Causality in Biomedicine Lecture Series: Lecture 4

Ava Khamseh (Biomedical AI Lab)

IGMM & School of Informatics

13 Nov, 2020

Causal Inference

- **Model** a causal inference problem with assumptions manifest in Causal Graphical Models **[Pearl]**

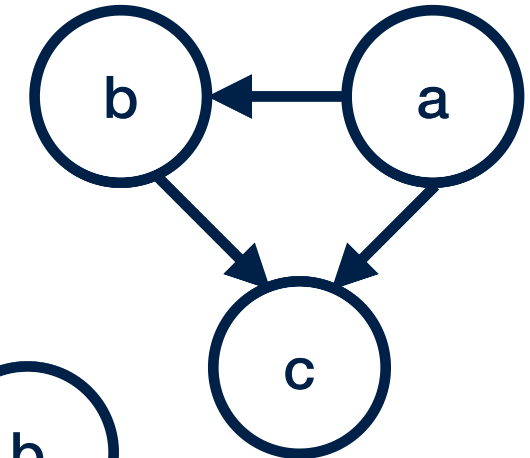
- **Identify** an expression for the causal effect under these assumptions ("causal estimand"), [**Pearl**]

- **Estimate** the expression using statistical methods such as matching or instrumental variables, [**Rubin**'s Potential Outcomes]

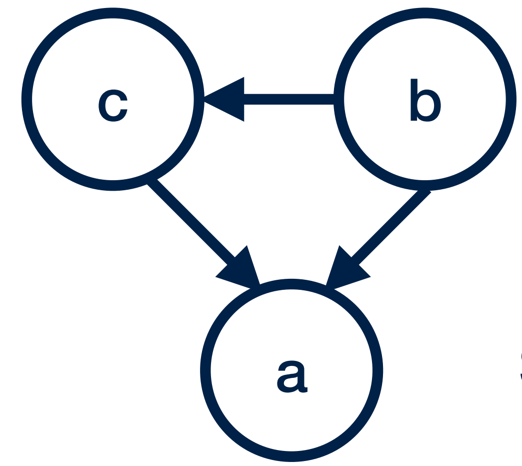
- **Verify** the validity of the estimate using a variety of robustness checks.

DAG contains more info than joint probability

$$p(a,b,c) = p(c|a,b)p(a,b) = p(c|a,b)p(b|a)p(a)$$



$$p(a,b,c) = p(a|b,c)p(b,c) = p(a|b,c)p(c|b)p(b)$$



Symmetric in a, b, c

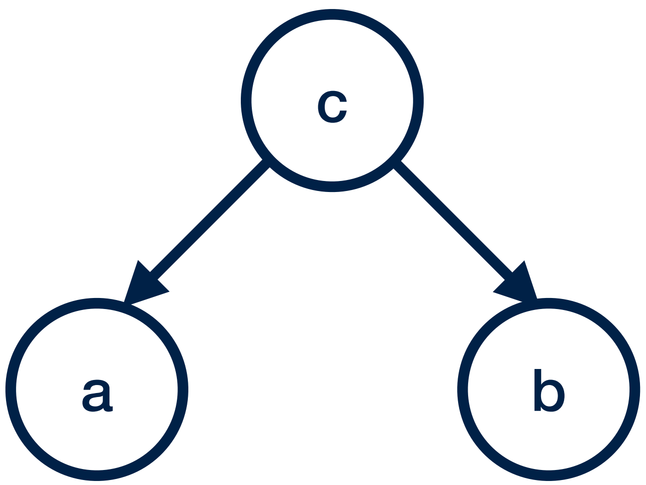
NOT symmetric

- Probabilistic notations are not enough to describe causal aspects
- Using repeated application of Bayes' rule, one can write any joint probability distribution in terms of its marginals
- A graph is **fully connected** if there is a link between every pair of nodes
- The interest lies in the **absence** of a link and link **direction**.

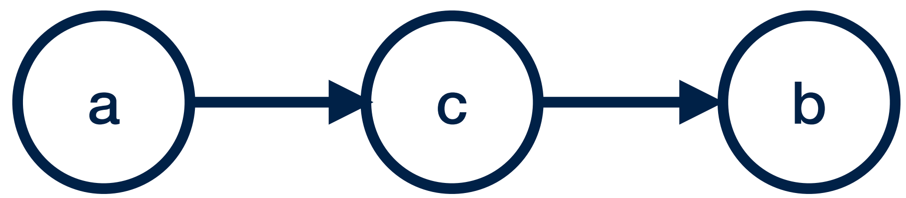
Bishop, Pattern Recognition and Machine Learning, Chapter 8

Recall:

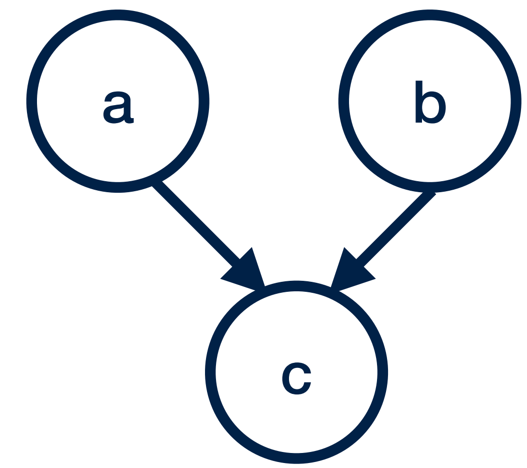
- Conditional independence via graphs and **D-separation**
- 3 main graph structures:



Fork



Chain



Collider

Fork:

$$a \not\perp\!\!\!\perp b | \emptyset$$
$$a \perp\!\!\!\perp b | c$$

Chain:

$$a \not\perp\!\!\!\perp b | \emptyset$$
$$a \perp\!\!\!\perp b | c$$

Collider:

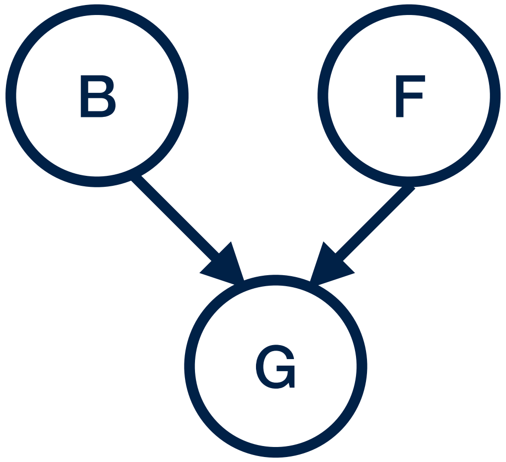
$$a \perp\!\!\!\perp b | \emptyset$$
$$a \not\perp\!\!\!\perp b | c$$

