Causality

Generative and Deep Learning (GDL)

Riccardo Massidda (riccardo.massidda@di.unipi.it)
Davide Bacciu (davide.bacciu@unipi.it)

IN SUPREMÆ DIGNITATIS 1343
UNIVERSITÀ DI PISA

Probabilistic and Causal Learning

- Bayesian Networks (Tuesday 24th)
- d-separation, Markov blankets (Wednesday 25th)
- Graphical Causal Models (Thursday 26th, **today**!)
  - Causation and Correlation
  - Causal Bayesian Networks
  - Causal Inference
  - Structural Causal Models
- Probabilistic and Causal Structure Learning (Tuesday 3rd)

Correlation, Dependence and Causation

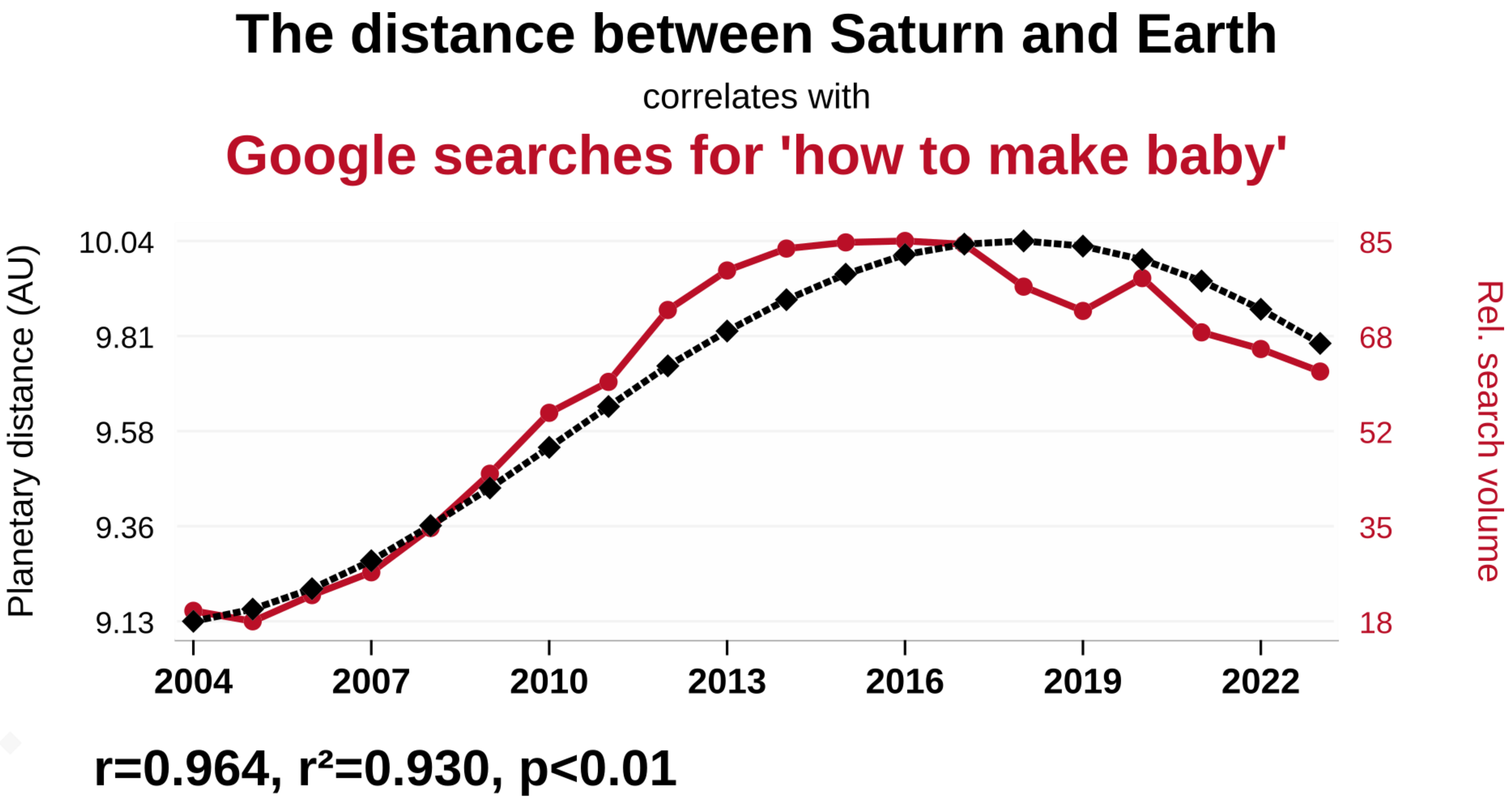
- A random variable is "**causing**" another random variable if a "**manipulation**" on the former alters the distribution of the latter.
- Correlation alone does not imply **direct causation**.
- In fact, completely **different causal structures** can entail the **same** set of conditional **independences** and dependences.

