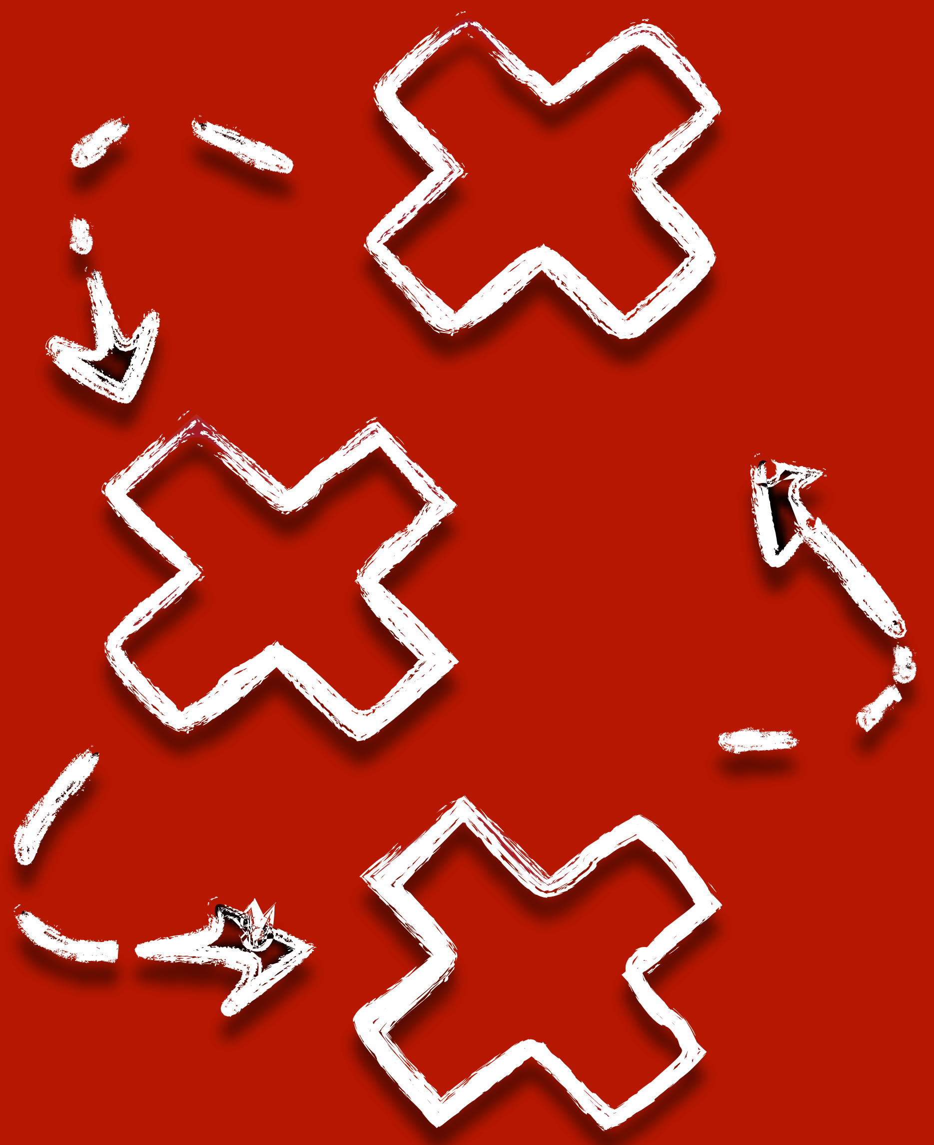


Causal Data Science

Lecture 7:1 Counterfactuals and potential outcomes

Lecturer: Sara Magliacane

UvA - Spring 2024

Last two classes: Identification strategies for causal effects

- Given a causal graph G, an **identification strategy** is a formula to estimate an interventional distribution from a combination of observational ones
- **Backdoor criterion, Adjustment criterion**

$$p(x_j \mid \mathrm{do}(x_i)) = \int_{x_{\mathbf{Z}}} p(x_j \mid x_i, x_{\mathbf{Z}}) p(x_{\mathbf{Z}}) dx_{\mathbf{Z}}$$

- **Frontdoor criterion**

$$p(x_j \mid \mathrm{do}(x_i')) = \int_{x_{\mathbf{M}}} p(x_{\mathbf{M}} \mid x_i') \int_{x_i} p(x_j \mid x_{\mathbf{M}}, x_i') p(x_i) dx_i$$

- **Instrumental variables**

Last two classes: Identification strategies for causal effects

- Given a causal graph G, an **identification strategy** is a formula to estimate an interventional distribution from a combination of observational ones
- **Backdoor criterion, Adjustment criterion**

$$p(x_j \mid \mathrm{do}(x_i)) = \int_{x_\mathbf{Z}} p(x_j \mid x_i, x_\mathbf{Z}) p(x_\mathbf{Z}) dx_\mathbf{Z}$$

- **Frontdoor criterion**

$$p(x_j \mid \mathrm{do}(x_i')) = \int_{x_\mathbf{M}} p(x_\mathbf{M} \mid x_i') \int_{x_i} p(x_j \mid x_\mathbf{M}, x_i) p(x_i) dx_i$$

- **Instrumental variables**

How do we estimate these reliably from data: **matching, IPW, …**
(Second part of today's class)

Last two classes: Identification strategies for causal effects

- Given a causal graph G, an **identification strategy** is a formula to estimate an interventional distribution from a combination of observational ones
- **Backdoor criterion, Adjustment criterion**

$$p(x_j \mid \mathrm{do}(x_i)) = \int_{x_{\mathbf{Z}}} p(x_j \mid x_i, x_{\mathbf{Z}}) p(x_{\mathbf{Z}}) dx_{\mathbf{Z}}$$

Introduce a different view on causality: **potential outcomes**

(First part of today's class)

How do we estimate these reliably from data: **matching, IPW, …**

(Second part of today's class)

$$p(x_j \mid \mathrm{do}(x_i)) = \int_{x_{\mathbf{M}}} p(x_{\mathbf{M}} \mid x_i') \int_{x_i} p(x_j \mid x_{\mathbf{M}}, x_i) p(x_i) dx_i$$

- **Instrumental variables**

From introduction: Causal Hierarchy [Pearl 2009]

Most ML

Causality

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

From introduction: Causal Hierarchy [Pearl 2009]

Most ML

Causality

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y \mid x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y \mid do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x \mid x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Counterfactual questions vs interventional questions

- Up to now we have been discussing how to estimate the **interventional distribution** $P(X_j \mid \mathrm{do}(X_i))$

Counterfactual questions vs interventional questions

- Up to now we have been discussing how to estimate the **interventional distribution** $P(X_j \mid \mathrm{do}(X_i))$

- But this does not tell us what would have happen to individuals under different interventions that the ones that were actually performed
    - For example: a patient was treated and they recovered, what would have happened if they were not treated? (**Retrospectively**)

Counterfactual questions vs interventional questions

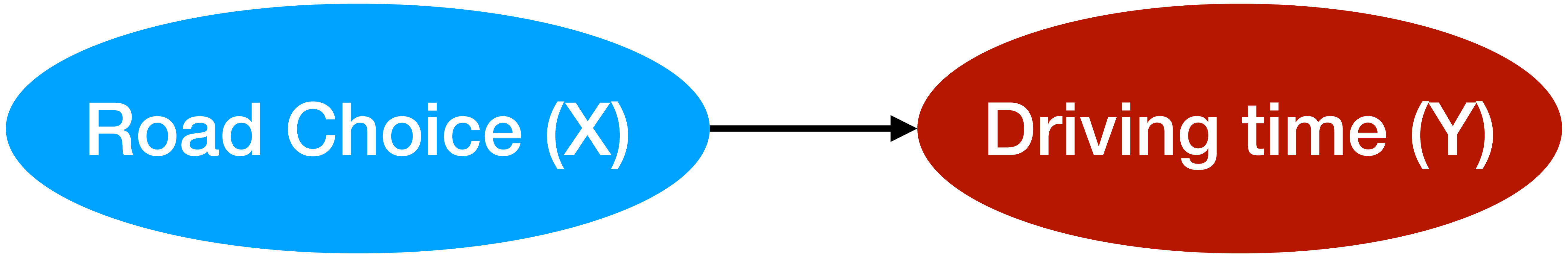
- Up to now we have been discussing how to estimate the **interventional distribution** $P(X_j \mid \mathrm{do}(X_i))$

- But this does not tell us what would have happen to individuals under different interventions that the ones that were actually performed
  - For example: a patient was treated and they recovered, what would have happened if they were not treated? (**Retrospectively**)

- This is an example of a **counterfactual question** (counter=against the fact)
  - https://youtu.be/iPBV3BlV7jk?t=41

Counterfactual vs interventional questions - example

- Driving home, two options:
  A. Take freeway
  B. Take normal road

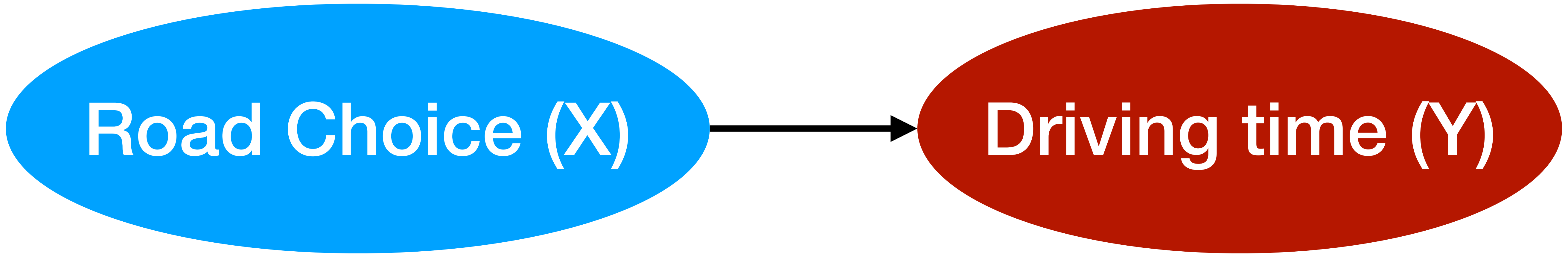


- I take option B and the driving time is 1 hour.

Adapted example from Primer (http://bayes.cs.ucla.edu/PRIMER/primer-ch4.pdf)

Counterfactual vs interventional questions - example

- Driving home, two options:
  A. Take freeway
  B. Take normal road



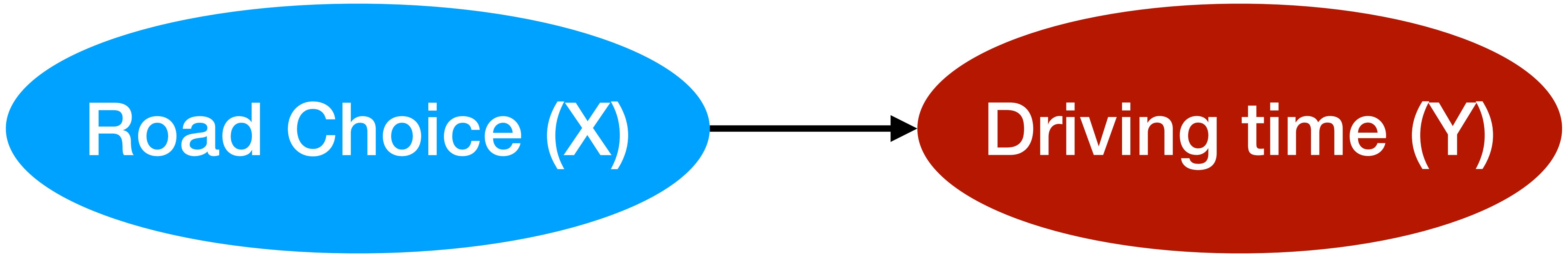
- I take option B and the driving time is 1 hour.
- Let's try to write down the counterfactual of taking the freeway:

$$E[Y \mid Y = 1\text{h}, \text{do}(X = A)]$$

Adapted example from Primer (http://bayes.cs.ucla.edu/PRIMER/primer-ch4.pdf)

Counterfactual vs interventional questions - example

- Driving home, two options:
  A. Take freeway
  B. Take normal road



- I take option B and the driving time is 1 hour.
- Let's try to write down the counterfactual of taking the freeway:

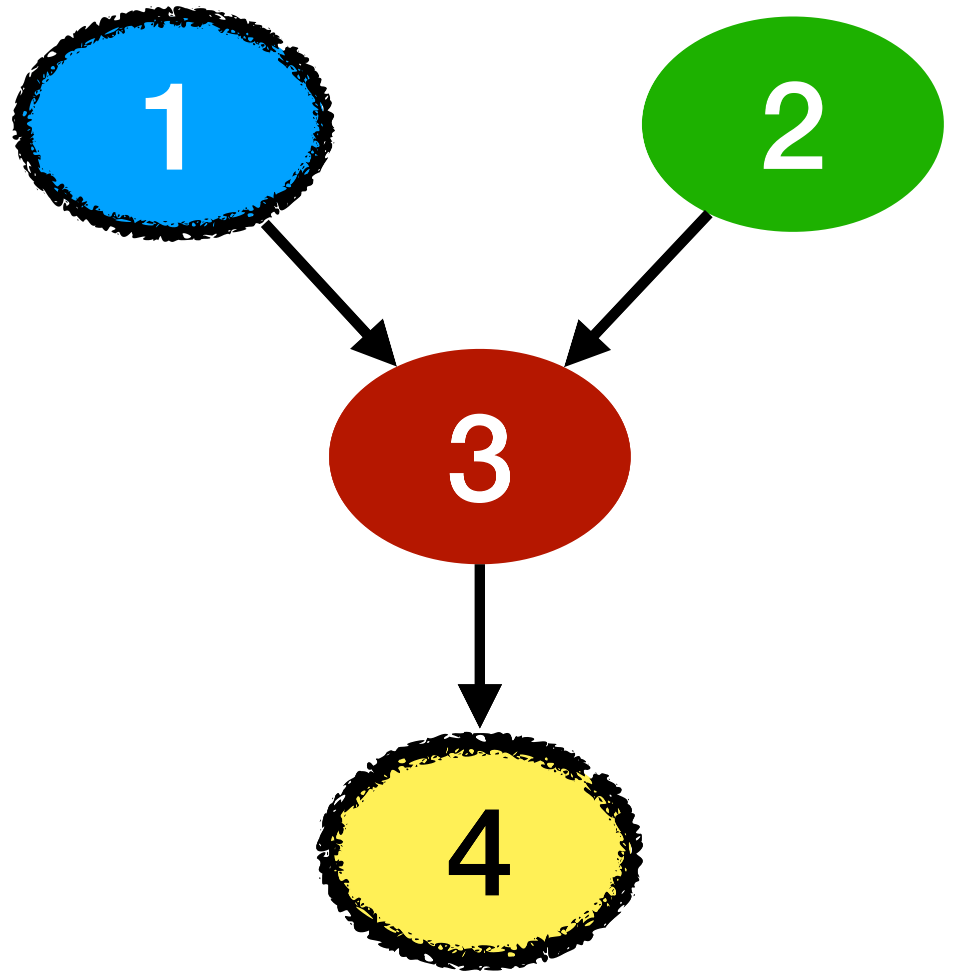
$$E[Y \mid Y = 1\text{h}, \text{do}(X = A)]$$

We need to distinguish the two Y, potential outcomes Y_A and Y_B

Adapted example from Primer (http://bayes.cs.ucla.edu/PRIMER/primer-ch4.pdf)

Counterfactuals example in a linear SCM

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \\ \forall i: \ \epsilon_i \sim N(0,1) \\ \forall i \neq j: \ \epsilon_i \perp\!\!\!\perp \epsilon_j \end{cases}$$