- **Do-calculus** and **causal identification**

Pearl, Causality (2009)

Collider example

B: State of battery, B=1 charged, B=0 flat
F: State of fuel tank, F=1 full, F=0 empty
G: State of electric fuel gauge, G=1 full, G=0 empty



Collider

Given Info:

$$p(B = 1) = 0.9$$

$$p(F = 1) = 0.9$$

$$p(G = 1 | B = 1, F = 1) = 0.8$$

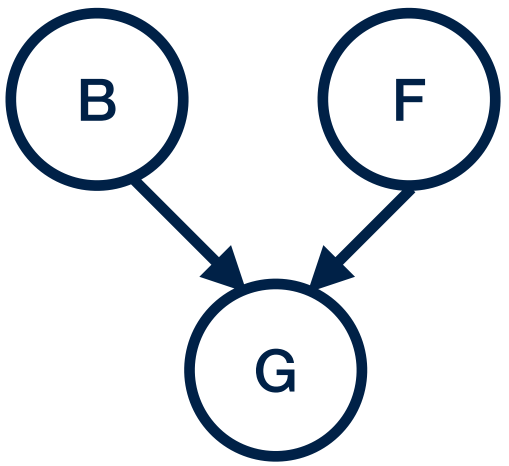
$$p(G = 1 | B = 1, F = 0) = 0.2$$

$$p(G = 0 | B = 0, F = 1) = 0.2$$

$$p(G = 1 | B = 0, F = 0) = 0.1$$

Collider example

B: State of battery, B=1 charged, B=0 flat
F: State of fuel tank, F=1 full, F=0 empty
G: State of electric fuel gauge, G=1 full, G=0 empty



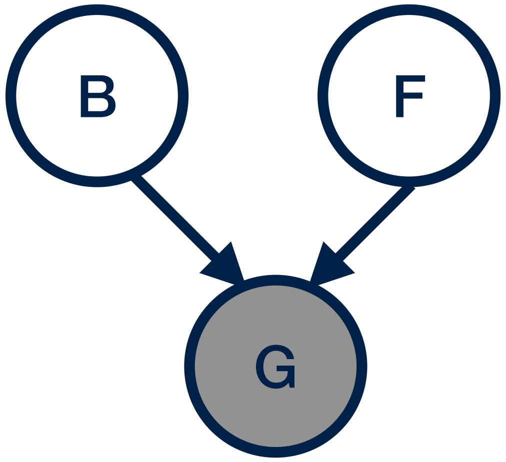
Collider

1 Before any conditioning (before observing):

$$p(F=0)=0.1$$

Collider example

B: State of battery, B=1 charged, B=0 flat
F: State of fuel tank, F=1 full, F=0 empty
G: State of electric fuel gauge, G=1 full, G=0 empty



Collider

1 Before any conditioning (before observing):

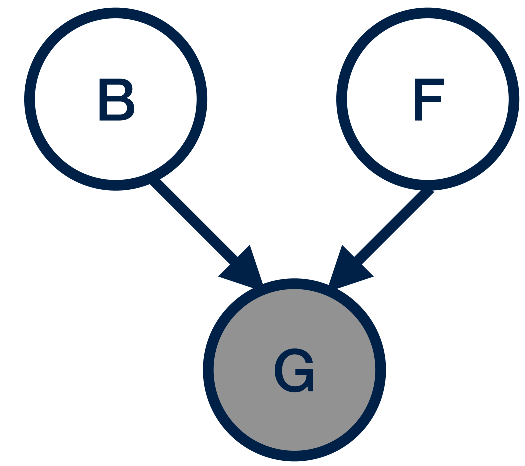
$$p(F=0)=0.1$$

2 Now suppose we observe G=0

$$p(F=0|G=0)=\frac{p(G=0|F=0)p(F=0)}{p(G=0)}$$

Collider example

B: State of battery, B=1 charged, B=0 flat
F: State of fuel tank, F=1 full, F=0 empty
G: State of electric fuel gauge, G=1 full, G=0 empty



Collider

1 Before any conditioning (before observing):

$$p(F=0)=0.1$$

2 Now suppose we observe G=0

$$p(F=0|G=0)=\frac{p(G=0|F=0)p(F=0)}{p(G=0)}$$

$$\sum_{B\in\{0,1\}} p(G=0|F=0,B)p(B)=0.81$$

$$\sum_{B,F\in\{0,1\}} p(G=0,B,F)$$

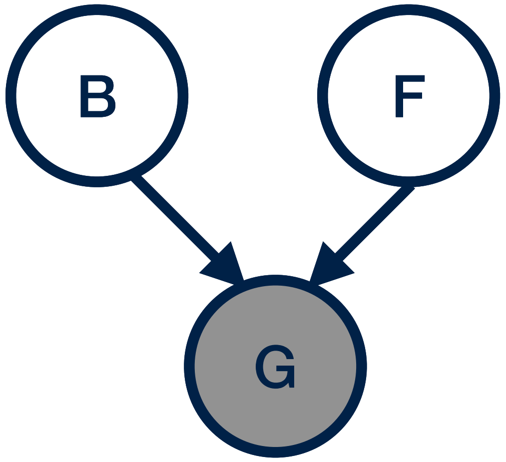
$$=\sum_{B,F\in\{0,1\}} p(G=0|B,F)p(B|F)p(F)$$

$$=\sum_{B,F\in\{0,1\}} p(G=0|B,F)p(B)p(F)=0.315$$

Since B and F are independent

Collider example

B: State of battery, B=1 charged, B=0 flat
F: State of fuel tank, F=1 full, F=0 empty
G: State of electric fuel gauge, G=1 full, G=0 empty



Collider

1 $p(F = 0) = 0.1$

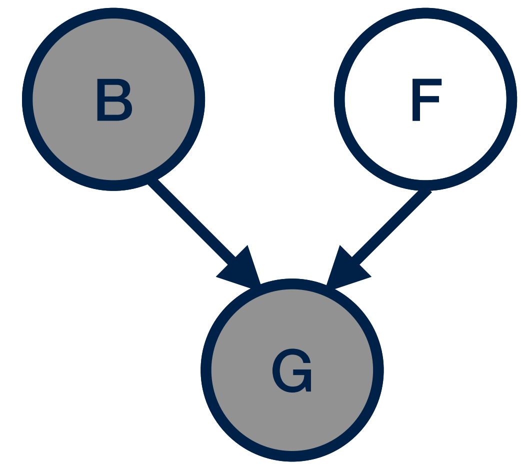
2 $p(F = 0 | G = 0) = 0.257$

$$p(F = 0) < p(F = 0 | G = 0)$$

Observing that gauge reads empty makes it more likely that the tank is indeed empty.

