An Introduction to Structural Causal Models

Alexander Reisach

15.11.2023

Université Paris Cité
Laboratoire MAP5

Why Structural Causal Models?

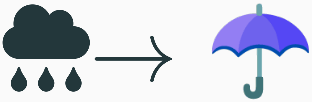
Causality

I will try to convince you that those things are fundamentally different.

1. Prediction
2. Prediction under **intervention**
3. (Prediction of Counterfactuals)

An **intervention** is the act of changing a component of an otherwise closed system from the outside.

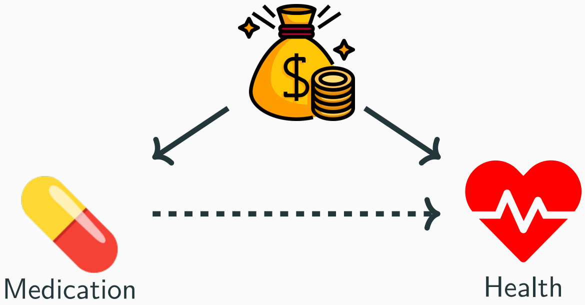
Causality is all about finding asymmetries in the relationship between variables that relate to interventions.



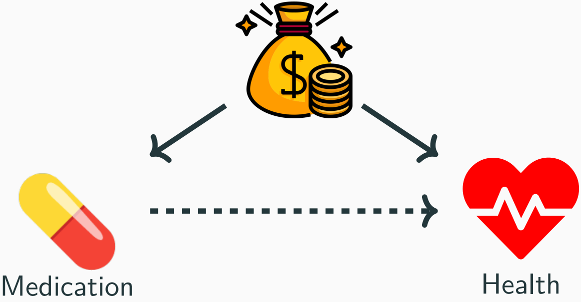
Confounder Bias

Motivating Example 1: Confounder Bias

Motivating Example 1: Confounder Bias



Motivating Example 1: Confounder Bias



Confounding bias

The relationship between intervention and outcome may be distorted by a confounding variable. Conditioning on the confounder removes the distortion.

Confounding Bias - Formalization

Let X, Y be random variables and x, y values in their respective domain.

- Target of prediction: $\mathbb{E}[Y \mid X]$
- Target of causal inference: $\mathbb{E}[Y \mid \mathrm{do}(X = x)]$

Confounding Bias - Formalization

Let X, Y be random variables and x, y values in their respective domain.

- Target of prediction: $\mathbb{E}[Y \mid X]$
- Target of causal inference: $\mathbb{E}[Y \mid \mathrm{do}(X = x)]$

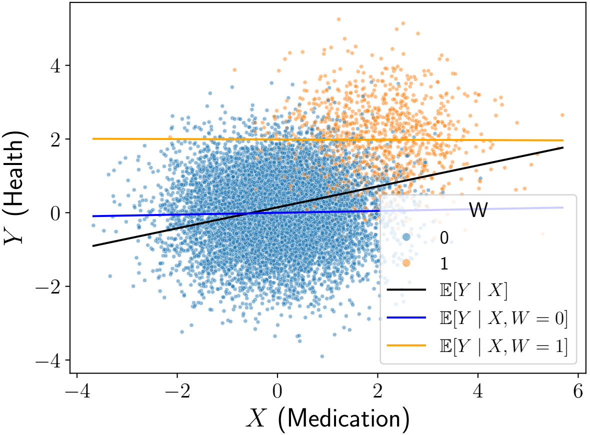
Confounder scenario:

$$W := N_W \text{ (Wealth)}$$
$$X := W + N_X \text{ (Medication)}$$
$$Y := W + N_Y \text{ (Health)}$$

with $N_X, N_Y \sim \mathcal{N}(0, 1)$ independently, and $N_W \sim \text{Bern}(0.1)$.

Figure 2: Structural Causal Model (SCM)

Confounder Bias - Simulation



What is $\mathbb{E}\left[Y \mid \mathrm{do}\left(X = x\right)\right]$?