Reichenbach's Principle

Reichenbach's Common Cause Principle

Let X and Y be two variables such that X and Y are **statistically dependent**, then it holds:

i. X is indirectly causing Y, or
ii. Y is indirectly causing X, or
iii. There is a possibly unobserved common cause Z that indirectly causes both X and Y.



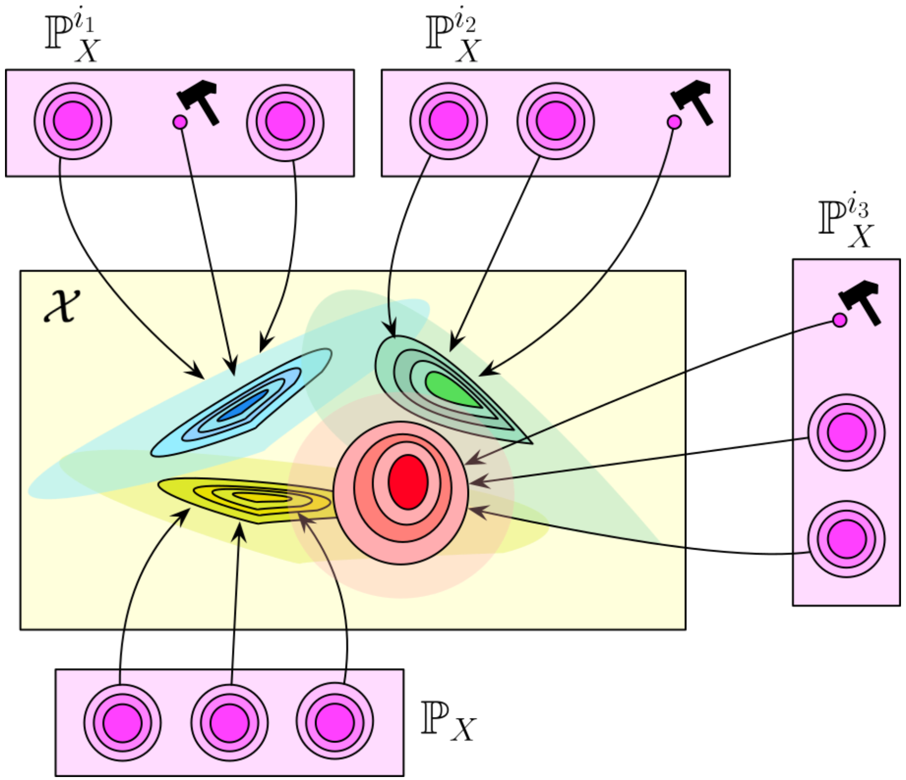
tylervigen.com/spurious/correlation/13099

Reichenbach's Principle

- The **principle** assumes that we can **perfectly** identify statistical dependence from data.
- In general, we need particular care:
  - Selection Bias
  - Small Size Datasets (Sampling Bias)
  - Common Trends
  - Data Manipulations
  - Measurement Errors
- PS: *Everything breaks apart in the quantum realm!* (checkout the SEP here, here, or here)

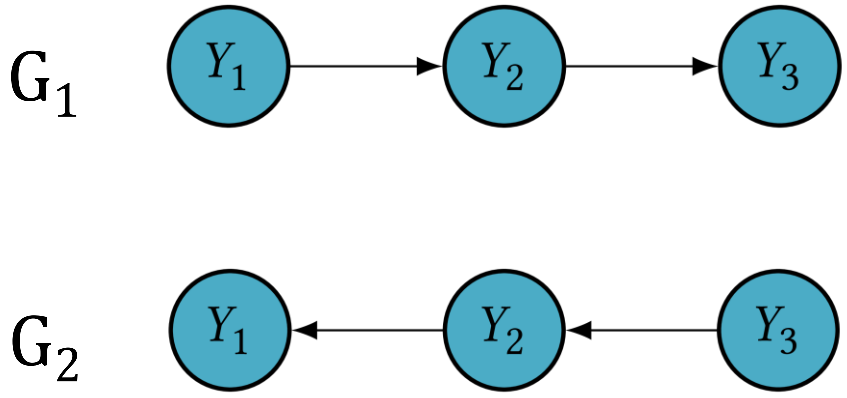
Causal Models

- **Causal models** represent causal relations and the results of **interventions** among random variables.
- While different probabilistic models can express the same conditional distributions, different causal models entail different **interventional distributions**.



