Disjunctive cause criterion (no need to know the full DAG)

22.04.21 05:35

- used to identify a sufficient set of variables to control for (2011, Van der Weele)

select the set of variables (observed variables) that are causes of the exposure (i.e. treatment) A, the outcome Y, or both (A and Y)

- if there is a set of observed variables satisfying the backdoor path criterion, then the variables selected based on the disjunctive cause criterion will be sufficient to control for confounding

Example:
- observed pre-treatment variables: {M, V, W}
- unobserved pre-treatment variables: {U_1, U_2}

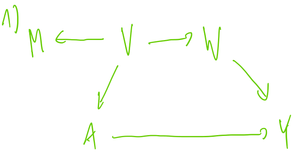
→ suppose, we know that V & W are causes of A, Y, or both
→ suppose M is not a cause of either A or Y

2 methods for selecting variables w/o knowledge of the DAG:

1. use all pre-treatment covariates: {M, V, W} (control for everything)
2. use variables based on the disjunctive cause criterion: {V, W}

Example scenarios with hypothetical DAGs:

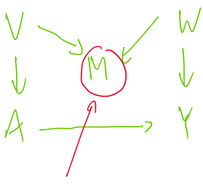
1)



- using method 1.): {M, V, W}
- using method 2.): {V, W}

} both sets satisfy the backdoor path criteria

2)

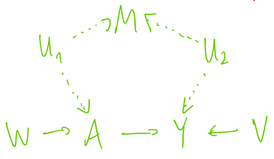


- method 1: {V, M, W}
- method 2: {V, W}

} both sets satisfy the backdoor path criteria

M is a collider: thus with knowledge of the DAG, we would not control for any variable i.e. use {} = ∅

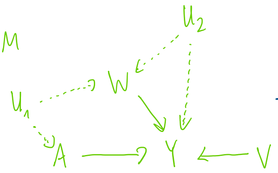
3)



- method 1: {M, V, W} X → violates backdoor path crit.
  ↳ controlling for M would create an unblocked backdoor path: $A \leftarrow U_1 \rightarrow U_2 \rightarrow Y$
- method 2: {V, W} → satisfies backdoor path criteria

U_i... unobserved vars.

4)



- method 1: {M, W, V}
- method 2: {W, V}

} none of these sets is sufficient blc the backdoor path $A \leftarrow U_1 \rightarrow W \leftarrow U_2 \rightarrow Y$ with the collider W is being unblocked with either method

Summary disjunctive cause criterion:

- doesn't select the smallest set of variables to control for
- conceptually simpler than
- guaranteed to select a sufficient set of variables to control for confounding iff:
  - such set exists in the observed variables of the dataset
  - we correctly identify all of the observed direct causes of A and Y.

- next lessons: once we identify set of variables to control for, how do we actually control for these variables??
  ↳ examples: 1) matching
  2) IPW (inverse probability weighting)