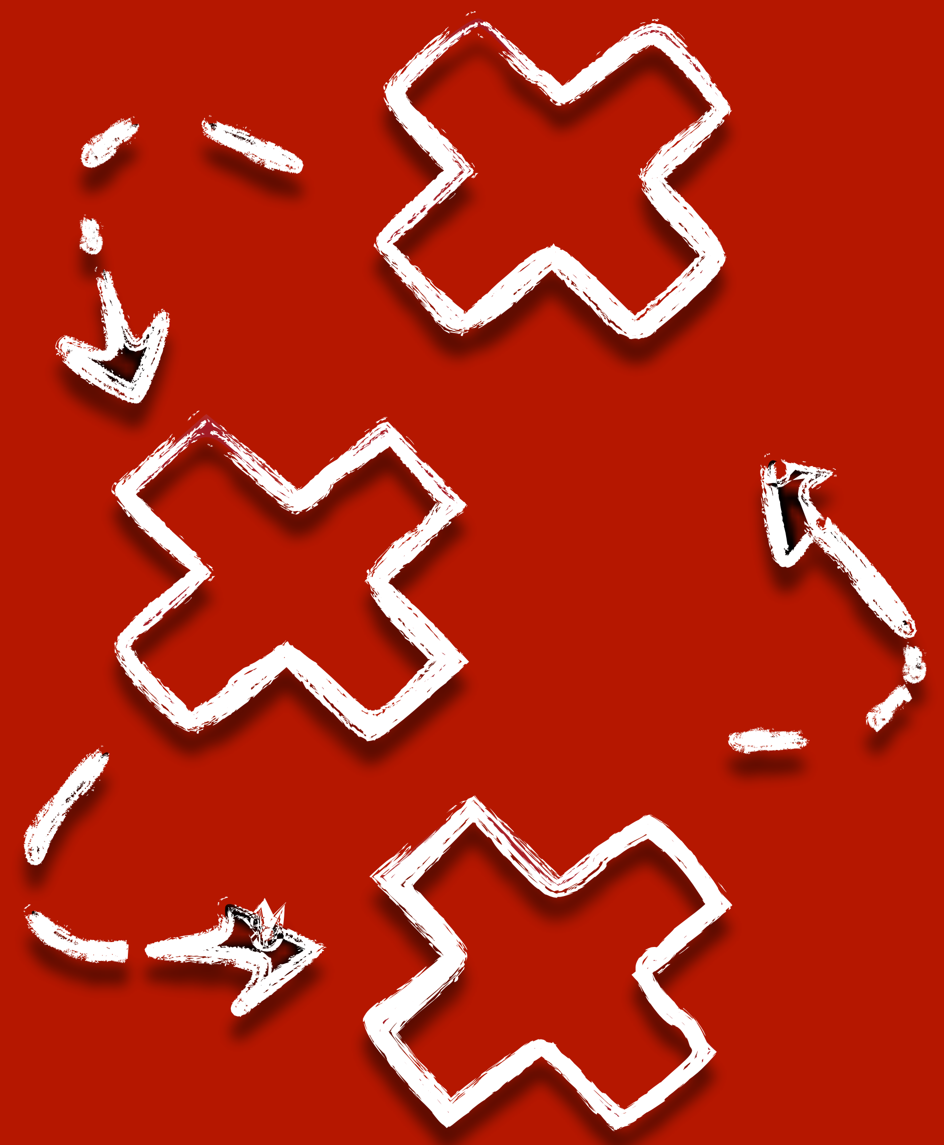


Causal Data Science

Lecture 5:1 Causal models and covariate adjustment

Lecturer: Sara Magliacane

UvA - Spring 2023

Last class: Causal Bayesian networks

- Given DAG $G = (\mathbf{V}, \mathbf{E})$ and distribution p, (G, p) is a Bayesian network if

$$p(X_1, \ldots, X_p) = \prod_{i \in V} p(X_i \mid \mathbf{X}_{\mathrm{pa}(i)})$$

- If for any $\mathbf{W} \subset \mathbf{V}$:

$$p(X_{\mathbf{V}} \mid \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \mid X_{pa(i)}) \cdot \mathbb{1}(X_{\bar{w}} = x_{\bar{w}})$$

(G, p) is a **causal Bayesian network**

$\boxed{\textbf{Parents are now direct causes}}$

Last class: a formal definition of causality

- X has a **causal effect** on Y iff

$$\exists x, x' : P(Y \mid \mathrm{do}(X = x)) \neq P(Y \mid \mathrm{do}(X = x'))$$

- In our case, we assume X causes Y iff:

$$\exists x : P(Y \mid \mathrm{do}(X = x)) \neq P(Y)$$

- In other words, $\forall x : P(Y \mid \mathrm{do}(X = x)) = P(Y) \iff X \text{ does not cause } Y$

- The effect of X on Y is **confounded** if $P(Y \mid \mathrm{do}(X = x)) \neq P(Y \mid X = x)$
    - Simpson's paradox Exercise<-Age->Cholesterol

Last class: Structural causal models (SCMs)

- Let (G, p) be a **causal** Bayesian network
- We can write each **endogenous** variable X_i for $i \in \mathbf{V}$ as a **function of its (endogenous) parents** in G and a **exogenous noise term** ϵ_i in a **structural equation:**

$$X_i \leftarrow h_i(X_{\mathrm{Pa}(i)}, \epsilon_i)$$

- We assume all exogenous noises are **independent of each other** $\forall i \neq j : \epsilon_i \perp\!\!\!\perp \epsilon_j$

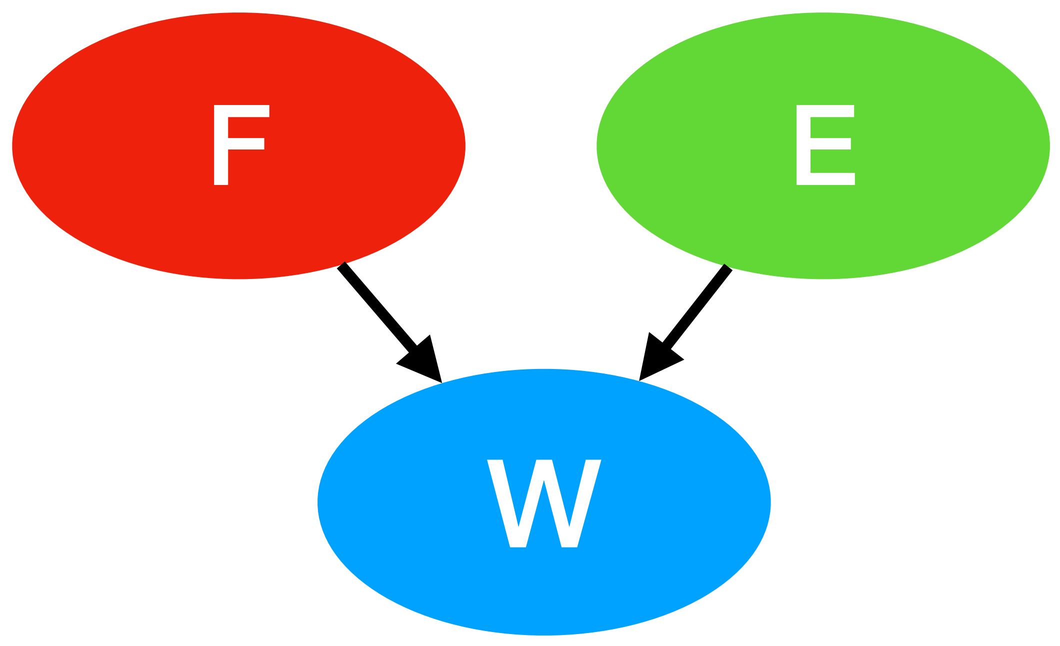
Last class: Interventions in SCMs

- An intervention $\mathrm{do}(X_j = x_j)$ can be modelled by replacing

$$X_i \leftarrow h_i(X_{\mathrm{Pa}(i)}, \epsilon_i) \text{ with } X_i \leftarrow x_j$$

$$\begin{cases} F \leftarrow 2000 + \epsilon_F \\ E \leftarrow 500 + \epsilon_E \\ W \leftarrow \dfrac{F - E - 1500}{7700} + 0.1\epsilon_W \end{cases}$$

