ECE 175B: Probabilistic Reasoning and Graphical Models: Lecture 7: Confounding in Causality and The Basic Junction Patterns – Part II

Florian Meyer & Ken Kreutz-Delgado

Electrical and Computer Engineering Department
University of California San Diego

UC San Diego
JACOBS SCHOOL OF ENGINEERING

Fork Junction

$$x \leftarrow z \rightarrow y \iff P(x,y,z) = P(x|z)P(y|z)P(z)$$

- Note that $P(x,y) = \sum_z P(x,y,z) \neq P(x)P(y)$ except when we numerically specify $P(x|z) = P(x)$ or $P(y|z) = P(y)$, i.e., we have

$$x \leftarrow z \rightarrow y \implies \Diamond(x \top\!\!\top y) \text{ (likely, not guaranteed)}$$

- On the other hand we always have $P(x,y,z) = P(x,z)P(y|z) \implies P(x,y|z) = P(x|z)P(y|z)$, i.e., now when conditioned on z, we have

$$x \leftarrow z \rightarrow y \implies \Box(x \perp\!\!\!\perp y \mid z) \text{ (guaranteed)}$$

Fork Junction

- Note that the Fork and Chain are Markov Equivalent (ME)

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

which can be shown directly:

$$P(x,y,z) = \underbrace{P(x|z)P(y|z)P(z)}_{\substack{x \quad\;\; z \quad\;\; y \\ \bigcirc \longleftarrow \bigcirc \longrightarrow \bigcirc}} = P(y|z)P(x,z)$$

$$= \underbrace{P(y|z)P(z|x)P(x)}_{\substack{x \quad\;\; z \quad\;\; y \\ \bigcirc \longrightarrow \bigcirc \longrightarrow \bigcirc}}$$

- Of course the Fork and Chain are NOT causally equivalent

Collider Junction

$$x \longrightarrow z \longleftarrow y \iff P(x,y,z) = P(z|x,y)P(x)P(y)$$

- $P(x,y) = \sum_z P(x,y,z) = P(x)P(y)\sum_z P(z|x,y) = P(x)P(y)$

  Thus, $x \longrightarrow z \longleftarrow y \implies \Box(x \perp\!\!\!\perp y)$ (unconditional independence)

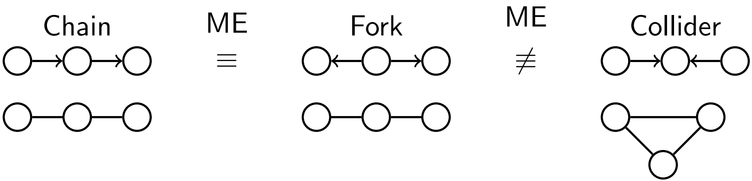
- On the other hand

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

  Except for special numerical specification, e.g., $P(z|x,y) = P(z|x)$
  Thus, when conditioned on z

$$x \longrightarrow z \longleftarrow y \implies \Diamond(x \top\!\!\!\top y|z) \text{ (likely, not guaranteed)}$$

Junction Properties



Graph World			Probability World
Type	Structure	d-connected	
Chain	$x \rightarrow z \rightarrow y$	$[x\|\emptyset\|y]_d$	$\Diamond(x \not\perp\!\!\!\perp y\|\emptyset)$
Fork	$x \leftarrow z \rightarrow y$	$[x\|\emptyset\|y]_d$	$\Diamond(x \not\perp\!\!\!\perp y\|\emptyset)$
Collider	$x \rightarrow z \leftarrow y$	$[x\|z\|y]_d$	$\Diamond(x \not\perp\!\!\!\perp y\|z)$

Information flow is likely unblocked (open)

Notation: $x \perp\!\!\!\perp y|\emptyset = x \perp\!\!\!\perp y$

Junction Properties

Graph World			Probability World
Type	Structure	d-separated	
Chain	$x \rightarrow z \rightarrow y$	$\langle x \| z \| y \rangle_d$	$\Box(x \perp\!\!\!\perp y \| z)$
Fork	$x \leftarrow z \rightarrow y$	$\langle x \| z \| y \rangle_d$	$\Box(x \perp\!\!\!\perp y \| z)$
Collider	$x \rightarrow z \leftarrow y$	$\langle x \| \emptyset \| y \rangle_d$	$\Box(x \perp\!\!\!\perp y \| \emptyset)$

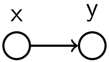
Information flow is guaranteed to be blocked

The Collider junction has the opposite behaviour compare to the Chain and Fork junctions

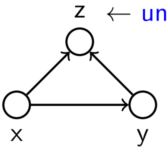
Junctions and Confounding

$\boxed{\text{Definition of Confounding: } P(y|\text{do}(x)) \neq P(y|x)}$ (seeing $\neq$ doing)

Consider the following situations that all assume a direct causal link from x to y

1. 

$P(y|\text{do}(x)) = P(y|x) \implies$ NO confounding

2. 

$\leftarrow$ unconditioned collider node

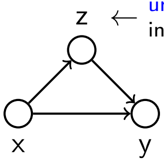
$$P(x,y,z) = P(z|x,y)P(y|x)P(x)$$

$$P(x,y) = \left(\sum_z P(z|x,y)\right) P(y|x)P(z) = P(y|x)P(x)$$

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

$\implies$ NO confounding

Junctions and Confounding

3. 

