Causality in Biomedicine Lecture Series: Lecture 6

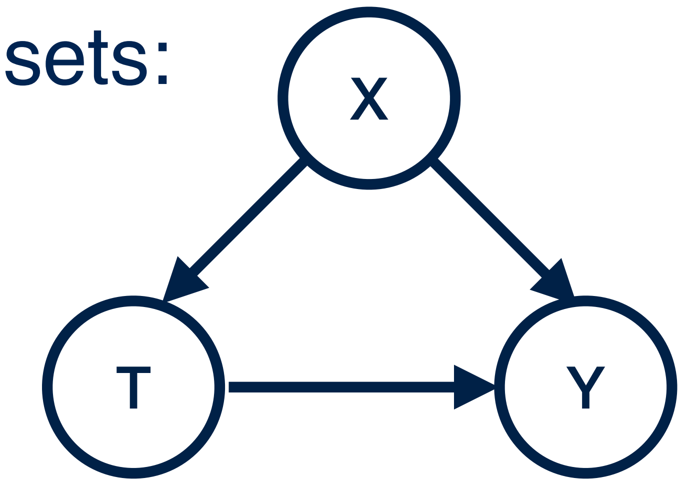
Ava Khamseh (Biomedical AI Lab)

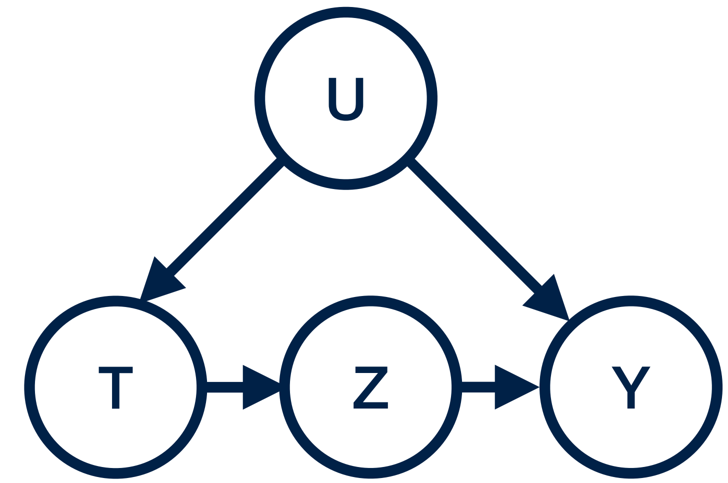
IGMM & School of Informatics

27 Nov, 2020

Pearl's Do-Calculus

- Do-calculus: Contains, as subsets:
  - Back-door criterion
  - Front-door criterion