Causal Bayesian Networks

- A **Causal Bayesian Network** is a Bayesian Network where each edge $Y_1 \rightarrow Y_2$ represents that Y_1 **directly causes** a variable Y_2.
- The two models G_1 and G_2 denote equivalent Bayesian Networks but distinct Causal Bayesian Networks.

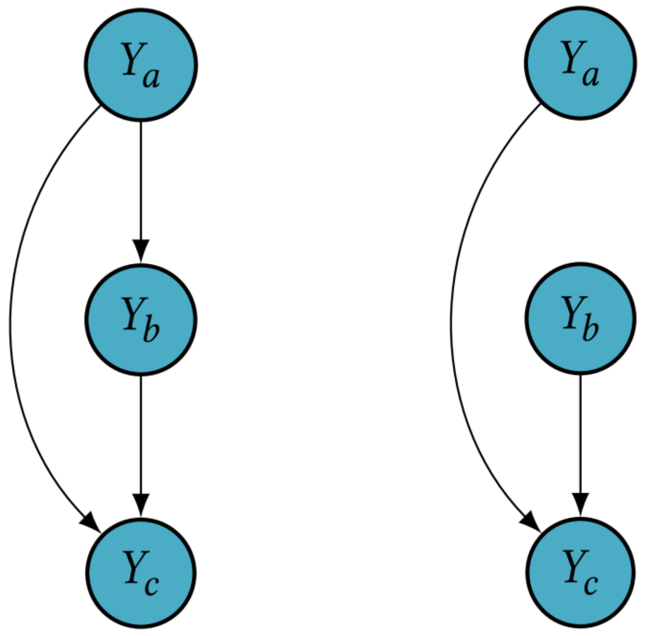


Hard Interventions

- Given a variable Y and a value k, we denote an **hard intervention**, also known as *ideal* or *perfect*, as

$$\mathrm{do}(Y \coloneqq k)$$

- The intervention replaces the variable of the model with the constant value.
- In general,
  $P(Y_2 \mid Y_1 = k) \neq P(Y_2 \mid \mathrm{do}(Y_1 \coloneqq k))$



$P(Y_a, Y_b, Y_c)$ $\quad$ $P(Y_a, Y_b, Y_c \mid \mathrm{do}(Y_b \coloneqq k))$

Truncated Factorization

- Let V be a set of variables and k a set of values.
- Then, the intervention $\mathrm{do}(V := k)$ assigns a value k_j to each $Y_j \in V$.
- Then, the **joint interventional distribution** factorizes as follows

$$P(Y_1, Y_2, \dots, Y_n \mid \mathrm{do}(V := k)) = \prod_{Y_i \notin V} P(Y_i \mid \mathrm{Pa}(Y_i)) \cdot \prod_{Y_j \in V} \mathbb{I}(Y_j = k_j)$$