Collider example

B: State of battery, B=1 charged, B=0 flat
F: State of fuel tank, F=1 full, F=0 empty
G: State of electric fuel gauge, G=1 full, G=0 empty



Collider

2 $p(F=0|G=0) = 0.257$

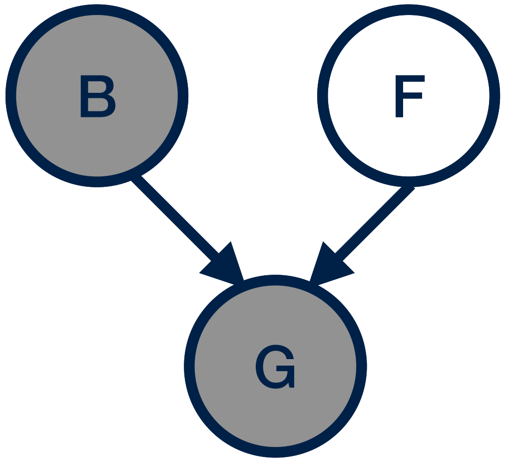
3 Now we **also** observe B=0

$$p(F=0|G=0,B=0) = \frac{p(F=0,G=0,B=0)}{p(G=0,B=0)}$$

$$= \frac{p(G=0|B=0,F=0)p(F=0)\cancel{p(B=0|F=0)}}{\sum_{F\in\{0,1\}} p(G=0|B=0,F)p(F)\cancel{p(B=0|F)}} = 0.111$$

Collider example

B: State of battery, B=1 charged, B=0 flat
F: State of fuel tank, F=1 full, F=0 empty
G: State of electric fuel gauge, G=1 full, G=0 empty



Collider

2 $$p(F=0|G=0)=0.257$$

3 Now we **also** observe B=0

$$p(F=0|G=0,B=0)=\frac{p(F=0,G=0,B=0)}{p(G=0,B=0)}$$

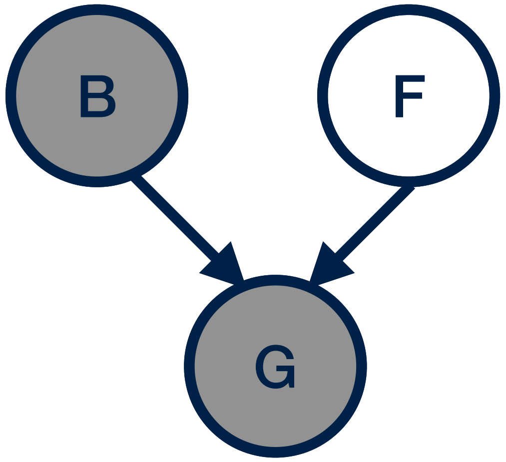
$$=\frac{p(G=0|B=0,F=0)p(F=0)\cancel{p(B=0|F=0)}}{\sum_{F\in\{0,1\}}p(G=0|B=0,F)p(F)\cancel{p(B=0|F)}}=0.111$$

$$p(F=0|G=0)>p(F=0|G=0,B=0)$$

Probability that tank is empty F=0 has decreased with extra information on the state of the battery

Collider example

B: State of battery, B=1 charged, B=0 flat
F: State of fuel tank, F=1 full, F=0 empty
G: State of electric fuel gauge, G=1 full, G=0 empty



Collider

1. $p(F=0)=0.1$
2. $p(F=0|G=0)=0.257$
3. $p(F=0|G=0,B=0)=0.111$

Conditioning on G, then finding out the battery is flat, ‘explains away’ the observation that the fuel gauge reads empty. The state of the fuel tank and the battery have become dependent:

$$p(F=0|G=0) \neq p(F=0|G=0,B=0)$$

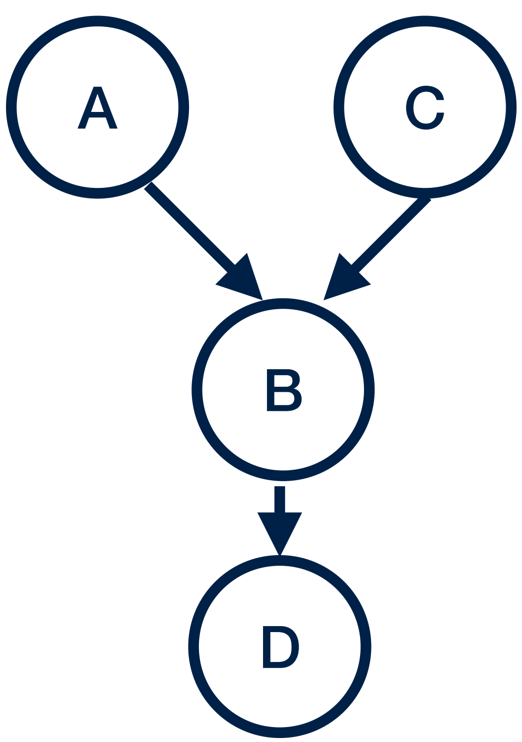
(Even though: $p(F)=p(F|B)$)

D-separation

A path p is **blocked** by a set of nodes Z if and only if:

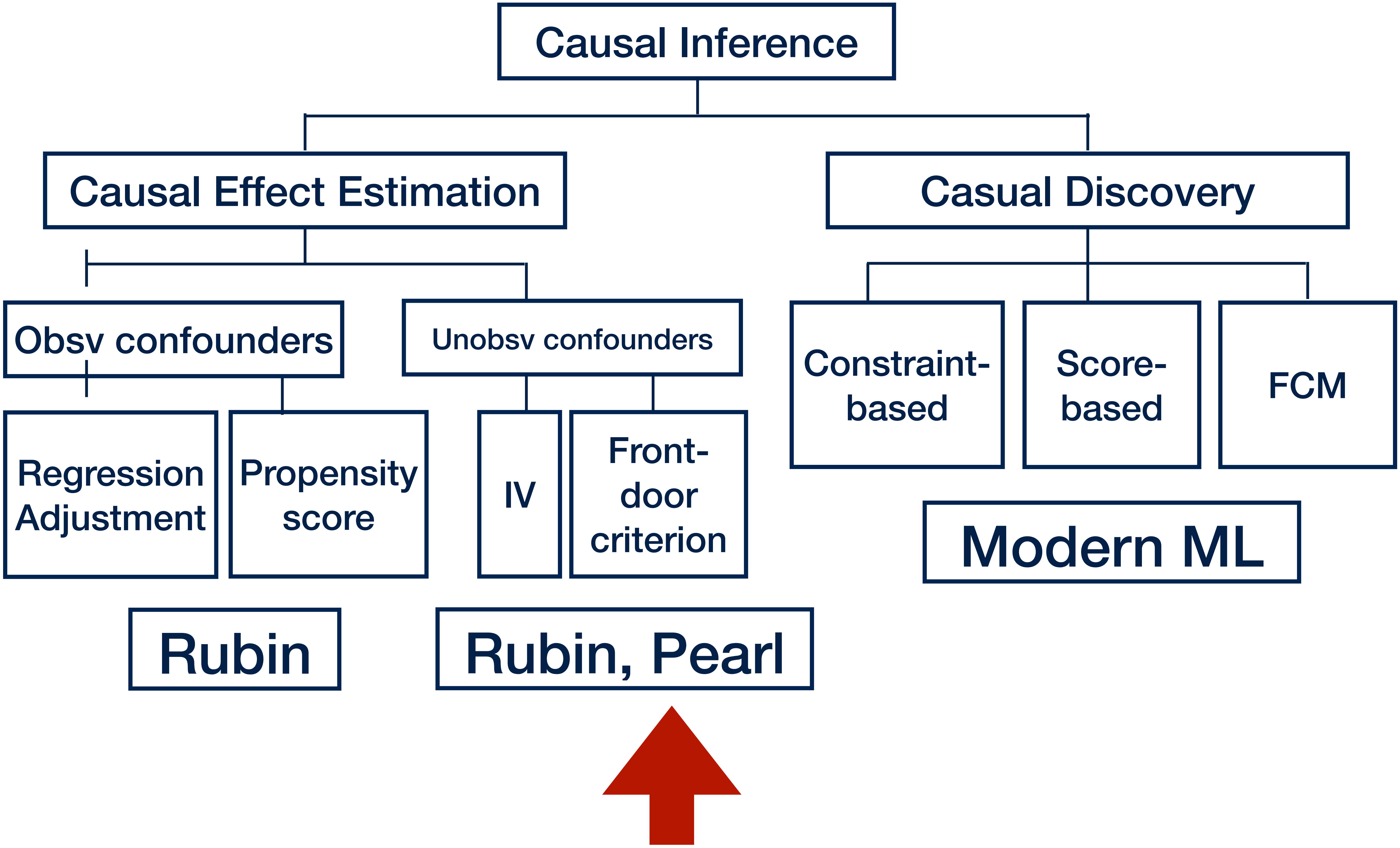
1) p contains a **chain** of nodes A -> B -> C or a **fork** A <- B -> C such that the middle node B is in Z (i.e. B is conditioned on), or

2) p contains a **collider** A -> B <- C such that the collision node B is not in Z, and no descendant of B is in Z.



Pearl, Causality (2009)

Overview of the course

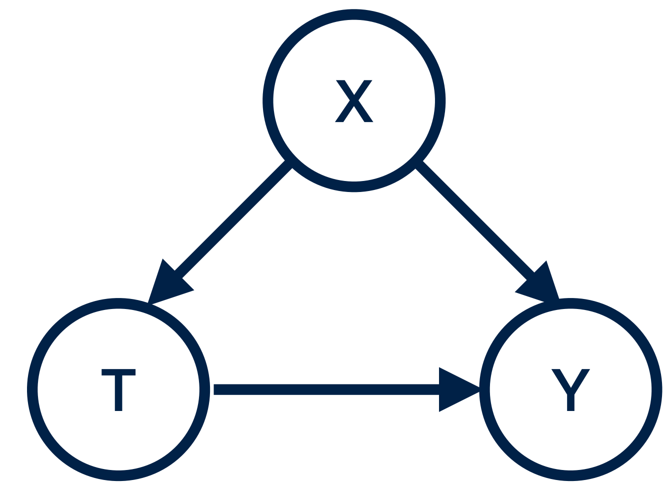


Pearl's framework

Graphical models & Do-calculus

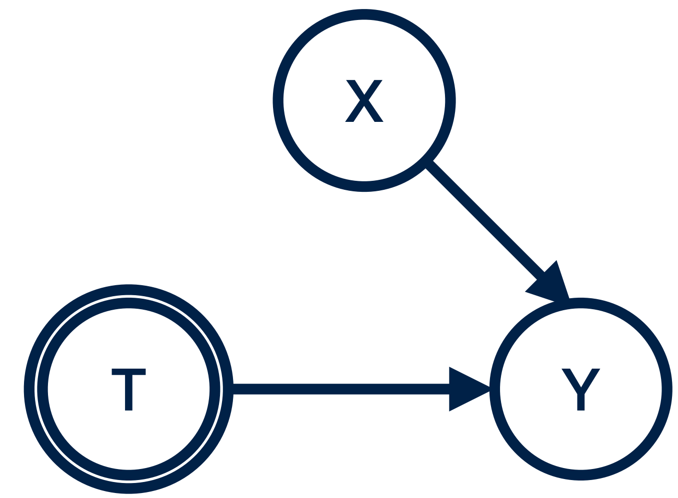
Observation (conditioning) vs intervention

Distinguish between: a variable T takes a value t naturally and cases where we **fix** T=t by denoting the latter do(T=t)



$$p(Y = y|T = t)$$

Probability that Y=y **conditional** on finding T=t
i.e., population distribution of Y among individuals whose T value is t (subset)



$$p(Y = y|do(T = t))$$

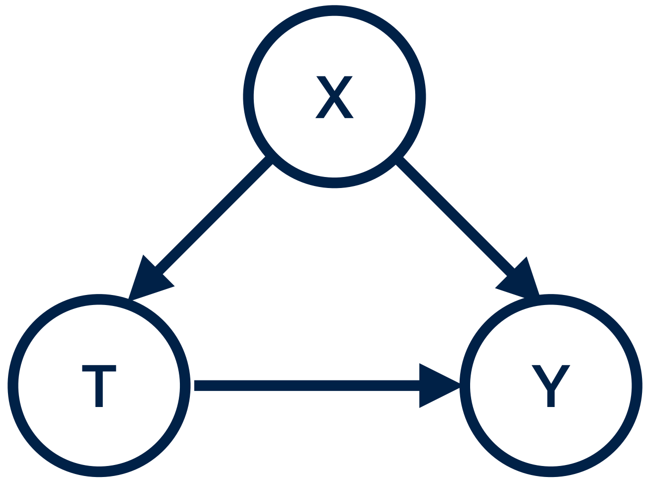
Probability that Y=y when we **intervene** to make T=t
i.e., population distribution of Y if **everyone in the population** had their T value fixed at t.