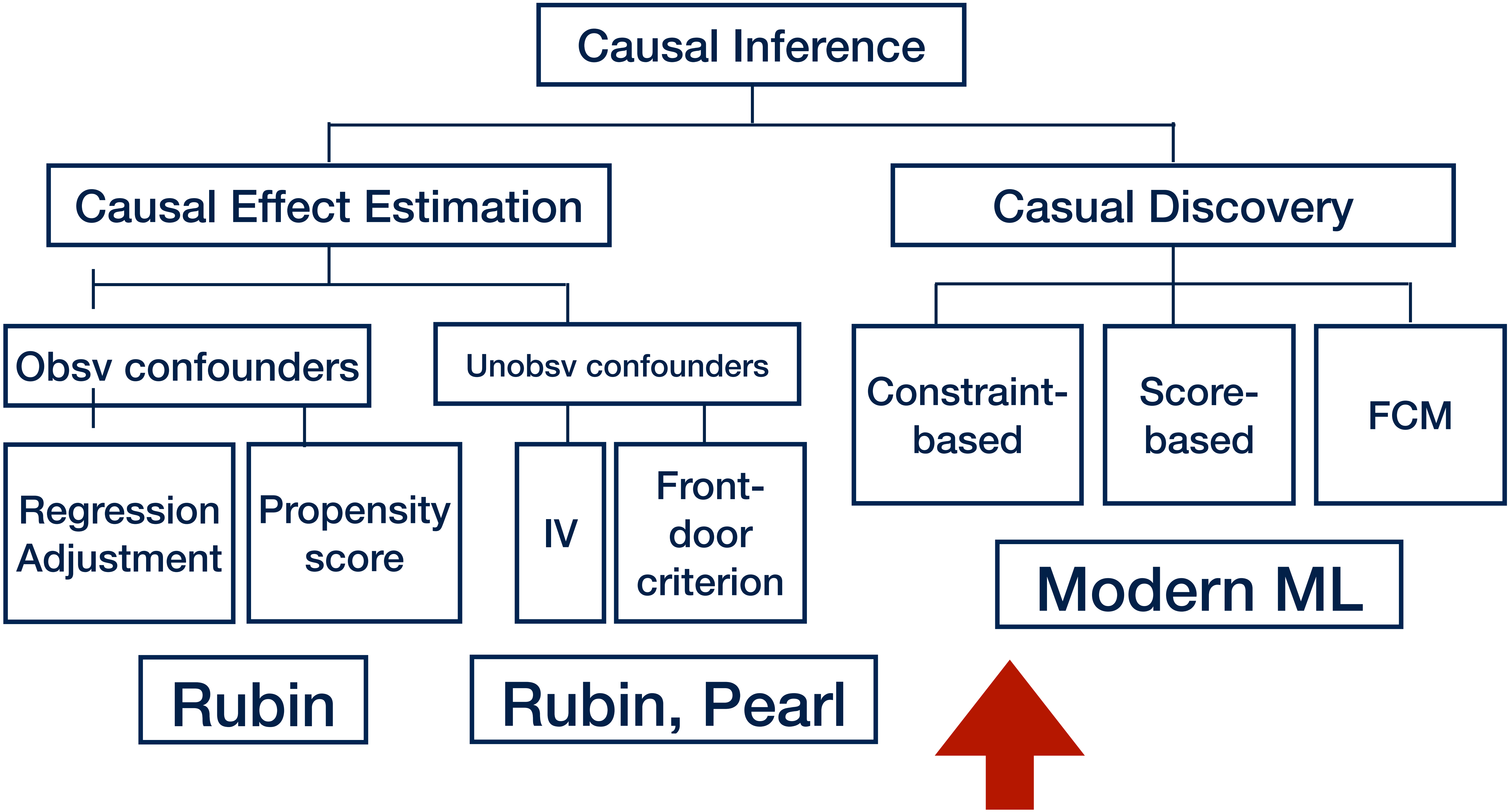


- Allows analysis of more intricate structure beyond back- and front-door

- Uncovers **all** causal effects that can be identified from a given causal graph

- Power of causal graphs is not just representation but actually **discovery** of causal information

Overview of the course



Causal Discovery
(Generally Pearl)

Learning causal relationships: Learn set of edges

- Causal axioms guide us in how a causal structure **constrains** the possible types of probability distribution that can be generated from that structure.

Learning causal relationships: Learn set of edges

- Causal axioms guide us in how a causal structure **constrains** the possible types of probability distribution that can be generated from that structure.
- Reverse: Obtain causal structures from probability distributions via causal inference

Learning causal relationships: Learn set of edges

- Causal axioms guide us in how a causal structure **constrains** the possible types of probability distribution that can be generated from that structure.
- Reverse: Obtain causal structures from probability distributions via causal inference
- Types of constraints: **Conditional independencies** (all parametric distributions), Vanishing determinants of partial covariance matrices (linear Gaussian with unobserved confounders), **Unequal dependence on residuals** (Non-linear additive noise, or linear non-Gaussian), **interventions/ perturbations**, time-series …

Causal Discovery Methods (Based on Graphical Models)

Class of Algorithm	Name	Assumptions	Short comings	Input
Constraint-based	PC (oldest)	Any distribution, No unobsv. confounders, Markov cond, faithfulness	Causal info only up to equivalence classes, Non bivariate	Complete undirected graph
	FCI	Any distribution, Asymptotically correct with confounders, Markov cond, faithfulness		
Score-based	GES	No unobsv. confounders	Non-bivariate	Empty graph, adds edges, removes some
Functional Causal Models (FCMs)	LinGAM/ ANM	Asymmetry in data	Requires additional assumptions (not general), harder for discrete data	Structural Equation Model

Glymour et al. (2019)

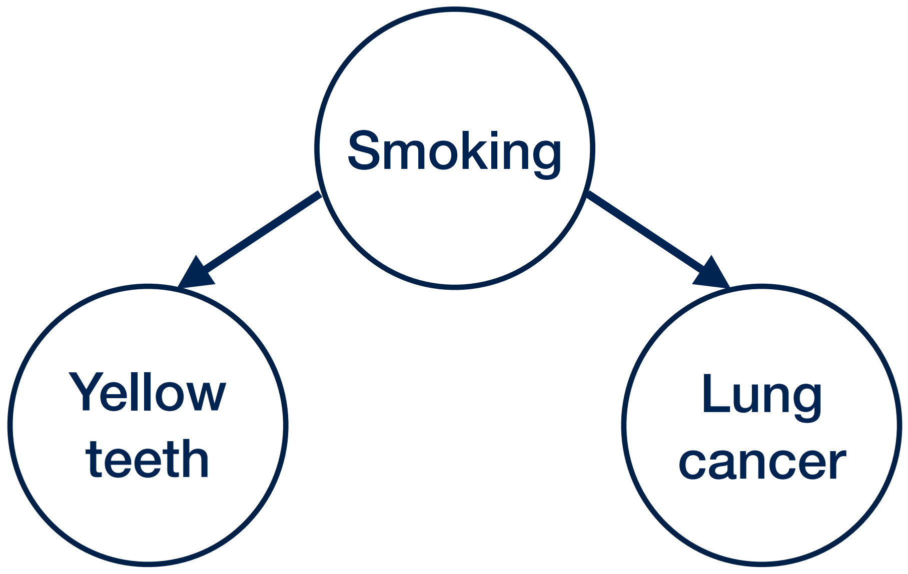
Assumptions 1: The Markov Condition

Any variable X is independent of all other variables, conditional on its parents (PA) and unobserved variables (noise):

$$P(x_1, \cdots, x_n) = \prod_{j=1}^{J} P(x_j | PA_j, \epsilon_j)$$

- Absent edge implies conditional independence (**CI**)
- Observing conditional dependence implies an edge

