ECE 175B: Probabilistic Reasoning and Graphical Models: Lecture 6: Confounding in Causality and The Basic Junction Patterns – Part I

Florian Meyer & Ken Kreutz-Delgado

Electrical and Computer Engineering Department
University of California San Diego

UC San Diego
JACOBS SCHOOL OF ENGINEERING

Forced Node Values - The "do−operator"

- Instead of passively observing the value of a node in a graph, we can intervene and force the value of a node x to be pinned to a specified value $\mathrm{x} \equiv x$, denoted by $\mathrm{do}(\mathrm{x} = x)$

- A "regular" conditioned probability is more clearly written as:

$$P(y|x) = P(y|\mathrm{see}(\mathrm{x} = x)) \qquad \text{(by definition)}$$

- Important questions are:
  - How to compute $P(y|\mathrm{do}(\mathrm{x} = x))$ in general
  - When, if ever, are we lucky enough to have

$$P(y|\mathrm{do}(\mathrm{x} = x)) = P(y|\mathrm{see}(\mathrm{x} = x)) = P(y|x)$$

The do−operator

- These questions can be clarified by reviewing what $P\big(y|\text{see}(\mathrm{x}=x)\big) = P\big(y|x\big)$ means
- We can do this by running n independent experiments, and taking $n \to \infty$ and invoking the law of large numbers (LLN)
- For each experiment, we passively observe a joint outcome $(\mathrm{x}(\omega_j) = x_j, \mathrm{y}(\omega_j) = y_j)$, $j = 1, \ldots, n$
- We have the following counts for the outcomes $x = \mathrm{x}(\omega_j), y = \mathrm{y}(\omega_j)$

$$n_x = \sum_{j=1}^{n} \mathbb{I}\big(\mathrm{x}(\omega_j) = x\big) \qquad n_y = \sum_{j=1}^{n} \mathbb{I}\big(\mathrm{y}(\omega_j) = y\big)$$

$$n_{x,y} = \sum_{j=1}^{n} \mathbb{I}\big(\mathrm{x}(\omega_j) = x \wedge \mathrm{y}(\omega_j) = y\big)$$

Note that the indicator function $\mathbb{I}(\cdot)$ is defined as $\mathbb{I}(\mathit{true}) = 1$ and $\mathbb{I}(\mathit{false}) = 0$

The do−operator

- LLN: As $n \to \infty$

$$P(x) = \frac{n_x}{n} \; ; \; P(y) = \frac{n_y}{n}$$

$$P(x, y) = \frac{n_{x,y}}{n} \; ;$$ fraction of times (x, y) occurs (in all measurements)

$$P(y|x) = \frac{n_{x,y}}{n_x} \; ;$$ fraction of times y occurs in the measurements where x occurs

Note that $P(y|x) = \frac{n_{x,y}}{n_x} = \frac{n_{x,y}/n}{n_x/n} = \frac{P(x,y)}{P(x)}$ as expected

- Probabilities give the "long run" $(n \to \infty)$ relative frequencies of occurrence of situations that are passively observed

The do−operator

Example: Suppose we observe the following values:

- We note that x, y are functionally related as $y = 2x$
- We consider the following take the BN model (M)

$$x \longrightarrow y$$

$$\Leftrightarrow P(x, y) = \underbrace{P(y|x)}_{\mathbb{I}(y=2x)} P(x)$$

- We intervene and force $x \equiv 2$, i.e. we $do(x = 2)$
- We expect to see $y = 4$
- Instead, we observe $y = -2$. What happened?

x	y
1	2
0	0
-1	-2
1	2
3	6
4	8
0	0
2	4
-2	-4

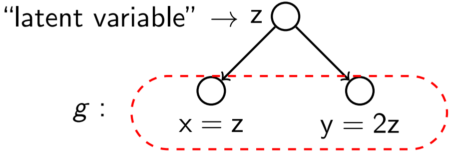
The do−operator

We have not verified that model M conforms to the actual causal structure of the "world"

- Therefore we are not "licensed" to use M for the causal control purposes
- E.g., Maybe the actual model is $x = \frac{1}{2}y$, $x \longleftarrow y$ or maybe not
- Maybe we have NOT captured all the variables and structure of interest

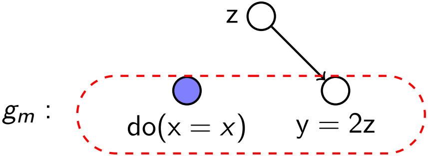
The do−operator

- Maybe the "Real World" is



No actual causal relationship between x and y

- The do−operator do(x = x) breaks a link in the graph g:



We have done "surgery" on g to create a modified graph g_m with modified distribution $P_m(x, y, z)$

The do−operator

- After this surgery, we have that

$$P\big(\mathrm{y}|\mathrm{do}(\mathrm{x}=x)\big) = P_m\big(y|x\big)$$

  Note that in our example $g_m : P(y|\mathrm{x}=x) = P_m(y)$, i.e. $\mathrm{y} \perp\!\!\!\perp \mathrm{x}$