$$\epsilon_E, \epsilon_F, \epsilon_W \sim \mathcal{N}(0,100)$$



Last class: Interventions in SCMs

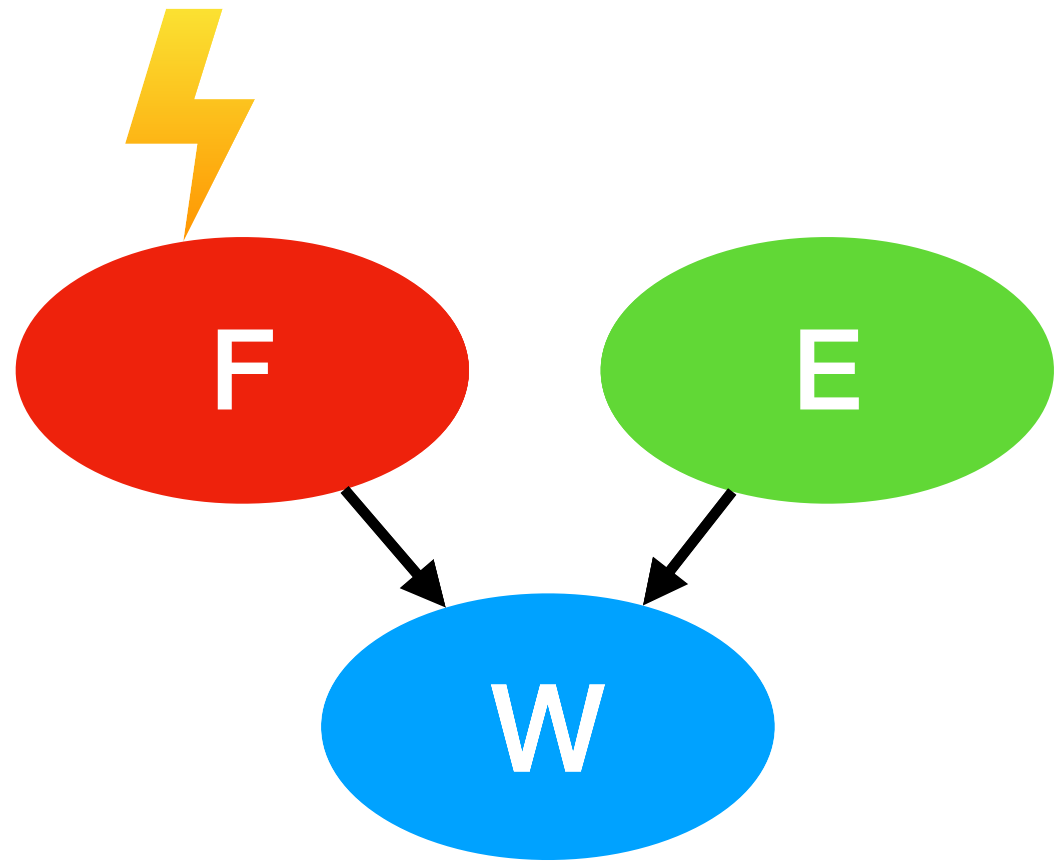
- An intervention $\mathrm{do}(X_j = x_j)$ can be modelled by replacing

$$X_i \leftarrow h_i(X_{\mathrm{Pa}(i)}, \epsilon_i) \text{ with } X_i \leftarrow x_j$$

$\mathrm{do}(F=1200)$

$$\begin{cases} F \leftarrow \cancel{2000} + \epsilon_F \\ E \leftarrow 500 + \epsilon_E \\ W \leftarrow \dfrac{F - E - 1500}{7700} + 0.1\epsilon_W \end{cases}$$

$F \leftarrow 1200$



$$\epsilon_E, \epsilon_F, \epsilon_W \sim \mathcal{N}(0,100)$$

Example 3.2 in Elements of Causal Inference

$$\begin{cases} X \leftarrow \epsilon_x \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

$$P(X) = \mathcal{N}(0,1)$$

$$P(Y) = \mathcal{N}(\mu_Y, \sigma_Y^2)$$

Stats recap:

Linear combinations of Gaussians are also Gaussian

μ_Y, σ_Y^2 ?

Example 3.2 in Elements of Causal Inference

$$\begin{cases} X \leftarrow \epsilon_x \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

$$\mathbb{E}[aA + bB] = a\mathbb{E}[A] + b\mathbb{E}[B]$$

$$P(X) = \mathcal{N}(0,1)$$

$$\mathrm{Var}[aA + bB] = a^2\mathrm{Var}[A] + b^2\mathrm{Var}[B] \quad \text{if} \quad A \perp\!\!\!\perp B$$

$$P(Y) = \mathcal{N}(\mu_Y, \sigma_Y^2)$$

$$\mu_Y = \mathbb{E}[Y] = \mathbb{E}[4 \cdot X + \epsilon_Y] = 4 \cdot \mathbb{E}[X] + \mathbb{E}[\epsilon_Y] = 4 \cdot 0 + 0 = 0$$

Example 3.2 in Elements of Causal Inference

$$\begin{cases} X \leftarrow \epsilon_x \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

$$\mathbb{E}[aA + bB] = a\mathbb{E}[A] + b\mathbb{E}[B]$$

$$P(X) = \mathcal{N}(0,1)$$

$$\mathrm{Var}[aA + bB] = a^2\mathrm{Var}[A] + b^2\mathrm{Var}[B] \quad \text{if} \quad A \perp\!\!\!\perp B$$

$$P(Y) = \mathcal{N}(\mu_Y, \sigma_Y^2)$$

$X \perp\!\!\!\perp \varepsilon_Y$

$\mu_Y = 0$ $\quad$ $\sigma^2_Y = \mathrm{Var}[Y] \overset{\downarrow}{=} 4^2 \cdot \mathrm{Var}[X] + \mathrm{Var}[\varepsilon_Y] = 17$

$$P(Y) = \mathcal{N}(0,17)$$

Example 3.2 in Elements of Causal Inference

$$\begin{cases} X \leftarrow \epsilon_x \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

$$P(X) = \mathcal{N}(0,1)$$

$$P(Y) = \mathcal{N}(0,17)$$

$do(X=2)$:

$$\begin{cases} X \leftarrow 2 \\ Y \leftarrow 4 \cdot \cancel{X} + \epsilon_Y \end{cases}$$

$4 \cdot 2 + \varepsilon_Y$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

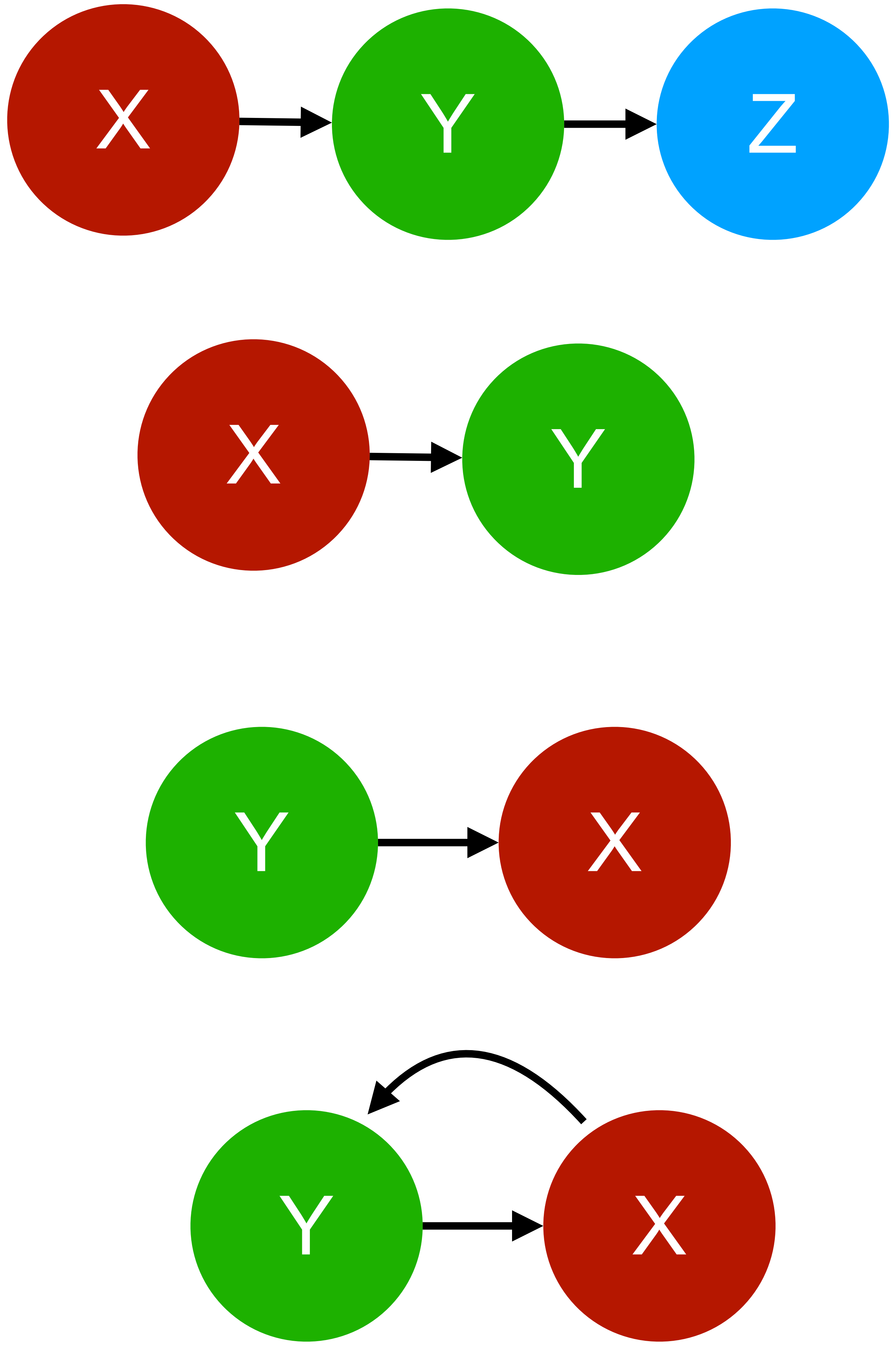
$$P(X \mid do(X=2)) = \begin{cases} 1 & X=2 \\ 0 & X \neq 2 \end{cases}$$