For example: Yellow teeth, lung cancer, smoking



An edge is wrongly inferred, when parent is omitted

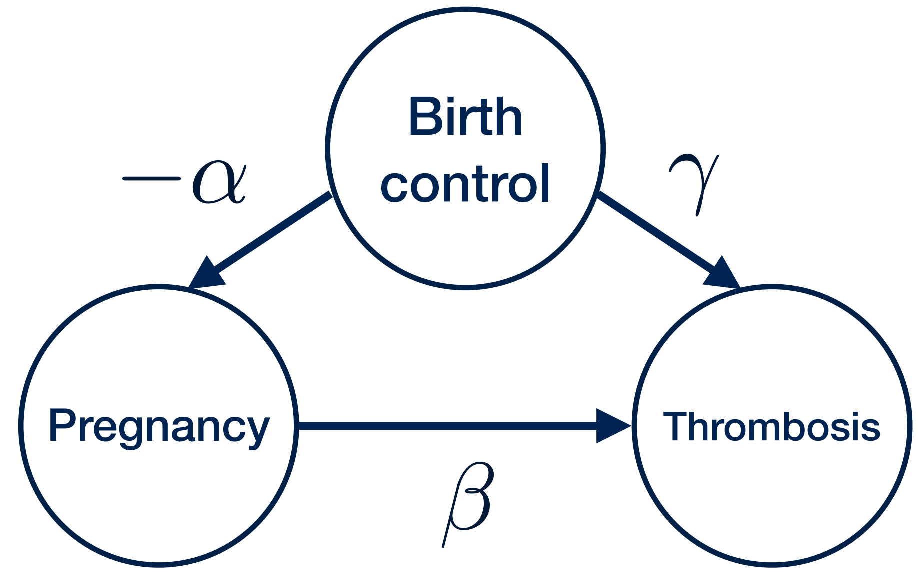
Assumptions 3: Faithfulness

It **fails** when distributions are set up in such a way that paths exactly cancel:

$$P = -\alpha B + U_P$$

$$T = \beta P + \gamma B + U_T$$

$$\Rightarrow T = (-\alpha\beta + \gamma)B + U$$



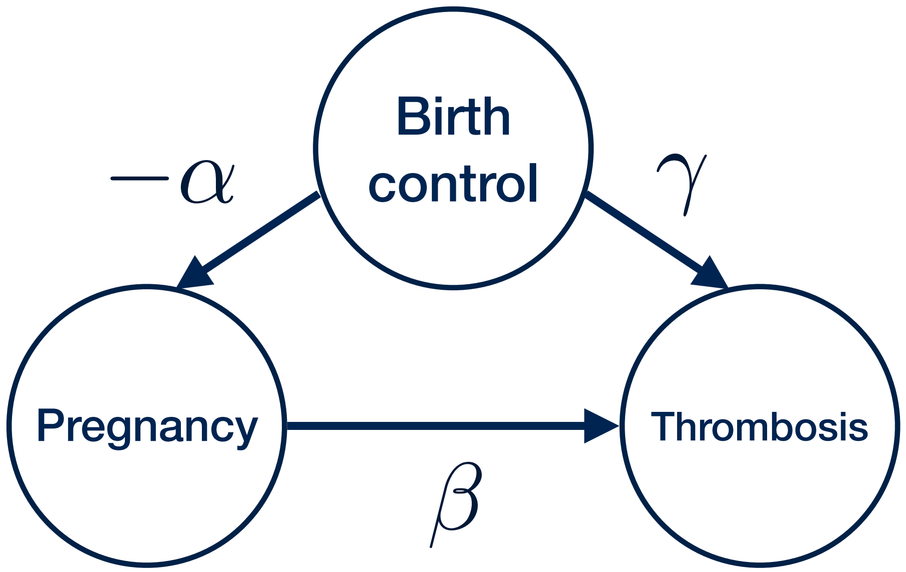
Assumptions 3: Faithfulness

It **fails** when distributions are set up in such a way that paths exactly cancel:

$$P = -\alpha B + U_P$$

$$T = \beta P + \gamma B + U_T$$

$$\Rightarrow T = (-\alpha\beta + \gamma)B + U$$



So if $\gamma = \alpha\beta$, no dependency between T and B will be observed!

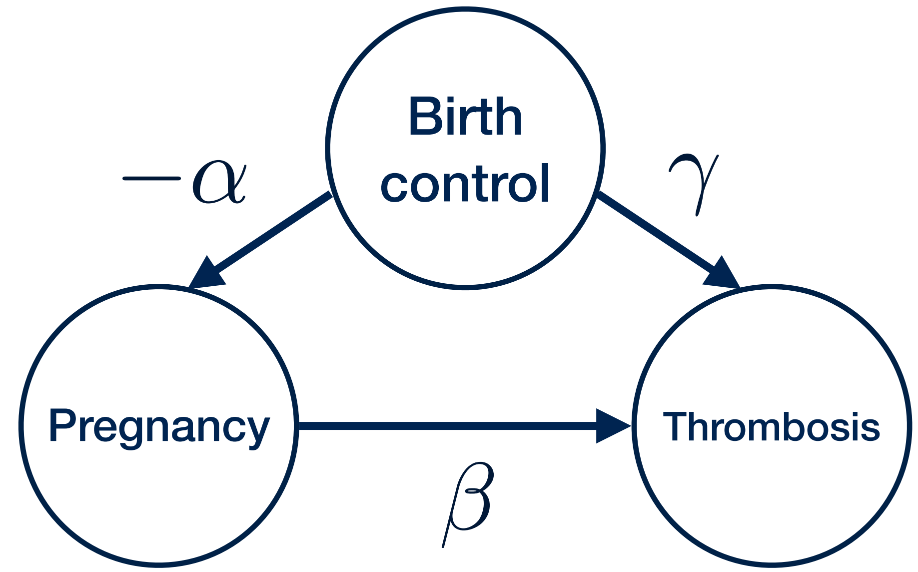
Assumptions 3: Faithfulness

It **fails** when distributions are set up in such a way that paths exactly cancel:

$$P = -\alpha B + U_P$$

$$T = \beta P + \gamma B + U_T$$

$$\Rightarrow T = (-\alpha\beta + \gamma)B + U$$



So if $\gamma = \alpha\beta$, no dependency between T and B will be observed!