- Thus $\mathrm{do}(\mathrm{x}=x)$ has no effect on y

- Thus our original model is wrong in the sense that it can not be used for control

- However, our model $\mathrm{x} \longrightarrow \mathrm{y}$ is OK for passive prediction of y from x since $\mathrm{y} = 2\mathrm{x}$ passively

Prediction versus Control

- Consider $\mathcal{X} = (\mathrm{x}, \mathrm{y})$ with $P(x, y) = P_{\mathrm{x,y}}(x, y)$

  We can model it in three ways:

  - $P(x, y)$ $\quad \mathrm{x} - \mathrm{y}$
  - $P(y|x)P(x)$ $\quad \mathrm{x} \longrightarrow \mathrm{y}$
  - $P(x|y)P(y)$ $\quad \mathrm{x} \longleftarrow \mathrm{y}$

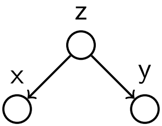
- All three are equivalent for prediction

- Can any of these models be used for control?
  Only if the model captures how the world "really is" regarding causal behaviour

- Consider $\mathrm{x} \longrightarrow \mathrm{y}$: Is $P(y|\mathrm{do}(x)) = P(y|x)$?
  Yes, if this is truly how the world is

Prediction versus Control

- But what if the world is actually, g:



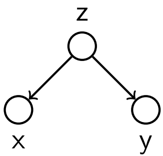
Say, we have $x = z$ and $y = 2z$

I.e. with $P(x, y, z) = \underbrace{P(x|z)}_{\mathbb{I}(x=z)} \underbrace{P(y|z)}_{\mathbb{I}(y=2z)} P(z)$

This yields $P(y|x) = \mathbb{I}(y = 2x)$, which allows perfect prediction, i.e., from observational data, we see and infer that $y = 2x$; this does NOT mean we can control y from x

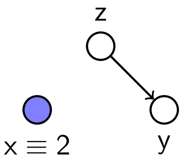
- Suppose we "do(x = 2)", i.e., we intervene and force x = 2

g :



Intervene
$\Longrightarrow$
do(x = 2)

g_m :



$$\begin{aligned} P(x,y,z) &= P(x|z)P(y|z)P(z) \\ P(x|z) &= \mathbb{I}(x=z) \\ P(y|z) &= \mathbb{I}(y=2z) \\ P(z) &= \text{anything} \\ P(y|x) &= \underbrace{\mathbb{I}(y=2x)} : \mathrm{y} \top\!\!\top \mathrm{x} \end{aligned}$$

functional relationship for seeing is y = 2x

$$\begin{aligned} P_m(x,y,z) &= P(x|z)P(y|z)P(z) \\ P_m(x|z) &= P_m(x) = \mathbb{I}(x=2) \\ P_m(y|z) &= P(y|z) = \mathbb{I}(y=2z) \\ P_m(z) &= P(z) = P_\mathrm{z}(z) \\ P_m(y|x) &= P_m(y) = P_\mathrm{z}(y/2) \end{aligned}$$

since y $\perp\!\!\!\perp$ x in g_m

Prediction or Control?

$$
\begin{aligned}
P_m(x,y) &= \sum_z P_m(x,y,z) \\
&= \sum_z P_m(x)P_m(y|z)P_m(z) \\
&= \sum_z \underbrace{\mathbb{I}(x=2)}_{P_m(x)} P_m(y,z) \\
&= P_m(x)\sum_z P_m(y,z) \\
&= P_m(x)P_m(y)
\end{aligned}
$$

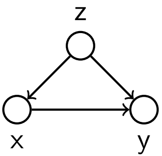
$\therefore P_m(x,y) = P_m(x)P_m(y) \Leftrightarrow \mathrm{y} \perp\!\!\!\perp \mathrm{x} \Leftrightarrow P_m(y|x) = P_m(y)$

Thus, y doesn't depend on what x is pinned to at all!

$\mathrm{y} = 2\mathrm{z}$ and $\mathrm{z} \sim P_\mathrm{z}(z) \implies P_m(y) = P_\mathrm{z}(y/2)$

Control or Prediction?

- Consider the more general situation, g :



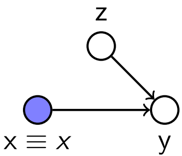
$$P(x, y, z) = P(y|x, z)\; P(x|z)\, \underbrace{P(x)}_{\text{anything}}$$

$$\begin{aligned} \text{Let} \qquad \mathrm{x} = \mathrm{z} &\Leftrightarrow P(x|z) = \mathbb{I}(x = z) \\ \mathrm{y} = \mathrm{x} + 2\mathrm{z} &\Leftrightarrow P(y|x, z) = \mathbb{I}(y = x + 2z) \\ P(y|z) &= \mathbb{I}(y = 3z) \qquad (y = 3z) \\ P(y|x) &= \mathbb{I}(y = 3x) \qquad (y = 3x) \end{aligned}$$

Because we have the functional relationship $y = 3x$ for seeing, we can do perfect prediction of y from x

Control or Prediction?