Average Treatment Effect

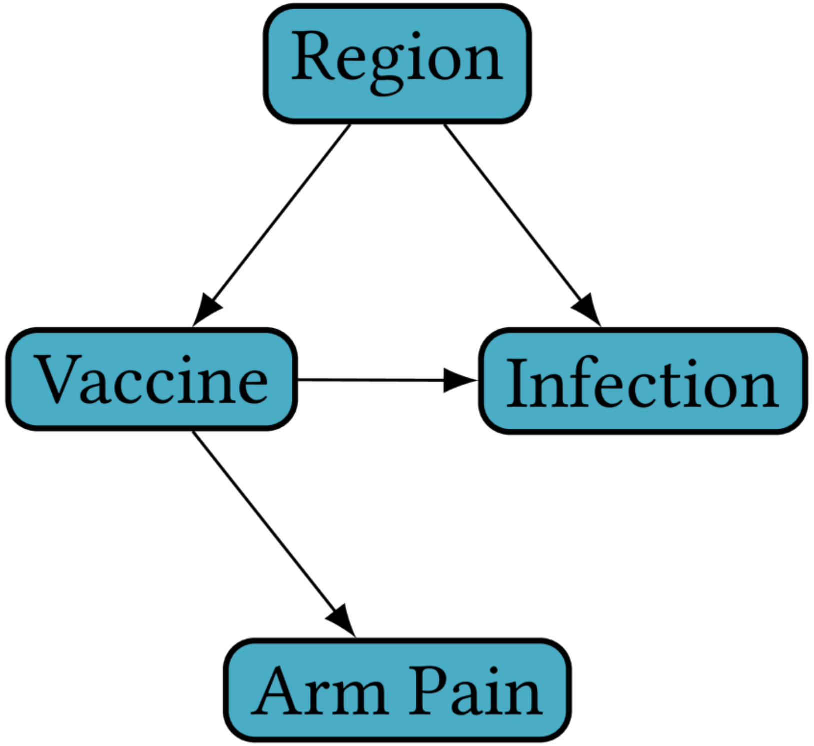
- Interventions are fundamental to study **causal effects**.
  - Does smoking causes cancer?
  - Will the vaccine avoid long-term infection?
  - How does the education level influence the average salary?
- Given a binary treatment variable Y_1 and an outcome variable Y_2, the **average treatment effect** of Y_1 on Y_2 is

$$\mathrm{ATE}(Y_1, Y_2) = \mathbb{E}\left[Y_2 \mid \mathrm{do}(Y_1 := T)\right] - \mathbb{E}\left[Y_2 \mid \mathrm{do}(Y_1 := F)\right]$$

Conditioning ≠ Intervening

- To estimate the ATE of a treatment of an outcome is fundamental to **distinguish** between **conditioning** and **intervening** on random variables.
- Conditioning is **not** a measure of causal effect.

$P(R) = 0.4$
$P(V \mid R = 1) = 0.2$
$P(V \mid R = 0) = 0.8$
$P(I \mid R = 0, V = 0) = 0.6$
$P(I \mid R = 0, V = 1) = 0.55$
$P(I \mid R = 1, V = 0) = 0.42$
$P(I \mid R = 1, V = 1) = 0.38$
$P(A \mid V = 0) = 0.1$
$P(A \mid V = 1) = 0.9$



Conditioning ≠ Intervening

- Does **observing** a painful arm increase the probability of observing an infection? Yes!

$$\mathbb{E}[I \mid A = 1] - \mathbb{E}[I \mid A = 0] > 0$$

- Does punching an arm increase the probability of infection? No!

$$\mathbb{E}[I \mid \mathrm{do}(A := 1)] - \mathbb{E}[I \mid \mathrm{do}(A := 0)] = 0$$

Conditioning ≠ Intervening

- Does **observing** a vaccine increase the probability of observing an infection? Yes!

$$\mathbb{E}[I \mid V = 1] - \mathbb{E}[I \mid V = 0] > 0$$

- Does taking a vaccine increase the probability of being infected?
- To answer, we need to **identify** the causal effect.

$$\mathbb{E}[I \mid \mathrm{do}(V := 1)] - \mathbb{E}[I \mid \mathrm{do}(V := 0)] = ?$$

Causal Effect Identifiability

- The **causal effect** of a treatment Y_1 on an outcome Y_2 is **identifiable** whenever there exists an adjustment set Z such that

$$P(Y_2 \mid \mathrm{do}(Y_1)) = \sum_z P(Y_2 \mid Y_1, Z = z) P(Z = z)$$

- The **do-calculus** is a complete system to find an adjustment set.
- From do-calculus, we can derive two fundamental adjustments:
  - The **back-door** criterion to handle observable confounders, and
  - The **front-door** adjustment to handle latent confounders.