- Fails in **regulatory systems**, e.g. home temperature, outside temp, thermostat: By design, thermostat keeps the inside temp independent of outside, always fixed at T*

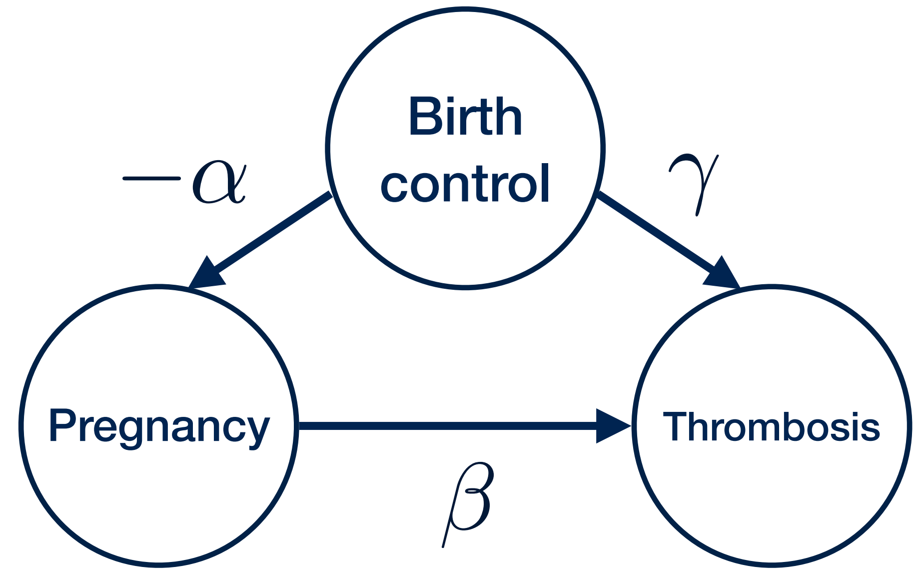
Assumptions 3: Faithfulness

It **fails** when distributions are set up in such a way that paths exactly cancel:

$$P = -\alpha B + U_P$$

$$T = \beta P + \gamma B + U_T$$

$$\Rightarrow T = (-\alpha\beta + \gamma)B + U$$



So if $\gamma = \alpha\beta$, no dependency between T and B will be observed!

- Fails in **regulatory systems**, e.g. home temperature, outside temp, thermostat: By design, thermostat keeps the inside temp independent of outside, always fixed at T*
- **Biology and homeostasis!**

Often keep the assumption and argue that most distributions are multimodal and will not cancel each other exactly …

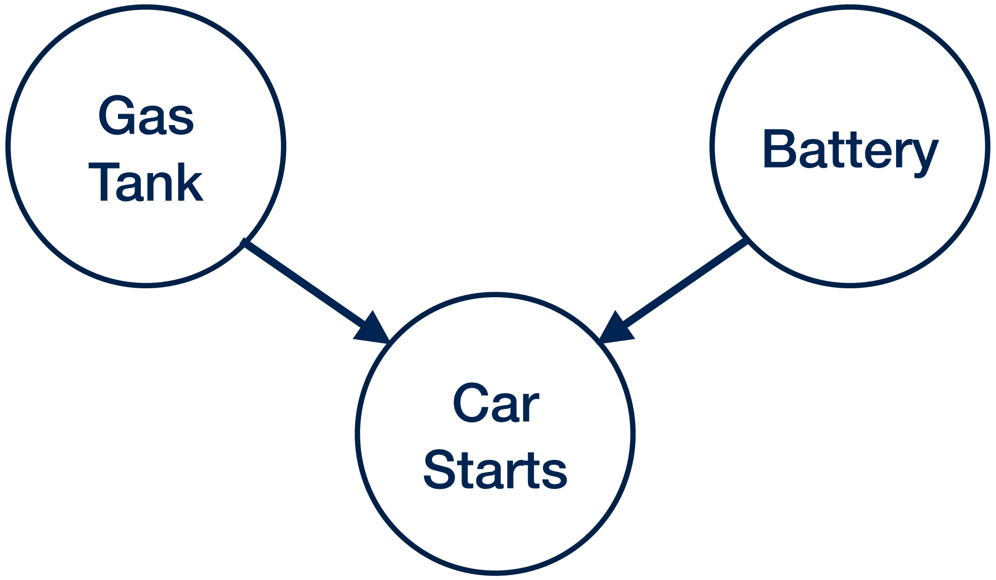
Silver 2018

Distinguishing causal structures: V-structures

- Recall collider example:

