Interventions

Source: (a) Elements of Causal Inference, Peters et al., Chap 6.
(b) Causality, J. Pearl, 2009.[2] → (aka "text")

Recall in our motivation that an experiment (aka intervention) to test the direction of causation (example 1 – wet vs. rain) or to correctly interpret data (example 2 – kidney stones and Simpson's paradox) required us to conduct additional (thought or real) experiments. These experiments led to an altered graphical model.

Goals: (1) Defining interventions through a family of SCMs and their corresponding DAGs.

(2) Calculus for computing interventional distributions from observational distribution.

The Observational SCM (repeated from Notes 3.)

$(X_1, X_2, \ldots, X_d)$ a collection of random variables, defined by the SCM $\mathfrak{C}$:

structural assignment

$$X_j := f_j(PA_j, N_j) \quad , \; j = 1, 2, \ldots, d.$$

where $PA_j \equiv$ Parents of $X_j \equiv$
subset of $\{X_1, \dots, X_d\} \setminus X_j$

$f_j \equiv$ deterministic function.
$N_j \equiv$ random variable, with
$\{N_1, N_2, \cdots, N_d\}$ mutually independent

The associated graphical model is a DAG.

Further, we denote the prob. measure induced on $(X_1, \dots, X_d)$ by $P^{\mathfrak{C}}(\cdot)$, and the joint dist. on them by $p^{\mathfrak{C}}(\cdot)$. We will abuse notation to use the same for marginals as well.

The Intervention SCM and Intervention Distribution

The intervention SCM, denoted by $\tilde{\mathfrak{C}}$ associated with an SCM $\mathfrak{C}$ is given by replacing one or more structural equations associated with $\mathfrak{C}$, i.e.,

replace $\quad X_j := f_j(PA_j, N_j) \quad$ with

$$X_j := \tilde{f}_j(\tilde{PA}_j, \tilde{N}_j)$$

where $\tilde{f}_j$, $\widetilde{PA}_j$, $\tilde{N}_j$ are new function / parents / noise, and with

$\tilde{N}_j$ mutually indep. of $(N_1, \dots, N_{j-1}, N_{j+1}, \dots N_d)$

Remarks:

① The original SCM $\mathfrak{C} = (S, N)$ is replaced by $\tilde{\mathfrak{C}} = (\tilde{S}, \tilde{N})$. This induces a new joint dist. over $(X_1, X_2, \dots, X_d)$.

② We denote the measure induced on $(X_1, \dots, X_d)$ due to $\tilde{\mathfrak{C}}$ to be $P^{\tilde{\mathfrak{C}}}$, and the joint dist. to be $P^{\tilde{\mathfrak{C}}}$. Since this is related (but different) from $P^{\mathfrak{C}}$ (the observational SCM), we also denote this by:

do operation denoting intervention

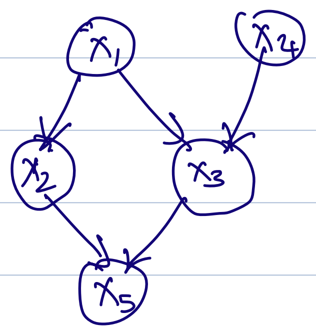
$$P^{\tilde{\mathfrak{C}}}(\cdot) = P^{\mathfrak{C};\, do(X_j := \tilde{f}(\widetilde{PA}_j, \tilde{N}_j))}(\cdot)$$

③ We require that the graph associated with $\tilde{\mathfrak{C}}$ be a DAG. But other than that, this allows quite general interventions — deterministic, stochastic, simultaneously on multiple variables, etc.

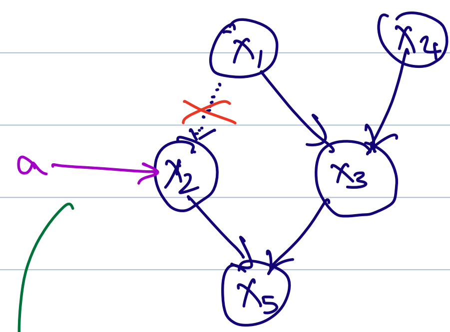
Special cases:

Atomic intervention: $X_j := a$, meaning deterministically set X_j to be 'a'.