We want to estimate the effect of the treatment X_1 on the outcome X_4

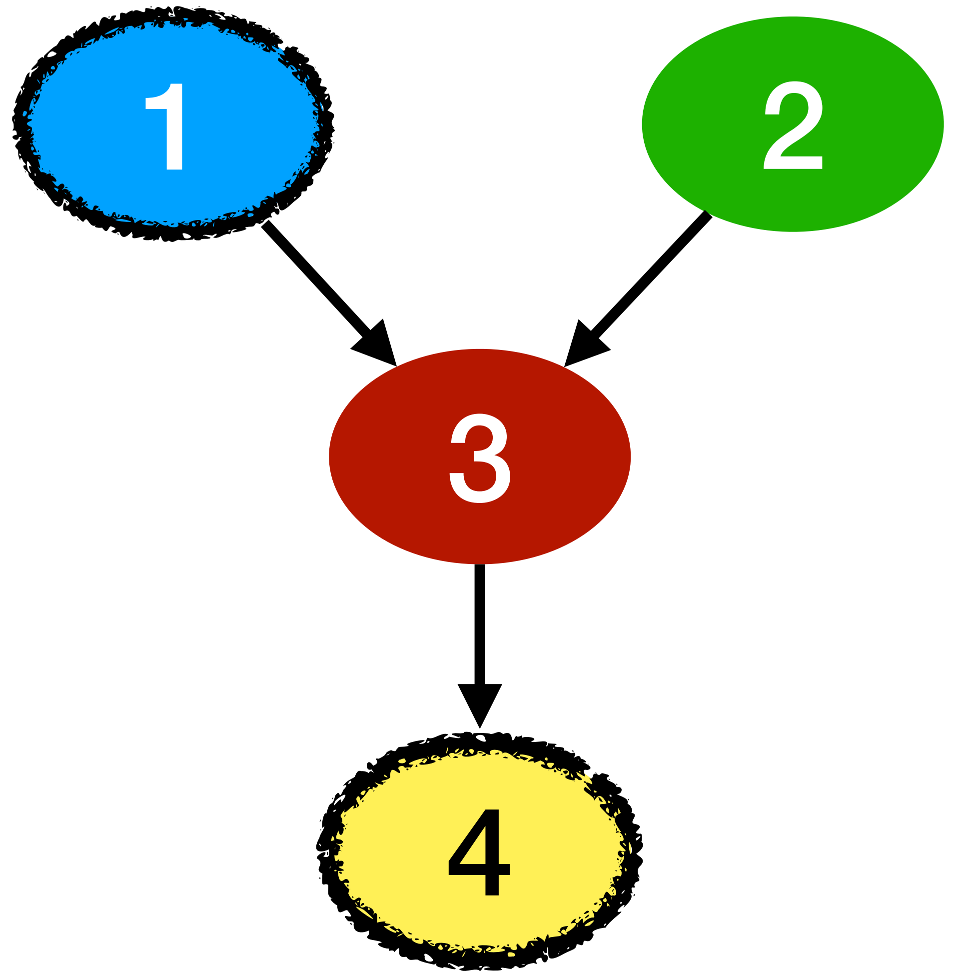
Counterfactuals example in a linear SCM

$do(X_1 = 0):$

$$\begin{cases} X_1 = 0 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \\ \forall i: \ \epsilon_i \sim N(0,1) \\ \forall i \neq j: \ \epsilon_i \perp\!\!\!\perp \epsilon_j \end{cases}$$

$do(X_1 = 1):$

$$\begin{cases} X_1 = 1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \\ \forall i: \ \epsilon_i \sim N(0,1) \\ \forall i \neq j: \ \epsilon_i \perp\!\!\!\perp \epsilon_j \end{cases}$$



We want to estimate the effect of the treatment X_1 on the outcome X_4

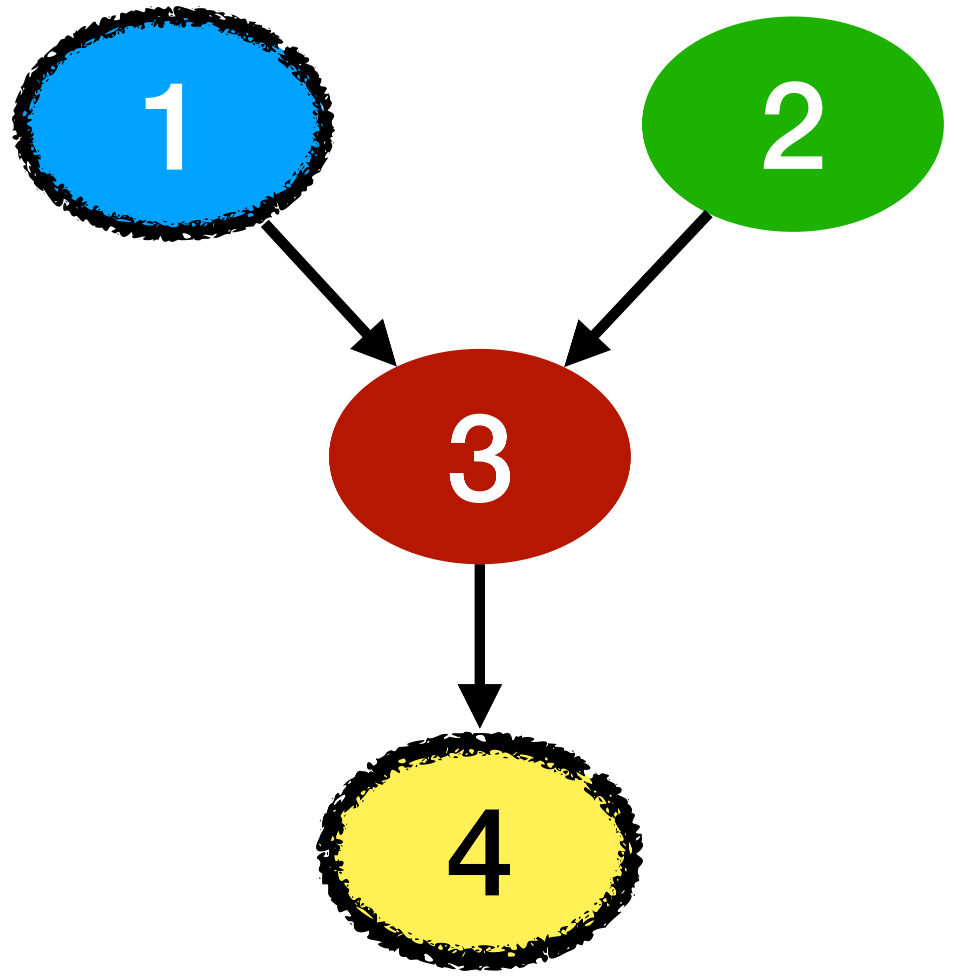
Counterfactuals example in a linear SCM

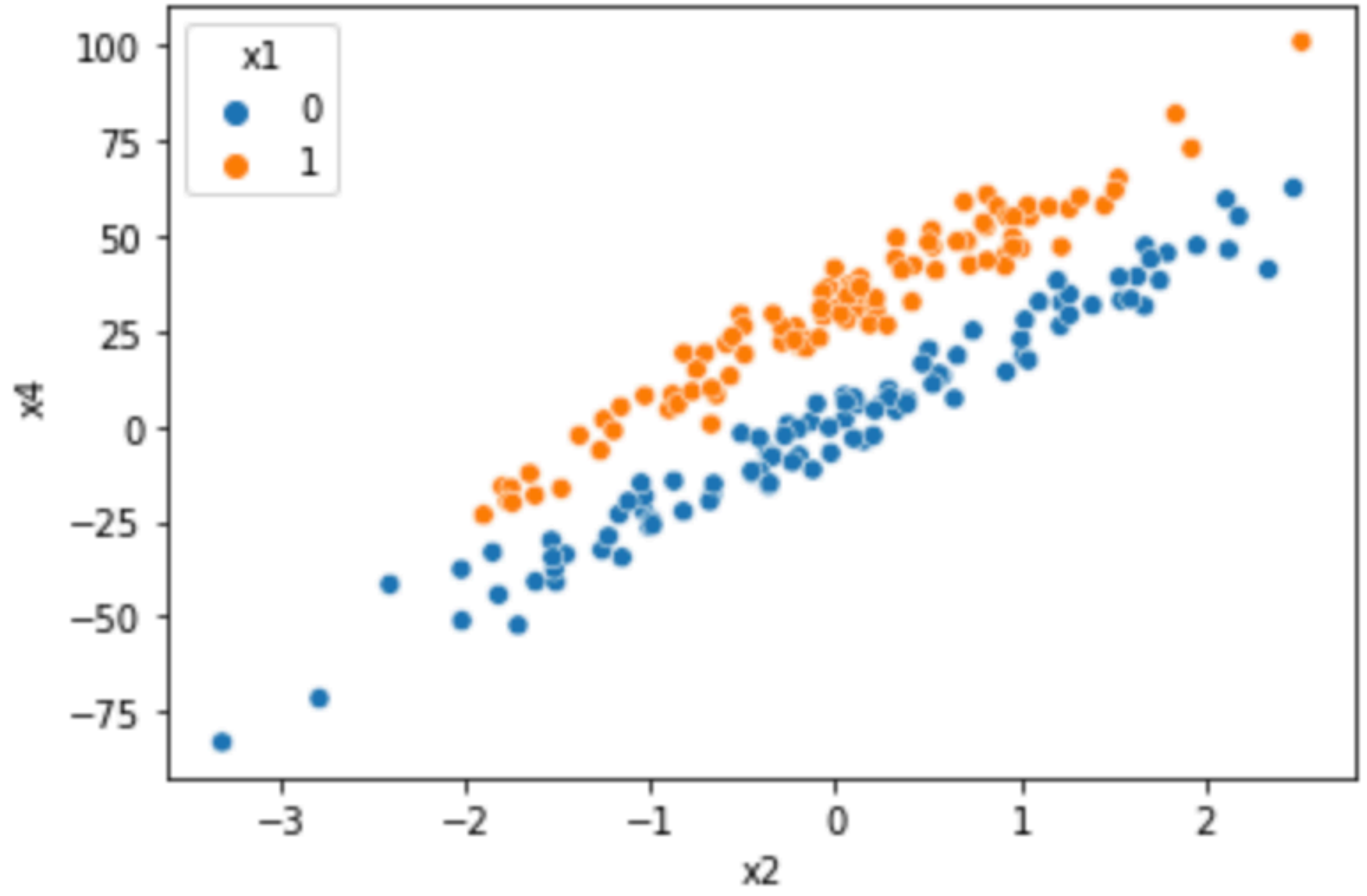
$do(X_1 = 0)$:

$$\begin{cases} X_1 = 0 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$

$do(X_1 = 1)$:

$$\begin{cases} X_1 = 1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$





$P(X_4 \mid X_2, do(X_1 = 1))$

$P(X_4 \mid X_2, do(X_1 = 0))$

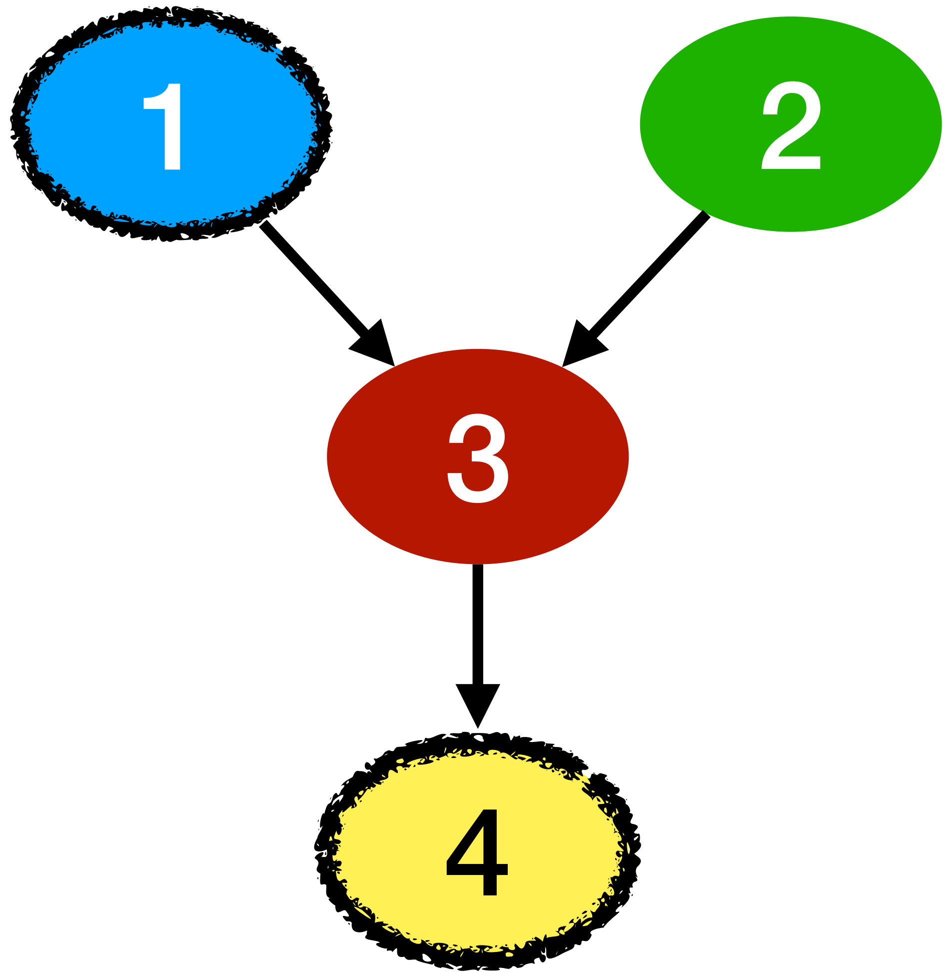
Counterfactuals example in a linear SCM

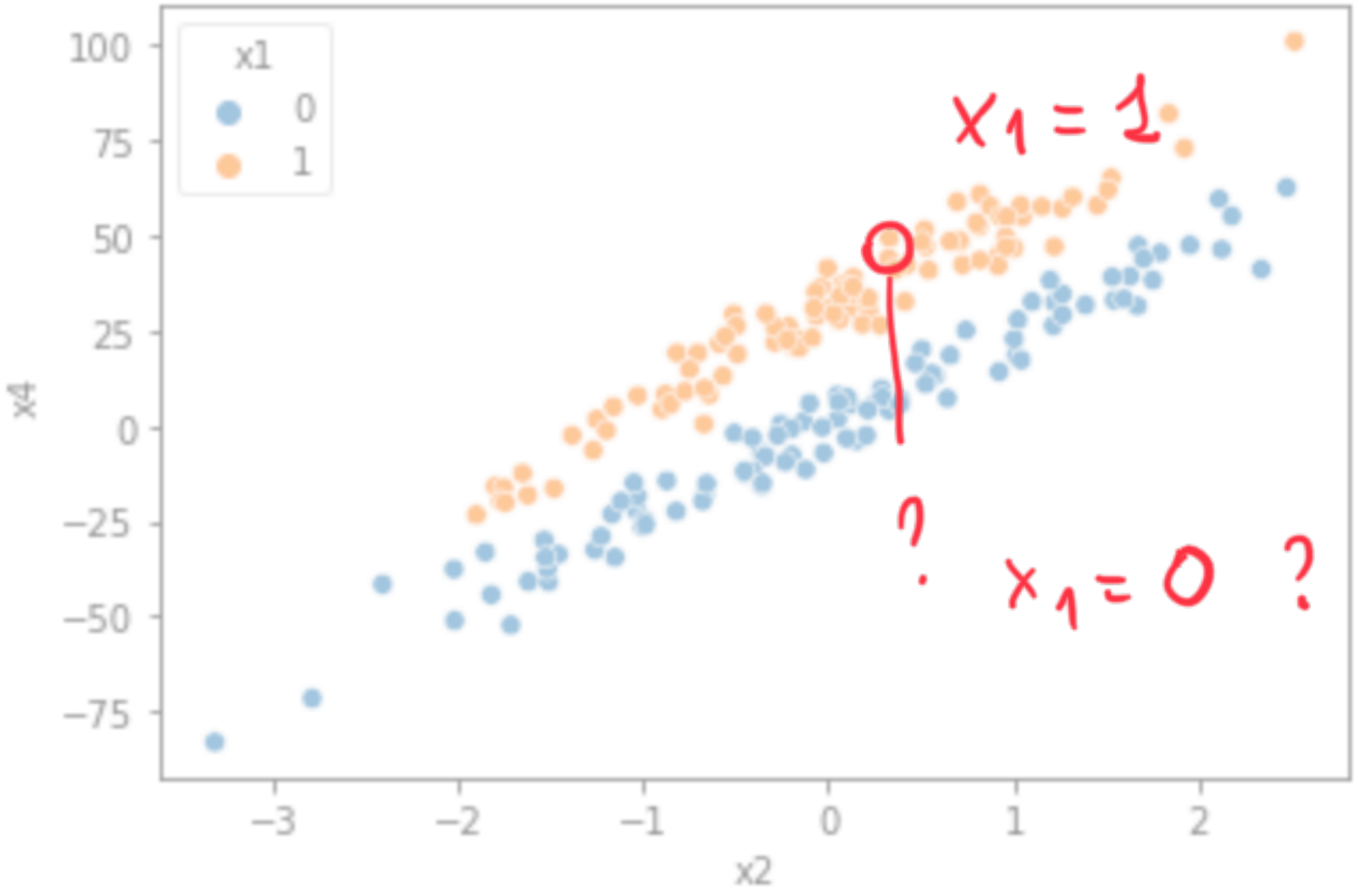
$do(X_1 = 0)$:

$$\begin{cases} X_1 = 0 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$

$do(X_1 = 1)$:

$$\begin{cases} X_1 = 1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$





$P(X_4 \mid X_2, do(X_1 = 1))$

$P(X_4 \mid X_2, do(X_1 = 0))$

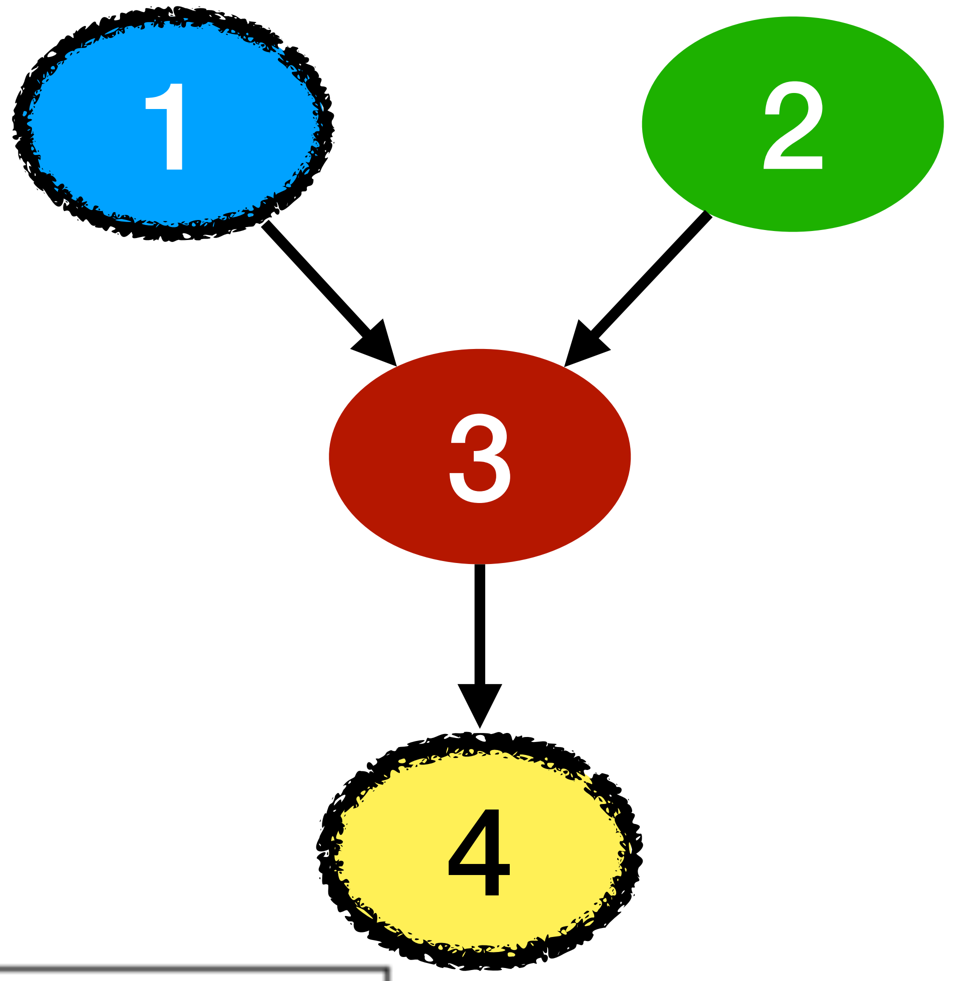
Counterfactuals example in a linear SCM

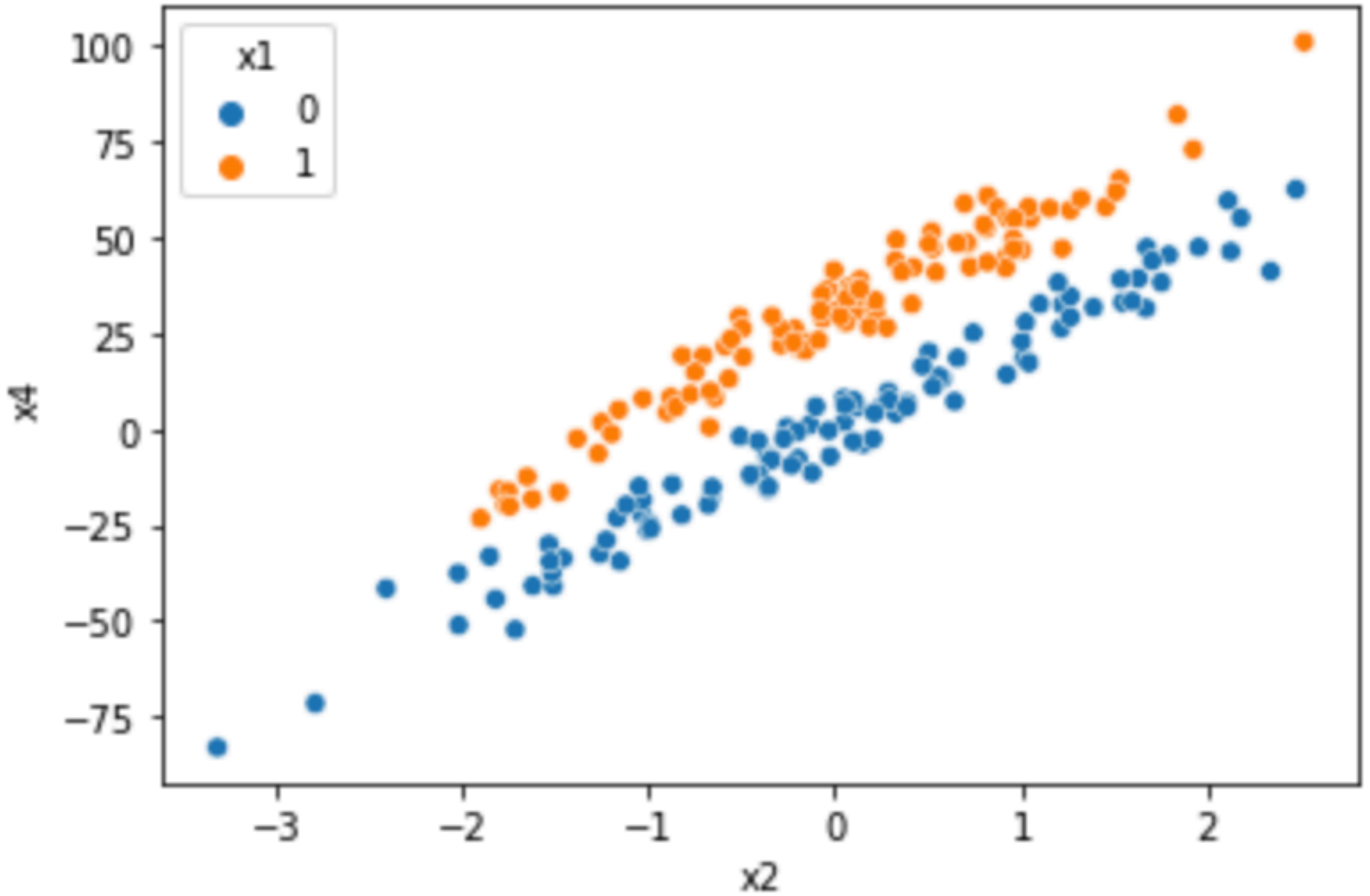
$do(X_1 = 0):$

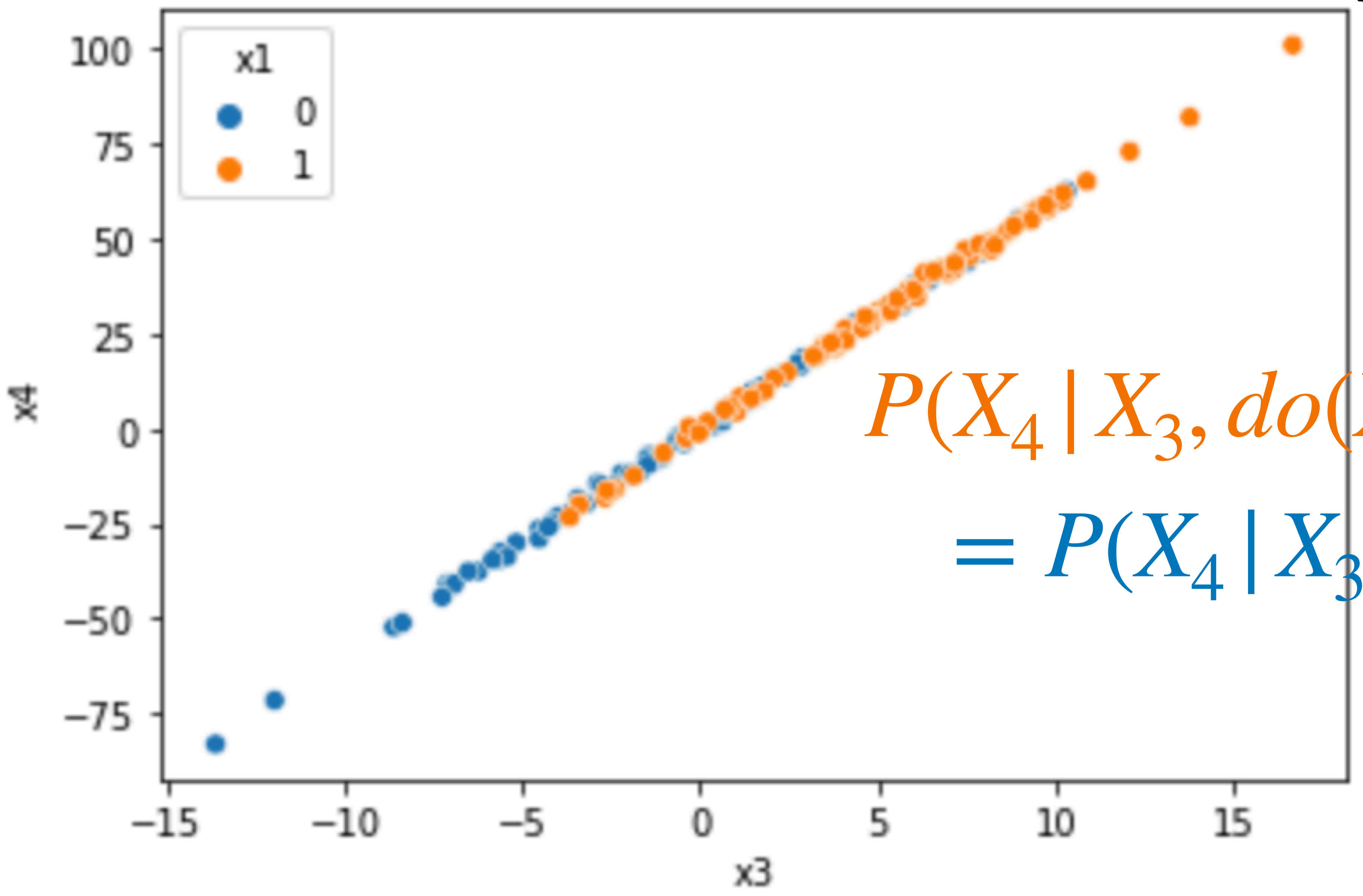
$$\begin{cases} X_1 = 0 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$

