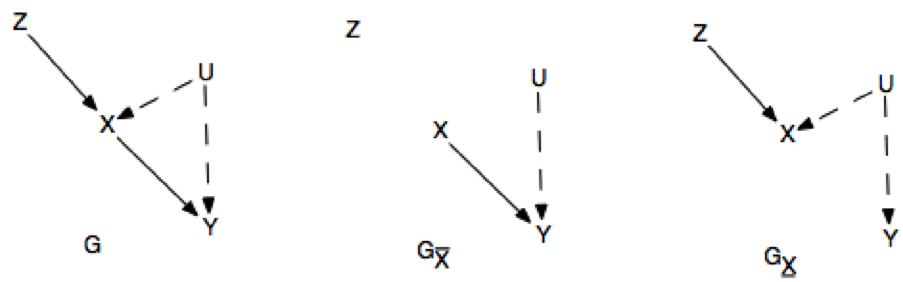
Ralph's partially-identified DAG

Ralph wishes to identify causal effects associated with the DAG below where Z is treatment assigned, X is treatment received, Y is outcome or response, and U is latent (unobserved) factors.



DAG G and its subgraphs

Suppose Y, X, and Z are binary and Ralph focuses on the aggregate or average causal effect[1]

$$ACE\left(X \rightarrow Y\right) = \Pr\left(y_1 \mid do\left(x_1\right)\right) - \Pr\left(y_1 \mid do\left(x_0\right)\right)$$

Suggested:

1. Suppose $W = U$ where W is observable. Can Ralph employ a back-door adjustment (see Ralph's back-door adjustment) to point-identify causal effects of X on Y? In this W observable setting, is it necessary for Ralph to condition on Z to identify causal effects of X on Y?

2. With U unobservable, can Ralph point-identify causal effects? Hint: can the back-door adjustment be implemented?

As a result of the challenges associated with latent factors U, Ralph decides an alternative frame is in order. Let R_x be a variable describing compliance behavior where $r_x = 0, 1, 2, 3$ denotes, respectively, a never-taker, complier, defier, and always-taker. Then,

$$x = \begin{array}{lll} x_0 & \text{if} & r_x = 0, \\ x_0 & \text{if} & r_x = 1 \text{ and } Z = z_0, \\ x_1 & \text{if} & r_x = 1 \text{ and } Z = z_1, \\ x_1 & \text{if} & r_x = 2 \text{ and } Z = z_0, \\ x_0 & \text{if} & r_x = 2 \text{ and } Z = z_1, \\ x_1 & \text{if} & r_x = 3 \end{array}$$

[1] Since Y is binary $E\left[Y\right] = \Pr\left(y_1\right)$.

Likewise, let the variable R_y convey response to treatment where $r_y = 0, 1, 2, 3$ represents never-recover, helped, harmed, always-recover, respectively.

$$y = \begin{array}{lll} y_0 & \text{if} & r_y = 0, \\ y_0 & \text{if} & r_y = 1 \text{ and } X = x_0, \\ y_1 & \text{if} & r_y = 1 \text{ and } X = x_1, \\ y_1 & \text{if} & r_y = 2 \text{ and } X = x_0, \\ y_0 & \text{if} & r_y = 2 \text{ and } X = x_1, \\ y_1 & \text{if} & r_y = 3 \end{array}$$

3. Determine $\Pr(y_1 \mid do(x_1))$, $\Pr(y_1 \mid do(x_0))$, and $ACE(X \to Y)$ in terms of r_y.

Further, Ralph recognizes that causal effect identification demands that he work with distributions over observables. In this case, the conditional distributions, $\Pr(y, x \mid z_0)$ and $\Pr(y, x \mid z_1)$, are observable. For simplicity, Ralph denotes the conditional distributions $p_{ij.k}$ where $i, j, k = 0, 1$ for y_i, x_j, z_k.

Ralph knows Pearl describes natural bounds on the causal effects.

$$p_{11.1} \leq \Pr(y_1 \mid do(x_1)) \leq 1 - p_{01.1}$$

and

$$p_{10.0} \leq \Pr(y_1 \mid do(x_0)) \leq 1 - p_{00.0}$$

4. Write an expression for the natural bounds on $ACE(X \to Y)$ in terms of $p_{ij.k}$.

Suppose the *DGP* (data generating process) is

$$\begin{array}{cc} p_{00.0} = 0.919 & p_{00.1} = 0.315 \\ p_{01.0} = 0.000 & p_{01.1} = 0.139 \\ p_{10.0} = 0.081 & p_{10.1} = 0.073 \\ p_{11.0} = 0.000 & p_{11.1} = 0.473 \\ \Pr(z_1) = 0.500 & \Pr(z_0) = 0.500 \end{array}$$

5. What is the compliance rate, $\Pr(x_1 \mid z_1)$?

6. What is the encouragement or intent to treat effect, $\Pr(y_1 \mid z_1) - \Pr(y_1 \mid z_0)$?[2]

7. What is the mean (observation not action) difference $\Pr(y_1 \mid x_1) - \Pr(y_1 \mid x_0)$?

[2] Policy decisions are sometimes based on this effect due to partial compliance complications.

