Statistical analysis for the discrete instrumental variable model

Thomas Richardson

We consider causal instrumental variable (IV) models containing an instrument (Z), a treatment (X) and a response (Y) in the case where X and Y are binary, while Z is categorical taking k levels. We assume that the instrument Z is randomized and has no direct effect on the outcome Y, except through X.

In the first part of the talk we consider the problem of characterizing those distributions over potential outcomes for Y that are compatible with a given observed distribution P(X,Y | Z). We show that this analysis of identification may be simplified by viewing the observed distribution as arising from a series of observational studies on the same population. We also show that this approach naturally leads to the restrictions imposed on the observed distribution by the IV model.

In the second part of the talk we consider statistical inference for this model. We first show that our characterization of the model for the observables leads to a 'transparent' approach to Bayesian inference under which identified and non-identified parameters are clearly distinguished. We contrast this with the 'direct' approach that puts priors directly on the distribution of potential outcomes.

Finally, time permitting, we will describe a frequentist approach to inference for the IV model via a new approach to constructing confidence regions for multinomial data with (non-asymptotic) coverage guarantees via a Chernoff-type tail bound.

Joint work with Robin J. Evans (Oxford), F. Richard Guo (University of Michigan), James M. Robins (Harvard), Yilin Song (University of Washington) and Gary Chan (University of Washington).