$$P(Y \mid do(X=2)) = N(8,1)$$

Exercise in Canvas: Structural causal models

$$\begin{cases} X = \varepsilon_X \\ Y = 5 \cdot X + 2 \cdot \varepsilon_Y \end{cases}$$

$$\varepsilon_X, \varepsilon_Y \sim N(0,1)$$



Linear SCMs

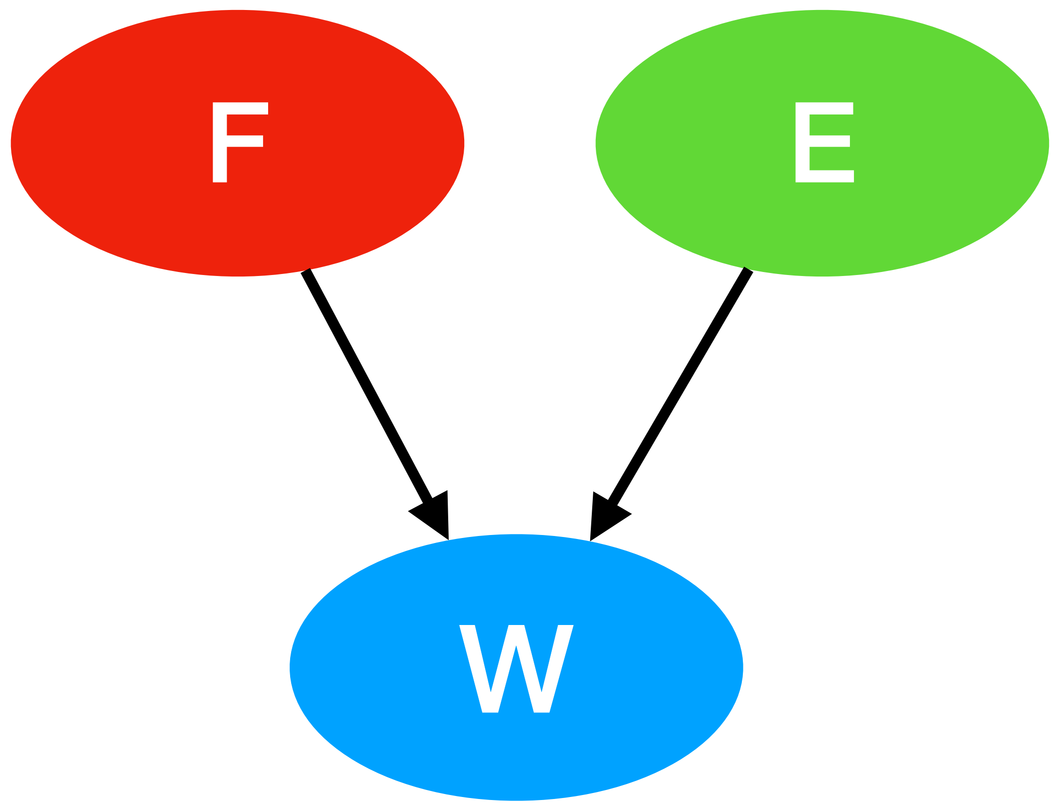
- SCMs where all functions h_i are linear (the noise is additive):

$$X_i \leftarrow h_i(X_{\mathrm{Pa}(i)}, \epsilon_i)$$

- The **direct average causal effect** are the coefficient

$$\begin{cases} F \leftarrow 2000 + \epsilon_F \\ E \leftarrow 500 + \epsilon_E \\ W \leftarrow \frac{F}{7700} - \frac{E}{7700} - \frac{1500}{7700} + 0.1\epsilon_W \end{cases}$$

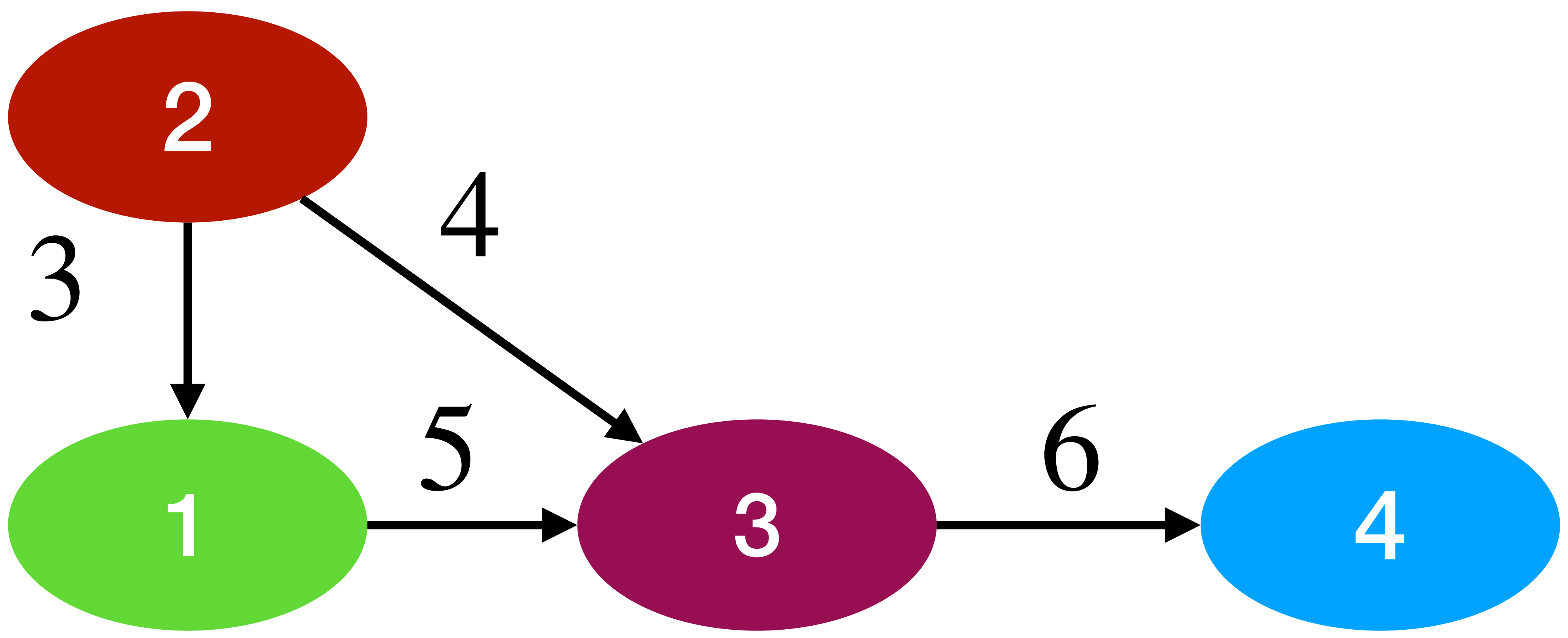
$$\epsilon_E, \epsilon_F, \epsilon_W \sim \mathcal{N}(0,100)$$