$do(X_1 = 1):$

$$\begin{cases} X_1 = 1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$



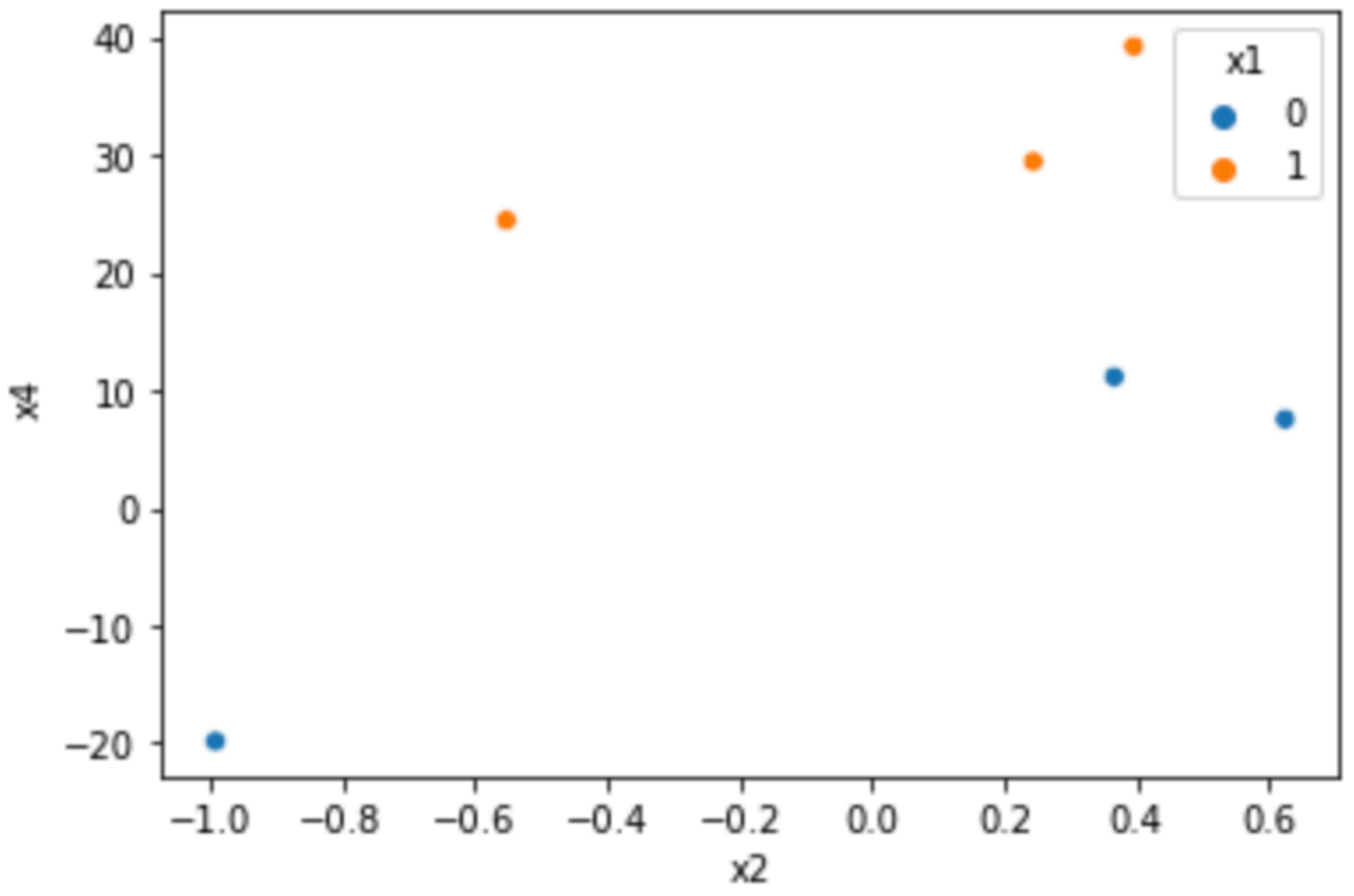


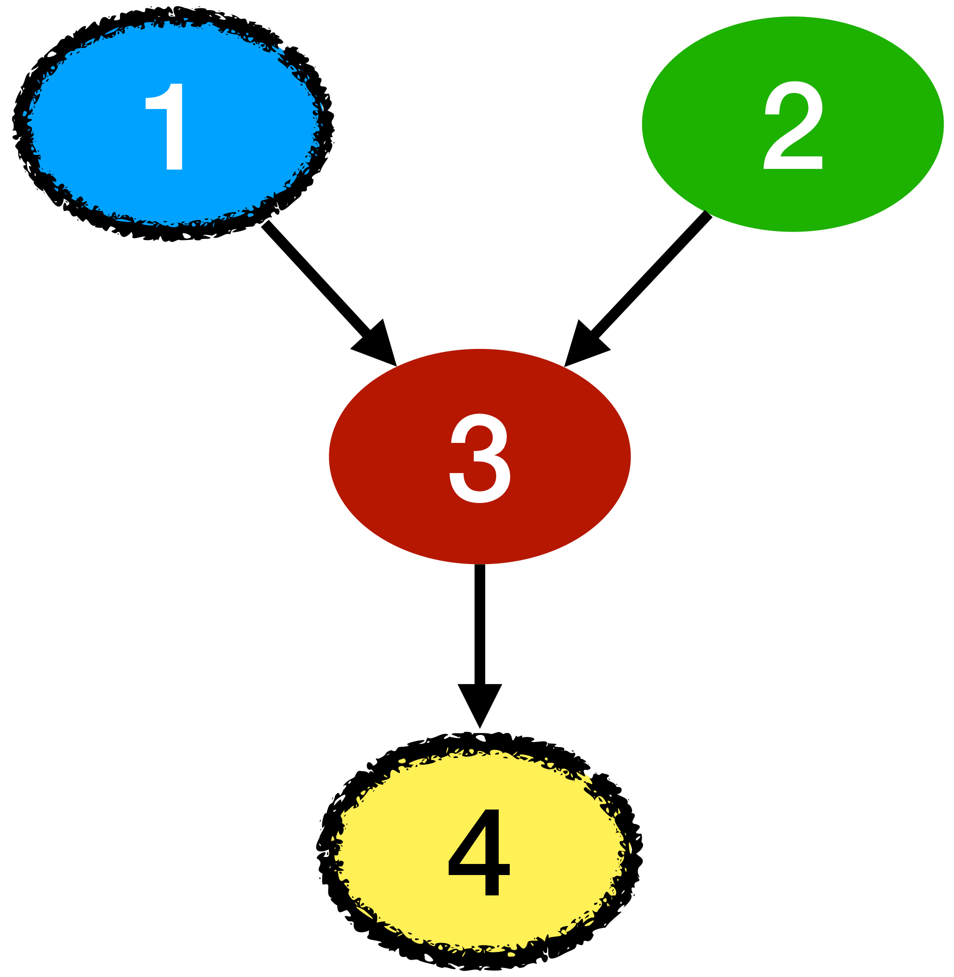


$$P(X_4 \mid X_3, do(X_1 = 1)) = P(X_4 \mid X_3, do(X_1 = 0))$$

Counterfactuals example in a linear SCM

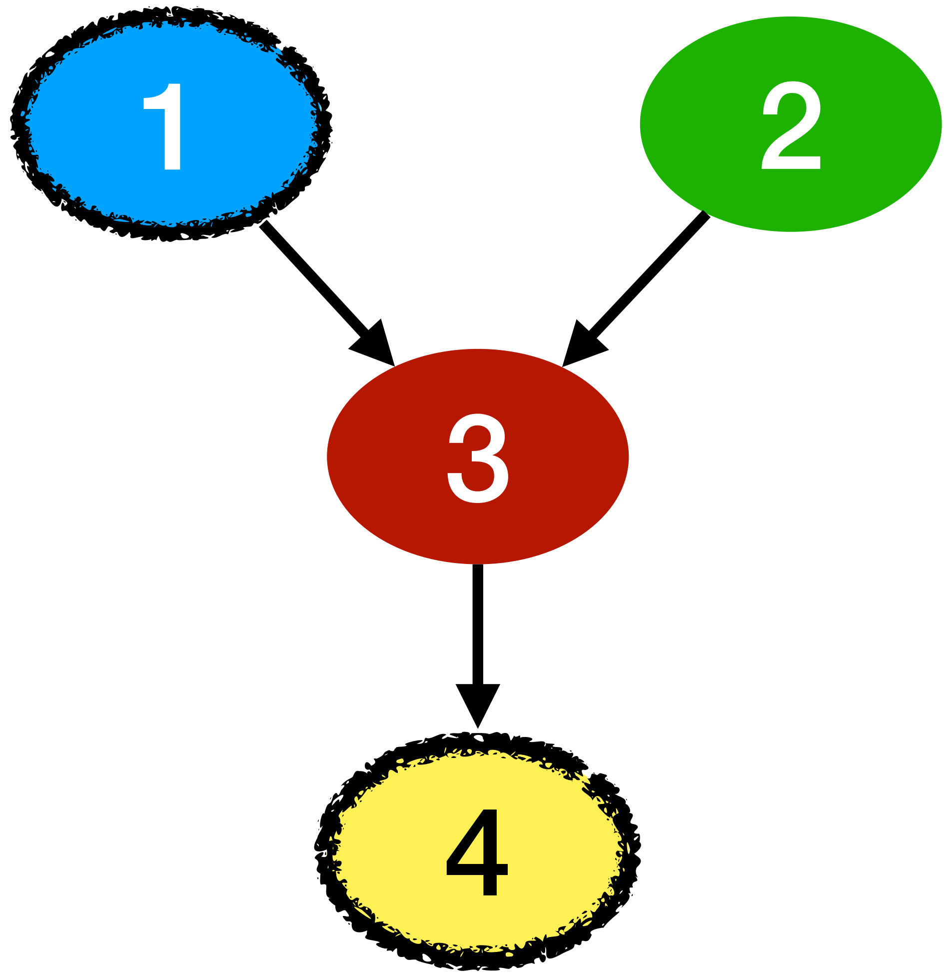
x1	x2	x3	x4
0	0.624457	1.447194	7.546849
0	-0.991089	-3.136825	-19.832298
0	0.366517	1.953662	11.142927
1	-0.551737	4.191562	24.444811
1	0.244229	4.923147	29.444147
1	0.396024	6.844468	39.227687

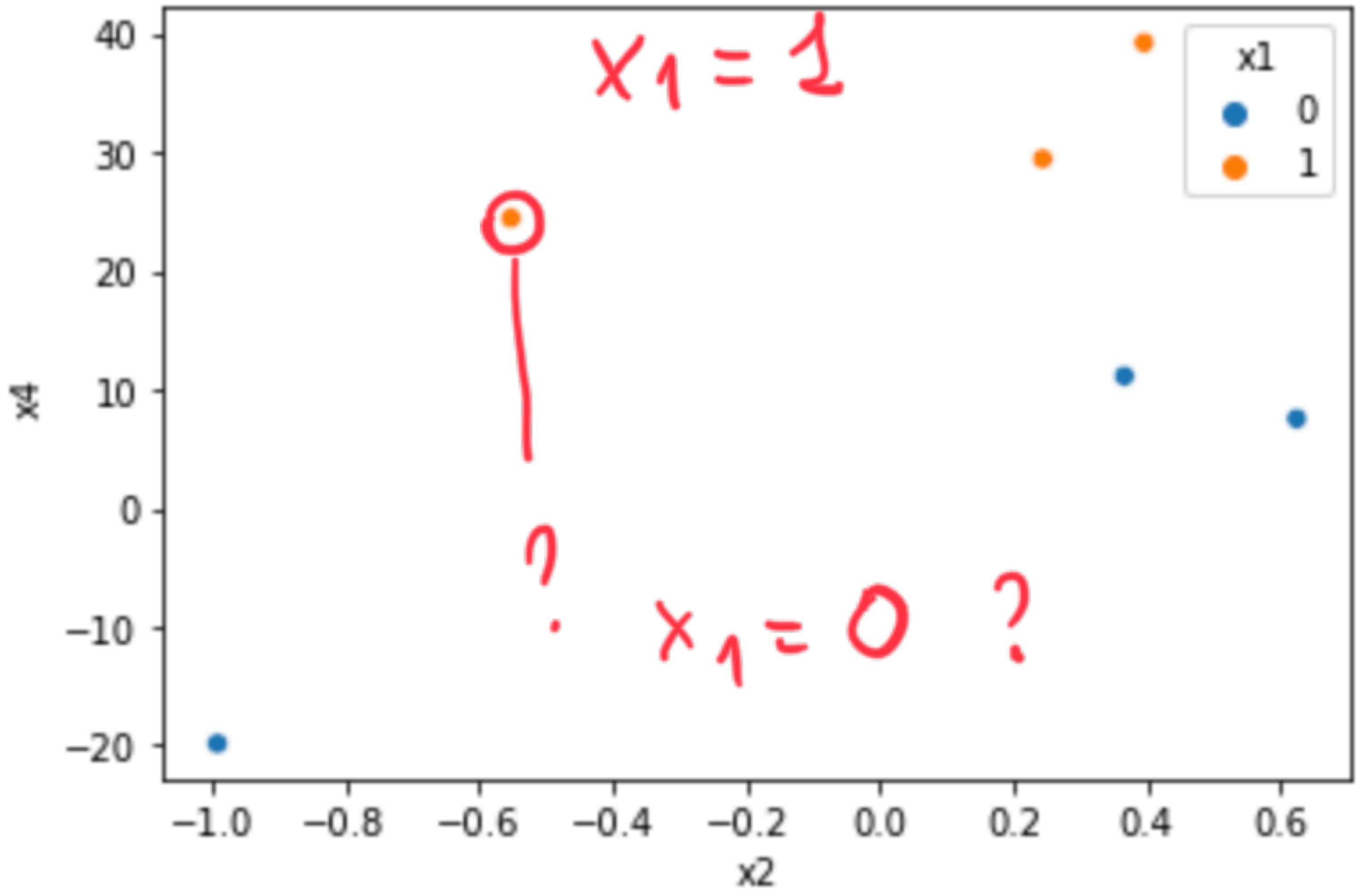




Counterfactuals example in a linear SCM

x1	x2	x3	x4
0	0.624457	1.447194	7.546849
0	-0.991089	-3.136825	-19.832298
0	0.366517	1.953662	11.142927
1	-0.551737	4.191562	24.444811
1	0.244229	4.923147	29.444147
1	0.396024	6.844468	39.227687



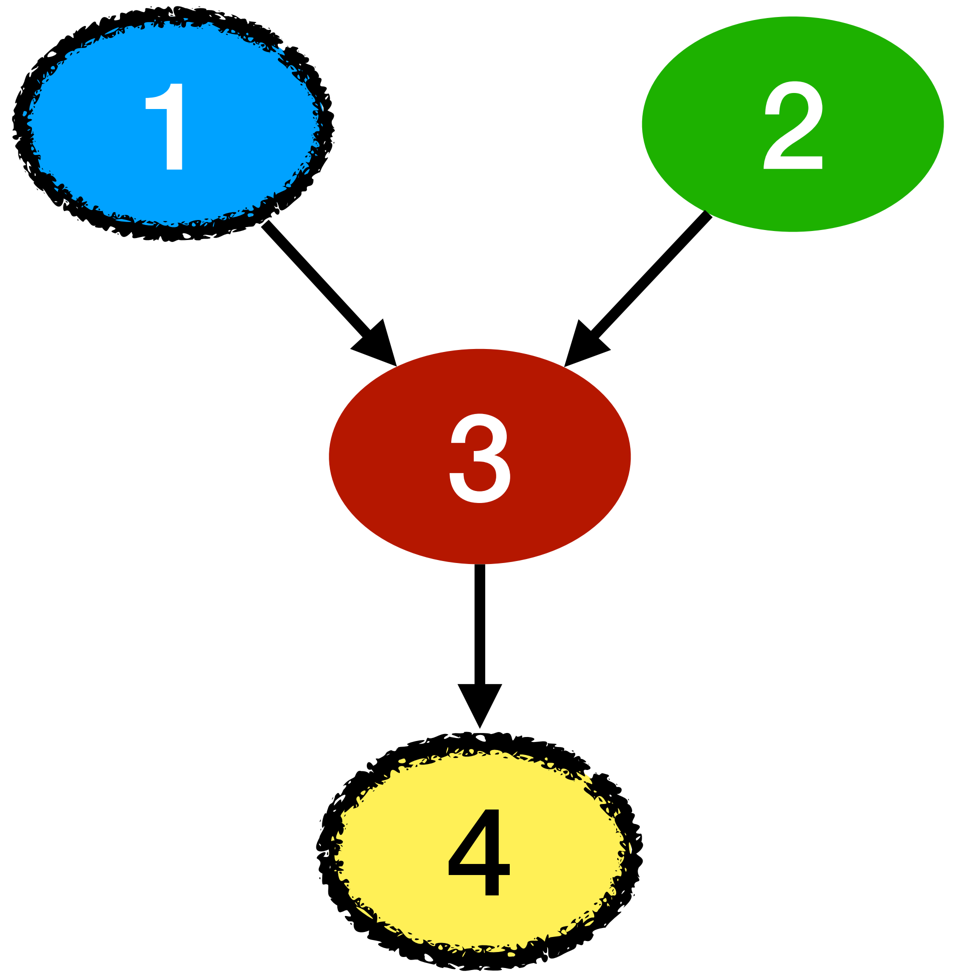


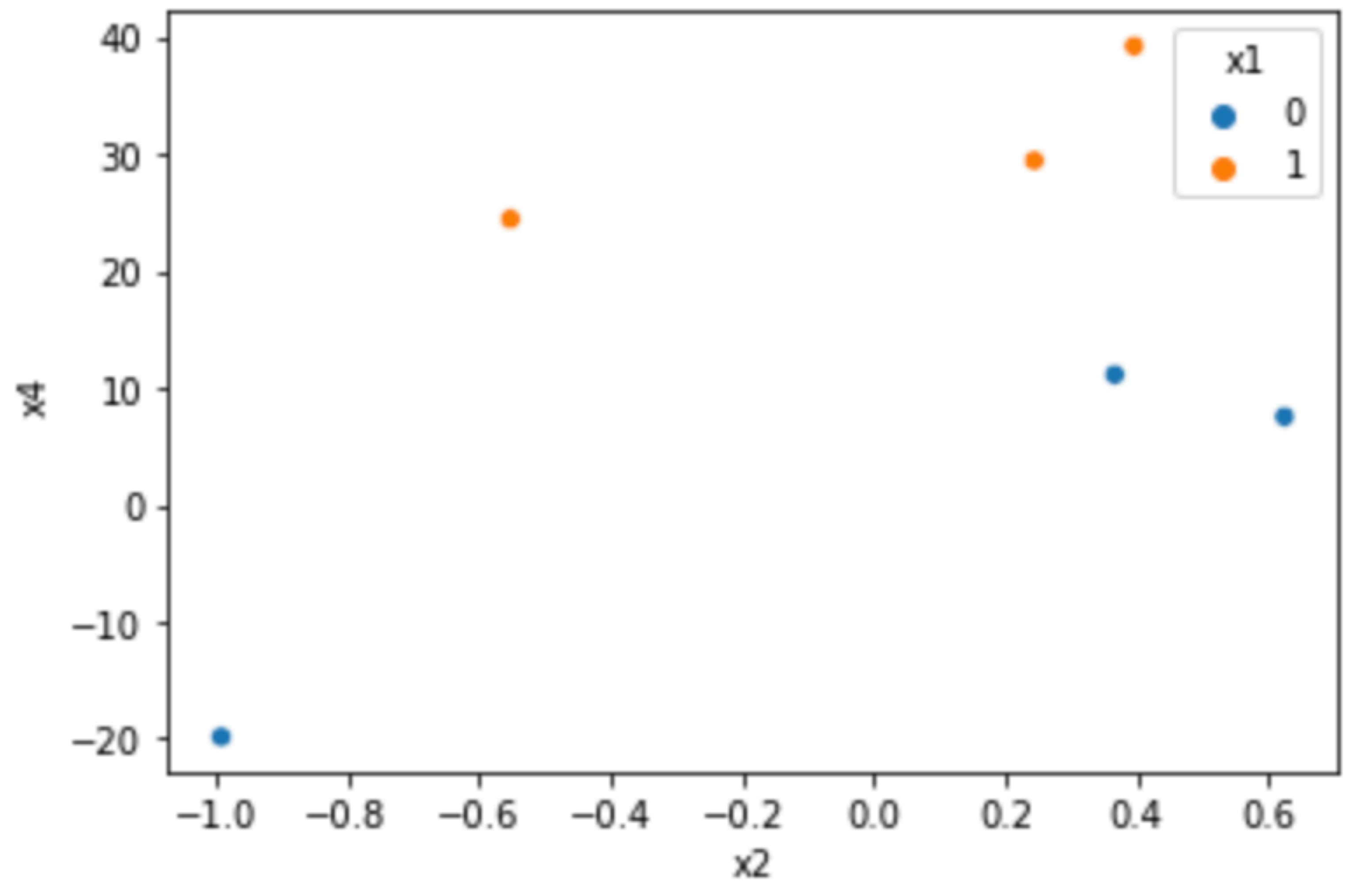
With fewer points it's even more difficult to estimate…

Counterfactuals example in a linear SCM

x1	x2	x3	x4
0	0.624457	1.447194	7.546849
0	-0.991089	-3.136825	-19.832298
0	0.366517	1.953662	11.142927
1	-0.551737	4.191562	24.444811
1	0.244229	4.923147	29.444147
1	0.396024	6.844468	39.227687

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$



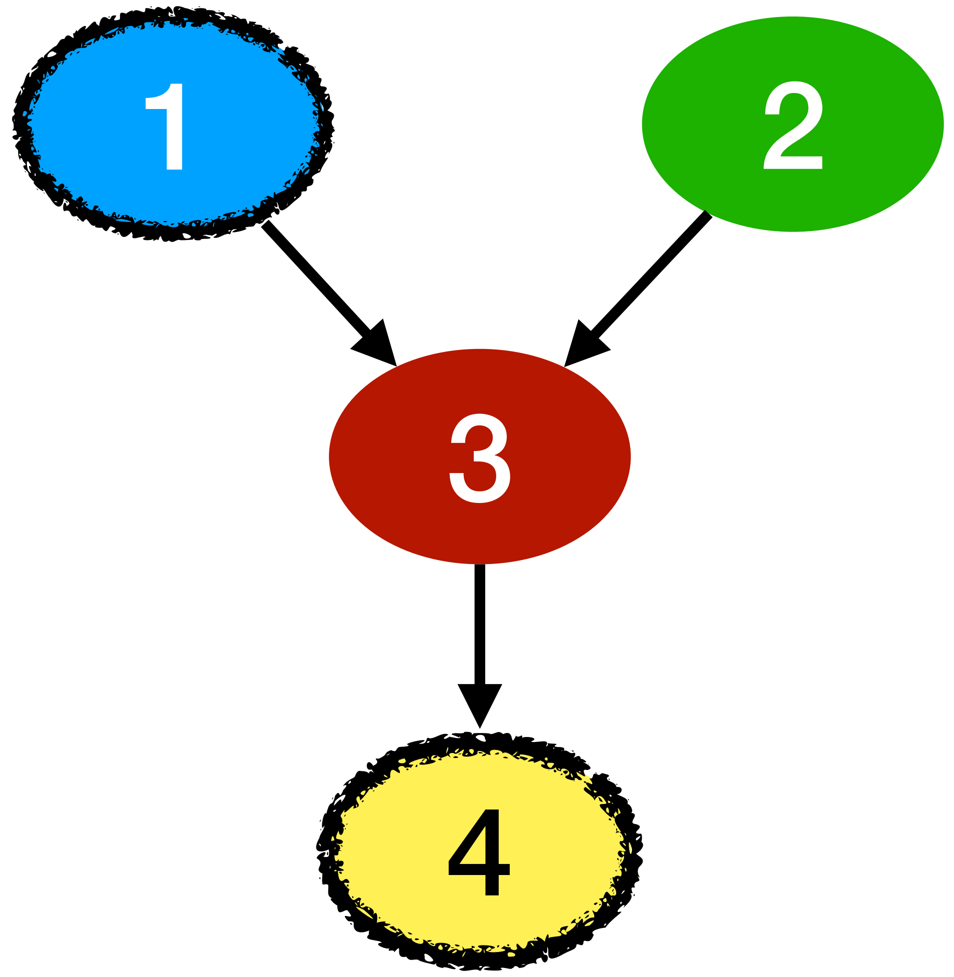


We know the true SCM, can we reconstruct the value of ϵ_2 for each sample?