Causal Sufficiency

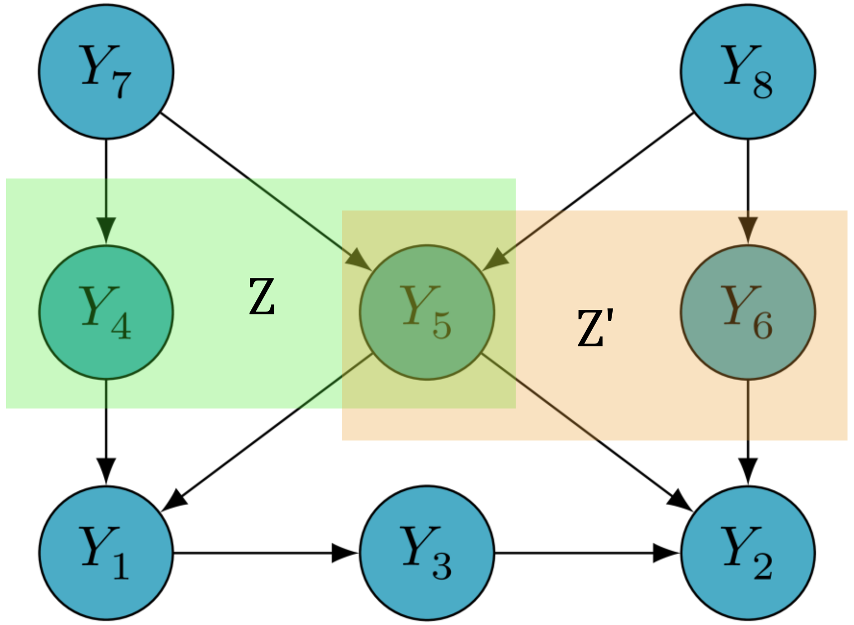
- A causal model is **causally sufficient** whenever:
  - All **confounders are observed**, and
  - There is **no selection bias** in the data.
- Causal sufficiency is sufficient but **not necessary** for causal identification.
- As for Markov property and faithfulness, it is an **assumption** on our model.

Back-Door Adjustment

- A set of variables Z satisfies the **back-door** criterion for the causal effect of Y_1 on Y_2 if:
  - No node in Z descends from Y_1, and
  - Z **blocks** every path between Y_1 and Y_2 that contains an edge entering Y_1.
- Then, it holds

$$P(Y_2 \mid \mathrm{do}(Y_1)) = \sum_z P(Y_2 \mid Y_1, Z = z) P(Z = z)$$

- Both Z={Y4, Y5} and Z'={Y5, Y6} are valid back-door adjustments.



Back-Door and Optimal Adjustment

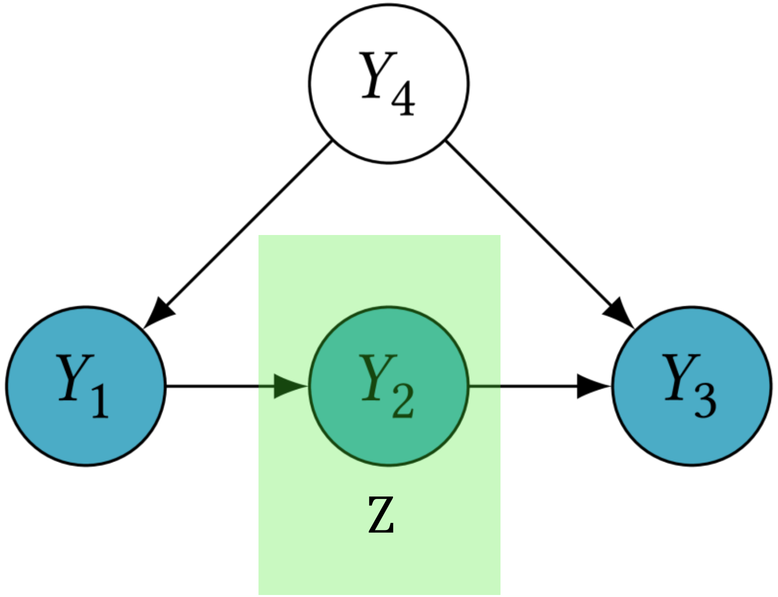
- The **back-door** criterion defines some but not all adjustment sets, i.e., it's **correct but not complete**.
- Different adjustments lead to different **asymptotic variance** of the ATE estimator.
- In a known graph, the **optimal adjustment set** is

$$\mathbf{O}(\mathrm{X}, \mathrm{Y}) = \mathrm{pa}(\mathrm{med}(\mathrm{X}, \mathrm{Y})) \setminus (\mathrm{med}(\mathrm{X}, \mathrm{Y}) \cup \{\mathrm{X}\})$$

  where $\mathrm{med}(\mathrm{X}, \mathrm{Y})$ is the set of mediators from X to Y, including Y but excluding X.
- More details? Look here: Witte et al. (2020).