An example of a linear SCM

$$\begin{cases} X_1 \leftarrow 3 \cdot X_2 + \epsilon_{X_1} \\ X_2 \leftarrow \epsilon_{X_2} \\ X_3 \leftarrow 5 \cdot X_1 + 4 \cdot X_2 + \epsilon_{X_3} \\ X_4 \leftarrow 6 \cdot X_3 + \epsilon_{X_4} \end{cases}$$

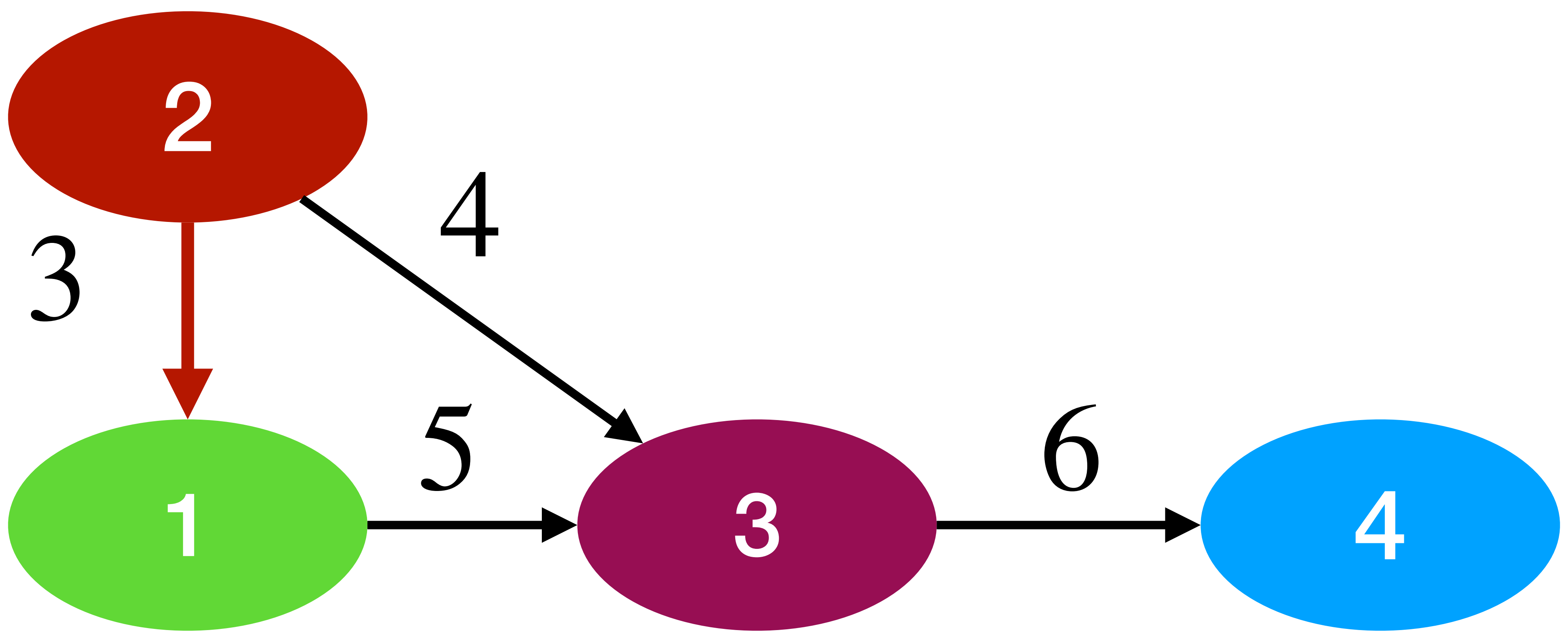
$$\epsilon_{X_1}, \epsilon_{X_2}, \epsilon_{X_3}, \epsilon_{X_4} \sim \mathcal{N}(0,1)$$



An example of a linear SCM

$$\begin{cases} X_1 \leftarrow 3 \cdot X_2 + \epsilon_{X_1} \\ X_2 \leftarrow \epsilon_{X_2} \\ X_3 \leftarrow 5 \cdot X_1 + 4 \cdot X_2 + \epsilon_{X_3} \\ X_4 \leftarrow 6 \cdot X_3 + \epsilon_{X_4} \end{cases}$$

$$\epsilon_{X_1}, \epsilon_{X_2}, \epsilon_{X_3}, \epsilon_{X_4} \sim \mathcal{N}(0,1)$$



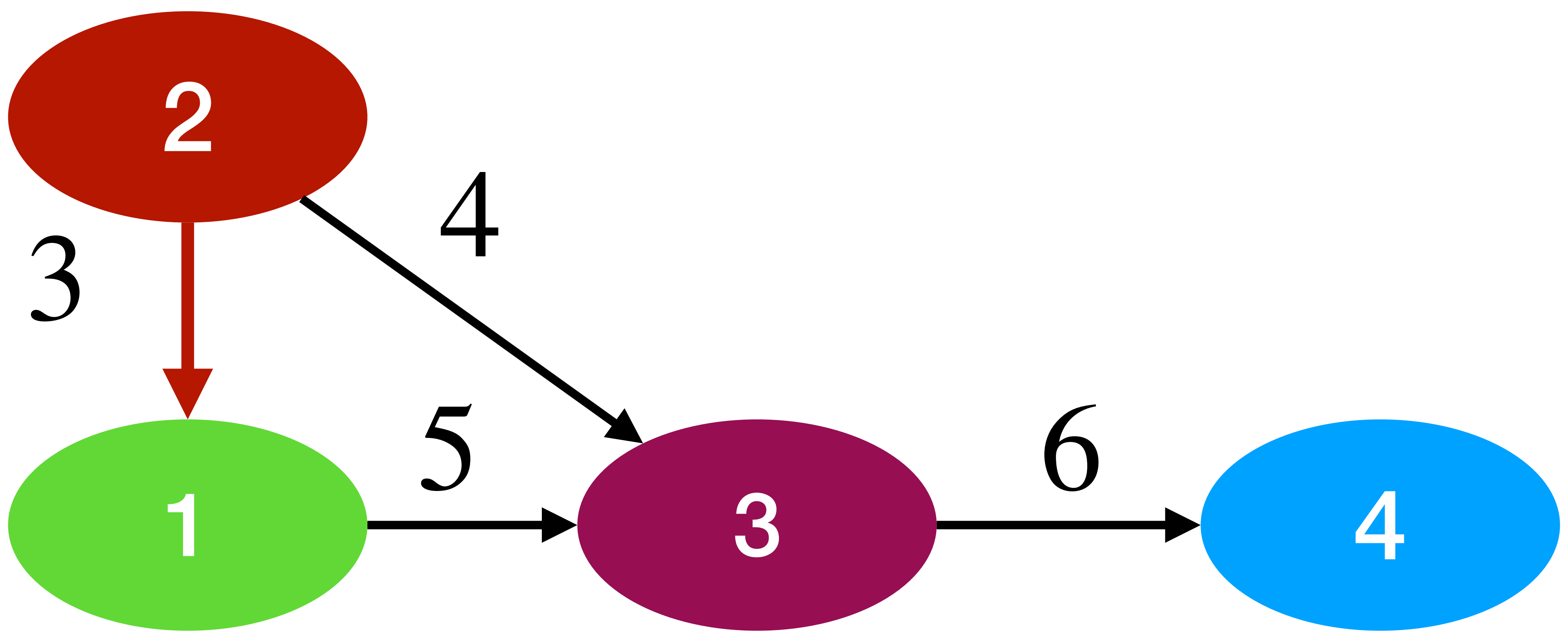
$$E[X_1 \mid do(X_2 = 1)] - E[X_1 \mid do(X_2 = 0)]?$$

