## Causal Inference

Causal inference covers problems where we want to predict the outcome of an intervention given some known causal structure

- Assume that the true causal structure is a directed graphical model where $X \to Y$ means X causes Y. We know the structure of the network but may not be able to observe the values of all the variables.

- Intervening to set $X = x$, denoted $do(X = x)$ in a causal network, has the effect of removing all links coming into the variable X (since its value is no longer determined by its parents) and setting its value to x.

- We can use d-separation to derive the 3 rules of the do-calculus that allow us write queries about interventions in terms of the pre-interventional distribution.

- These rules are complete. If an expression containing do's can be converted into one that does not by repeated application of the rules then you can determine the post-interventional distribution from observational data. If you can't, its not possible without additional assumptions.

The rules (simplified to cover only the case where we intervene on a single variable) can all be illustrated by the graph in figure 1. Assume our query is $P(Y|do(X))$. $G^\dagger$ is the original graph, G, augmented to include a variable $\hat{X}$ that acts as a switch determining if X takes its value as determined by its parents or from an external intervention. $G_{\overline{X}}$ is G after the intervention.

Figure 1: The do calculus

(a) $G^\dagger$ (b) $G_{\overline{X}}$
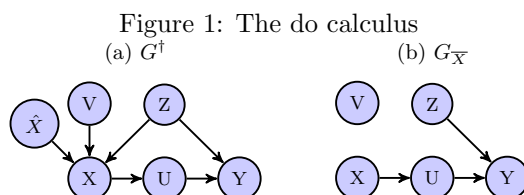
1. D-separation still applies after an intervention. $Y \perp\!\!\!\perp V | X$ in $G_{\overline{X}} \implies P(Y|do(X), V) = P(Y|do(X))$

2. If the outcome variable is independent of the way in which the intervened on variable takes its value, intervention = observation. $Y \perp\!\!\!\perp \hat{X} | X, Z$ in $G^\dagger \implies P(Y|do(X), Z) = P(Y|X, Z)$

3. If there is no direct path from the intervened on variable to the outcome variable, the intervention does not effect the distribution of the outcome variable. $Y \perp\!\!\!\perp \hat{X} | U, Z$ in $G^\dagger \implies P(Y|do(X), U, Z) = P(Y|U, Z)$ (note: we condition on Z because conditioning on U opens the path $\hat{X} \leftarrow Z \to Y$)

A single application of the rules show we can identify the causal effect of X on Y if we can observe Z. By repeated application of the rules the query is also identifiable if we can observe only U.
Key general references: [5, 4]

Causal Structure Learning - learn causal structure from observational data

Conditional independence approach

- Assume that the data was generated by an unknown causal directed graphical model

- Assume that all dependencies in the graph are reflected by conditional dependencies in the distribution of the data (ie no dependencies exactly cancel one-another out) ← faithfullness assumption

- Finding the true causal structure is equivalent to finding the set of graphs that are perfect maps for the data distribution. In general, result is not unique but some links may have the same direction in all graphs in the set.

- Approach can be extended to allow for latent variables

- Key general reference: [6]

Functional form approach

Conditional independence based methods cannot distinguish between graphs with the same dependency structure, for example, $X \rightarrow Y$ vs $Y \rightarrow X$. We can infer causality in such cases if we make braod assumptions about the relationship between the function mapping cause to effect and any noise. Suppose $X \rightarrow Y$:

- If we assume noise is additive in the causal direction, $Y = f(X) + \epsilon$ cannot generally be inverted to give an additive model in the other direction (linear f and gaussian ϵ are an unfortunate exception)[2].

- If f is deterministic and invertible, for most input distributions $p_X(x)$ the distribution of $p_Y(y)$ will be higher where f' is small and a large region of X maps to similar values of Y. We expect that f and p_X are independent but f' and p_Y are correlated [1].

- In general, $P(X)$ and $P(Y|X)$ should be independent but $P(Y)$ and $P(X|Y)$ will not be [3]

References

[1] Povilas Daniusis, Dominik Janzing, Joris Mooij, Jakob Zscheischler, Bastian Steudel, Kun Zhang, and Bernhard Schölkopf. Inferring deterministic causal relations. In *UAI*, 2010.

[2] Patrick Hoyer, Dominik Janzing, and Joris Mooij. Nonlinear causal discovery with additive noise models. In *NIPS*, 2009.

[3] Dominik Janzing, Jonas Peters, Eleni Sgouritsa, Kun Zhang, Joris M. Mooij, and Bernhard Schölkopf. On causal and anticausal learning. In *ICML*, 2012.

[4] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT Press, 2009.

[5] Judea Pearl. *Causality: models, reasoning and inference*. MIT Press, Cambridge, 2000.

[6] Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*, volume 81. MIT press, 2000.