Counterfactuals example in a linear SCM

x1	x2	x3	x4
0	0.624457	1.447194	7.546849
0	-0.991089	-3.136825	-19.832298
0	0.366517	1.953662	11.142927
1	-0.551737	4.191562	24.444811
1	0.244229	4.923147	29.444147
1	0.396024	6.844468	39.227687

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$



x1	x2	x3	x4	epsilon_x2
0	0.624457	1.447194	7.546849	0.624457
0	-0.991089	-3.136825	-19.832298	-0.991089
0	0.366517	1.953662	11.142927	0.366517
1	-0.551737	4.191562	24.444811	-0.551737
1	0.244229	4.923147	29.444147	0.244229
1	0.396024	6.844468	39.227687	0.396024

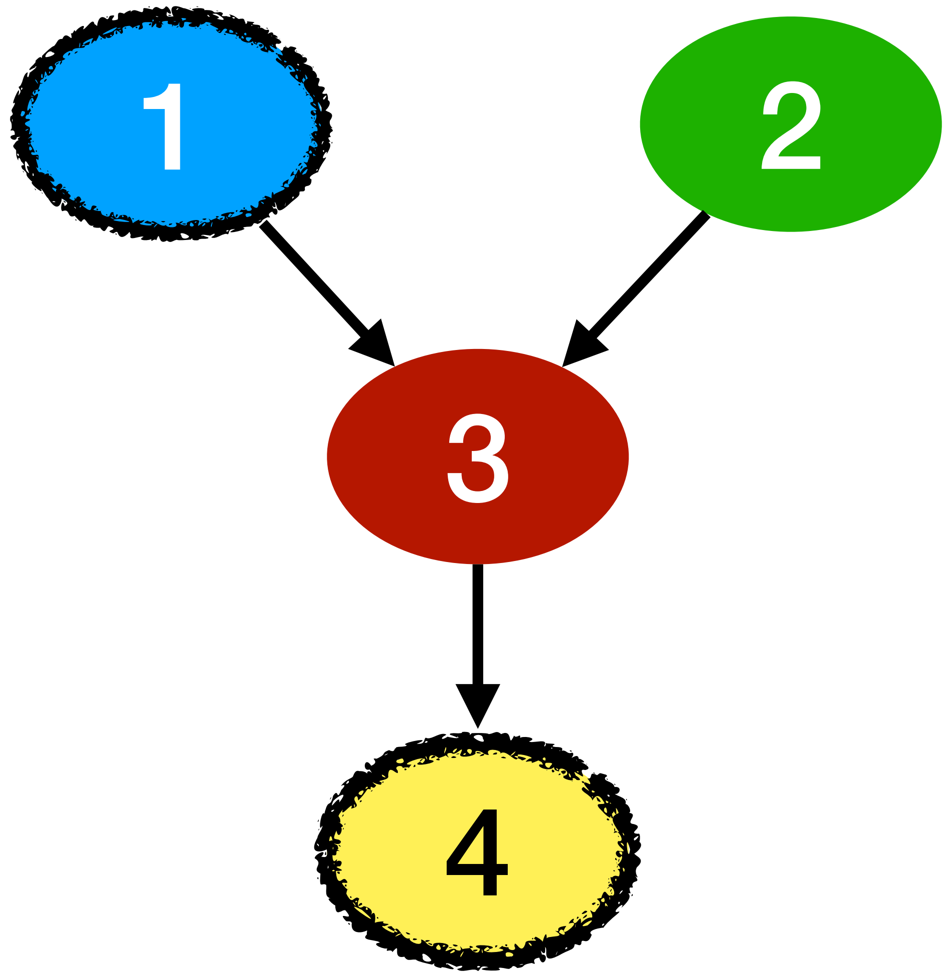
$$\epsilon_2 = X_2$$

What about the other epsilons?

Counterfactuals example in a linear SCM

x1	x2	x3	x4
0	0.624457	1.447194	7.546849
0	-0.991089	-3.136825	-19.832298
0	0.366517	1.953662	11.142927
1	-0.551737	4.191562	24.444811
1	0.244229	4.923147	29.444147
1	0.396024	6.844468	39.227687

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$



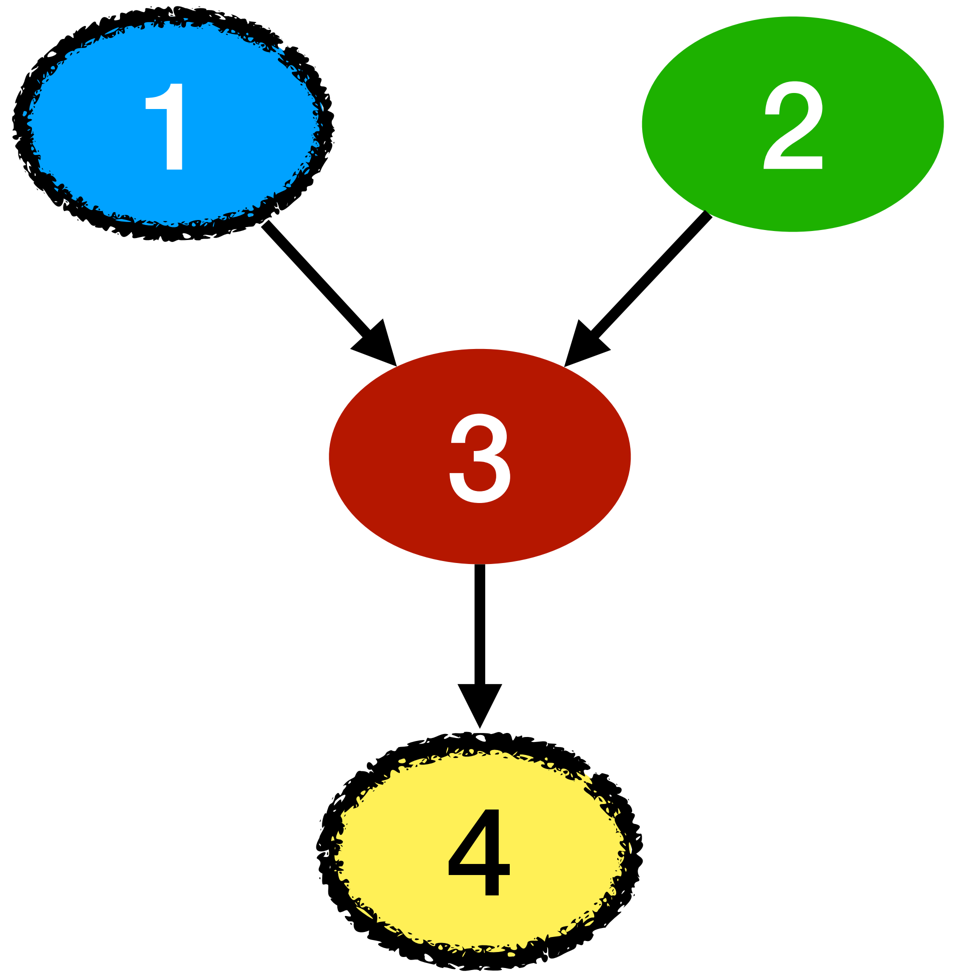
x1	x2	x3	x4	epsilon_x2	epsilon_x3	epsilon_x4
0	0.624457	1.447194	7.546849	0.624457	-1.050634	-1.136316
0	-0.991089	-3.136825	-19.832298	-0.991089	0.827529	-1.011348
0	0.366517	1.953662	11.142927	0.366517	0.487593	-0.579048
1	-0.551737	4.191562	24.444811	-0.551737	1.398509	-0.704563
1	0.244229	4.923147	29.444147	0.244229	-1.053768	-0.094737
1	0.396024	6.844468	39.227687	0.396024	0.260372	-1.839122

$$\begin{cases} \epsilon_2 = X_2 \\ \epsilon_3 = -5X_1 - 4X_2 + X_3 \\ \epsilon_4 = -6X_3 + X_4 \end{cases}$$

Step 1: Abduction

Counterfactuals example in a linear SCM

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$



x1	x2	x3	x4	epsilon_x2	epsilon_x3	epsilon_x4	x1c
0	0.624457	1.447194	7.546849	0.624457	-1.050634	-1.136316	1
0	-0.991089	-3.136825	-19.832298	-0.991089	0.827529	-1.011348	1
0	0.366517	1.953662	11.142927	0.366517	0.487593	-0.579048	1
1	-0.551737	4.191562	24.444811	-0.551737	1.398509	-0.704563	0
1	0.244229	4.923147	29.444147	0.244229	-1.053768	-0.094737	0
1	0.396024	6.844468	39.227687	0.396024	0.260372	-1.839122	0

$$X_1^c = 1 - \epsilon_1 = \begin{cases} 1 \text{ if } \epsilon_1 = 0 \\ 0 \text{ if } \epsilon_1 = 1 \end{cases}$$

The rest of the SCM stays the same

Step 2: Action

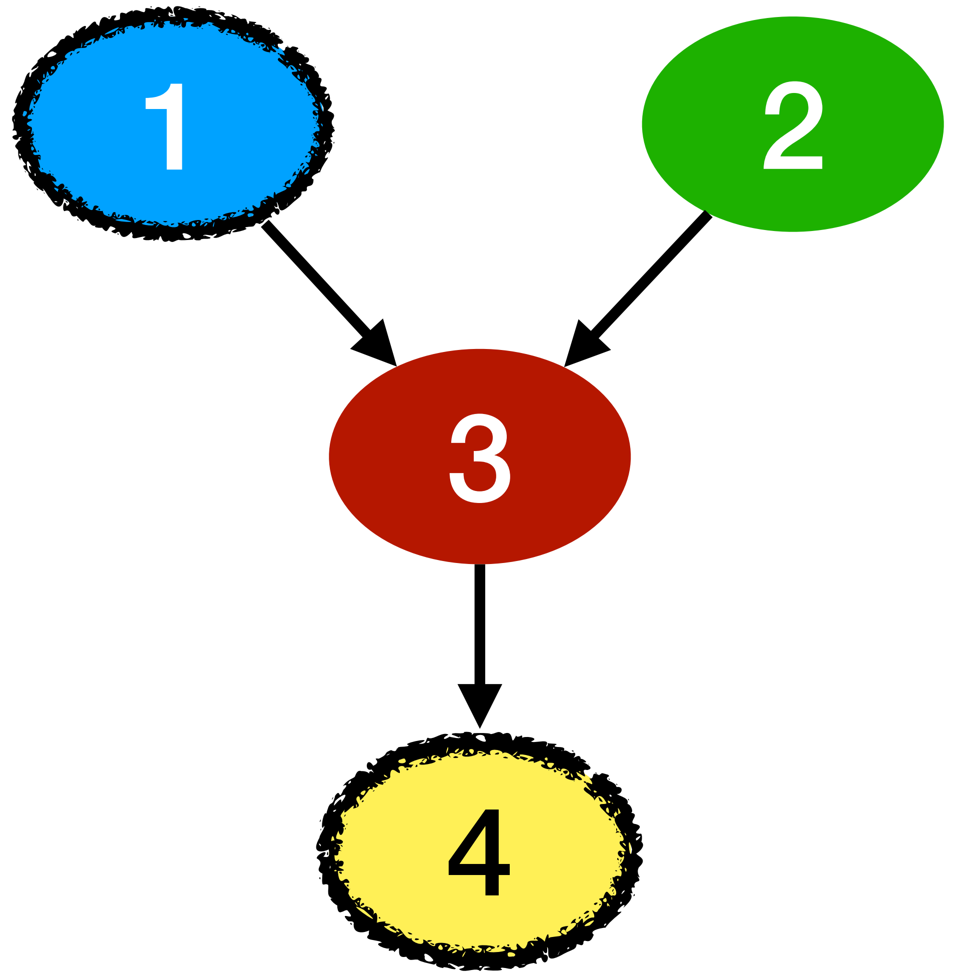
Counterfactuals example in a linear SCM

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$

Factual SCM

$$\begin{cases} X_1^c = 1 - \epsilon_1 \\ X_2^c = \epsilon_2 \\ X_3^c = 5X_1^c + 4X_2^c + \epsilon_3 \\ X_4^c = 6X_3^c + \epsilon_4 \end{cases}$$

Counterfactual SCM



x1	x2	x3	x4	epsilon_x2	epsilon_x3	epsilon_x4	x1c
0	0.624457	1.447194	7.546849	0.624457	-1.050634	-1.136316	1
0	-0.991089	-3.136825	-19.832298	-0.991089	0.827529	-1.011348	1
0	0.366517	1.953662	11.142927	0.366517	0.487593	-0.579048	1
1	-0.551737	4.191562	24.444811	-0.551737	1.398509	-0.704563	0
1	0.244229	4.923147	29.444147	0.244229	-1.053768	-0.094737	0
1	0.396024	6.844468	39.227687	0.396024	0.260372	-1.839122	0

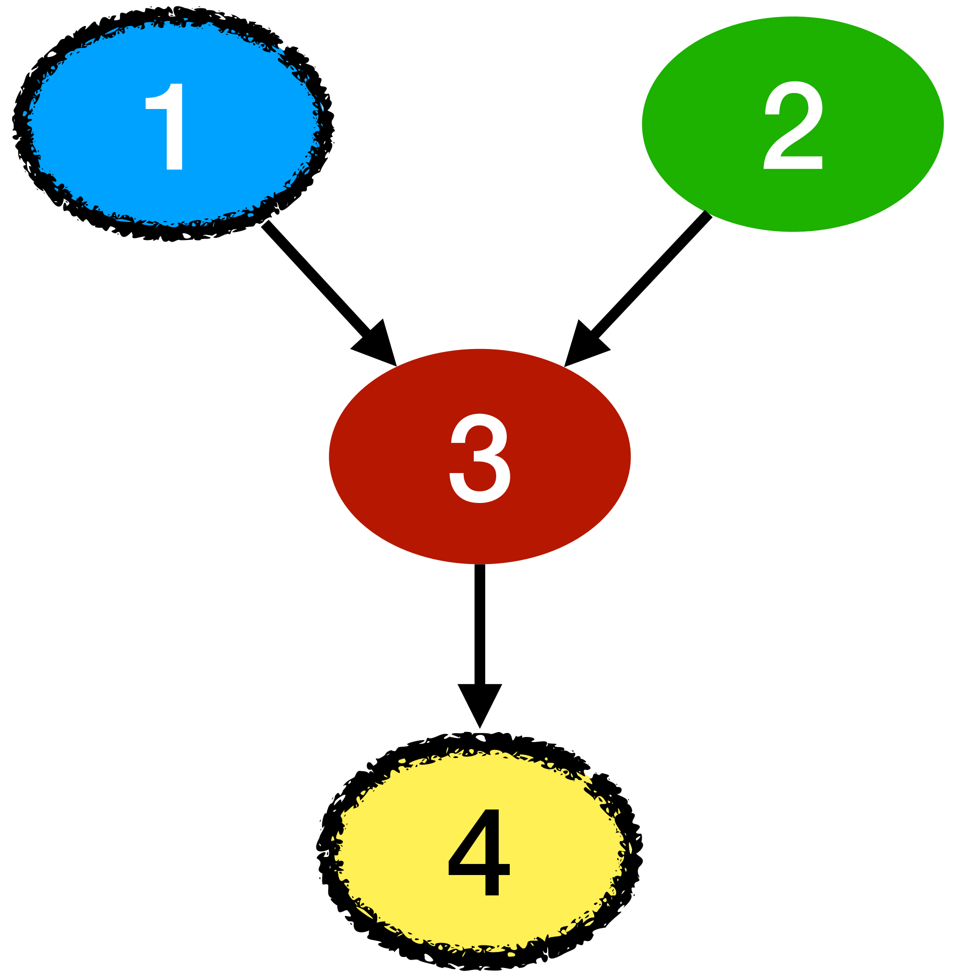
Counterfactuals example in a linear SCM

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$

Factual SCM

$$\begin{cases} X_1^c = 1 - \epsilon_1 \\ X_2^c = \epsilon_2 \\ X_3^c = 5X_1^c + 4X_2^c + \epsilon_3 \\ X_4^c = 6X_3^c + \epsilon_4 \end{cases}$$

Counterfactual SCM



x1	x2	x3	x4	epsilon_x2	epsilon_x3	epsilon_x4	x1c
0	0.624457	1.447194	7.546849	0.624457	-1.050634	-1.136316	1
0	-0.991089	-3.136825	-19.832298	-0.991089	0.827529	-1.011348	1
0	0.366517	1.953662	11.142927	0.366517	0.487593	-0.579048	1
1	-0.551737	4.191562	24.444811	-0.551737	1.398509	-0.704563	0
1	0.244229	4.923147	29.444147	0.244229	-1.053768	-0.094737	0
1	0.396024	6.844468	39.227687	0.396024	0.260372	-1.839122	0

We can predict the counterfactual values with the Counterfactual SCM and the data