$\boxed{\textbf{Graph surgery}}$

Pearl, Causal Inference in Statistics (2016)

The adjustment formula

T: Drug usage
X: Gender
Y: Recovery



To know how effective the drugs is in the population, compare the **hypothetical interventions** by which

(i) the drug is administered uniformly to the entire population do(T=1) **vs**

(ii) complement, i.e., everyone is prevented from taking the drug do(T=0)

Aim: Estimate the difference (**Average Causal Effect ACE**)

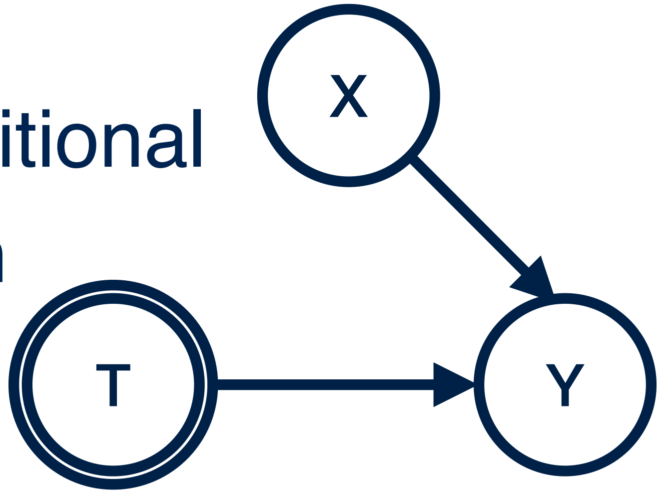
$$\boxed{p(Y = 1|do(T = 1)) - p(Y = 1|do(T = 0))}$$

Pearl, Causal Inference in Statistics (2016)

The adjustment formula

Using a **causal theory**, we aim to write $p(Y = y | do(T = t))$ in terms of quantities we can compute from the data, i.e., conditional probabilities.

The causal effect $p(Y = y | do(T = t))$ is equal to conditional probability $p_m(Y = y | T = t)$ in the manipulated graph



Key observation: p_m shares 2 properties with p :

(i) $p_m(X = x) = p(X = x)$ is **invariant** under the intervention, X is not affected by removing the arrow from X to T, i.e. the proportion of males and females remain the same before and after the intervention

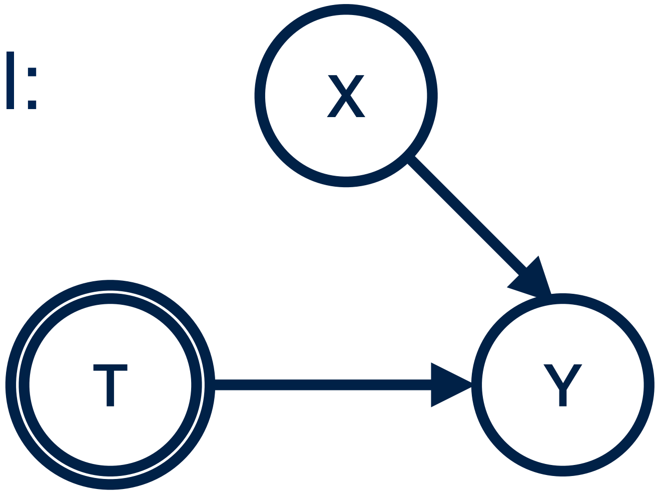
(ii) $p_m(Y = y | X = x, T = t) = p(Y = y | X = x, T = t)$ is **invariant**

Pearl, Causal Inference in Statistics (2016)

The adjustment formula

Moreover, T and X are d-separated in the modified model:

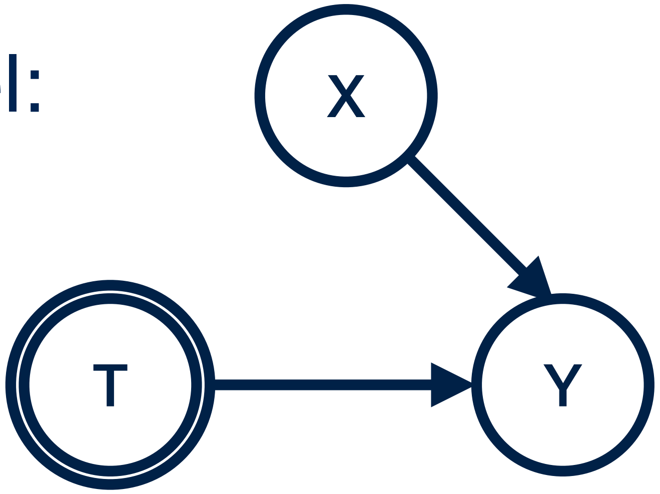
$$p_m(X = x | T = t) = p_m(X = x) = p(X = x) \quad *$$



The adjustment formula

Moreover, T and X are d-separated in the modified model:

$$p_m(X = x|T = t) = p_m(X = x) = p(X = x) \quad *$$



Putting these together:

$$p(Y = y|do(T = t)) = p_m(Y = y|T = t) \quad \text{by definition}$$

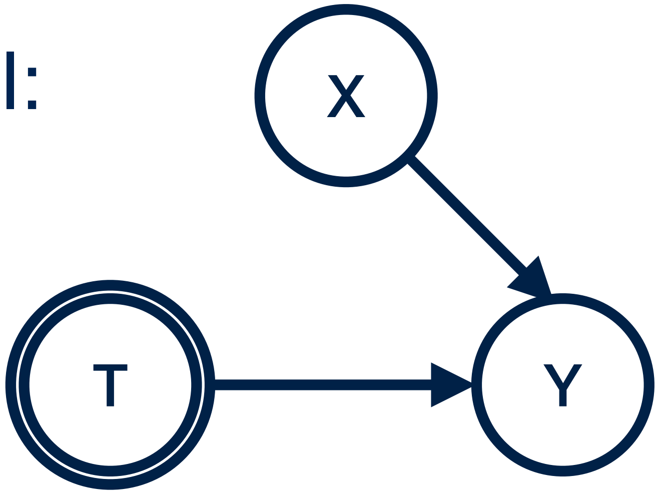
$$\sum_x p_m(Y = y|T = t, X = x)p_m(X = x|T = t) \quad \text{law of total prob}$$

$$\sum_x p_m(Y = y|T = t, X = x)p_m(X = x) \quad *$$

Pearl, Causal Inference in Statistics (2016)