This means that in the DAG $\mathcal{G}$, we delete incoming edges into X_j, and set $X_j := a$. The outgoing edges are not altered.



$\mathcal{G}$



$\tilde{\mathcal{G}}$

Observational distribution: $P^{\mathcal{G}}(\cdot)$

the incoming edge into X_2 has been deleted, and X_2 is set to the constant a. This operation induces a new joint dist. over $(X_1, X_2, X_3, X_4, X_5)$.

Intervention distribution: $P^{\tilde{\mathcal{G}}}(\cdot) = P^{\mathcal{G}; do(X_2 := a)}(\cdot)$.

In general, $\boxed{p^{\mathfrak{C}}(\cdot \mid x_2 = a) \neq p^{\mathfrak{C};\, do(x_2 := a)}(\cdot).}$

i.e., 'do' is NOT 'conditioning'.

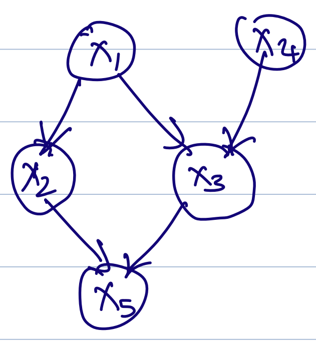
Conditioning affects ALL variables — both upstream and downstream of the intervention
do affects only downstream variables, as we are working with an ALTERED DAG.

Stochastic Intervention / Soft Intervention: $\widetilde{PA}_j = PA_j$, and X_j has strictly positive variance.

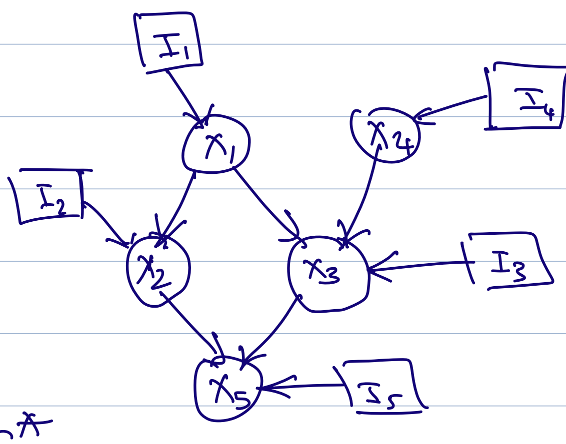
Alternative way to formalize Interventions using Intervention Variables

SCM $\mathfrak{C} = (S, P_N)$ over $(X_1, \ldots, X_d)$

For each node X_j, associate a new, additional parent node I_j, $j = 1, 2, \ldots d$ called intervention variables.



$\mathfrak{C}$



$\mathfrak{C}^*$

where $$X_j = \begin{cases} f_j(PA_j, N_j) & \text{if } I_j = \text{idle} \\ I_j & \text{otherwise.} \end{cases}$$

I_j (when active) encodes the intervention, i.e., I_j can take values $\{x_{i_1}, \dots x_{i_d}\}$, with an intervention pmf.

Then, $$P_Y^{\mathfrak{C};\, do(X_j := x_{i_j})}(\cdot) = P_{Y \mid I_j = x_{i_j}}^{\mathfrak{C}^*}(\cdot)$$

measure on some target variable Y.

i.e., interventions can be computed using the usual conditional dist. on the augmented $\mathfrak{C}^*$ model.

Note: that this does NOT mean 'do' is the same as 'conditioning'. We still have to reason on a new DAG. However, this representation is useful for proofs later.

Example 3 (6.11 in text ; Myopia).

