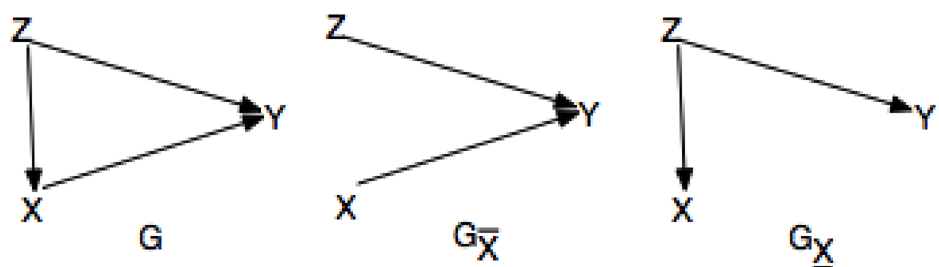
Ralph's back-door adjustment

Ralph wishes to nonparametrically assess the causal effect of X on Y. That is, evaluate how setting X to x impacts the probability of Y.[1] Ralph recognizes this involves combining a thought experiment with data where the effect may be confounded by other variables, Z. After pondering the causal connections amongst the variables, Ralph envisions the following DAG (directed acyclic graph), G, and its subgraphs, $G_{\overline{X}}$ and $G_{\underline{X}}$.



DAG G and subgraphs

In other words, Z causes X and Y, and X causes Y (via some unknown and unspecified functional relations) with all variables potentially impacted by unobservables (not made explicit in the graphs).[2]

Ralph recognizes considerable work has been completed regarding DAGs including the relation between d-separation (d denotes directional) and conditional independence and the do-calculus theorems.

Definition 1 (d-separation) *a path p is d-separated (blocked) by a set of nodes Z (including the null set $\emptyset$) if and only if*
1. p contains a chain $i \to m \to j$ or fork $i \leftarrow m \to j$ such that the middle node m is in Z, or
2. p contains an inverted fork (collider) $i \to m \leftarrow j$ such that the middle node m is not in Z and no descendant of m is in Z.

A set Z d-separates X and Y if and only if Z blocks every path from X to Y.

[1] Setting X to x iinvolves active intervention (as opposed to passive observation) and is usually referred to by the operation $do\,(X = x)$.

[2] If the same unobservable is connected to two or more observables it's typically important to include the unobservable in the graph. See Ralph's front-door adjustment for an illustration.

Theorem 1 (d-separation and conditional independence) *If sets X and Y are d-separated by Z in a DAG G, then X is independent of Y conditional on Z in every distribution consistent with G. Conversely, if X and Y are not d-separated by Z in a DAG G, then X and Y are dependent conditional on Z in at least one distribution consistent with G.*

The converse part is actually much stronger. If X and Y are not blocked then they are dependent in almost all distributions consistent with G. Independence of unblocked paths requires precise parameter tuning that is unlikely.

Theorem 2 (do-calculus) [3]*Let G be the DAG associated with a causal model and let $\Pr(\cdot)$ be the probability distribution induced by the model. For any disjoint set of variables X, Y, Z, and W the following rules apply.*

Rule 1 (insertion/deletion of observations):

$$\Pr(y \mid do(x), z, w) = \Pr(y \mid do(x), w) \quad if \;\; (Y \perp Z \mid X, W)_{G_{\overline{X}}}$$

where $\perp$ refers to stochastic independence or d-separation in the graph.

Rule 2 (action/observation exchange):

$$\Pr(y \mid do(x), do(z), w) = \Pr(y \mid do(x), z, w) \quad if \;\; (Y \perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$$

Rule 3 (insertion/deletion of actions):

$$\Pr(y \mid do(x), do(z), w) = \Pr(y \mid do(x), w) \quad if \;\; (Y \perp Z \mid X, W)_{G_{\overline{X}, \overline{Z(W)}}}$$

where $Z(W)$ is the set of Z-nodes that are not ancestors of any W-nodes in $G_{\overline{X}}$.

Observation and action are fundamentally different, the former is passive while the latter involves active intervention to set the value of a variable as in $do(X = x)$. Intervention effectively eliminates paths into the variable actively set (if Z is a parent to X, $\Pr(do(x) \mid z) = 1$). Algebraically, this alters the Bayesian chain rule. For example,

$$\begin{aligned} \Pr(Y = y, do(X = x), Z = z) &= \Pr(y \mid do(x), z) \Pr(do(x) \mid z) \Pr(z) \\ &= \Pr(y \mid do(x), z)(1) \Pr(z) \end{aligned}$$

The action/observation distinction is essential to nonparametric identification of causal effects.

Definition 2 (causal effect of X on Y) *The causal effect of X on Y is*

$$\Pr(Y = y \mid do(X = x))$$

[3] The proof is in the appendix to Pearl [1995], *Biometrika*, "Causal diagrams for empirical research," December 82(4), 669-710.

The causal effect is identified only when the quantity can be expressed entirely in terms of observation (combining the thought experiment of the DAG with data). When the causal effect is identified but confounded by Z, the causal effect can often derived by conditioning on Z via the back-door adjustment.

$$\Pr(y \mid do(x)) = \sum_{z} \Pr(y \mid x, z) \Pr(z) \qquad \text{(back-door adj)}$$

Definition 3 (back-door) *a set of variables Z is a back-door to the ordered pair (X, Y) if*
(i) no node in Z is a descendant of X, and
(ii) Z blocks every path between X and Y that contains an arrow into X.

Suggested:

1. Is the causal effect of X on Y identified for the setting depicted by DAG G? Explain by utilizing the Bayesian chain rule along with the do-calculus theorem.

Suppose X, Y, and Z are binary with DGP (data generating process):

	0	1
$\Pr(y \mid x = 0, z = 0)$	0.95	0.05
$\Pr(y \mid x = 0, z = 1)$	0.1	0.9
$\Pr(y \mid x = 1, z = 0)$	0.05	0.95
$\Pr(y \mid x = 1, z = 1)$	0.9	0.1
$\Pr(x \mid z = 0)$	0.35	0.65
$\Pr(x \mid z = 1)$	0.6	0.4
$\Pr(z)$	0.4	0.6

2. Does the back-door adjustment apply? If so, derive the causal effects: $\Pr(y = 1 \mid do(X = 0))$ and $\Pr(y = 1 \mid do(X = 1))$.

3. Compare action in 2 to observation: $\Pr(y = 1 \mid X = 0)$ and $\Pr(y = 1 \mid X = 1)$. Hint: you may find it instructive to write out the joint distribution $\Pr(y, x, z)$.

4. Suppose there is no path from $Z \to X$ in G (G is $G_{\overline{X}}$). Does Z confound the causal effect of X on Y? (hint: explore rule 2 in this setting.) For this setting, derive the causal effects, $\Pr(y = 1 \mid do(X = 0))$ and $\Pr(y = 1 \mid do(X = 1))$, where the only change in the DGP is

	0	1
$\Pr(x \mid z = 0)$	0.475	0.525
$\Pr(x \mid z = 1)$	0.475	0.525

Hint: if Z is not confounding, you can still apply the back-door adjustment but the conditioning variables are the null set, $\emptyset$, in place of Z. This changes

$$\Pr\left(y \mid do\left(x\right)\right) = \sum_{z} \Pr\left(y \mid x, z\right) \Pr\left(z\right)$$

to

$$\Pr\left(y \mid do\left(x\right)\right) = \Pr\left(y \mid x\right)$$