Controlling (Adjusting) for a Variable

In our confounder example, we have that

$$\mathbb{E}\left[Y \mid X\right] \neq \mathbb{E}\left[Y \mid \mathrm{do}\left(X = x\right)\right].$$

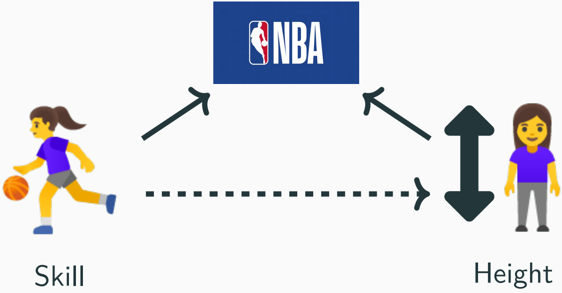
Controlling for W removes the bias, and

$$\begin{aligned} & \mathbb{E}_W\left[\mathbb{E}\left[Y \mid X, W\right]\right] \\ =& \mathbb{E}\left[Y \mid X, W = 0\right] P(W = 0) + \mathbb{E}\left[Y \mid X, W = 1\right] P(W = 1) \\ =& \mathbb{E}\left[Y \mid \mathrm{do}\left(X = x\right)\right]. \end{aligned}$$

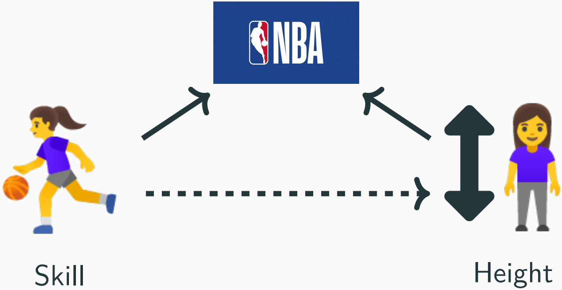
Collider Bias

Motivating Example 2: Collider Bias

Motivating Example 2: Collider Bias



Motivating Example 2: Collider Bias



Collider bias

Conditioning on the collider may result in a spurious relationship where really there is none.

Collider Bias - Formalization

Let X, Y be random variables and x, y values in their respective domain.

- Target of prediction: $\mathbb{E}[Y \mid X]$
- Target of causal inference: $\mathbb{E}[Y \mid \mathrm{do}(X = x)]$

Collider Bias - Formalization

Let X, Y be random variables and x, y values in their respective domain.

- Target of prediction: $\mathbb{E}[Y \mid X]$
- Target of causal inference: $\mathbb{E}[Y \mid \mathrm{do}(X = x)]$

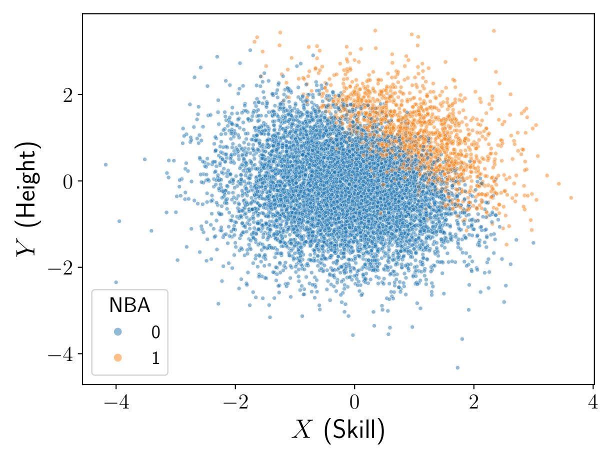
Collider scenario:

$$
\begin{aligned}
X &:= N_X \text{ (Skill)} \\
Y &:= N_Y \text{ (Height)} \\
\text{NBA} &:= \begin{cases} 1 & \text{if } 2X + 2Y + N_{\text{NBA}} > 3 \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}
$$

with $N_X, N_Y, N_{\text{NBA}} \sim \mathcal{N}(0, 1)$ iid.