An example of a linear SCM

Direct causal effect (parent to child)

$$\begin{cases} X_1 \leftarrow 3 \cdot X_2 + \epsilon_{X_1} \\ X_2 \leftarrow \epsilon_{X_2} \\ X_3 \leftarrow 5 \cdot X_1 + 4 \cdot X_2 + \epsilon_{X_3} \\ X_4 \leftarrow 6 \cdot X_3 + \epsilon_{X_4} \end{cases}$$

$$\epsilon_{X_1}, \epsilon_{X_2}, \epsilon_{X_3}, \epsilon_{X_4} \sim \mathcal{N}(0,1)$$



$$P(X_1 \mid do(X_2 = 1)) = N(3,1)$$

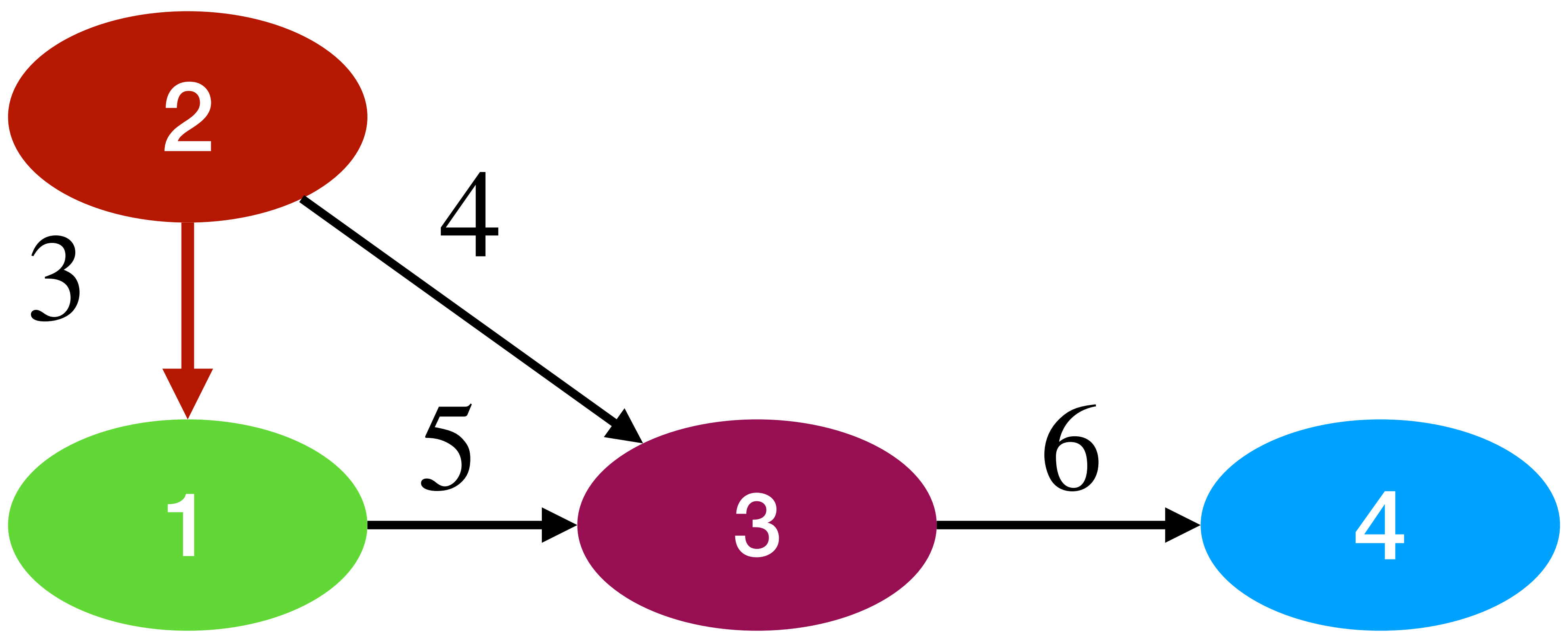
$$P(X_1 \mid do(X_2 = 0)) = N(0,1)$$

$$E[X_1 \mid do(X_2 = 1)] - E[X_1 \mid do(X_2 = 0)] = 3 - 0 = 3$$

An example of a linear SCM

$$\begin{cases} X_1 \leftarrow 3 \cdot X_2 + \epsilon_{X_1} \\ X_2 \leftarrow \epsilon_{X_2} \\ X_3 \leftarrow 5 \cdot X_1 + 4 \cdot X_2 + \epsilon_{X_3} \\ X_4 \leftarrow 6 \cdot X_3 + \epsilon_{X_4} \end{cases}$$

$$\epsilon_{X_1}, \epsilon_{X_2}, \epsilon_{X_3}, \epsilon_{X_4} \sim \mathcal{N}(0,1)$$



$$P(X_1 \mid do(X_2 = 1)) = N(3,1)$$

$$P(X_1 \mid do(X_2 = 2)) = N(?,1)$$

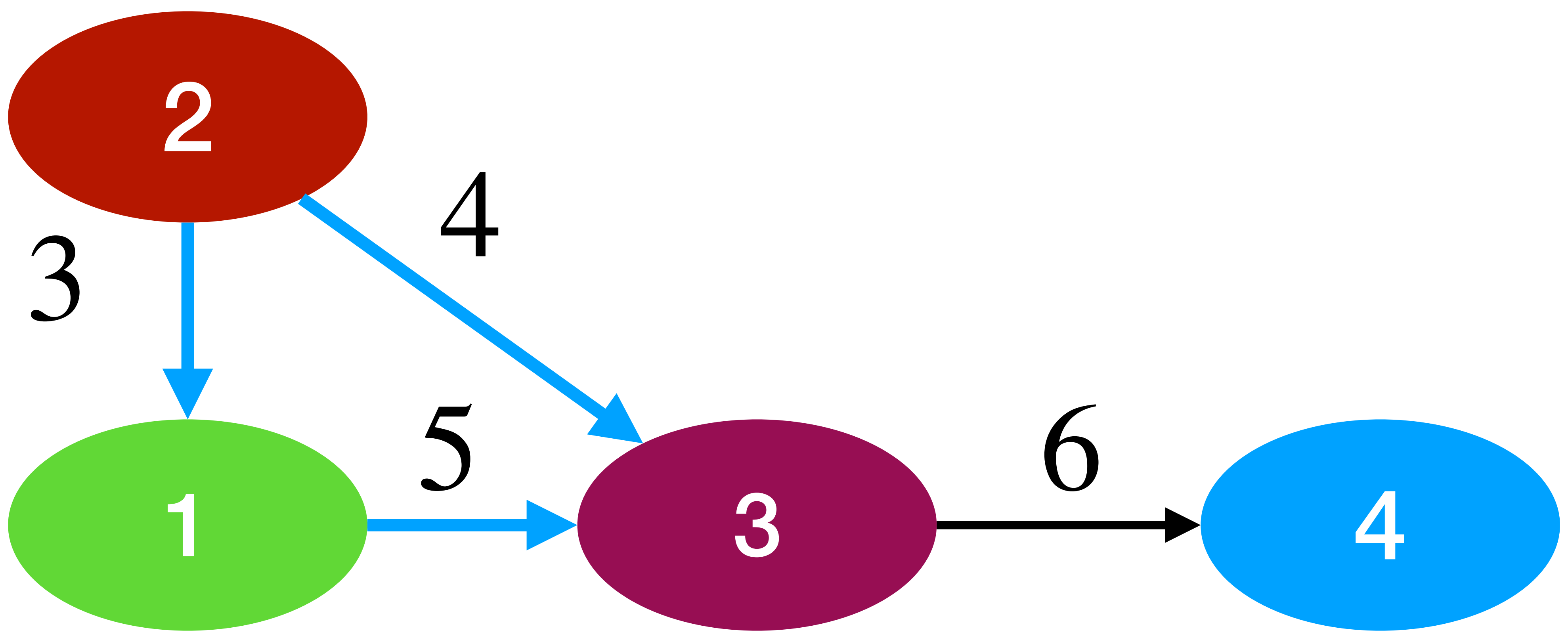
$$E[X_1 \mid do(X_2 = 2)] - E[X_1 \mid do(X_2 = 1)]?$$

An example of a linear SCM

Total causal effect (both direct and indirect)

$$\begin{cases} X_1 \leftarrow 3 \cdot X_2 + \epsilon_{X_1} \\ X_2 \leftarrow \epsilon_{X_2} \\ X_3 \leftarrow 5 \cdot X_1 + 4 \cdot X_2 + \epsilon_{X_3} \\ X_4 \leftarrow 6 \cdot X_3 + \epsilon_{X_4} \end{cases}$$

$$\epsilon_{X_1}, \epsilon_{X_2}, \epsilon_{X_3}, \epsilon_{X_4} \sim \mathcal{N}(0,1)$$