Gas tank $\perp\!\!\!\perp$ Battery

Gas tank $\not\!\perp\!\!\!\perp$ Battery | Car starts = 0

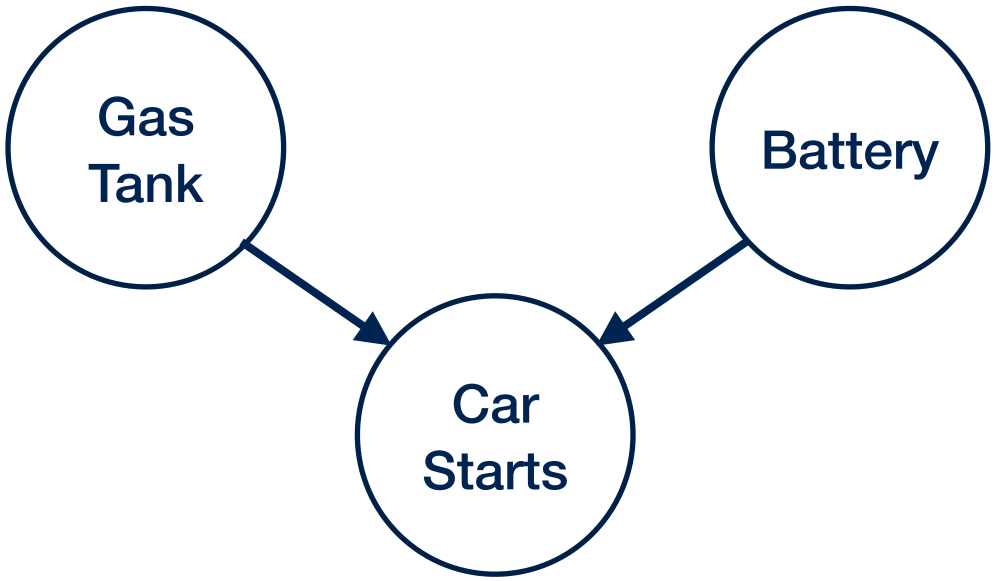


Distinguishing causal structures: V-structures

- Recall collider example:

Gas tank $\perp\!\!\!\perp$ Battery

Gas tank $\not\perp\!\!\!\perp$ Battery | Car starts = 0



- **Markov Equivalence Class (MEC)**: Two graphs G and G' belong to the same equivalence class iff each conditional independence implied by G is also implied by G' and vice versa.

- We can learn edges/directions using MEC and d-separation.
- D-separations gives all CI implied by graph

Markov Equivalence Class (MEC)

True DAG	$A \rightarrow B \rightarrow C$	$A \rightarrow B \leftarrow C$
All Observed CIs	$A \perp\!\!\!\perp C \mid B$	$A \perp\!\!\!\perp C \mid \emptyset$
Set of DAGs in MEC	$A \rightarrow B \rightarrow C$ $A \leftarrow B \leftarrow C$ $A \leftarrow B \rightarrow C$	$A \rightarrow B \leftarrow C$
CPDAG (complete partially DAG)	$A - B - C$	$A \rightarrow B \leftarrow C$

Glymour et al. (2019)

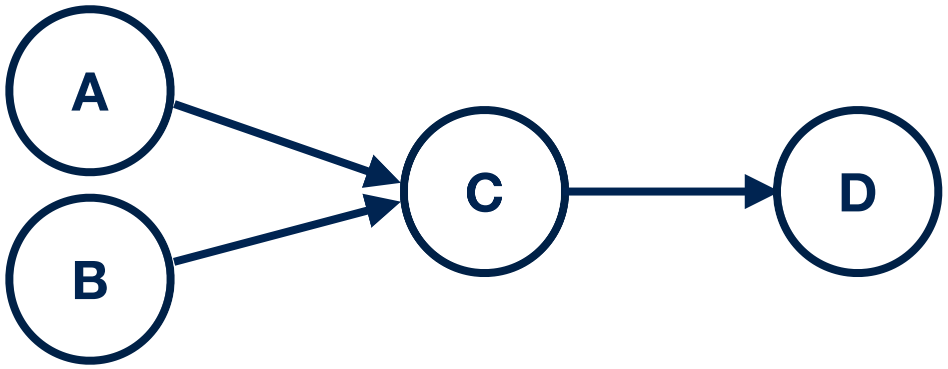
The Search Space of Causal Graphs

- For |V|=n nodes there are $\binom{n}{2} = \frac{1}{2}(n-1)n$ distinct pairs of variables
- There are at least $2^{\frac{1}{2}(n-1)n}$ possible graphs where between any two pairs there is either an edge or no edge.
- There are at most $3^{\frac{1}{2}(n-1)n}$ possible graphs since we may have either of: $A \rightarrow B, \ A \leftarrow B, \ A \qquad B$
- Grows super exponentially in the number of nodes
- Requires efficient causal discovery algorithms: PC algorithm

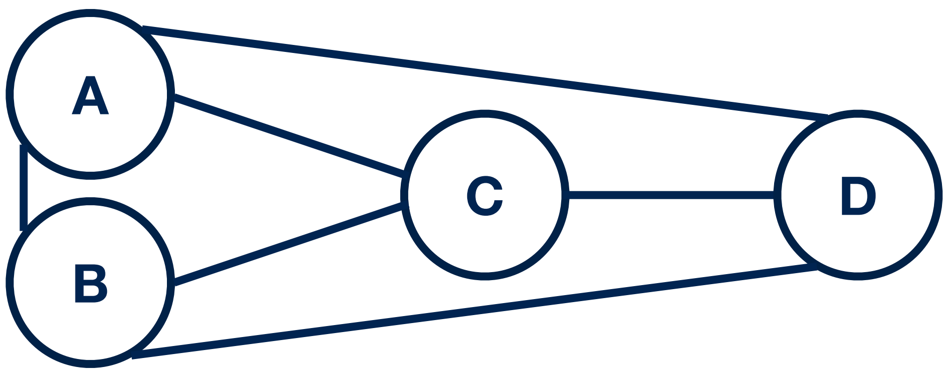
Silver 2018

Peter-Clark (PC) Algorithm

True causal graph:

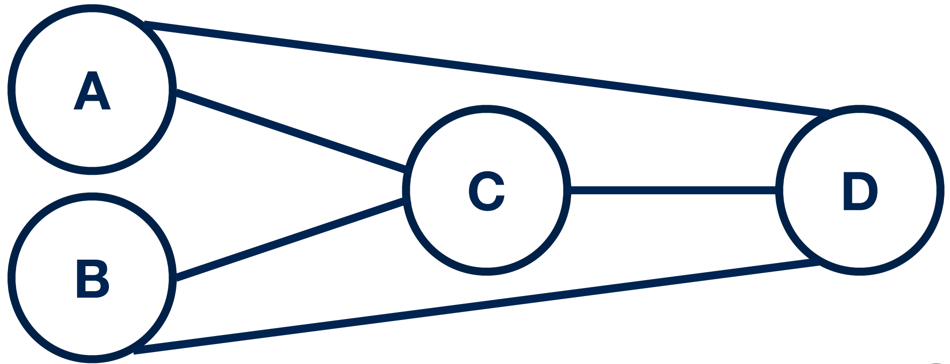


1. Start with the complete graph



2. Zeroth order CI, $A \perp\!\!\!\perp B$, by faithfulness:

See later for statistical independence tests.

