Causal Inference, Homework 2

Junyi Liao

1. Confounder Identification

Question. Compare the backdoor criterion and traditional statistical approach for identifying confounders. In what case, the traditional method would lead mistake? Think how to fix the traditional definition of confounders.

Solution. For convenience, we denote the treatment or exposure of interest as A and the outcome as Y.

Traditional Statistical Approach

A variable is identified as a confounder if will satisfies the following conditions:
(i) It is associated with the treatment;
(ii) It is associated with the outcome conditioning on the treatment;
(iii) It does not lie on a causal pathway between the treatment and the outcome.

Backdoor Criterion

Conceptually, a **backdoor path** is a path that links A and Y with an edge into A, e.g., $A \leftarrow L \rightarrow Y$. The backdoor criterion states that, given treatment A and outcome Y, if a set of variables Z contains no descendent of A and blocks all backdoor paths between A and Y, then conditioning on Z suffices to control for confounding. In that sense, such a set of variables is identified as confounders.

Comparison and Failure Cases of Traditional Approach

According to VanderWeele et al., 2013[1], the causal inference literature has not produced a clear formal definition of a confounder, as it has given priority to the concept of confounding over that of a confounder. Here we compare the two methods for identifying confounders. The **traditional approach** relies on only **statistical association** and identifies all non-causal variables that is associated to both the treatment and the outcome as confounders, while the **backdoor criterion** is established in the context of **causal DAG**, which may relies on some **prior causal knowledge**. The backdoor criterion attempts to find a minimal but sufficient set to control for confounding. Hence, confounder becomes a relative concept since there can exist multiple set meeting the criterion.

In general, the traditional approach does not really distinguish the "cause" and "effect", more specifically, the mechanism of events. Here we give some failure cases of traditional approach. Consider the DAG $A \rightarrow L \leftarrow U \rightarrow Y$. The variable L meets all (i)-(iii) above and is identified as a **traditionally defined confounder**. However, enlightened by the **backdoor criterion**, the path from A to Y is naturally blocked since L is a collider, and controlling on L, conversely, introduces bias.
Another typical case is the **collider**. In a causal DAG, a collider is a variable that lies on a path from A to Y with both arrows into it, e.g., $A \rightarrow C \leftarrow Y$. The collider C meets all (i)-(iii) listed above, hence is identified as a confounder by traditional approach. However, if we controlled for C, A would be conditionally dependent on

Y according to causal DAG theory, hence introduce the **collider selection bias.**

To fix the traditional definition of confounders, we can impose restriction on **timing** of the variables of interest. Since conditioning on a collider or any its descendant introduces bias, we restrict the candidate confounders to the covariates observed before treatment or exposure, i.e. we only identify confounders from **pre-treatment or pre-exposure factors**. Such a restriction is consistent with the definition of **backdoor path** (which has an edge into A).
Based on the timing rule, we can go a step further. Here we borrow the candidate definition for a confounder from Miettinen, 1974[2]:

Definition 1. A pre-exposure covariate C is a confounder for the effect of A on Y if there exists a set of pre-exposure covariates X such that $C \not\perp A \mid X$ and $C \not\perp Y \mid \{A, X\}$.

Under this definition, the failure cases we put forward can be fixed since C and L are both post-exposure. Furthermore, this definition implies the following property:

Property 1. If S consists of the set of all confounders for the effect of A on Y, then there is no confounding of the effect of A on Y conditional on S, i.e. $Y^a \perp A \mid S$.

A detailed proof is given in the appendix of VanderWeele et al., 2013.

References

[1] VanderWeele TJ, Shpitser I. On the definition of a confounder. Annals of Statistics. Feb 2013; 41(1):196-220.
[2] Miettinen OS. Confounding and effect modification. American Journal of Epidemiology. 1974; 100:350–353.