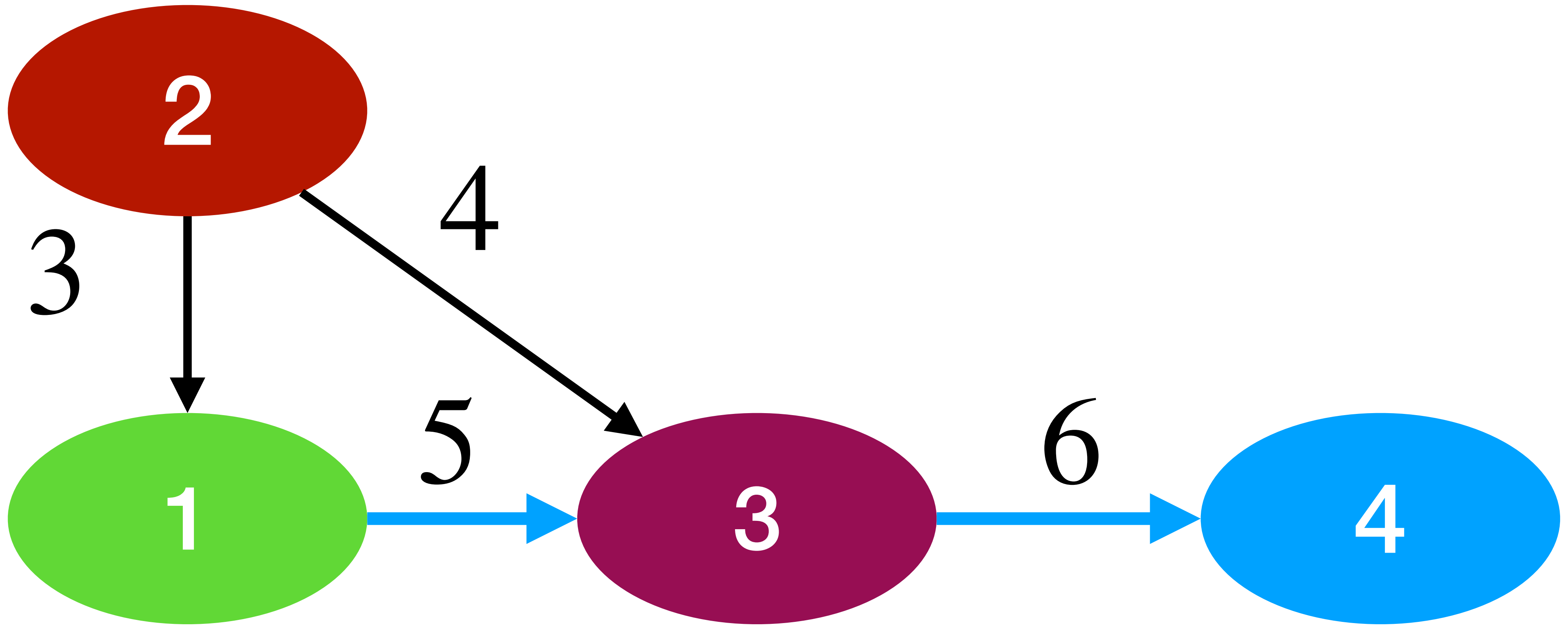
$$E[X_3 \,|\, do(X_2 = 0)] = 0$$

$$E[X_3 \,|\, do(X_2 = 1)] = 5 \cdot 3 \cdot 1 + 4 \cdot 1 = 19$$

$$E[X_3 \,|\, do(X_2 = 1)] - E[X_3 \,|\, do(X_2 = 0)]?$$

Path method for estimating causal effects in linear SCMs

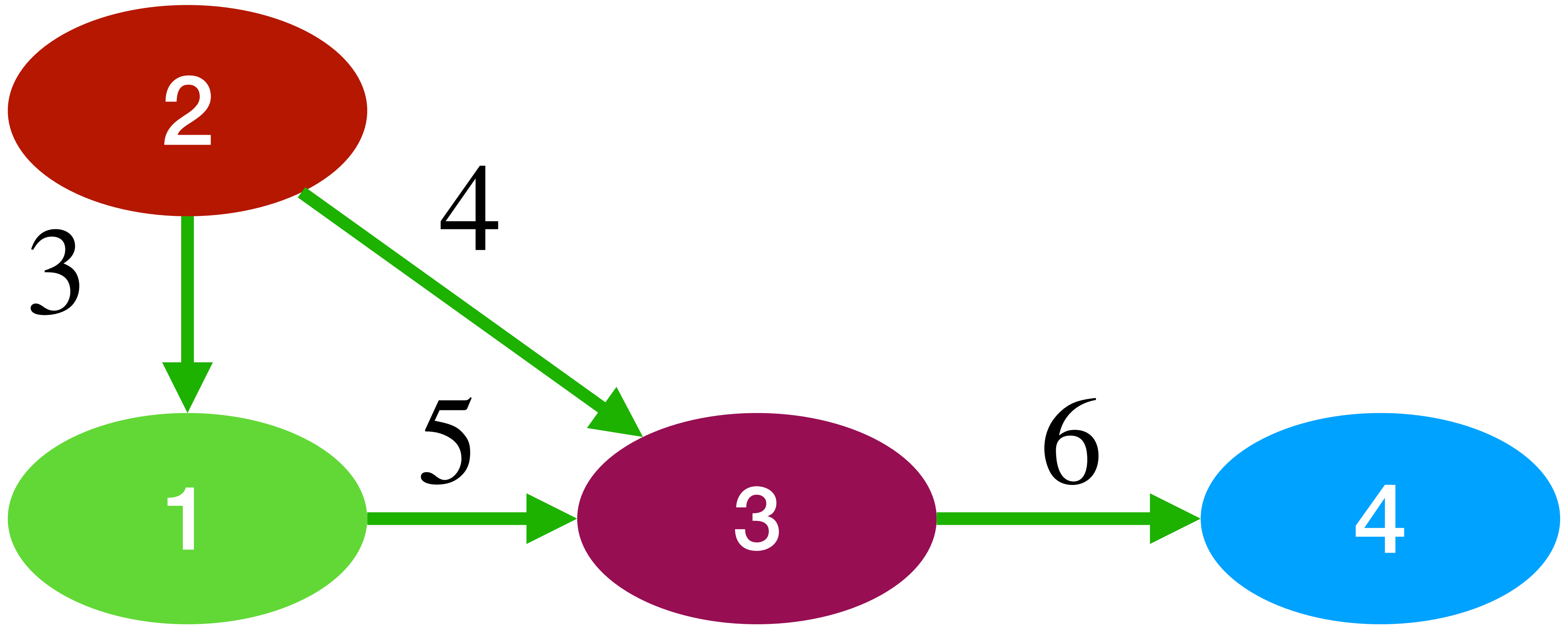
- In a linear SCM we estimate the **total average causal effect** of X_i on X_j :
  - For each **directed path from X_i to X_j**, multiply the edge weights
  - Sum the weights from all paths



$$E[X_4 \mid do(X_1 = 1)] - E[X_4 \mid do(X_1 = 0)] = 5 \cdot 6 = 30$$

Path method for estimating causal effects in linear SCMs

- In a linear SCM we estimate the **total average causal effect** of X_i on X_j :
  - For each **directed path from X_i to X_j**, multiply the edge weights
  - Sum the weights from all paths

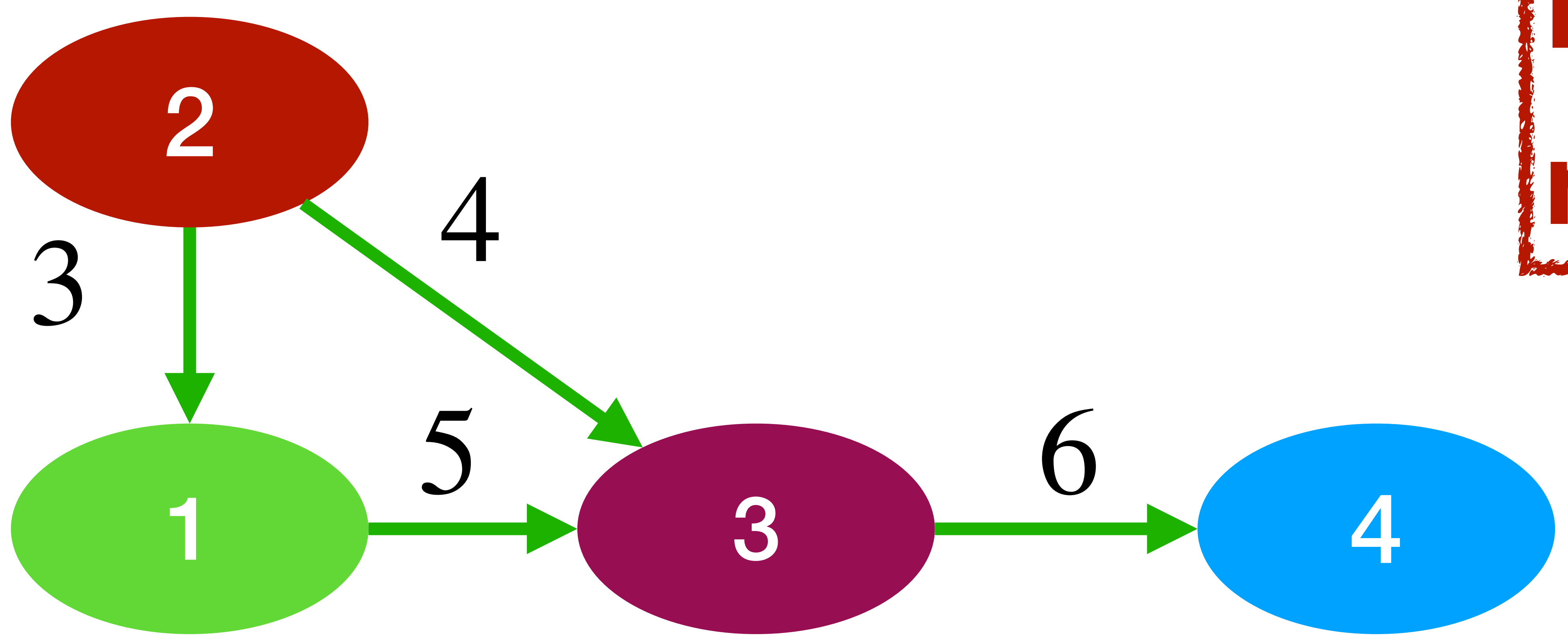


$$E[X_4 \mid do(X_2 = 1)] - E[X_4 \mid do(X_2 = 0)] = 3 \cdot 5 \cdot 6 + 4 \cdot 6 = 114$$