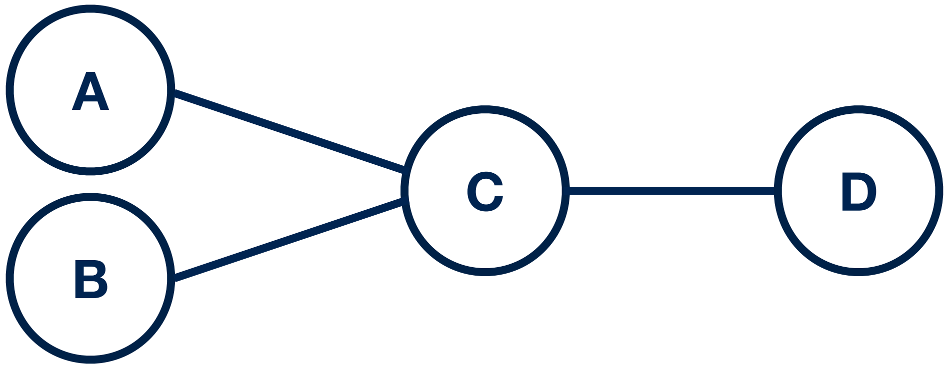


Glymour et al. (2019)

Peter-Clark (PC) Algorithm

3. 1st order CI, $A \perp\!\!\!\perp D | C$, by faithfulness:

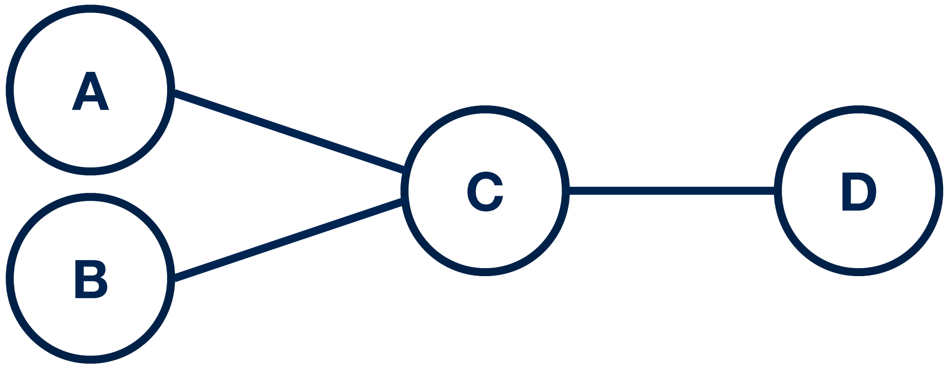
$B \perp\!\!\!\perp D | C$



Peter-Clark (PC) Algorithm

3. 1st order CI, $A \perp\!\!\!\perp D | C$, by faithfulness:

$B \perp\!\!\!\perp D | C$

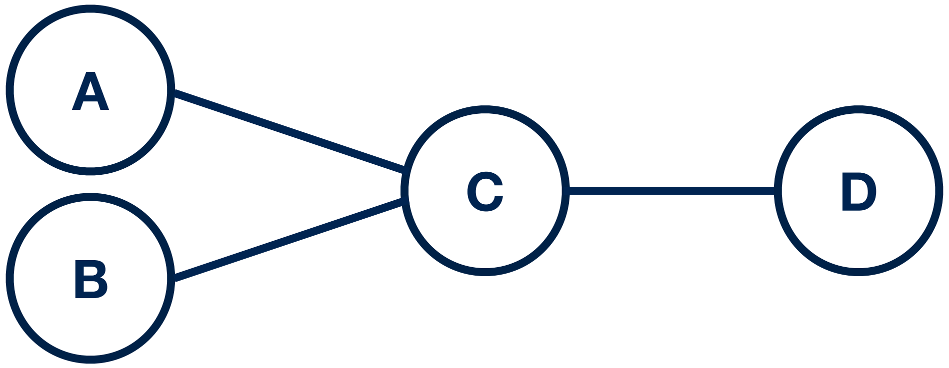


4. No higher order CI observed. Notice that conditioning sets only need to contain **neighbours** for the two nodes due to the Markov condition. We do not know the parents but parents are a subsets of neighbours. As the graph becomes sparser, the number of tests to be performed decreases. This makes PC very efficient.