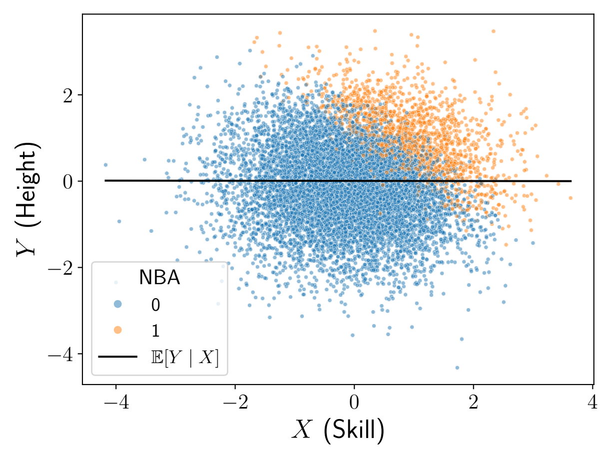
Figure 3: Structural Causal Model (SCM)

Collider Bias - Simulation



What is $\mathbb{E}[Y \mid X]$?

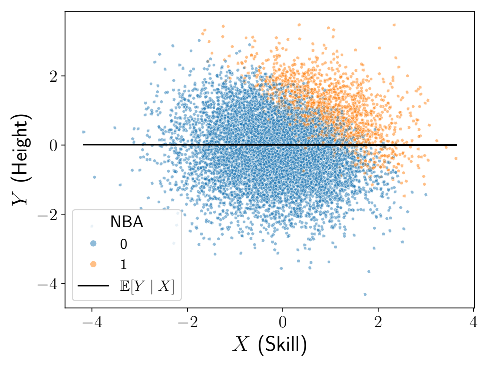
Collider Bias - Simulation



What happens if we control for NBA?

In other words, what is $\mathbb{E}_{\text{NBA}}\left[\mathbb{E}\left[Y \mid X, \text{NBA}\right]\right]$?

Collider Bias - Simulation

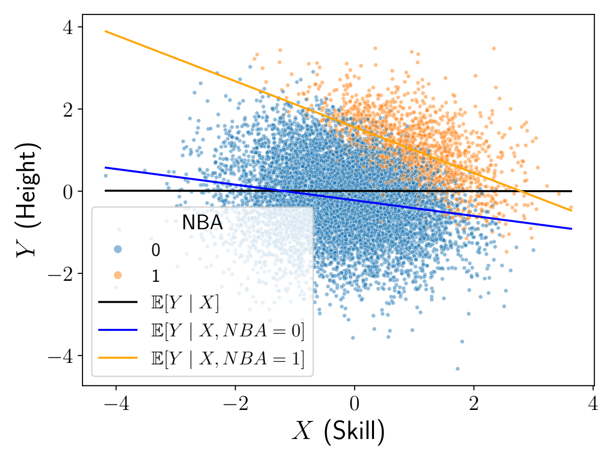


Controlling for NBA gives

$$\begin{aligned}\mathbb{E}_{\mathrm{NBA}}\left[\mathbb{E}\left[Y \mid X, \mathrm{NBA}\right]\right] = & \mathbb{E}\left[Y \mid X, \mathrm{NBA} = 0\right] P(\mathrm{NBA} = 0) + \\ & \mathbb{E}\left[Y \mid X, \mathrm{NBA} = 1\right] P(\mathrm{NBA} = 1).\end{aligned}$$

What are $\mathbb{E}\left[Y \mid X, NBA = 0\right]$ and $\mathbb{E}\left[Y \mid X, NBA = 1\right]$?

Collider Bias - Simulation



What is $\mathbb{E}\left[Y \mid \text{do}\left(X = x\right)\right]$?

Controlling (Adjusting) for a Variable

In our collider example, we have that

$$\mathbb{E}\left[Y \mid X\right] = \mathbb{E}\left[Y \mid \mathrm{do}\left(X = x\right)\right].$$

Controlling for NBA introduces bias, meaning that

$$\mathbb{E}_{\mathrm{NBA}}\left[\mathbb{E}\left[Y \mid X, \mathrm{NBA}\right]\right] \neq \mathbb{E}\left[Y \mid \mathrm{do}\left(X = x\right)\right].$$

(The opposite was true in the confounding case!)

Takeaway

We need to take the causal structure into account when estimating causal quantities. Causal Inference is all about finding valid sets of control variables (also called adjustment sets).

Structural Causal Models

Structural Causal Models

A Structural Causal Model (SCM) is a triple $(N, X, \mathcal{F})$ where

- N is a set of exogenous random variables $\{N_1, N_2, \ldots\}$
- X is a set of endogenous random variables $\{X_1, X_2, \ldots\}$
- $\mathcal{F}$ is a set of functions $\mathcal{F} = \{f_{X_1}, f_{X_2}, \ldots\}$ defining an endogenous variable X_i in terms of N_i and other endogenous variables X_{-i}.