Path method for estimating causal effects in linear SCMs

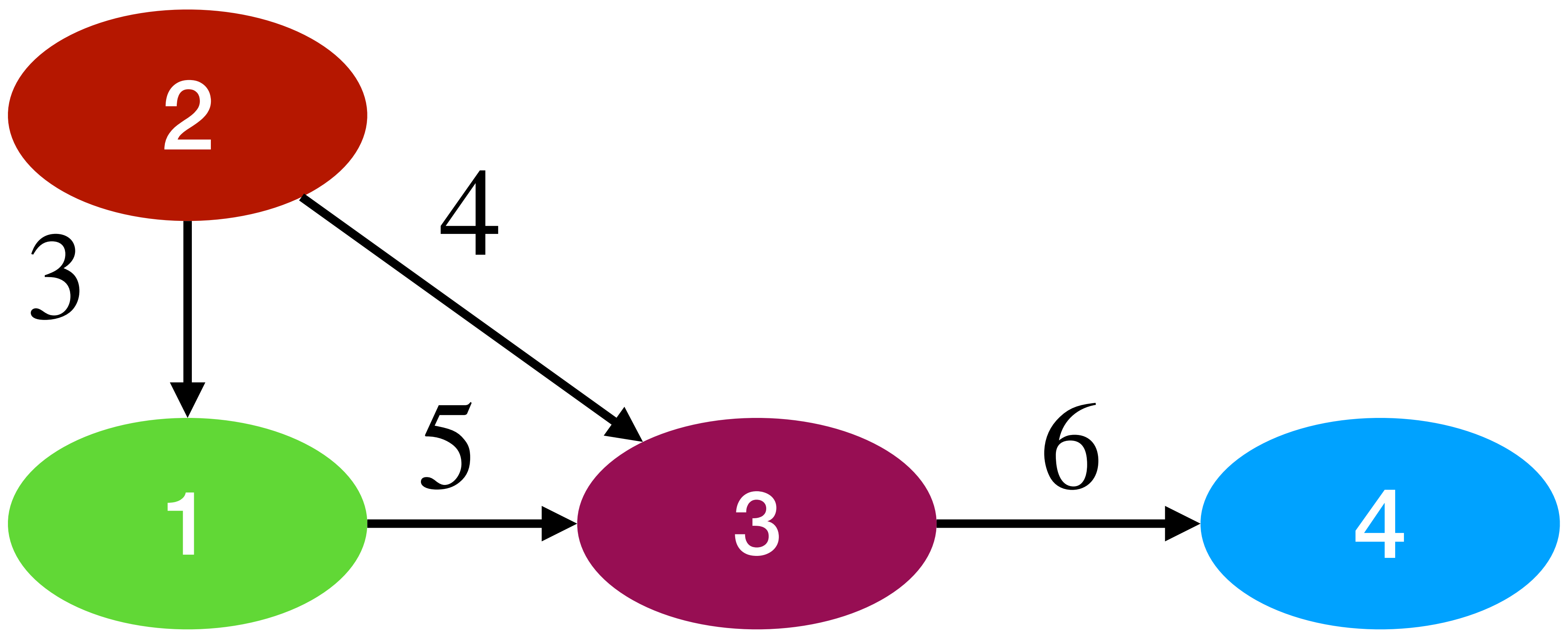
- In a linear SCM we estimate the **total average causal effect** of X_i on X_j :
  - For each **directed path from X_i to X_j**, multiply the edge weights
  - Sum the weights from all paths

For non-linear cases this does not work!



$$E[X_4 \mid do(X_2 = 1)] - E[X_4 \mid do(X_2 = 0)] = 3 \cdot 5 \cdot 6 + 4 \cdot 6 = 114$$

Example in Jupyter notebook Linear SCM Example



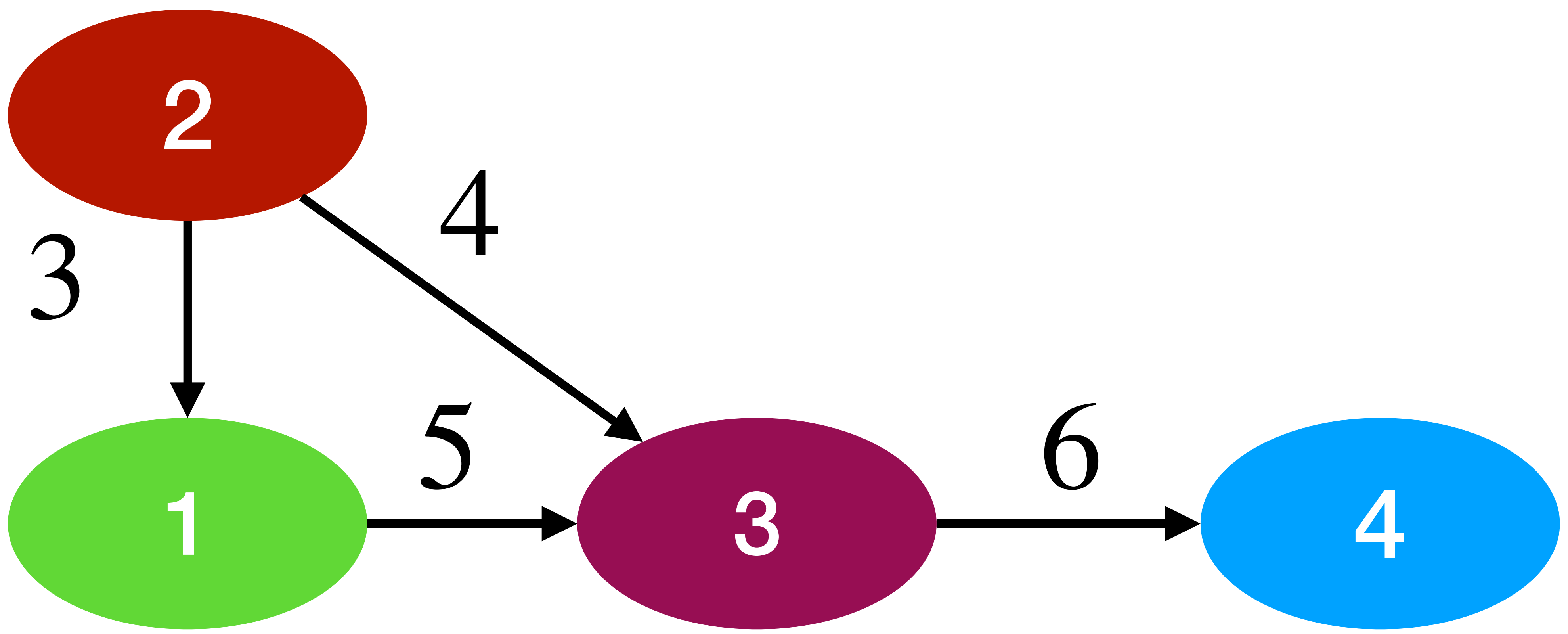

```
x2_1 = randn(n_samples)
x1_1 =  1
x3_1 = 5 * x1_1 + 4 * x2_1 + randn(n_samples)
x4_1 = 6 * x3_1 + randn(n_samples)

x2_0 = randn(n_samples)
x1_0 =  0
x3_0 = 5 * x1_0 + 4 * x2_0 + randn(n_samples)
x4_0 = 6 * x3_0 + randn(n_samples)
diff = np.mean(x4_1) - np.mean(x4_0)
print(diff)
```

```
30.514748479180785
```