8. What are the natural bounds on $\Pr(y_1 \mid do(x_1))$, $\Pr(y_1 \mid do(x_0))$, and $ACE(X \rightarrow Y)$?

The natural bounds can be tightened via usage of a dual linear program. The lower bound on $\Pr(y_1 \mid do(x_1))$ is

$$\Pr(y_1 \mid do(x_1)) \geq \max \left\{ \begin{array}{c} p_{11.1}, \\ p_{11.0}, \\ -p_{00.0} - p_{01.0} + p_{00.1} + p_{11.1}, \\ -p_{01.0} - p_{10.0} + p_{10.1} + p_{11.1} \end{array} \right\}$$

The upper bound on $\Pr(y_1 \mid do(x_1))$ is

$$\Pr(y_1 \mid do(x_1)) \leq \min \left\{ \begin{array}{c} 1 - p_{01.1}, \\ 1 - p_{01.0}, \\ p_{00.0} + p_{11.0} + p_{10.1} + p_{11.1}, \\ p_{10.0} + p_{11.0} + p_{00.1} + p_{11.1}, \end{array} \right\}$$

The lower bound on $\Pr(y_1 \mid do(x_0))$ is

$$\Pr(y_1 \mid do(x_0)) \geq \max \left\{ \begin{array}{c} p_{10.1}, \\ p_{10.0}, \\ p_{10.0} + p_{11.0} - p_{00.1} - p_{11.1}, \\ p_{01.0} + p_{10.0} - p_{00.1} - p_{01.1} \end{array} \right\}$$

The upper bound on $\Pr(y_1 \mid do(x_0))$ is

$$\Pr(y_1 \mid do(x_0)) \leq \min \left\{ \begin{array}{c} 1 - p_{00.1}, \\ 1 - p_{00.0}, \\ p_{01.0} + p_{10.0} + p_{10.1} + p_{11.1}, \\ p_{10.0} + p_{11.0} + p_{01.1} + p_{10.1}, \end{array} \right\}$$

9. What are the linear programming bounds on $\Pr(y_1 \mid do(x_1))$, $\Pr(y_1 \mid do(x_0))$, and $ACE(X \rightarrow Y)$? Interpret the results in light of the degree of noncompliance.

10. Ralph is also interested in the effect of the existing program under its current incentive system and current participants which is treatment on treated. Average treatment on the treated is quantified as

$$\begin{aligned} ETT(X \rightarrow Y) &= \Pr(Y_{\widehat{x}_1} = y_1 \mid x_1) - \Pr(Y_{\widehat{x}_0} = y_1 \mid x_1) \\ &= \sum_u [\Pr(y_1 \mid x_1, u) - \Pr(y_1 \mid x_0, u)] \Pr(u \mid x_1) \end{aligned}$$

where $Y_{\widehat{x}_j}$ refers to outcome when intervening with action x_j and conditioning on x_1 refers to the subpopulation observed in the treatment regime. Again, this utilizes unobservable U so Ralph employs partial identification once again. Natural bounds on ETT are

$$\begin{aligned} & \frac{P(y_1) - P(x_1 \mid z_0) - P(y_1, x_0 \mid z_0)}{P(x_1)} \\ \leq \quad & ETT(X \to Y) \leq \\ & \frac{P(y_1) - P(y_1, x_0 \mid z_0)}{P(x_1)} \end{aligned}$$

Natural bounds on the untreated (say, $ETUT$) complement natural bounds on the treated to give the natural bounds on ACE as

$$ACE(X \to Y) = \Pr(x_1)\, ETT(X \to Y) + \Pr(x_0)\, ETUT(X \to Y)$$

where

$$\begin{aligned} & \frac{-P(y_1) + P(y_1, x_1 \mid z_1)}{P(x_0)} \\ \leq \quad & ETUT(X \to Y) \leq \\ & \frac{-P(y_1) + P(x_0 \mid z_1) + P(y_1, x_1 \mid z_1)}{P(x_0)} \end{aligned}$$

Determine the natural bounds for ETT, and $ETUT$ and verify their complementarity with ACE.

Alternatively, suppose the DGP is

$$\begin{array}{cc} p_{00.0} = 0.55 & p_{00.1} = 0.45 \\ p_{01.0} = 0.45 & p_{01.1} = 0. \\ p_{10.0} = 0. & p_{10.1} = 0. \\ p_{11.0} = 0. & p_{11.1} = 0.55 \\ \Pr(z_1) = 0.50 & \Pr(z_0) = 0.50 \end{array}$$

11. Repeat 5 through 10 for this DGP.