The adjustment formula

Moreover, T and X are d-separated in the modified model:



$$p_m(X = x | T = t) = p_m(X = x) = p(X = x) \;\; *$$

Putting these together:

$$p(Y = y | do(T = t)) = p_m(Y = y | T = t) \quad \text{by definition}$$

$$\sum_x p_m(Y = y | T = t, X = x) p_m(X = x | T = t) \quad \text{law of total prob}$$

$$\sum_x p_m(Y = y | T = t, X = x) p_m(X = x) \;\; *$$

Using the two invariance relations, we have the **adjustment formula**:

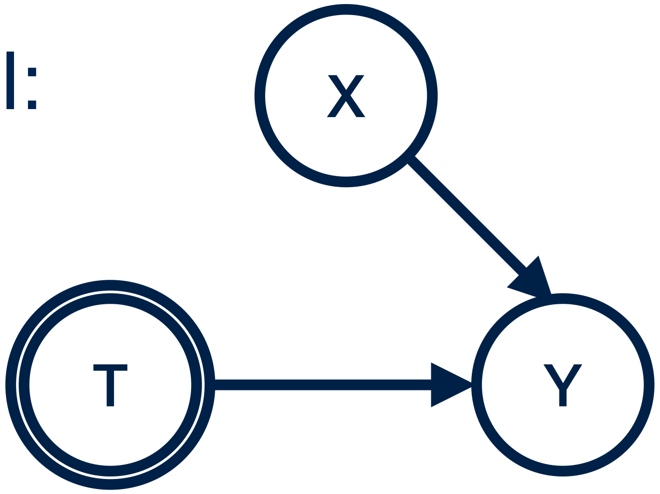
$$\boxed{p(Y = y | do(T = t)) = \sum_x p(Y = y | T = t, X = x) p(X = x)}$$

Pearl, Causal Inference in Statistics (2016)

The adjustment formula

Moreover, T and X are d-separated in the modified model:

$$p_m(X = x|T = t) = p_m(X = x) = p(X = x) \;\; *$$



Putting these together:

$$p(Y = y|do(T = t)) = p_m(Y = y|T = t) \quad \text{by definition}$$

$$\sum_x p_m(Y = y|T = t, X = x)p_m(X = x|T = t) \quad \text{law of total prob}$$

$$\sum_x p_m(Y = y|T = t, X = x)p_m(X = x) \;\; *$$

Use Pm as an intermediate tool

Using the two invariance relations, we have the **adjustment formula**:

$$\boxed{p(Y = y|do(T = t)) = \sum_x p(Y = y|T = t, X = x)p(X = x)}$$

Pearl, Causal Inference in Statistics (2016)

The adjustment formula

$$\boxed{p(Y = y | do(T = t)) = \sum_x p(Y = y | T = t, X = x) p(X = x)}$$

Adjusting for X (controlling for X) … **seen before**?

Example: T=1 taking the drug, X=1 male, Y=1 recovery

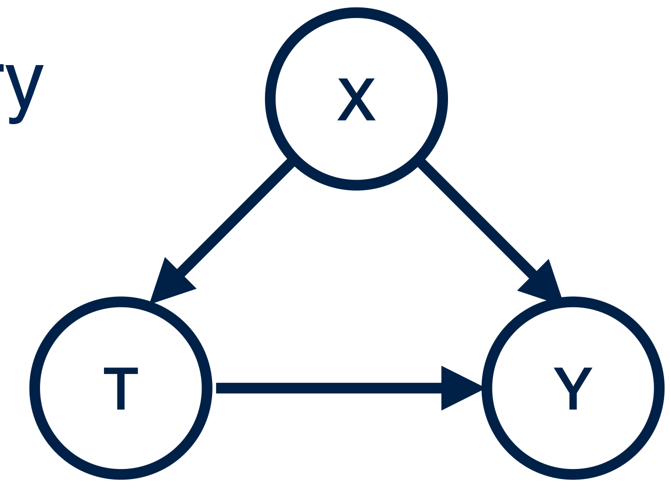


Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Pearl, Causal Inference in Statistics (2016)

The adjustment formula

$$\boxed{p(Y=y|do(T=t)) = \sum_x p(Y=y|T=t, X=x)p(X=x)}$$

T=1 taking drug
X=1 male
Y=1 recovery

$$p(Y=y|do(T=1)) = p(Y=1|T=1, X=1)p(X=1) + p(Y=1|T=1, X=0)p(X=0)$$

$$p(Y=1|do(T=1)) = \frac{0.93(87+270)}{700} + \frac{0.73(263+80)}{700} = 0.832$$

$$p(Y=1|do(T=0)) = \frac{0.87(87+270)}{700} + \frac{0.69(263+80)}{700} = 0.7818$$

$$ACE : p(Y=1|do(T=1)) - p(Y=1|do(T=0)) = 0.832 - 0.7818 = 0.0505$$ ✓

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Pearl, Causal Inference in Statistics (2016)

The adjustment formula

$$\boxed{p(Y=y|do(T=t)) = \sum_x p(Y=y|T=t, X=x)p(X=x)}$$

T=1 taking drug
X=1 male
Y=1 recovery

$$p(Y=y|do(T=1)) = p(Y=1|T=1, X=1)p(X=1) + p(Y=1|T=1, X=0)p(X=0)$$

$$p(Y=1|do(T=1)) = \frac{0.93(87+270)}{700} + \frac{0.73(263+80)}{700} = 0.832$$

$$p(Y=1|do(T=0)) = \frac{0.87(87+270)}{700} + \frac{0.69(263+80)}{700} = 0.7818$$

Stratification!