Front-Door Adjustment

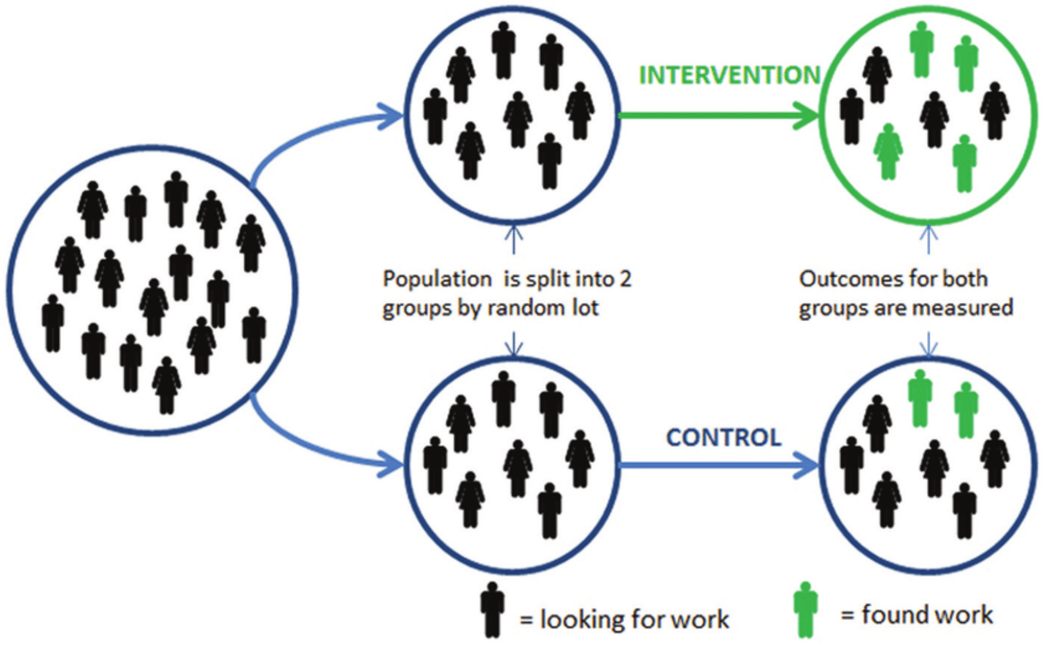
- A set of variables Z satisfies the **front-door** criterion for the causal effect of Y_1 on Y_2 if:
  - Z intercepts all directed paths from Y_1 to Y_2, and
  - there is no unblocked back-door path from Y_1 to Z, and
  - all back-door paths from Z to Y_2 are blocked by Y_1.
- Then, it holds



$$P(Y_2 \mid \mathrm{do}(Y_1)) = \sum_z P(Z = z \mid Y_1) \sum_{y_1'} P(Y_2 \mid Y_1 = y_1', Z = z) P(Y_1 = y_1')$$

Randomized Control Trials

- Causal effects might not be identifiable from observational data.
- Randomized Control Trials (**RCTs**) produce **interventional data**, but require careful **experimental design**.
- Observational studies are still fundamental for **practical** and **ethical** reasons.



Counterfactual Reasoning

- **Counterfactual** queries naturally occurs when we retrospectively reason on alternative outcomes **after** an intervention.
  - If the patient had received a placebo instead, would their recovery have been the same?
  - If the student had not studied the night before, would they still have passed the exam?
- Causal Bayesian Networks **cannot answer** counterfactual queries.