Night Light (NL) in child's room
Child Myopia (CM)
Parent Myopia (PM)

Original study (Quinn et al, 1999) — Showed dependence between NL and CM. With caveats, they stated that:

" statistical strength of association .... suggests absence of night light .... factor in development of myopia"

(Please see Example 6.11 on pp. 90-91 in text by Peters et. al. for a more carful ellipsis).

Question: Is there a causal effect between NL and CM?

Answer: No! Future studies showed that PM is the true source of the correlation. Specifically, parents with myopia more likely to put night light in child's room.

The "correct" DAG is:

$$\begin{tikzcd}
 & PM \arrow[dl] \arrow[dr] & \\
CM & & NL
\end{tikzcd}$$

(Quinn, 99): $CM \not\perp\!\!\!\perp NL$.

$\Longrightarrow$ dependence / correlation between CM and NL

Causal relation: Address using interventions and do(·) DAG. Conduct the following three experiments:

NL : night light $\in \begin{cases} \text{dark} & \text{— } a \\ \text{low light} & \text{— } b \\ \text{room light} & \text{— } c \end{cases}$

$$P_{CM}^{\mathfrak{C};\, do(NL:=a)}(\cdot) \overset{?}{=} P_{CM}^{\mathfrak{C};\, do(NL:=b)}(\cdot) \overset{?}{=} P_{CM}^{\mathfrak{C};\, do(NL:=c)}(\cdot)$$

measure induced on CM by doing NL:=a.

If so, then there is no causal effect between NL and CM.

Equivalently, set noise $\tilde{N}_{NL} = \begin{cases} a & \text{w.p. } 1/3 \\ b & \text{w.p. } 1/3 \\ c & \text{w.p. } 1/3 \end{cases}$

and consider the GM with measure $P^{\mathfrak{C};\, do(NL:=\tilde{N}_{NL})}$

The there is no causal effect if $NL \perp\!\!\!\perp CM$ in $P^{\mathfrak{C};\, do(NL:=\tilde{N}_{NL})}$.

Definition: (Total Causal Effect) A total causal effect exists from X_i to X_j if and only if $\exists$ a r.v. $\tilde{N}_{X_i}$ (i.e., $X_i := \tilde{N}_{X_i}$), s.t.

$$X_i \not\perp\!\!\!\perp X_j \quad \text{in} \quad P^{\mathfrak{C}; do(X_i := \tilde{N}_{X_i})}$$

Remark: This definition formally connects causality with independence with respect to an intervention measure/distribution. Note also that this relation is directional, because the intervention is on X_i.

Proposition (6.13 in text): (Equivalent forms to check for Total Causal Effect)

For an SCM $\mathfrak{C}$, the following four conditions are equivalent.

(i) There is a total causal effect from X_i to X_j

(ii) $\exists\ x_1, x_2$ s.t. $P_{X_j}^{\mathfrak{C}; do(X_i := x_1)} \neq P_{X_j}^{\mathfrak{C}; do(X_i := x_2)}$

(iii) $\exists\ x$ s.t. $P_{X_j}^{\mathfrak{C}; do(X_i := x)} \neq P_{X_j}^{\mathfrak{C}}$

(iv) $X_i \not\perp\!\!\!\perp X_j$ in $P_{X_i, X_j}^{\mathfrak{C}; do(X_i := \tilde{N}_i)}$ for any $\tilde{N}_i$ that has full support over the alphabet of X_i.

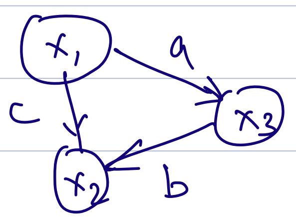
Pf: See text. (Appendix C.4).

Finally connecting graph structure with Total Causal Effect (TCE):

Proposition (6.14 in text) (Structure and TCE)

If $\nexists$ directed path from X_i to X_j in the DAG associated with $\mathfrak{C}$, then there is no TCE from X_i to X_j

Note: The converse is not true:



There is no TCE from $X_1 \to X_2$ if $c + ab = 0$

and the noise is iid $N(0,1)$ at each node.