Glymour et al. (2019)

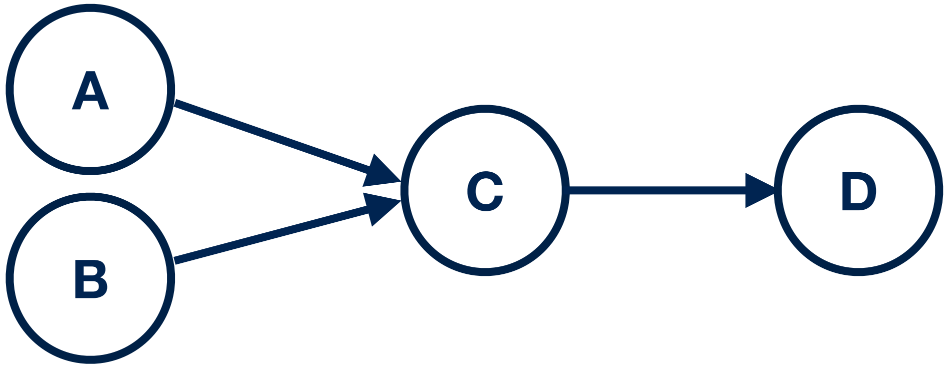
Peter-Clark (PC) Algorithm

3. 1st order CI, $A \perp\!\!\!\perp D|C$, by faithfulness:
$B \perp\!\!\!\perp D|C$



4. No higher order CI observed. Notice that conditioning sets only need to contain **neighbours** for the two nodes due to the Markov condition. We do not know the parents but parents are a subsets of neighbours. As the graph becomes sparser, the number of tests to be performed decreases. This makes PC very efficient.

5. Orient V-structures (colliders): take triplets where 2 nodes are connected to the 3rd: $A \not\perp\!\!\!\perp B|C$ only.



Note $C \leftarrow D$ cannot be as it would have been a collider (not detected in 5)

Glymour et al. (2019)

Remarks

- Missing/unobserved variables could lead to wrong/biased graphs
- Conditional independence tests are subject of active research
- Parallelised PC
- PC for heterogeneous data etc.

Glymour et al. (2019)

Structural Causal Models (SCM)

An SCM consists of d structural assignments

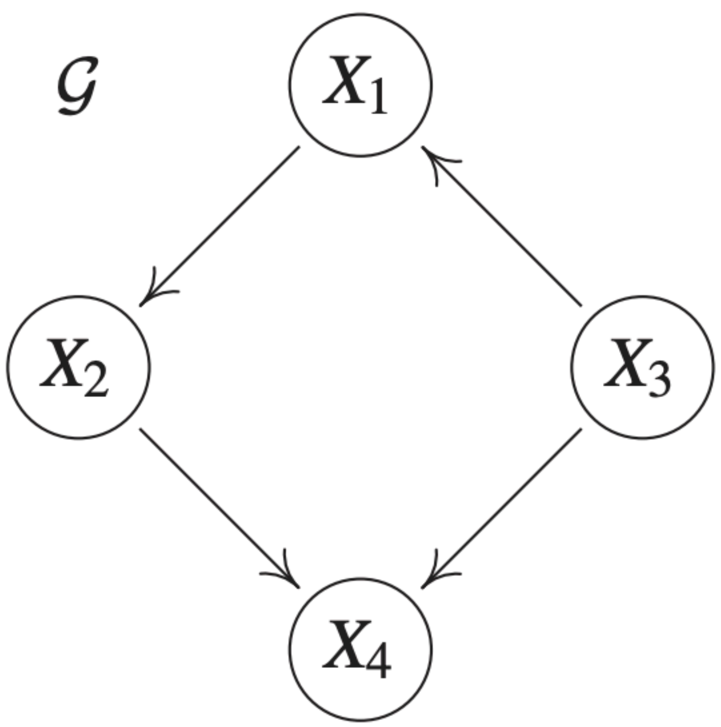
$$X_j := f_j(PA_j, N_j) \quad , \quad j = 1, \cdots, d$$

Parents of X_j , i.e., direct causes of X_j

Jointly independent noise variables

$$X_1 := f_1(X_3, N_1)$$
$$X_2 := f_2(X_1, N_2)$$
$$X_3 := f_3(N_3)$$
$$X_4 := f_4(X_2, X_3, N_4)$$

- $N_1, \ldots, N_4$ jointly independent
- $\mathcal{G}$ is acyclic



Jonas Peters et al, Elements of Causal Inference (2017)

Intervention vs observation

- Consider the following causal model with structure equations:

Random Variables

$$C := N_C$$
$$E := 4 \cdot C + N_E$$

where, $N_C, N_E \sim \mathcal{N}(0, 1)$, are independent and iid.

Intervention vs observation

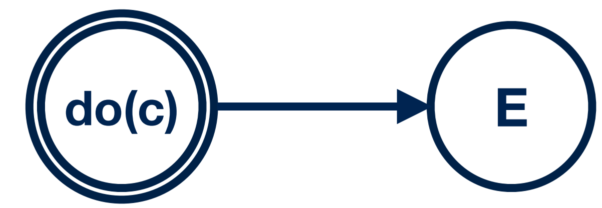
- Consider the following causal model with structure equations:

Random Variables

$$C := N_C$$
$$E := 4 \cdot C + N_E$$

where, $N_C, N_E \sim \mathcal{N}(0,1)$, are independent and iid. **We expect**:

- Apply do(C):
  - The new distribution $p(E|do(C)) \neq p(E)$
  - Since there are no other confounders: $p(E|do(C)) = p(E|C)$



Intervention vs observation

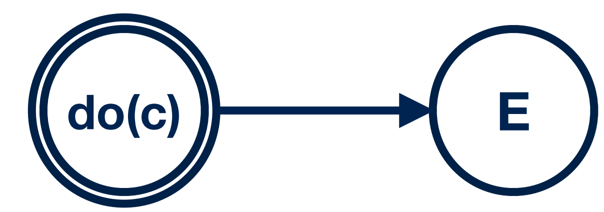
- Consider the following causal model with structure equations:

Random Variables

$$C := N_C$$
$$E := 4 \cdot C + N_E$$

where, $N_C, N_E \sim \mathcal{N}(0,1)$, are independent and iid. **We expect**:

- Apply do(C):
  - The new distribution $p(E|do(C)) \neq p(E)$
  - Since there are no other confounders: $p(E|do(C)) = p(E|C)$



- Apply do(E):
  - The new distribution $p(C|do(E)) = p(C)$
  - Graph structure changes: $p(C|do(E)) \neq p(C|E)$

Jonas Peters et al, Elements of Causal Inference (2017)

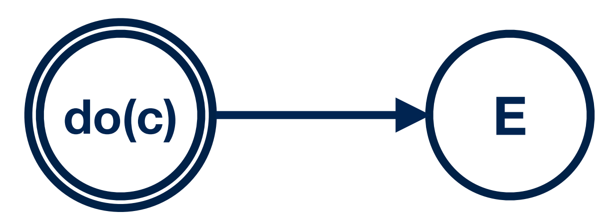
Intervention vs observation: Analytical computation

$$C := N_C$$
$$E := 4 \cdot C + N_E$$
$$N_C, N_E \sim \mathcal{N}(0,1), N_C \perp\!\!\!\perp N_E$$

Using $\mathrm{Var}[aX] = a^2\mathrm{Var}[X]$, $4C \sim \mathcal{N}(0,16)$.

Using, $4C \perp\!\!\!\perp N_E$, and the sum of two normally distributed random variables is another normally distributed random variable (by **convolution**):

$$E \sim \mathcal{N}\left(\mu_{4C} + \mu_{N_E}, \sigma^2_{4C} + \sigma^2_{N_E}\right)$$
$$\Rightarrow E \sim \mathcal{N}\left(0, 17\right)$$



A fixed number

$$p(E) = \mathcal{N}\left(0,17\right) \neq \mathcal{N}\left(8,1\right) = p(E|do(C=2)) = p(E|C=2)$$
$$\neq \mathcal{N}\left(12,1\right) = p(E|do(C=3)) = p(E|C=3)$$

Jonas Peters et al, Elements of Causal Inference (2017)

Intervention vs observation: Analytical computation

$$C := N_C$$
$$E := 4 \cdot C + N_E$$
$$N_C, N_E \sim \mathcal{N}(0,1), N_C \perp\!\!\!\perp N_E$$

$$p(C|do(E=2)) = \mathcal{N}(0,1) = p(C|do(E = \text{Any } r > 0)) = p(C)$$

$\neq p(C|E=2)$ in the original distribution above

Proof: Use product rule: $p(C|E) = \dfrac{p(C,E)}{p(E)}$

For a bivariate normal distribution (2 joint normal distributions), the marginal:

$$p(C|E) = \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2) \;\; \text{s.t.} \;\; \tilde{\mu} = \mu_C + \rho\frac{\sigma_C}{\sigma_E}(E - \mu_E), \;\; \tilde{\sigma}^2 = \sigma_C^2\left(1 - \rho^2\right)$$

Jonas Peters et al, Elements of Causal Inference (2017)

Intervention vs observation: Analytical computation

$$C := N_C$$
$$E := 4 \cdot C + N_E$$
$$N_C, N_E \sim \mathcal{N}(0,1), N_C \perp\!\!\!\perp N_E$$

Proof (Cont.): Use $\mathrm{Cov}(aX, bY + cZ) = ab\ \mathrm{Cov}(X,Y) + ac\ \mathrm{Cov}(X,Z)$

$$\Rightarrow \rho = \frac{\mathrm{Cov}(C,E)}{\sigma_C \sigma_E} = \frac{4\mathrm{Cov}(N_C,N_C) + \mathrm{Cov}(N_C,N_E)}{\sigma_C \sigma_E} = \frac{4}{\sqrt{17}}$$

$$\Rightarrow p(C|E=2) = \mathcal{N}\left(\frac{8}{17}, \sigma^2 = \frac{1}{17}\right) \Rightarrow \quad p(C|do(E)) \neq p(C|E)$$

Jonas Peters et al, Elements of Causal Inference (2017)

Next time

- **Functional Causal Models (FCMs):** Utilising asymmetry in data for causal discovery
- **LiNGAMs:** Linear non-gaussian acyclic models, allow for new approaches for causal learning from observational data
- **ANM:** Additive noise models and **causal identifiablity**
- **IGCI:** Information Geometric Causal Inference

Causality in Biomedicine Lecture Series: Lecture 6

Ava Khamseh (Biomedical AI Lab)

IGMM & School of Informatics

27 Nov, 2020

Convolution of probability distributions

Random Variables

$$C := N_C$$

$$N_C, N_E \sim \mathcal{N}(0,1), N_C \perp\!\!\!\perp N_E$$

$$E := 4 \cdot C + (\text{intercept} = 0) + N_E$$

'Residual'

Convolution of probability distributions

Random Variables

$$C := N_C$$

$$E := 4 \cdot C + (\text{intercept} = 0) + N_E$$

$$N_C, N_E \sim \mathcal{N}(0,1), N_C \perp\!\!\!\perp N_E$$

'Residuals'

- C, E, N_C, N_E, are **random variables** and the above relation is **NOT** an algebraic equation (in general)
- Linear operations on **random variables** in **Structural Causal Models (SCMs)** can only be understood in terms of operations on their **corresponding probability distributions,** e.g., for Z = X + Y:

$$P_{X+Y}(Z=z) = \int P_{XY}(x, z-x)dx$$

- Key **independence statements,** $X \perp\!\!\!\perp Y$

allow factorisation to the well-known **convolution** of probabilities:

$$P_{X+Y}(Z=z) = \int P_X(x)P_Y(z-x)dx$$