Structural Causal Models

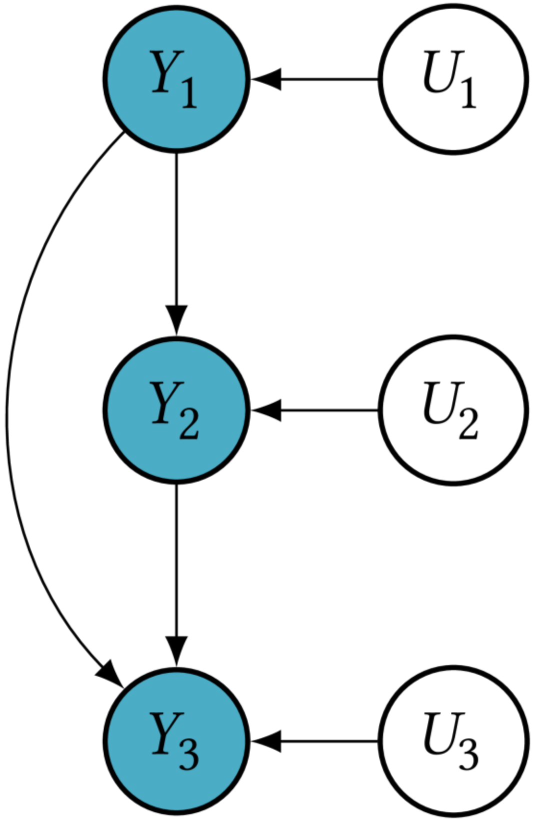
A Structural Causal Model (**SCM**)

$$M=(Y, U, f, P(U))$$

specifies the **deterministic mechanisms** f of a set of endogenous variables Y given a set of exogenous variables U with distribution P(U).

Formally,

$$Y_j = f_j(Y_{\text{pa}(Y_j)}, U_j)$$

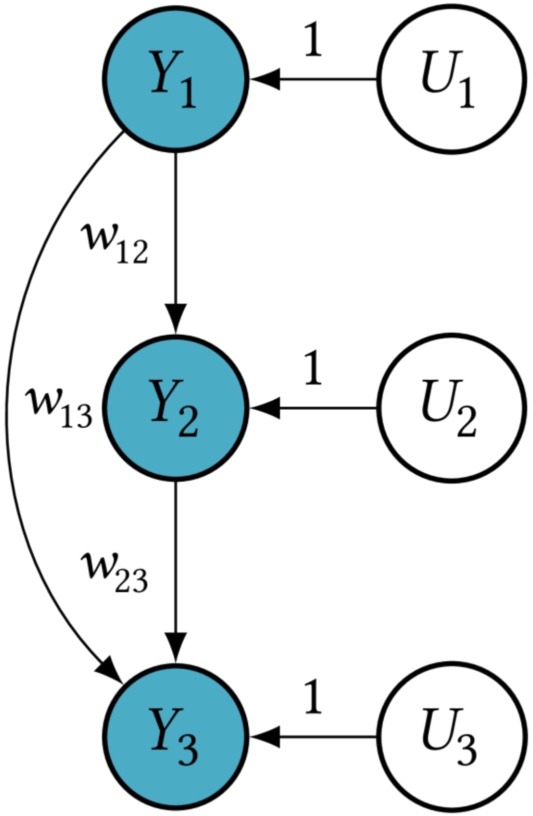


Linear Additive Noise Model

A Linear Additive Noise Model (**ANM**) is a structural causal model where the **functional mechanisms** are **linear**.

Formally, given a matrix $W \in \mathbb{R}^{n \times n}$,

$$Y_j = \sum_{Y_i \in \mathrm{pa}(Y_j)} w_{ij} Y_i + U_j$$



Computing Counterfactuals in SCMs

Given a fully specified SCM, we can directly compute counterfactuals using the following three-step procedure:

1. **Abduction**. Update the exogenous distribution $\mathrm{P}(\mathrm{U}|\mathrm{Y})$ given the evidence Y.
2. **Action**. Intervene on the treatment applying $\mathrm{do}(\mathrm{Y}_1)$ on the SCM.
3. **Prediction**. Infer the probability of the outcome given the new treatment as in $\mathrm{P}(\mathrm{Y}_2|\mathrm{do}(\mathrm{Y}_1),\mathrm{U}) \cdot \mathrm{P}(\mathrm{U}|\mathrm{Y})$.

Wrap-Up

- The "**ladder of causation**" determines the relation between models and queries on a system:
  - **Probabilistic** Queries $P(Y_2|Y_1) \rightarrow$ **Bayesian** Networks
  - **Interventional** Queries $P(Y_2|do(Y_1)) \rightarrow$ **Causal Bayesian** Networks
  - **Counterfactual** Queries $P(Y_2|do(Y_1), Y) \rightarrow$ **Structural Causal** Models
- When they are **identifiable**, different causal models provides a solution to **answer causal queries**.
- How to **learn** Bayesian Networks, Causal Bayesian Networks and Structural Causal Models from data? **Next week!**