Note equivalence to Rubin's FW

$$ACE : p(Y=1|do(T=1)) - p(Y=1|do(T=0)) = 0.832 - 0.7818 = 0.0505$$

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Pearl, Causal Inference in Statistics (2016)

Pearl & Rubin

Pearl

$$\mathbb{E}(Y|do(T=1)) = \mathbb{E}(Y|T=1, X=1)p(X=1) + \mathbb{E}(Y|T=1, X=0)p(X=0)$$
$$\mathbb{E}(Y|do(T=0)) = \mathbb{E}(Y|T=0, X=1)p(X=1) + \mathbb{E}(Y|T=0, X=0)p(X=0)$$
$$\mathbb{E}(Y|do(T=1)) - \mathbb{E}(Y|do(T=0))$$

Rubin recall potential outcomes $y_0^{(i)}$ and $y_1^{(i)}$ and ATE:

$$\tau = \hat{\mathbb{E}}[\tau^{(i)}] = \hat{\mathbb{E}}[y_1^{(i)} - y_0^{(i)}] = \frac{1}{N}\sum_{i=0}^{N}\left(y_1^{(i)} - y_0^{(i)}\right)$$

Pearl & Rubin

Pearl

$$\mathbb{E}(Y|do(T=1)) = \mathbb{E}(Y|T=1, X=1)p(X=1) + \mathbb{E}(Y|T=1, X=0)p(X=0)$$
$$\mathbb{E}(Y|do(T=0)) = \mathbb{E}(Y|T=0, X=1)p(X=1) + \mathbb{E}(Y|T=0, X=0)p(X=0)$$
$$\mathbb{E}(Y|do(T=1)) - \mathbb{E}(Y|do(T=0))$$

Rubin recall potential outcomes $y_0^{(i)}$ and $y_1^{(i)}$ and ATE:

$$\tau = \hat{\mathbb{E}}[\tau^{(i)}] = \hat{\mathbb{E}}[y_1^{(i)} - y_0^{(i)}] = \frac{1}{N}\sum_{i=0}^{N}\left(y_1^{(i)} - y_0^{(i)}\right)$$
$$= \frac{1}{N}\left(\sum_{i \in \text{males}}\left(y_1^{(i)} - y_0^{(i)}\right) + \sum_{i \in \text{females}}\left(y_1^{(i)} - y_0^{(i)}\right)\right)$$

Pearl: To adjust or not to adjust

The previous example may give the impression that X-specific analysis, as compared to nonspecific, is the correct way forward. This is not the case. For example, let T=drug, Y=recovery, X= blood pressure **post-treatment,** i.e., important to take into account **how** the data is generated. Here, we know:

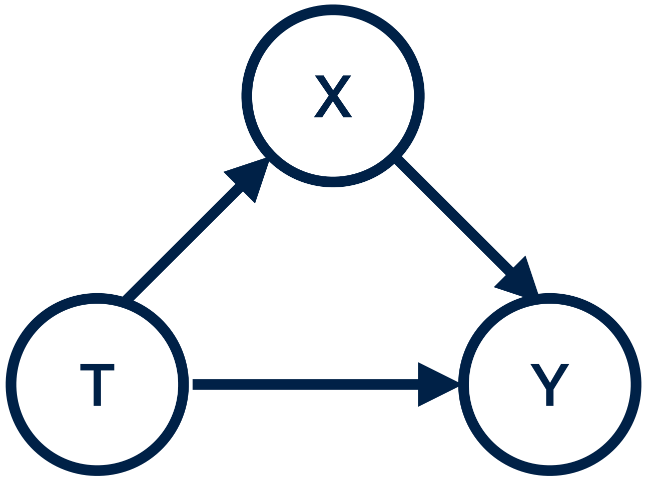
(i) the drug affects recovery by lowering the blood pressure

(ii) but it has a toxic effect for those who take it

NB: Data (numbers) in this table are identical to those in Table 1.1.

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No drug	Drug
Low BP	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High BP	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)



Pearl, Causal Inference in Statistics (2016)

Pearl: To adjust or not to adjust

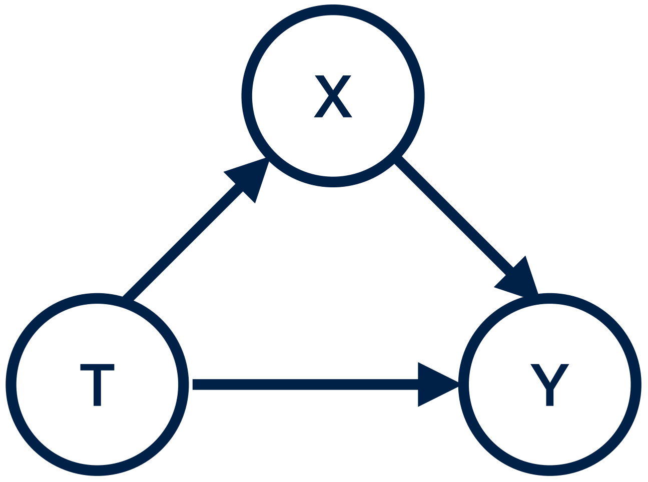
For general population, the drug might improve recovery rates because of its effect on blood pressure. But in low BP/high BP **post-treatment** subpopulations, we only observe the toxic effect of the drug.

Aim, as before, to gauge the overall causal effect of the drug on recovery.
Unlike before, it does **not** make sense to separate results by blood pressure as treatment affect recovery via reducing BP.
Contrast this with the a situation per BP is measure **before** treatment and direction of arrow from T to X is reversed.

Therefore, we **should** recommend treatment in this case because 78% < 83% .