Step 3: Prediction

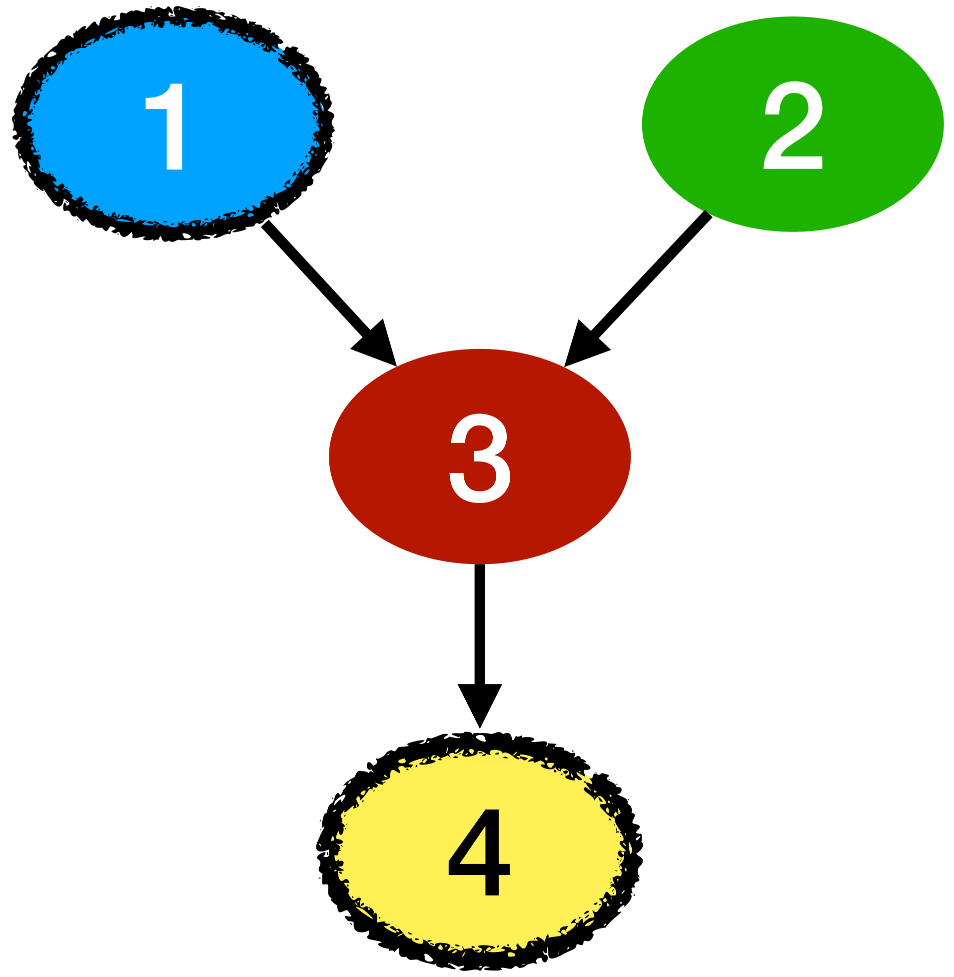
Counterfactuals example in a linear SCM

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$

Factual SCM

$$\begin{cases} X_1^c = 1 - \epsilon_1 \\ X_2^c = \epsilon_2 \\ X_3^c = 5X_1^c + 4X_2^c + \epsilon_3 \\ X_4^c = 6X_3^c + \epsilon_4 \end{cases}$$

Counterfactual SCM



x1	x2	x3	x4	epsilon_x2	epsilon_x3	epsilon_x4	x1c
0	0.624457	1.447194	7.546849	0.624457	-1.050634	-1.136316	1
0	-0.991089	-3.136825	-19.832298	-0.991089	0.827529	-1.011348	1
0	0.366517	1.953662	11.142927	0.366517	0.487593	-0.579048	1
1	-0.551737	4.191562	24.444811	-0.551737	1.398509	-0.704563	0
1	0.244229	4.923147	29.444147	0.244229	-1.053768	-0.094737	0
1	0.396024	6.844468	39.227687	0.396024	0.260372	-1.839122	0

What are the values for X_2?

Step 3: Prediction

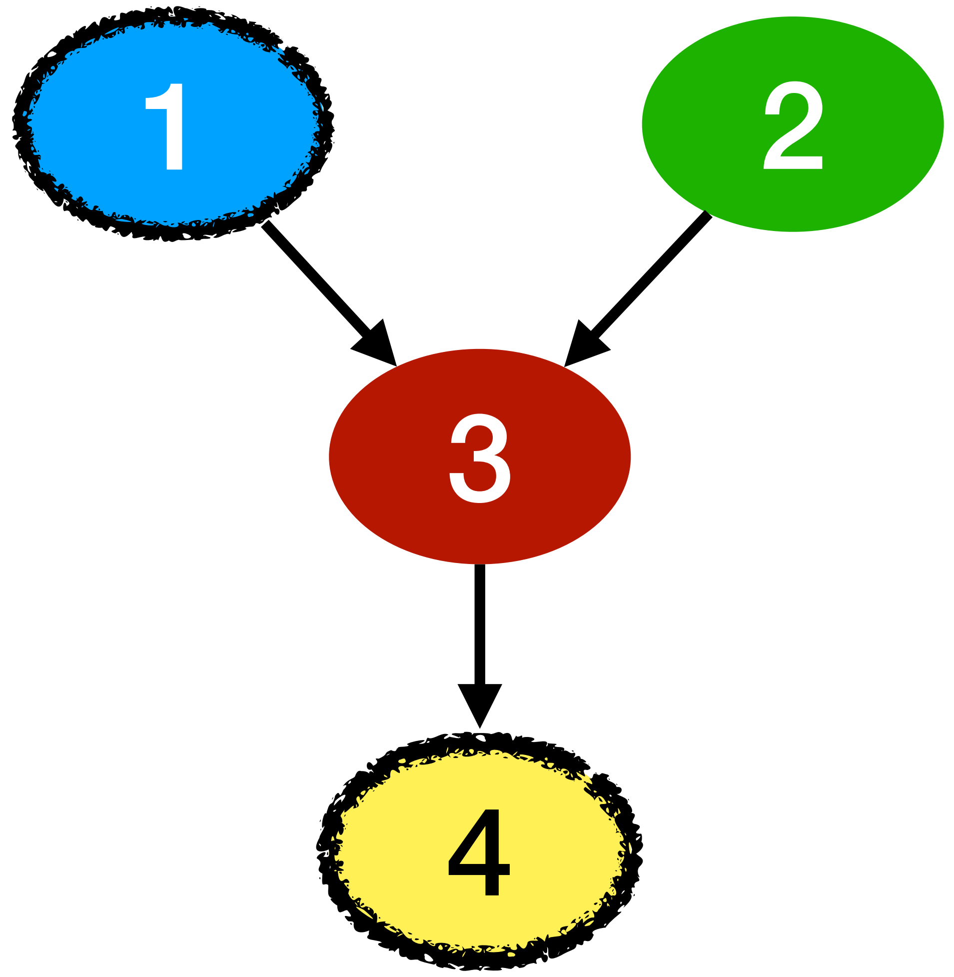
Counterfactuals example in a linear SCM

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$

Factual SCM

$$\begin{cases} X_1^c = 1 - \epsilon_1 \\ X_2^c = \epsilon_2 \\ X_3^c = 5X_1^c + 4X_2^c + \epsilon_3 \\ X_4^c = 6X_3^c + \epsilon_4 \end{cases}$$

Counterfactual SCM



x1	x2	x3	x4	epsilon_x2	epsilon_x3	epsilon_x4	x1c	x2c
0	0.624457	1.447194	7.546849	0.624457	-1.050634	-1.136316	1	0.624457
0	-0.991089	-3.136825	-19.832298	-0.991089	0.827529	-1.011348	1	-0.991089
0	0.366517	1.953662	11.142927	0.366517	0.487593	-0.579048	1	0.366517
1	-0.551737	4.191562	24.444811	-0.551737	1.398509	-0.704563	0	-0.551737
1	0.244229	4.923147	29.444147	0.244229	-1.053768	-0.094737	0	0.244229
1	0.396024	6.844468	39.227687	0.396024	0.260372	-1.839122	0	0.396024

Step 3: Prediction

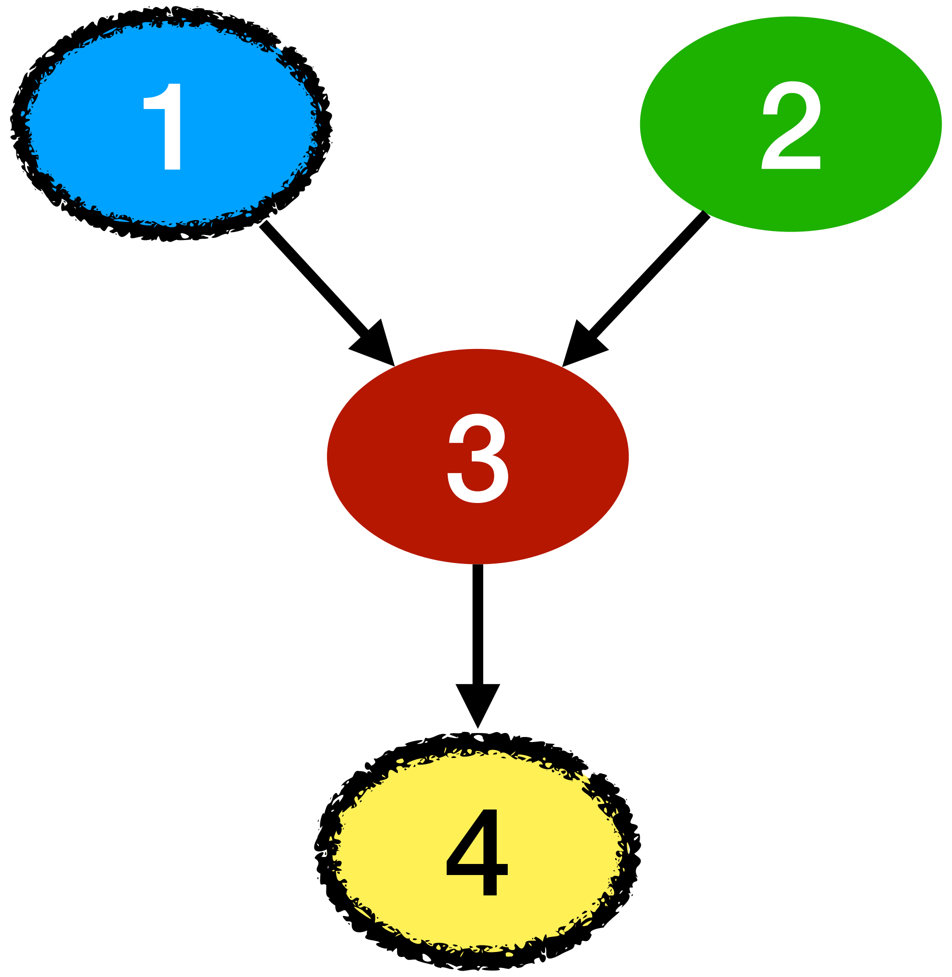
Counterfactuals example in a linear SCM

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$

Factual SCM

$$\begin{cases} X_1^c = 1 - \epsilon_1 \\ X_2^c = \epsilon_2 \\ X_3^c = 5X_1^c + 4X_2^c + \epsilon_3 \\ X_4^c = 6X_3^c + \epsilon_4 \end{cases}$$

Counterfactual SCM



x1	x2	x3	x4	epsilon_x2	epsilon_x3	epsilon_x4	x1c	x2c	x3c	x4c
0	0.624457	1.447194	7.546849	0.624457	-1.050634	-1.136316	1	0.624457	6.447194	37.546849
0	-0.991089	-3.136825	-19.832298	-0.991089	0.827529	-1.011348	1	-0.991089	1.863175	10.167702
0	0.366517	1.953662	11.142927	0.366517	0.487593	-0.579048	1	0.366517	6.953662	41.142927
1	-0.551737	4.191562	24.444811	-0.551737	1.398509	-0.704563	0	-0.551737	-0.808438	-5.555189
1	0.244229	4.923147	29.444147	0.244229	-1.053768	-0.094737	0	0.244229	-0.076853	-0.555853
1	0.396024	6.844468	39.227687	0.396024	0.260372	-1.839122	0	0.396024	1.844468	9.227687

Step 3: Prediction

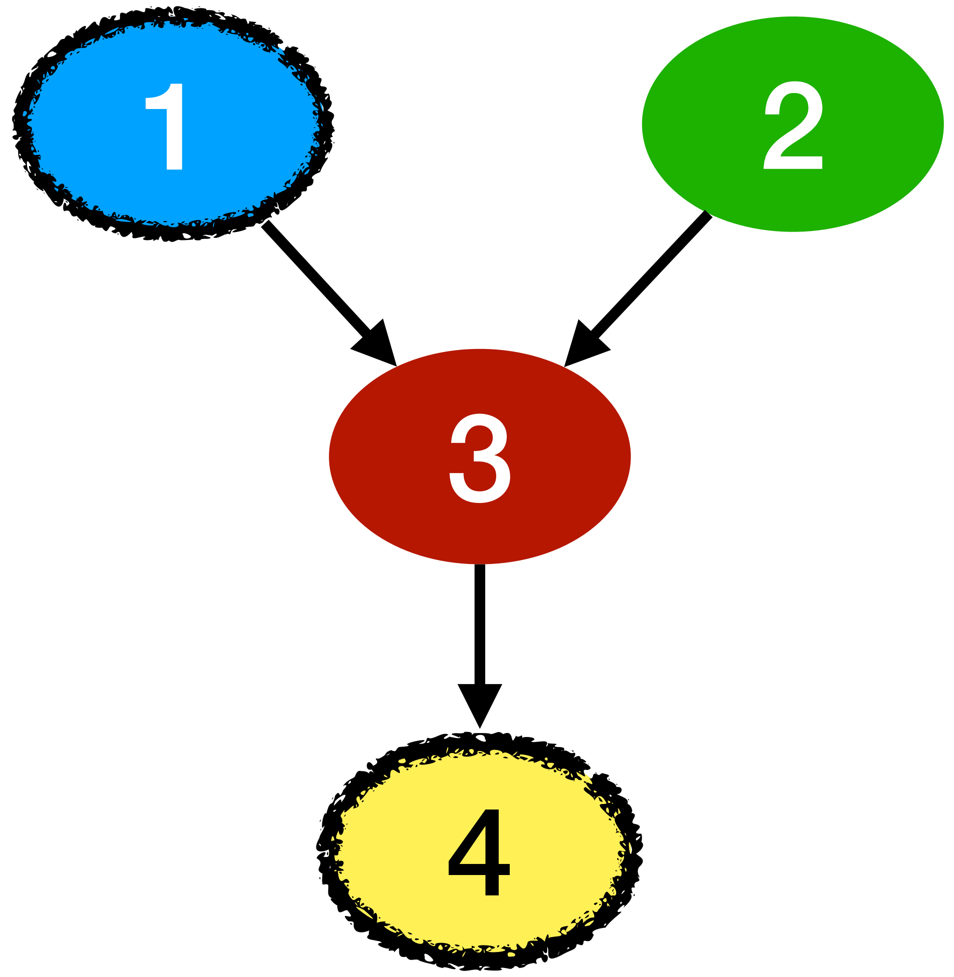
Counterfactuals example in a linear SCM

$$\begin{cases} X_1 = \epsilon_1 \\ X_2 = \epsilon_2 \\ X_3 = 5X_1 + 4X_2 + \epsilon_3 \\ X_4 = 6X_3 + \epsilon_4 \end{cases}$$

Factual SCM

$$\begin{cases} X_1^c = 1 - \epsilon_1 \\ X_2^c = \epsilon_2 \\ X_3^c = 5X_1^c + 4X_2^c + \epsilon_3 \\ X_4^c = 6X_3^c + \epsilon_4 \end{cases}$$

