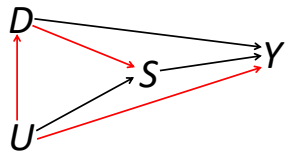
Analyzing Post-Treatment Proxy Conditioning with DAGs

Cyrus Samii
NYU Politics – Quant II
April 2018

Let's use DAGs to analyse post-treatment proxy control, as considered by Rosenbaum (1984) using potential outcomes. Consider the following DAG:

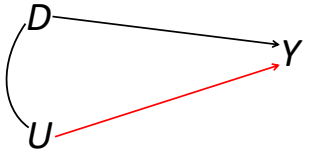


The black lines show causal relations that we are rather sure exist. The red ones show relations that we are not sure may exist. Suppose the goal is to estimate the effect of D on Y. We think that U confounds this relationship, and we know that at least part of this confounding relationship flows through S.

What happens when we "control for" S? What we mean is that we *condition* on S. By the theory of DAGs (e.g., Pearl 2009) we can represent this by a graph operation:

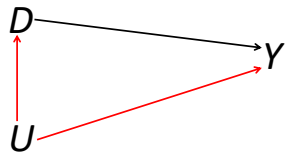
(1) If S is a collider, link all pairs of parents of S with an undirected edge, connoting induced dependency.
(2) For any ancestor of S, if this ancestor is itself a collider, link all pairs of parents of this ancestor with undirected arcs to connote induced dependencies.
(3) Erase S from the graph and all edges connected with S.

If the causal relationship $D \rightarrow S$ is active, then these operations yield the following:



Note that this holds *regardless* of whether the causal relationship $U \rightarrow D$ is active. This is how conditioning on post-treatment variables can *induce* confounding. That is, if $U \rightarrow D$ was *not* active (e.g., imagine that D was randomly assigned), but the causal relation $U \rightarrow Y$ (not mediated by S) was active, then conditioning on S would introduce confounding bias that would not have been there had we not conditioned at all.

If the causal relation $D \rightarrow S$ is not active, then conditioning on S yields the following:



This simply highlights the importance of the assumption that all effects of U must be mediated via S for the proxy conditioning strategy to work to identify the effect of D on Y. If not, then we have the possibility of a backdoor path via U.