SCMs and Directed Acyclic Graphs

The causal relationships between variables in an SCM can be encoded in a directed acyclic graph (DAG) $\mathcal{G}(X, E)$ where an edge $X_i \to X_j$ exists if and only if X_j is a function of X_i in $\mathcal{F}$.

Examples

Let $N_1, N_2, N_3 \sim \mathcal{N}(0, 1)$ iid.

$X_1 := N_1$	$X_1 := X_3 + N_1$	$X_1 := N_1$
$X_2 := N_2$	$X_2 := X_3 + N_2$	$X_2 := X_1 + N_2$
$X_3 := X_1 + X_2 + N_3$	$X_3 := N_3$	$X_3 := X_2 + N_3$

SCMs and Directed Acyclic Graphs

The causal relationships between variables in an SCM can be encoded in a directed acyclic graph (DAG) $\mathcal{G}(X, E)$ where an edge $X_i \to X_j$ exists if and only if X_j is a function of X_i in $\mathcal{F}$.

Examples

Let $N_1, N_2, N_3 \sim \mathcal{N}(0, 1)$ iid.

$X_1 := N_1$	$X_1 := X_3 + N_1$	$X_1 := N_1$
$X_2 := N_2$	$X_2 := X_3 + N_2$	$X_2 := X_1 + N_2$
$X_3 := X_1 + X_2 + N_3$	$X_3 := N_3$	$X_3 := X_2 + N_3$

$$\boldsymbol{X_1} \to \boldsymbol{X_3} \leftarrow \boldsymbol{X_2}$$

SCMs and Directed Acyclic Graphs

The causal relationships between variables in an SCM can be encoded in a directed acyclic graph (DAG) $\mathcal{G}(X, E)$ where an edge $X_i \rightarrow X_j$ exists if and only if X_j is a function of X_i in $\mathcal{F}$.

Examples

Let $N_1, N_2, N_3 \sim \mathcal{N}(0, 1)$ iid.

$X_1 := N_1$	$X_1 := X_3 + N_1$	$X_1 := N_1$
$X_2 := N_2$	$X_2 := X_3 + N_2$	$X_2 := X_1 + N_2$
$X_3 := X_1 + X_2 + N_3$	$X_3 := N_3$	$X_3 := X_2 + N_3$

$\boldsymbol{X_1} \rightarrow \boldsymbol{X_3} \leftarrow \boldsymbol{X_2}$ $\quad$ $\boldsymbol{X_1} \leftarrow \boldsymbol{X_3} \rightarrow \boldsymbol{X_2}$

SCMs and Directed Acyclic Graphs

The causal relationships between variables in an SCM can be encoded in a directed acyclic graph (DAG) $\mathcal{G}(X, E)$ where an edge $X_i \to X_j$ exists if and only if X_j is a function of X_i in $\mathcal{F}$.

Examples

Let $N_1, N_2, N_3 \sim \mathcal{N}(0, 1)$ iid.

$X_1 := N_1$	$X_1 := X_3 + N_1$	$X_1 := N_1$
$X_2 := N_2$	$X_2 := X_3 + N_2$	$X_2 := X_1 + N_2$
$X_3 := X_1 + X_2 + N_3$	$X_3 := N_3$	$X_3 := X_2 + N_3$
$\boldsymbol{X_1} \to \boldsymbol{X_3} \leftarrow \boldsymbol{X_2}$	$\boldsymbol{X_1} \leftarrow \boldsymbol{X_3} \to \boldsymbol{X_2}$	$\boldsymbol{X_1} \to \boldsymbol{X_2} \to \boldsymbol{X_3}$

Utilizing Directed Acyclic Graphs (DAGs)

DAGs and Joint Probability Distributions

Assume we have a causal model $(N, X, \mathcal{F})$ with DAG $\mathcal{G}(X, E)$.

We say the DAG has the **causal Markov property** if the joint probability distribution P factorizes according to the DAG:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}_{\mathcal{G}}(X_i)),$$

where $\mathrm{Pa}_{\mathcal{G}}(X_i)$ denotes the parents (direct causes) of X_i in $\mathcal{G}$.