Counterfactual SCM

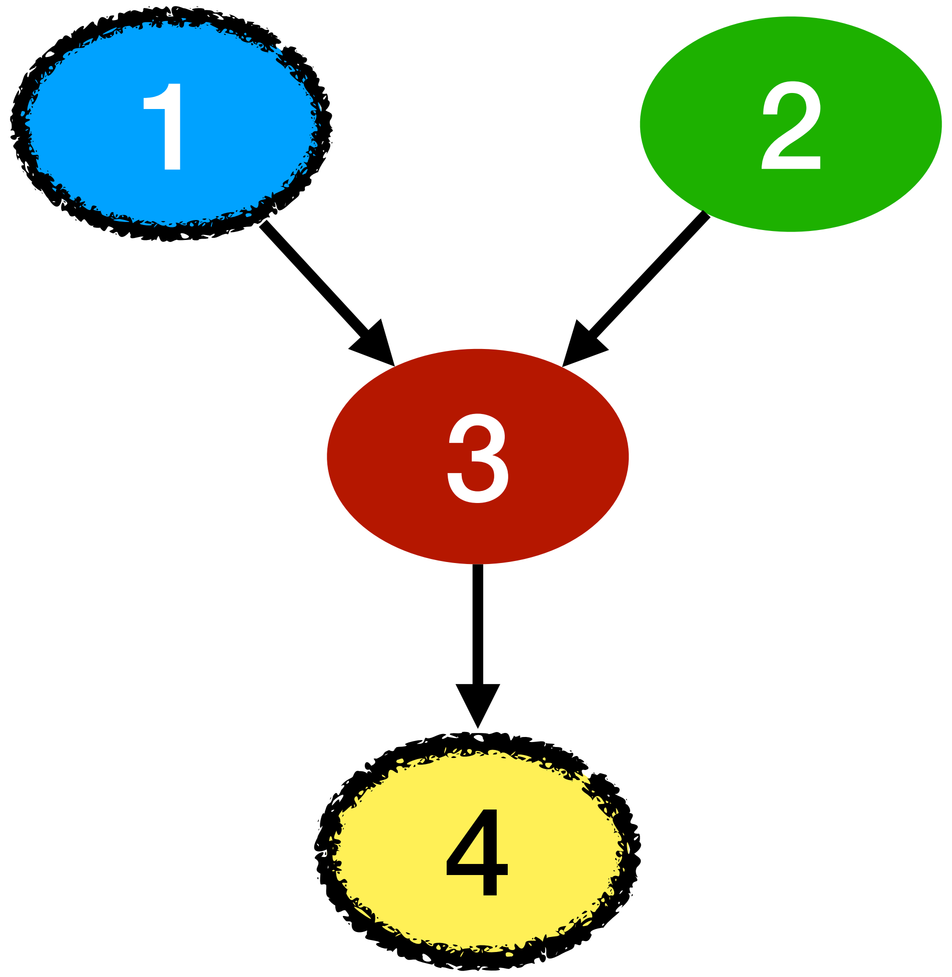


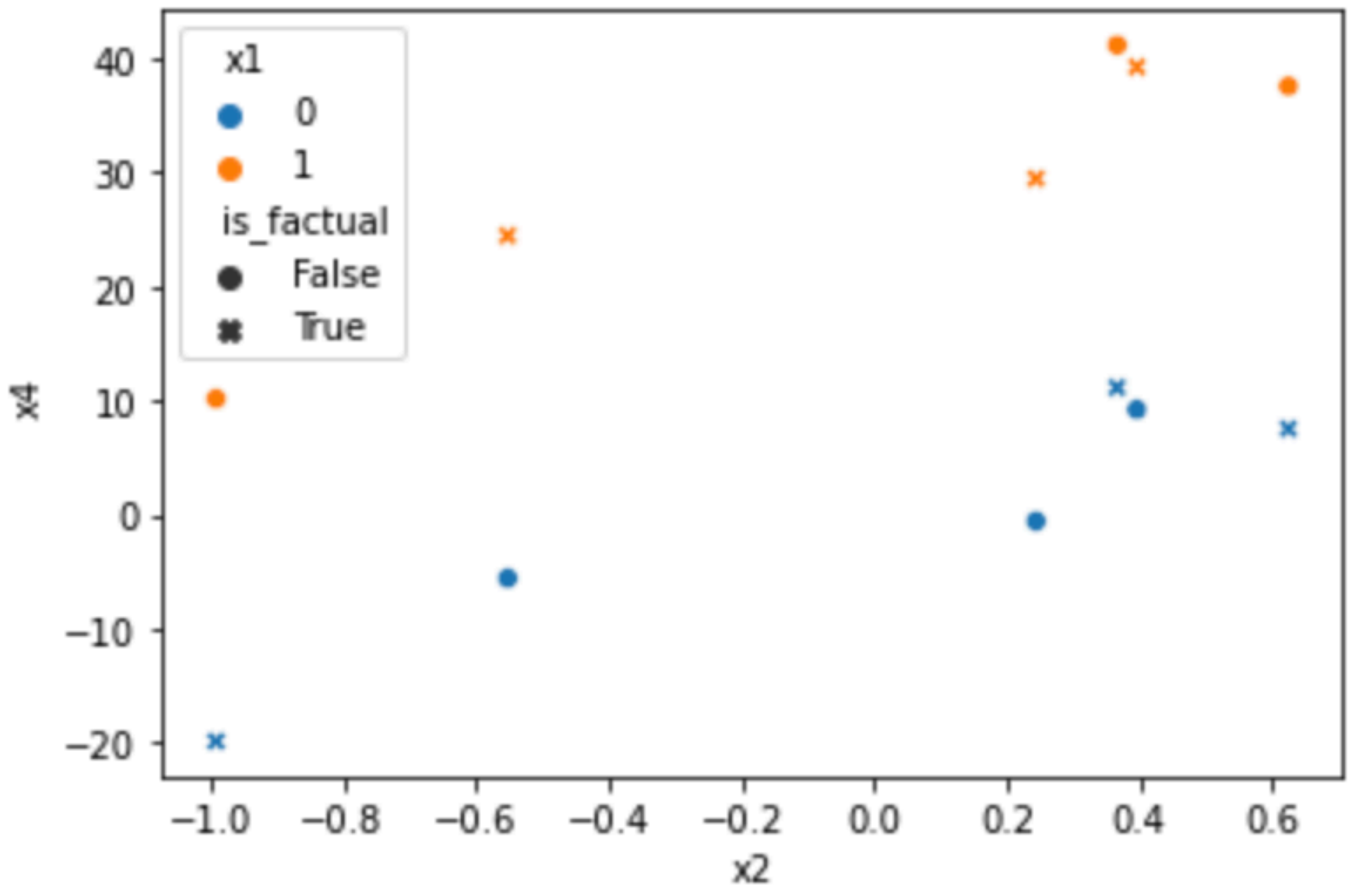
x1	x2	x3	x4	epsilon_x2	epsilon_x3	epsilon_x4	x1c	x2c	x3c	x4c
0	0.624457	1.447194	7.546849	0.624457	-1.050634	-1.136316	1	0.624457	6.447194	37.546849
0	-0.991089	-3.136825	-19.832298	-0.991089	0.827529	-1.011348	1	-0.991089	1.863175	10.167702
0	0.366517	1.953662	11.142927	0.366517	0.487593	-0.579048	1	0.366517	6.953662	41.142927
1	-0.551737	4.191562	24.444811	-0.551737	1.398509	-0.704563	0	-0.551737	-0.808438	-5.555189
1	0.244229	4.923147	29.444147	0.244229	-1.053768	-0.094737	0	0.244229	-0.076853	-0.555853
1	0.396024	6.844468	39.227687	0.396024	0.260372	-1.839122	0	0.396024	1.844468	9.227687

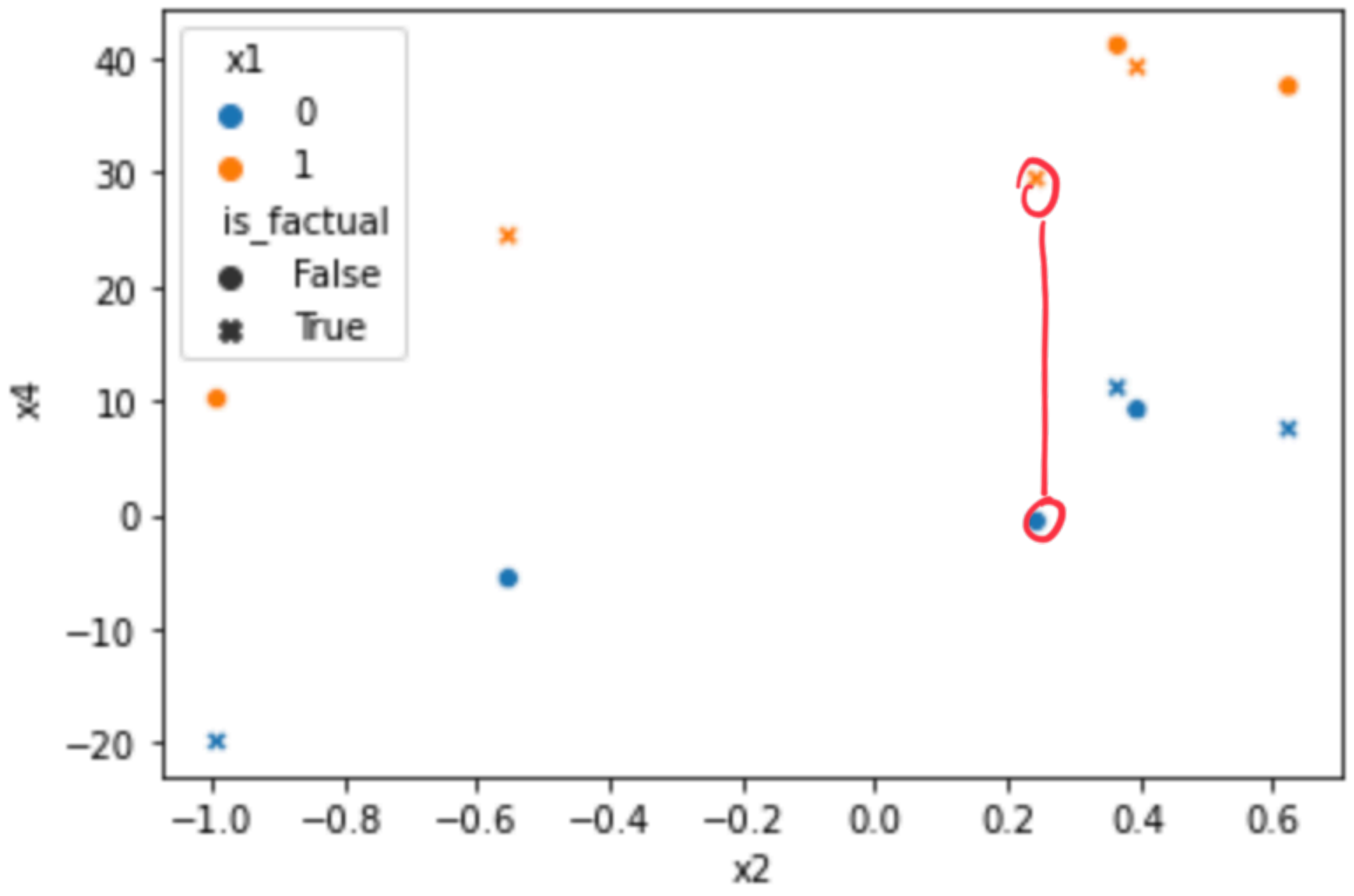
Step 3: Prediction

Counterfactuals example in a linear SCM

x1	x2	x3	x4	epsilon_x2	epsilon_x3	epsilon_x4	x1c	x2c	x3c	x4c
0	0.624457	1.447194	7.546849	0.624457	-1.050634	-1.136316	1	0.624457	6.447194	37.546849
0	-0.991089	-3.136825	-19.832298	-0.991089	0.827529	-1.011348	1	-0.991089	1.863175	10.167702
0	0.366517	1.953662	11.142927	0.366517	0.487593	-0.579048	1	0.366517	6.953662	41.142927
1	-0.551737	4.191562	24.444811	-0.551737	1.398509	-0.704563	0	-0.551737	-0.808438	-5.555189
1	0.244229	4.923147	29.444147	0.244229	-1.053768	-0.094737	0	0.244229	-0.076853	-0.555853
1	0.396024	6.844468	39.227687	0.396024	0.260372	-1.839122	0	0.396024	1.844468	9.227687



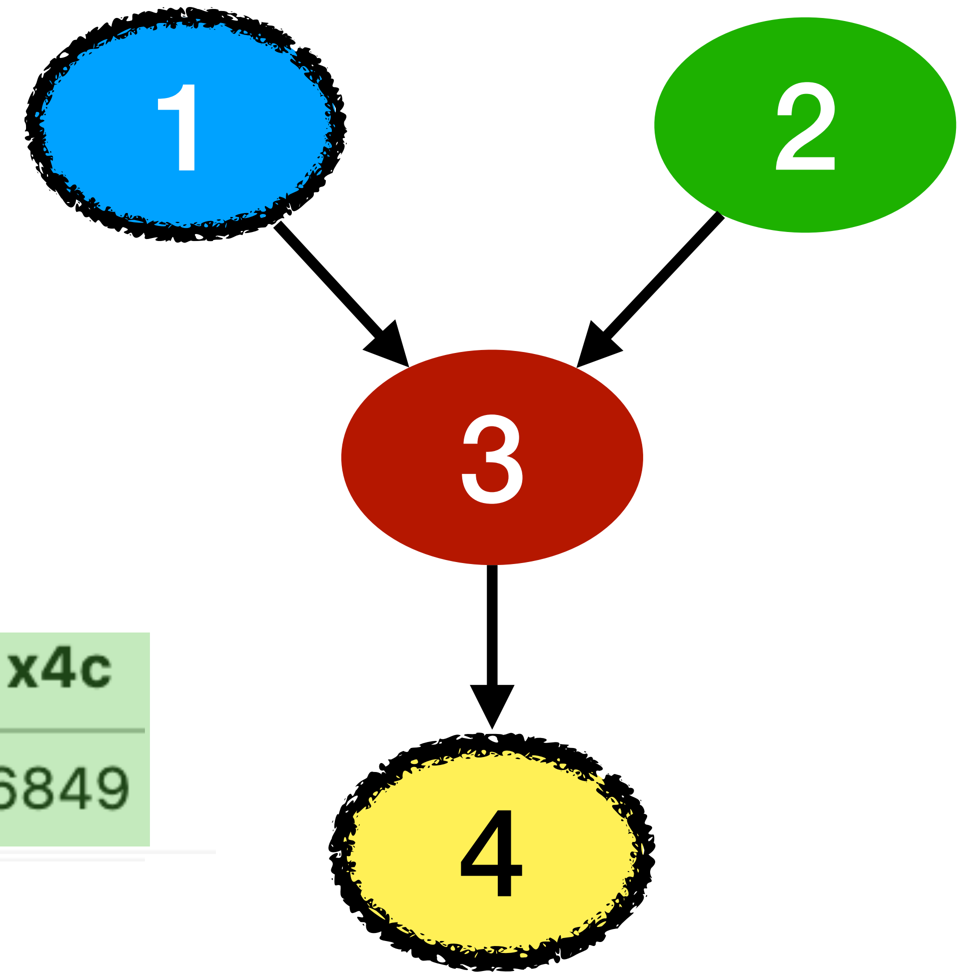




In the example: unit-level counterfactuals

- X_1 was set to 0 for one *unit* and we observed the values x_2, x_3, x_4
- **Counterfactual**: Had X_1 been set to 1, the value of X_4 would have been $\tilde{x}_4$ for the same *unit*.
- **Assumption**: the noise variables stay the same

x1	x2	x3	x4	epsilon_x2	epsilon_x3	epsilon_x4	x1c	x2c	x3c	x4c
0	0.624457	1.447194	7.546849	0.624457	-1.050634	-1.136316	1	0.624457	6.447194	37.546849



Unit-level counterfactuals recipe

- SCM S with observed variables $(X_1, \ldots, X_p)$ and noises $(\epsilon_{X_1}, \ldots, \epsilon_{X_p})$
- We can compute counterfactuals for $\mathrm{do}(X_j)$ and unit i with $(x_1^i, \ldots, x_p^i)$:

1. **Abduction:** reconstruct the noise variable values for i using S: $(\hat{\epsilon}^i_{X_1}, \ldots, \hat{\epsilon}^i_{X_p})$
2. **Action:** If $x_j^i = 0$ in the original data, change the equation for i to $x_j^i \leftarrow 1$, else if $x_j^i = 1$, change it to $x_j^i \leftarrow 0$ (the counterfactual assignment)
3. **Prediction:** Recompute $(\hat{x}_1^i, \ldots, \hat{x}_p^i)$ using S and $(\hat{\epsilon}^i_{X_1}, \ldots, \hat{\epsilon}^i_{X_p})$

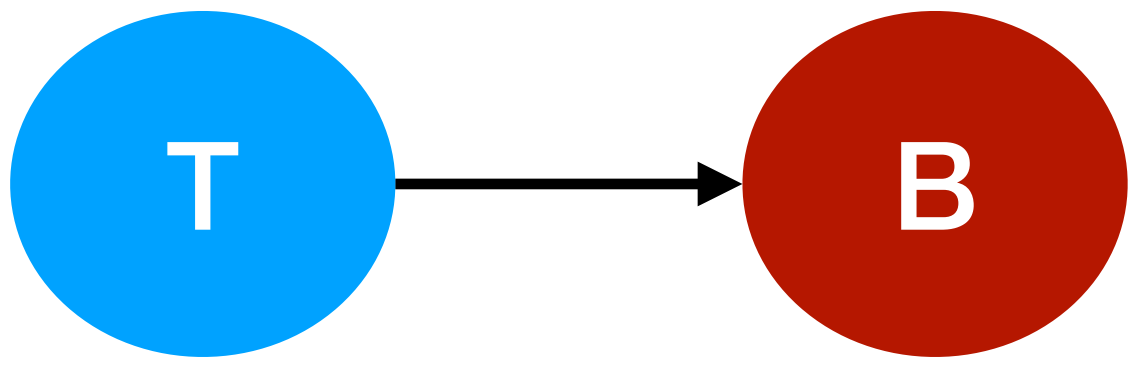
Example 3.4 in Elements of Causal Inference