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No drug	Drug
Low BP	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High BP	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)



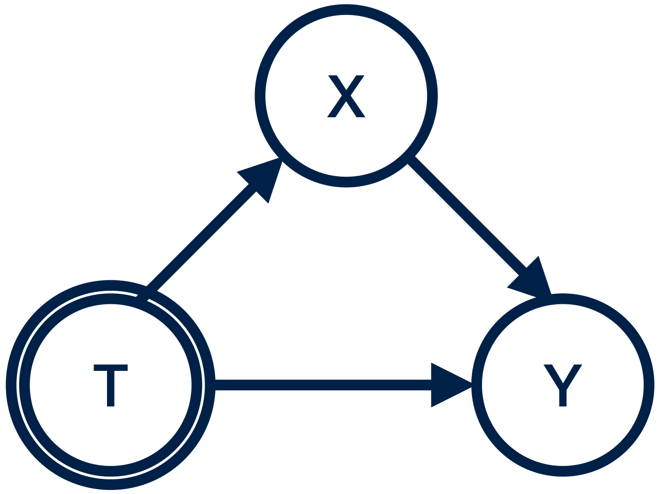
Pearl, Causal Inference in Statistics (2016)

Pearl: To adjust or not to adjust

Pearls algorithmic approach tells us to adjust or not. Starting with: $p(Y=1|do(T=1))$, intervene on T. But since no arrow is entering T, there will be no change in the graph: $p(Y=1|do(T=1)) = p(Y=1|T=1)$

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No drug	Drug
Low BP	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High BP	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)



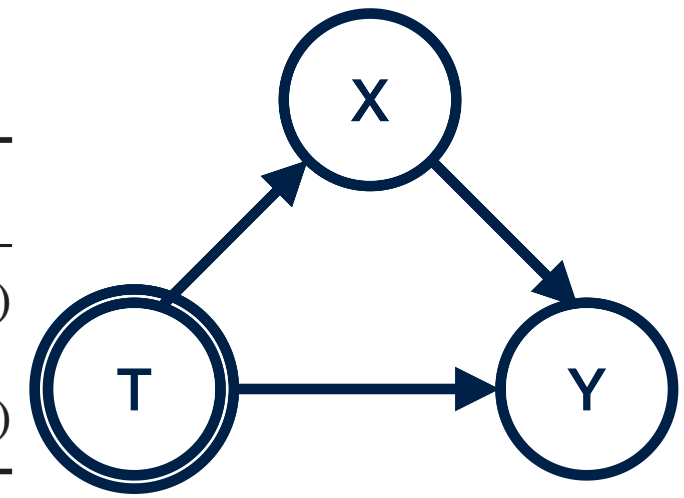
Pearl, Causal Inference in Statistics (2016)

Pearl: To adjust or not to adjust

Pearls algorithmic approach tells us to adjust or not. Starting with: $p(Y=1|do(T=1))$, intervene on T. But since no arrow is entering T, there will be no change in the graph: $p(Y=1|do(T=1)) = p(Y=1|T=1)$

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No drug	Drug
Low BP	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High BP	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)



The Causal Effect Rule: Given a graph G in which a set of variables PA are designated as the parents of T, the causal effect of T on Y is given by:

$$\boxed{p(Y=y|do(T=t)) = \sum_x p(Y=y|T=t, PA=X)p(PA=X)}$$

Pearl, Causal Inference in Statistics (2016)

The Backdoor Criterion

Under what conditions does a causal model permit computing the causal effect of one variable on another, from **data** obtained from **passive observations**, with **no intervention**?
i.e.,
Under what conditions is the structure of a causal graph sufficient of computing a causal effect from a given data set? **Identifiability**

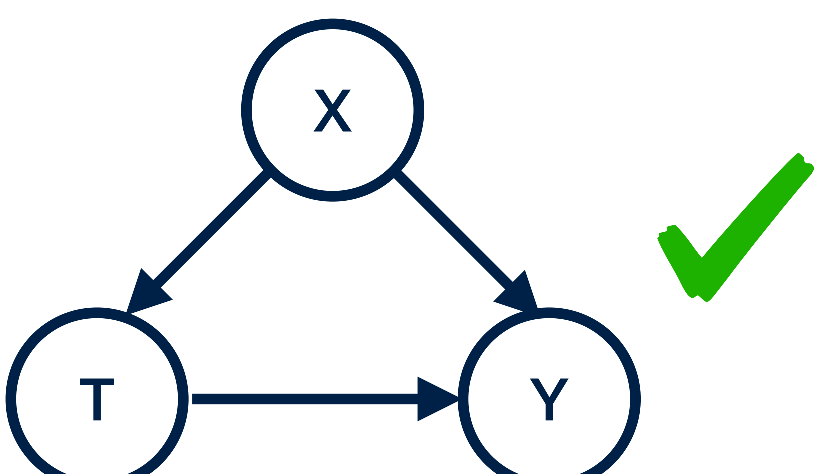
Backdoor Criterion: Given an ordered pair of variables (T,Y) in a DAG G, a set of variables X satisfies the backdoor criterion relative to (T,Y) if:

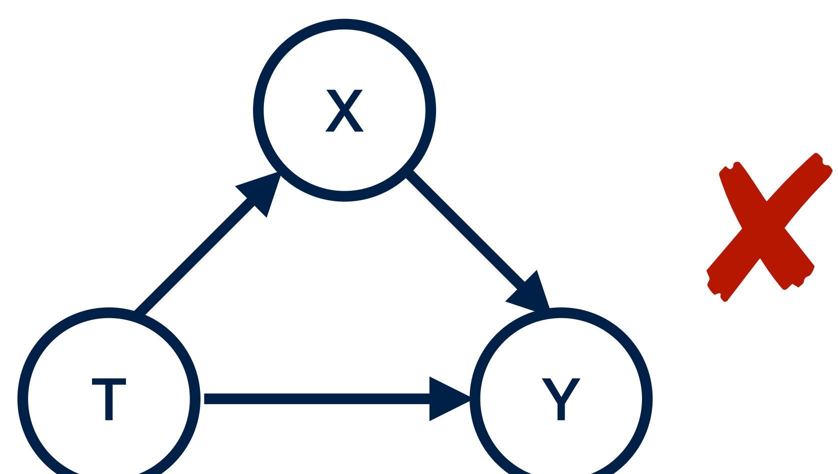
(i) no node in X is a descendent of T

(ii) X block every path between T and Y that contains an arrow into T

If X satisfies the backdoor criterion then the causal effect of T on Y is given by:

$$p(Y=y|do(T=t)) = \sum_x p(Y=y|T=t, X=x)p(X=x)$$





Pearl, Causal Inference in Statistics (2016)

Rubin vs Pearl

Rubin	Pearl
SUTVA	Implicit assumption of no interference between any pairs of individual
Unconfoundedness	Back-door criterion satisfied
Potential outcomes: $y_0^{(i)}$, $y_1^{(i)}$ Observed: $y_{\underline{0}}^{(i)}$, Unobserved: $y^*{}_1^{(i)}$	Counterfactuals are equivalent to individual unobserved outcomes in Rubin

Causality in Biomedicine Lecture Series: Lecture 4

Ava Khamseh (Biomedical AI Lab)

IGMM & School of Informatics

13 Nov, 2020