Takeaway

Every variable is independent of all others given its parents in the causal DAG.

Finding Variables to Control For

Valid Sets of Control Variables (Adjustment Sets)

An adjustment set $\mathcal{A} \subset X$ is valid for estimating the causal effect of X_i on X_j if

$$\mathbb{E}\left[X_j \mid \mathrm{do}\left(X_i = x\right)\right] = \mathbb{E}_{\mathcal{A}}\left[\mathbb{E}\left[X_j \mid X_i, \mathcal{A}\right]\right].$$

In general, an adjustment set is valid if it blocks all 'backdoor paths' from X_i to X_j.

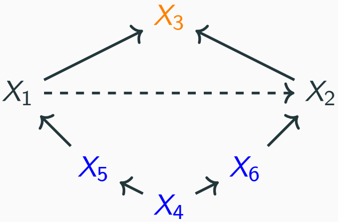
Backdoor Paths

A backdoor path is an undirected path from X_i to X_j starting with an edge pointing into X_i, that does not contain any collider (structure of the form $X_1 \to X_2 \leftarrow X_3$, with X_2 being the collider).

Due to the causal Markov property, we know that $\mathrm{Pa}_{\mathcal{G}}(X_i)$ is a valid adjustment set for estimating the causal effect of X_i on X_j.

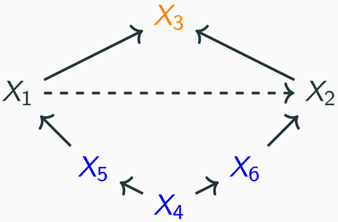
Backdoor Paths Example

What should we control for when estimating the causal effect $\mathbb{E}[X_2 \mid \mathrm{do}(X_1 = x)]$?



Backdoor Paths Example

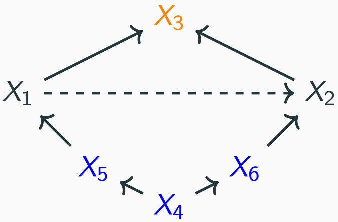
What should we control for when estimating the causal effect $\mathbb{E}\left[X_2 \mid \mathrm{do}\left(X_1 = x\right)\right]$?



Any one or combination of X_4, X_5, X_6 is a valid adjustment set.
For example: $\mathbb{E}\left[X_2 \mid \mathrm{do}\left(X_1 = x\right)\right] = \mathbb{E}_{X_4}\left[\mathbb{E}\left[X_2 \mid X_1, X_4\right]\right]$.

Backdoor Paths Example

What should we control for when estimating the causal effect $\mathbb{E}\left[X_2 \mid \mathrm{do}\left(X_1 = x\right)\right]$?



Any one or combination of X_4, X_5, X_6 is a valid adjustment set.
For example: $\mathbb{E}\left[X_2 \mid \mathrm{do}\left(X_1 = x\right)\right] = \mathbb{E}_{X_4}\left[\mathbb{E}\left[X_2 \mid X_1, X_4\right]\right]$.

Finding valid adjustment sets can be done algorithmically using the DAG, meaning that we can automate the process of identifying causal effects once we know the DAG underlying the SCM. 🎉

Summary Roadmap of Causal Reasoning

1. Experiments or Observation → Data
   May be costly, infeasible, or unethical
2. Expert Knowledge or Causal Discovery → SCM, DAG
   Requires domain expertise or strong assumptions
3. Causal Inference → Causal Estimates
   Relies on the first two steps

Further Reading

[1] Judea Pearl and Dana Mackenzie. ***The book of why: the new science of cause and effect.*** Basic books, 2018.

[2] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. ***Elements of Causal Inference: Foundations and Learning Algorithms.*** The MIT Press, 2017.

Causal Inference Course by Brady Neal:

https://www.bradyneal.com/causal-inference-course

Further Reading

[1] Judea Pearl and Dana Mackenzie. ***The book of why: the new science of cause and effect***. Basic books, 2018.

[2] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. ***Elements of Causal Inference: Foundations and Learning Algorithms***. The MIT Press, 2017.

Causal Inference Course by Brady Neal:

https://www.bradyneal.com/causal-inference-course

Thank you for your attention!