- T is a very effective treatment for blindness B

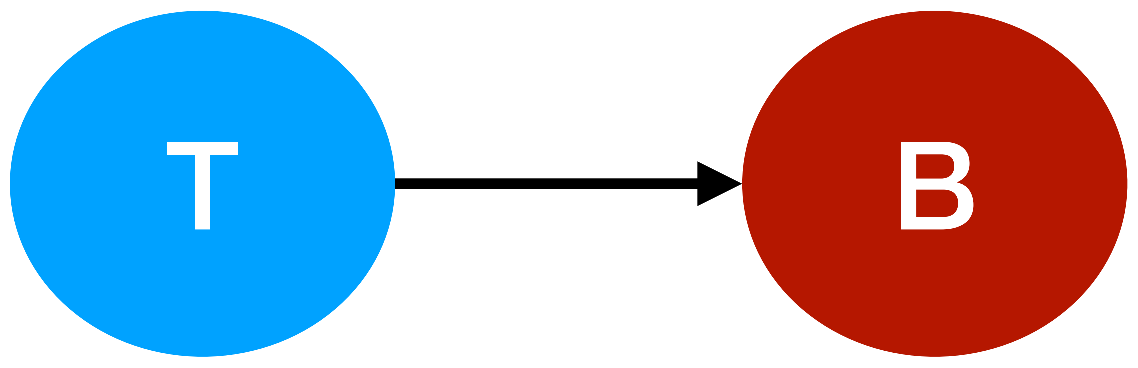
$$P(B = 0 \mid T = 1) = 0.99$$

- Without the treatment patient go blind very often

$$P(B = 1 \mid T = 0) = 0.99$$



Example 3.4 in Elements of Causal Inference



- T is a very effective treatment for blindness B

$$P(B = 0 \mid T = 1) = 0.99$$

- Without the treatment patient go blind very often

$$P(B = 1 \mid T = 0) = 0.99$$

- Very rarely the treatment does not work

$$P(B = 1 \mid T = 1) = 0.01$$

- Very rarely the patients get better on their own

$$P(B = 0 \mid T = 0) = 0.01$$

Let's assume this is because of a very rare and unobserved genetic condition

Example 3.4 in Elements of Causal Inference

- T is a very effective treatment for blindness B

$$P(B=0 \mid T=1) = 0.99$$

- Without the treatment patient go blind very often

$$P(B=1 \mid T=0) = 0.99$$

- Very rarely the treatment does not work

$$P(B=1 \mid T=1) = 0.01$$

- Very rarely the patients get better on their own

$$P(B=0 \mid T=0) = 0.01$$

- Let's assume we treat half of the patients

$$\begin{cases} T = \epsilon_T \\ B = T \cdot \epsilon_B + (1-T) \cdot (1-\epsilon_B) \\ \epsilon_B \sim \text{Ber}(0.01) \\ \epsilon_T \sim \text{Ber}(0.5) \end{cases}$$

$$P(\epsilon_T = 1) = 0.5$$

$$P(\epsilon_B = 1) = 0.01$$

$$P(\epsilon_B = 0) = 0.99$$

Example 3.4 in Elements of Causal Inference

$$\begin{cases} T = \epsilon_T \\ B = T \cdot \epsilon_B + (1 - T) \cdot (1 - \epsilon_B) \\ \epsilon_B \sim \mathrm{Ber}(0.01) \\ \epsilon_T \sim \mathrm{Ber}(0.5) \end{cases}$$

ϵ_B represents the unobserved genetic condition

- Bob goes to the doctor and gets assigned T=1.
- Unfortunately Bob goes blind B=1.
- What would have happened to Bob with T=0?

Example 3.4 in Elements of Causal Inference

$$\begin{cases} T = \epsilon_T \\ B = T \cdot \epsilon_B + (1 - T) \cdot (1 - \epsilon_B) \\ \epsilon_B \sim \mathrm{Ber}(0.01) \\ \epsilon_T \sim \mathrm{Ber}(0.5) \end{cases}$$

- Bob goes to the doctor and gets assigned T=1.
- Unfortunately Bob goes blind B=1.
- What would have happened to Bob with T=0?

Step 1: Abduction

We need to reconstruct the noise variables for Bob by substituting his data

$$\begin{cases} T = 1 = \epsilon_T \\ B = 1 = T \cdot \epsilon_B + (1 - T) \cdot (1 - \epsilon_B) \end{cases}$$

$$1 = 1 \cdot \epsilon_B + (1 - 1) \cdot (1 - \epsilon_B) = \epsilon_B$$

Example 3.4 in Elements of Causal Inference

$$\begin{cases} T = \epsilon_T \\ B = T \cdot \epsilon_B + (1 - T) \cdot (1 - \epsilon_B) \\ \epsilon_B \sim \mathrm{Ber}(0.01) \\ \epsilon_T \sim \mathrm{Ber}(0.5) \end{cases}$$

- Bob goes to the doctor and gets assigned T=1.
- Unfortunately Bob goes blind B=1.
- What would have happened to Bob with T=0?

Step 1: Abduction

We need to reconstruct the noise variables for Bob by substituting his data

$$\epsilon_T^{Bob} = 1$$

$$\epsilon_B^{Bob} = 1$$

Example 3.4 in Elements of Causal Inference

$$
\begin{cases}
T = \epsilon_T \\
B = T \cdot \epsilon_B + (1 - T) \cdot (1 - \epsilon_B) \\
\epsilon_B \sim \mathrm{Ber}(0.01) \\
\epsilon_T \sim \mathrm{Ber}(0.5)
\end{cases}
$$

- Bob goes to the doctor and gets assigned T=1.
- Unfortunately Bob goes blind B=1.
- What would have happened to Bob with T=0?

Step 2: Action

We assign the counterfactual treatment to Bob

$$\epsilon_T^{Bob} = 1$$

$$\epsilon_B^{Bob} = 1$$

$$T^{c,Bob} = 0$$

Example 3.4 in Elements of Causal Inference

$$\begin{cases} T = \epsilon_T \\ B = T \cdot \epsilon_B + (1 - T) \cdot (1 - \epsilon_B) \\ \epsilon_B \sim \mathrm{Ber}(0.01) \\ \epsilon_T \sim \mathrm{Ber}(0.5) \end{cases}$$

- Bob goes to the doctor and gets assigned T=1.
- Unfortunately Bob goes blind B=1.
- What would have happened to Bob with T=0?

Step 3: Prediction

We assign the counterfactual treatment to Bob

$$\epsilon_T^{Bob} = 1$$

$$\epsilon_B^{Bob} = 1$$

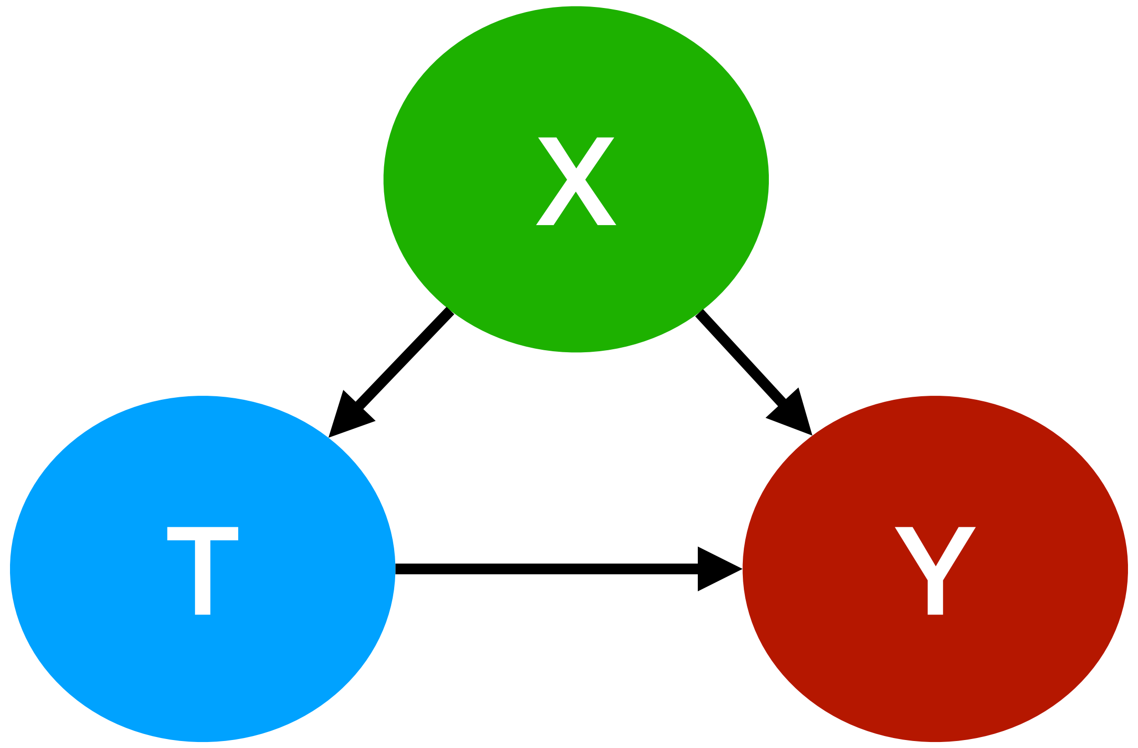
$$T^{c,Bob} = 0$$

$$\begin{cases} T = 0 \\ B = T \cdot 1 + (1 - T) \cdot (1 - 1) \end{cases}$$

$$T^{c,Bob} = 0, \; B^{c,Bob} = 0$$

Counterfactuals quiz in Canvas

$$\begin{cases} X = \epsilon_X \\ T = \begin{cases} \epsilon_T & \text{if } X = 1 \\ 1 - \epsilon_T & \text{if } X = 0 \end{cases} \\ Y = \begin{cases} \epsilon_Y & \text{if } X = T \\ 1 - \epsilon_Y & \text{if } X \neq T \end{cases} \\ \epsilon_X, \epsilon_Y, \epsilon_T \sim \text{Ber}(0.4) \end{cases}$$



$$X^{Carla} = 1,\ T^{Carla} = 0,\ Y^{Carla} = 1$$

Unit-level counterfactuals recipe

- SCM S with observed variables $(X_1, \ldots, X_p)$ and noises $(\epsilon_{X_1}, \ldots, \epsilon_{X_p})$
- We can compute counterfactuals for $\mathrm{do}(X_j)$ and unit i with $(x_1^i, \ldots, x_p^i)$:

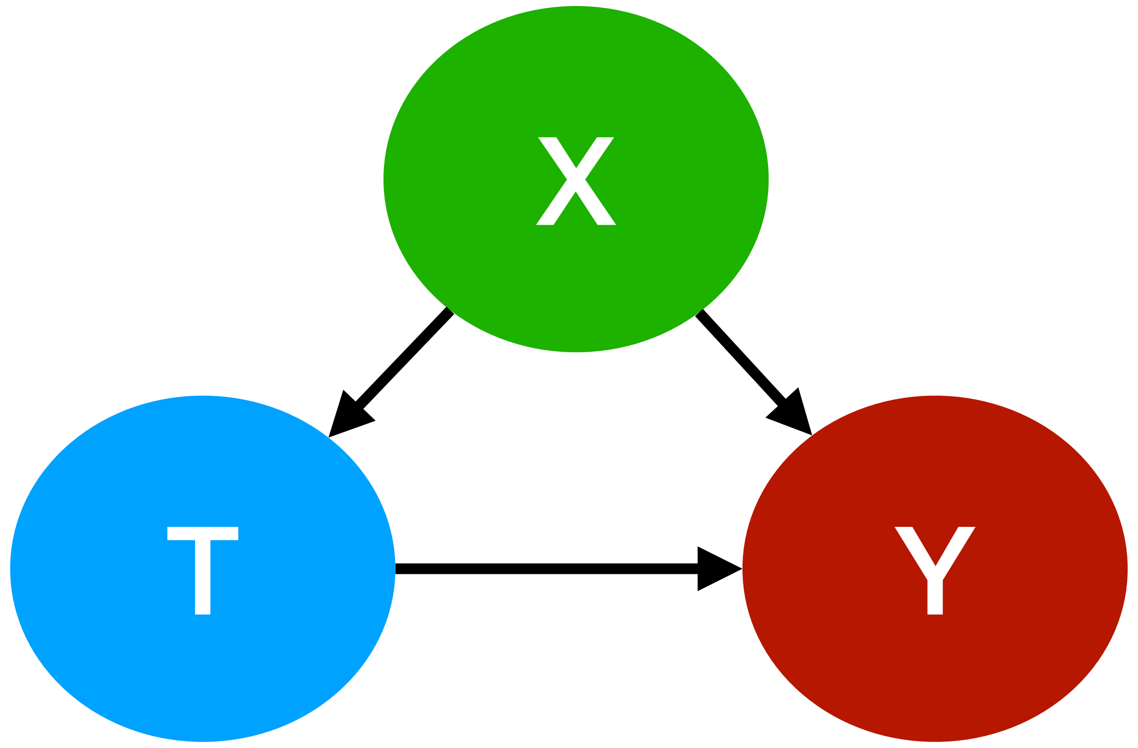
1. **Abduction:** reconstruct the noise variable values for i using S: $(\hat{\epsilon}^i_{X_1}, \ldots, \hat{\epsilon}^i_{X_p})$
2. **Action:** If $x_j^i = 0$ in the original data, change the equation for i to $x_j^i \leftarrow 1$, else if $x_j^i = 1$, change it to $x_j^i \leftarrow 0$ (the counterfactual assignment)
3. **Prediction:** Recompute $(\hat{x}_1^i, \ldots, \hat{x}_p^i)$ using S and $(\hat{\epsilon}^i_{X_1}, \ldots, \hat{\epsilon}^i_{X_p})$

Unit-level counterfactuals - issues

- We cannot use **unit-level counterfactuals** to **falsify a wrong causal model**
- We cannot test the (unit-level) counterfactual statements in the data
- **In general, we cannot observe the noise variables and we cannot always reconstruct them by inverting the SCM**
- We had to assume that the noise is constant, which is not in general true

Potential outcomes (Rubin causal model)

- **Potential outcomes:** an alternative, but equally popular, formulation of causal inference (as opposed to Pearl's graphs):
  - Biostatistics, economics
  - Usually no graphs
  - Usually treatment T, outcome Y and covariates X



- Unification framework, e.g. SWIGs (https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.644.1881&rep=rep1&type=pdf)

Potential outcomes with binary treatment

- **Unit-level outcome** Y_i or Y^i "outcome for patient i"
- The **potential** outcome for unit i if we assigned i to $t = 0$
  - $Y_i(t = 0) = Y_i(t_0) = Y_i(0) = Y^i(0)$
- The **potential** outcome for unit i if we assigned i to $t = 1$
  - $Y_i(t = 1) = Y_i(t_1) = Y_i(1) = Y^i(1)$
- Potential outcomes are regardless of what happened, **counterfactual outcome** is "against the fact"

Counterfactuals continuation

$$\begin{cases} X = \epsilon_X \\ T = \begin{cases} \epsilon_T & \text{if } X = 1 \\ 1 - \epsilon_T & \text{if } X = 0 \end{cases} \\ Y = \begin{cases} \epsilon_Y & \text{if } X = T \\ 1 - \epsilon_Y & \text{if } X \neq T \end{cases} \\ \epsilon_X, \epsilon_Y, \epsilon_T \sim \text{Ber}(0.4) \end{cases}$$

$$Y^{Carla}(t = 0) = 1$$

Factual (potential) outcome

$$Y^{Carla}(t = 1) = ?$$

Counterfactual (potential) outcome

$$X^{Carla} = 1,\ T^{Carla} = 0,\ Y^{Carla} = 1$$

Unit-level causal effects vs average causal effects

- **Unit-level causal effect:** $Y_i(t=1) - Y_i(t=0)$
- **Fundamental problem of causal inference:** we cannot observe a factual and a counterfactual outcome for each unit.

Unit-level causal effects vs average causal effects

- **Unit-level causal effect:** $Y_i(t = 1) - Y_i(t = 0)$

- **Fundamental problem of causal inference:** we cannot observe a factual and a counterfactual outcome for each unit.

  - In general the treated population and the untreated population are composed by individuals that are not exactly the same.

- **But:** we can estimate the effect from data at a population level

$$\text{ATE} = \mathbb{E}[Y(t = 1) - Y(t = 0)]$$

Potential outcomes with binary treatment and binary outcome

OBSERVED

Unit	$Y_i(t=0)$	$Y_i(t=1)$	T	Y_i
1	0	1	1	1
2	1	0	0	1
3	0	0	1	0
4	0	1	0	0
5	1	1	1	1
6	1	0	1	0

Potential outcomes with binary treatment and binary outcome

- Causal inference as a missing data problem:

Unit	$Y_i(t=0)$	$Y_i(t=1)$	T	Y_i
1	?	1	1	1
2	1	?	0	1
3	?	0	1	0
4	0	?	0	0
5	?	1	1	1
6	?	0	1	0

Potential outcomes typical assumptions

- **Stable Unit Treatment Value Assumption (SUTVA)**
  - Treatment assignment of a unit i should not affect others
  - **Opposite**: interference (e.g. contagion, network effects, etc)
- **Consistency:** $\forall t'$, if $T = t'$, $Y = Y(t = t')$ **vs Partial Compliance**
  - The actual outcome is the potential outcome of the assigned treatment
- **(Conditional) ignorability** (no unmeasured confounding):
  - $T \perp\!\!\!\perp Y(t=0) \,|\, \mathbf{X}$ and $T \perp\!\!\!\perp Y(t=1) \,|\, \mathbf{X}$ (or $T \perp\!\!\!\perp_{G_{\setminus\{T \to Y\}}} Y \,|\, \mathbf{X}$)
- **Positivity** $P(T = t \,|\, X = x) > 0 \ \forall x, t$ - no deterministic assignment