$\leftarrow$ unconditioned mediator (chain) node - allows an indirect path of x influencing y

direct influence of x on y

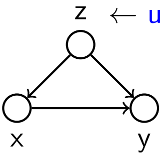
$$P(x,y,z) = P(y|x,z)\,P(z|x)\,P(x)$$

$$P(x,y) = P(x)\left(\sum_z P(y|x,z)\,P(z|x)\right) = P(y|x)\,P(x)$$

$$P(y|\mathrm{do}(x)) = P(y|x) = \sum_z \overbrace{P(y|x,z)}^{\text{z has an influence on y}}\ \underbrace{P(z|x)}_{\text{part of z due to x}}$$

$\implies$ NO confounding

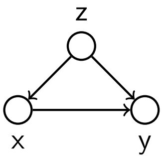
Junctions and Confounding

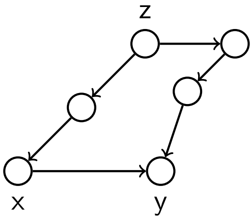
4. 

$\leftarrow$ unconditioned fork node $\Longrightarrow$ Confounding

- As we have seen, information flow between x and y "mixes up" prediction of y (for seeing z) and control of y (for doing x)
- Note that x $\leftarrow$ z $\rightarrow$ y is a confounding path, i.e., it implies that $P\big(y|\mathrm{do}(x)\big) \neq P(y|x)$,

"Backdoor Path" and Confounding



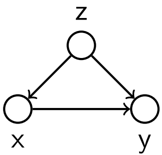


- Let us assume with have a directed edge x $\rightarrow$ y
- When another edge into x allows further information flow between x and y, we have a backdoor path that causes confounding
- Note that in the two cases shown, we can close the backdoor path by conditioning on z to block the information flow between x and y
- This is called controlling for confounding (caused by the backdoor path) and z is called a deconfounder variable

"Backdoor Path" and Confounding

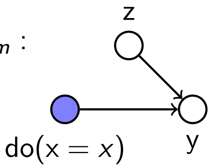
- Consider why this works for fork:

g :



$\overset{\text{Surgery}}{\Longrightarrow}$

g_m :



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

$$\overset{\text{Surgery}}{\Longrightarrow} P_m(x, y, z) = \overbrace{P_m(y|x, z)}^{P(y|x,z)} \underbrace{P_m(z)}_{P(z)} P_m(x)$$

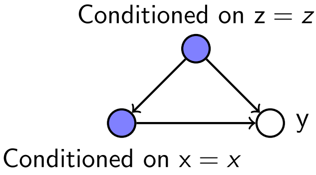
$$P_m(y|x) = \sum_z P_m(y, z|x) = \sum_z P_m(y|x, z)P_m(z)$$

$$P(y|\text{do}(x)) = \sum_z \underbrace{P(y|x, z)}_{\text{conditioned on z}} P(z) = \mathbb{E}_z\{P(y|x, z)\}$$

Fork Backdoor Path

- $P(y|\mathrm{do}(\mathrm{x}=x)) = \sum_z P(y|x,z)P(z) = \mathbb{E}_{\mathrm{z}}\{P(y|x,\mathrm{z})\}$
  By conditioning on $\mathrm{z} = z$ we determine the consequence of $\mathrm{do}(\mathrm{x} = x)$ for each "level" of z-value, $\mathrm{z} = z$

- In particular, we first determine



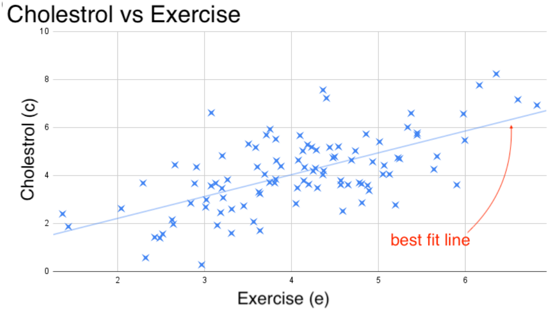
this gives $\underbrace{P(y|x,z)}_{\text{do for different values(levels) of z}}$

- We then get the average control effect of $\mathrm{x} = x$ by averaging $P(y|x,\mathrm{z}=z)$ over the different z-values

$$\boxed{\text{Backdoor Procedure: } P(y|\mathrm{do}(\mathrm{x}=x)) = \mathbb{E}_{\mathrm{z}}\{P(y|x,\mathrm{z})\}}$$

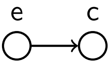
Confounding and Simpson's Paradox

- Simpson's "paradox" arises from confusing $P(y|x)$ with $P(y|\mathrm{do}(x))$
- Example: Doctors know that for a specific individual, exercise reduces cholesterol. A population study is done which yields the following scatter plot of cholesterol versus exercise

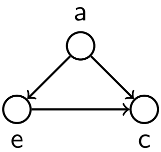


This seems paradoxical given what doctors see for each patient. What is happening?

Simpson's Paradox

- The assumed model is 

- However, in reality age a is a (de)confounder variable

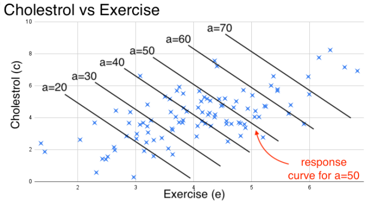


- We should determine the response $P(c|e, a)$ for each age stratum a $= a$, and then get the average of these responses to obtain

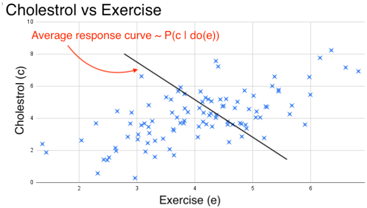
$$P(c|\mathrm{do}(e)) = E_{\mathrm{a}}\{P(c|e, a)\}$$

Simpson's Paradox

- This procedure yields



- We average over the age-dependent response to get



- It is apparent that $P(c|\mathrm{do}(e)) \neq P(c|e)$