- Perform do(x $= x$), g_m :



Then we have y $=$ do(x $= x) + 2z$ with x $\neq$ z : $P_m(x) = \mathbb{I}(\mathrm{x} = x)$

$$\begin{aligned} P_m(x, y, z) &= P_m(y|x, z) P_m(x) P_m(z) \\ &= P(y|x, z) P_m(x) P(z) \end{aligned}$$

- If we "see" x $= x$, then we can predict $y = 3x$
- If we do(x $= x$), then we can obtain $y = x + 2z, \; z \sim P(z)$
- This suggest that if we can observe z, then we can control for its influence on y

Basic Junction Patterns

- We want to understand and analyse how "information flows" through paths in a BN
- There are the following three basic ("atomic") 3-node/2-link paths or junction patterns
  - Chain: $\bigcirc \longrightarrow \bigcirc \longrightarrow \bigcirc$
  - Fork: $\bigcirc \longleftarrow \bigcirc \longrightarrow \bigcirc$
  - Collider: $\bigcirc \longrightarrow \bigcirc \longleftarrow \bigcirc$

Basic Junction Patterns

- Consider any path through a DAG between any two nodes A and B having at least two edges
- Along such a path, one will encounter one or more of the three (basic) junction patterns; there are only three such patterns

Pattern	Model
chain	$x \rightarrow z \rightarrow y$
fork	$x \leftarrow z \rightarrow y$
collider	$x \rightarrow z \leftarrow y$

- Note that a junction always involves two consecutive edges and three consecutive nodes, where the middle node is the junction node
- Every non-endpoint node of a path is a junction node for some junction

Basic Junction Patterns

Pattern	Model
chain	$x \rightarrow z \rightarrow y$
fork	$x \leftarrow z \rightarrow y$
collider	$x \rightarrow z \leftarrow y$

- We can look at any non-endpoint node and decide what kind it is: mediator (for a chain); forking (for a fork); or collider (for a collider)
- We want to know the information blocking/transmission properties of the junction

Chain Junction

- $g: \quad x \longrightarrow z \longrightarrow y$

  $P(x, y, z) = P(y|z)P(z|x)P(x)$

- Note that we can alternatively expand $P(x, y, z)$ as

  $$P(x, y, z) = \left(\frac{P(z|y)P(y)}{P(z)}\right) P(x, z) = P(x|z)P(z|y)P(y)$$

  by applying Bayes Rule for $P(y|z)$ term and $P(z|x)P(x) = P(x, z)$

- Thus, the graph g is probabilistically equivalent to the chain

  $g': \quad x \longleftarrow z \longleftarrow y$

- When two graphs g and g' encode the same probability structure; we say they are Markov Equivalent (ME) and write $g \overset{\text{ME}}{\equiv} g'$

Chain Junction

- Two junctions are ME if they have the same moral graph = drop the arrows and "marry the parents", i.e.,

$$\bigcirc \rightarrow \bigcirc \rightarrow \bigcirc \quad \overset{\text{moralize}}{\Longrightarrow} \quad \bigcirc - \bigcirc - \bigcirc \quad \overset{\text{moralize}}{\Longleftarrow} \quad \bigcirc \leftarrow \bigcirc \leftarrow \bigcirc$$

- Being ME means that g g' are equivalent for prediction purposes; it does NOT mean that they are equivalent for causal modeling (this should be obvious)

- Note that $P(x, y) = \sum_z P(x, y, z) = P(x) \underbrace{\sum_z P(y|z)P(z|x)}_{\text{In general, not } = \ P(y)}$

  We expect that $P(x, y) \neq P(x)P(y) \iff \mathrm{x} \top\!\!\top \mathrm{y}$
  (Notation: $\mathrm{x} \top\!\!\top \mathrm{y} = \neg(\mathrm{x} \perp\!\!\!\perp \mathrm{y})$)

Chain Junction

- But if the numerical specification values are chosen to satisfy $P(y|z) = P(y)$ or $P(z|x) = P(z)$ then we can have $P(x, y) = P(x)P(y) \iff x \perp\!\!\!\perp_P y$ but this is highly unlikely (Notation: $\perp\!\!\!\perp_P$ "special P" indicates that this is an independency that is not shown in the graph)

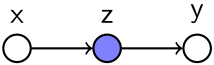
- We write: $\underset{}{\overset{x}{\bigcirc}} \longrightarrow \overset{z}{\bigcirc} \longrightarrow \overset{y}{\bigcirc} \implies \Diamond(x \top\!\!\!\top y)$

  $\Diamond$ denotes "likely", not guaranteed; it is likely that "information flows" from x to y

Chain Junction

- Now note that $P(x, y|z) = P(y|z)P(x|z)$ always structuraly true

  $\forall P(x, y, z)$ when conditioned on z, we have

  $\implies \Box(\mathrm{x} \perp\!\!\!\perp \mathrm{y} \mid \mathrm{z})$

  $\Box$ denotes "guaranteed, always necessarily"; it is guaranteed that information flow between x and y is blocked when conditioned on z

- This is the general theme: We can 100% guarantee independencies (they are structurally true) from graph separation, but NOT dependencies (they are likely but can be numerically untrue)