducted in various fields in the empirical sciences, including "external validity," "meta-analysis," "heterogeneity," "quasi-experiments," "transportability," and "sampling selection bias." I will conclude by presenting new challenges and opportunities opened by this research.

Keywords: causal inference; generalizability; recalibration; experimental design.