$$E[X_4 \mid do(X_1 = 1)] - E[X_4 \mid do(X_1 = 0)] = 5 \cdot 6 = 30$$

Example in Jupyter notebook Linear SCM Example



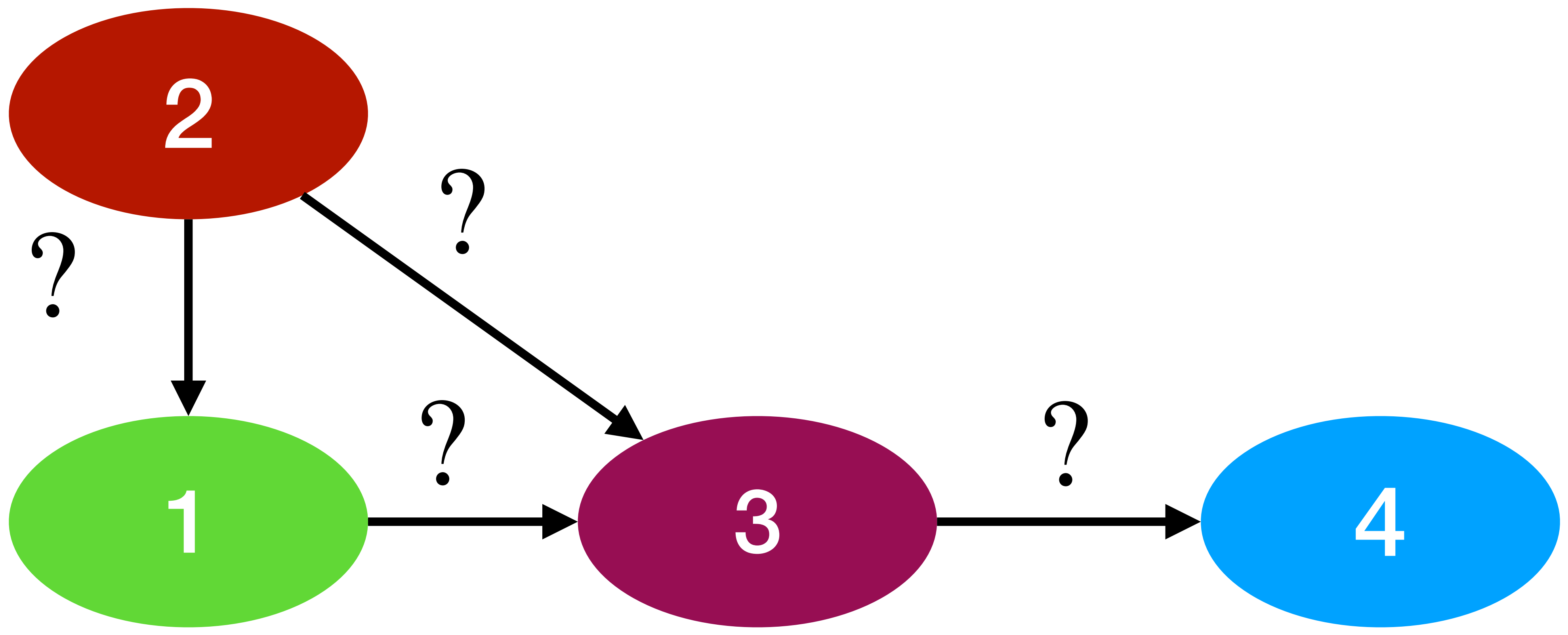

```
x2_1 = 1
x1_1 =  3 * x2_1 + randn(n_samples)
x3_1 = 5 * x1_1 + 4 * x2_1 + randn(n_samples)
x4_1 = 6 * x3_1 + randn(n_samples)

x2_0 = 0
x1_0 = 3 * x2_0 + randn(n_samples)
x3_0 = 5 * x1_0 + 4 * x2_0 + randn(n_samples)
x4_0 = 6 * x3_0 + randn(n_samples)
diff = np.mean(x4_1) - np.mean(x4_0)
print(diff)
```

```
115.57450550736193
```

$$E[X_4 \,|\, do(X_2 = 1)] - E[X_4 \,|\, do(X_2 = 0)] = 3 \cdot 5 \cdot 6 + 4 \cdot 6 = 114$$

Example in Jupyter notebook Linear SCM Example

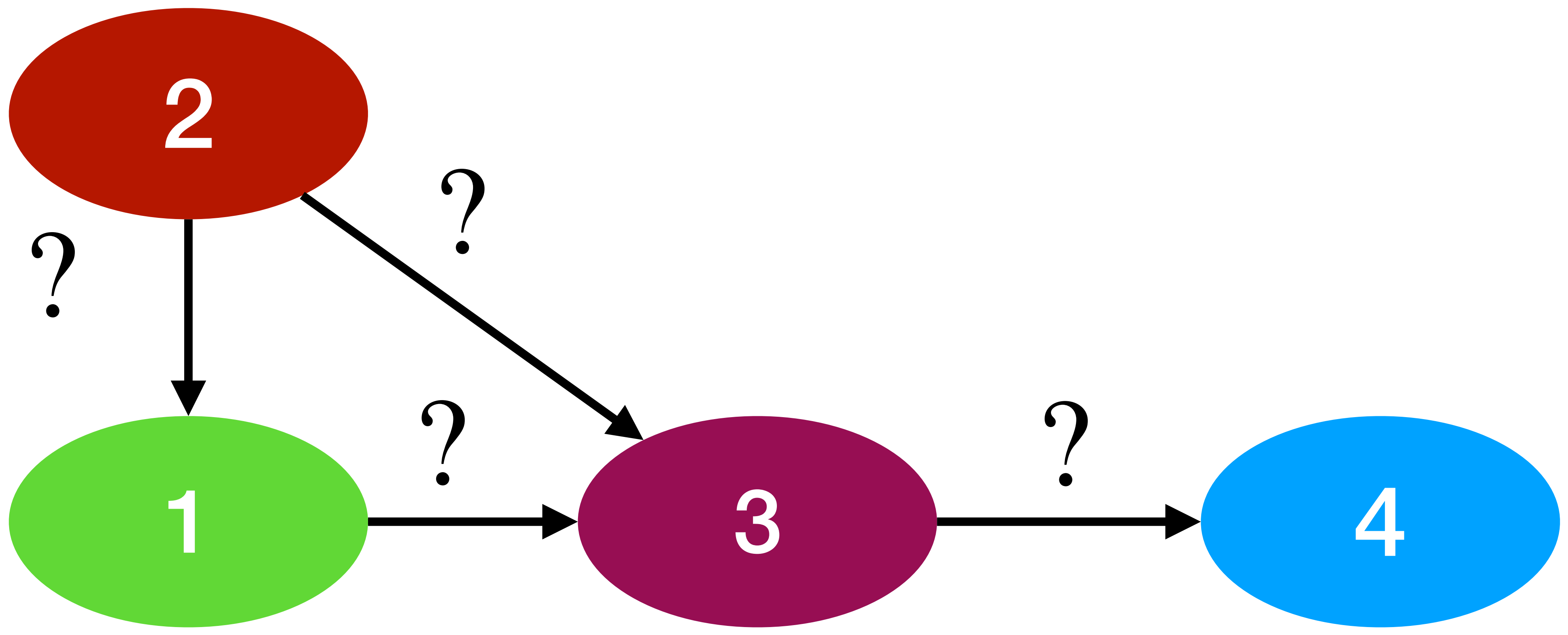


What if we don't know the coefficients and cannot simulate interventional data??

- Let's assume we have some observational data:

```
x2 = randn(n_samples)
x1 = 3 * x2 + randn(n_samples)
x3 = 5 * x1 + 4 * x2 + randn(n_samples)
x4 = 6 * x3 + randn(n_samples)
```

Example in Jupyter notebook Linear SCM Example



- We have observational data:

```
x2 = randn(n_samples)
x1 = 3 * x2 + randn(n_samples)
x3 = 5 * x1 + 4 * x2 + randn(n_samples)
x4 = 6 * x3 + randn(n_samples)
```

- Let's regress $lm(X_4 \sim X_1)$

```
linear_regressor = LinearRegression()
linear_regressor.fit(X1, Y)
linear_regressor.coef_
```

```
array([[37.15893506]])
```

- Let's regress $lm(X_4 \sim X_1, X_2)$

```
linear_regressorX12 = LinearRegression()
linear_regressorX12.fit(X21, Y)
linear_regressorX12.coef_[:,1]
```

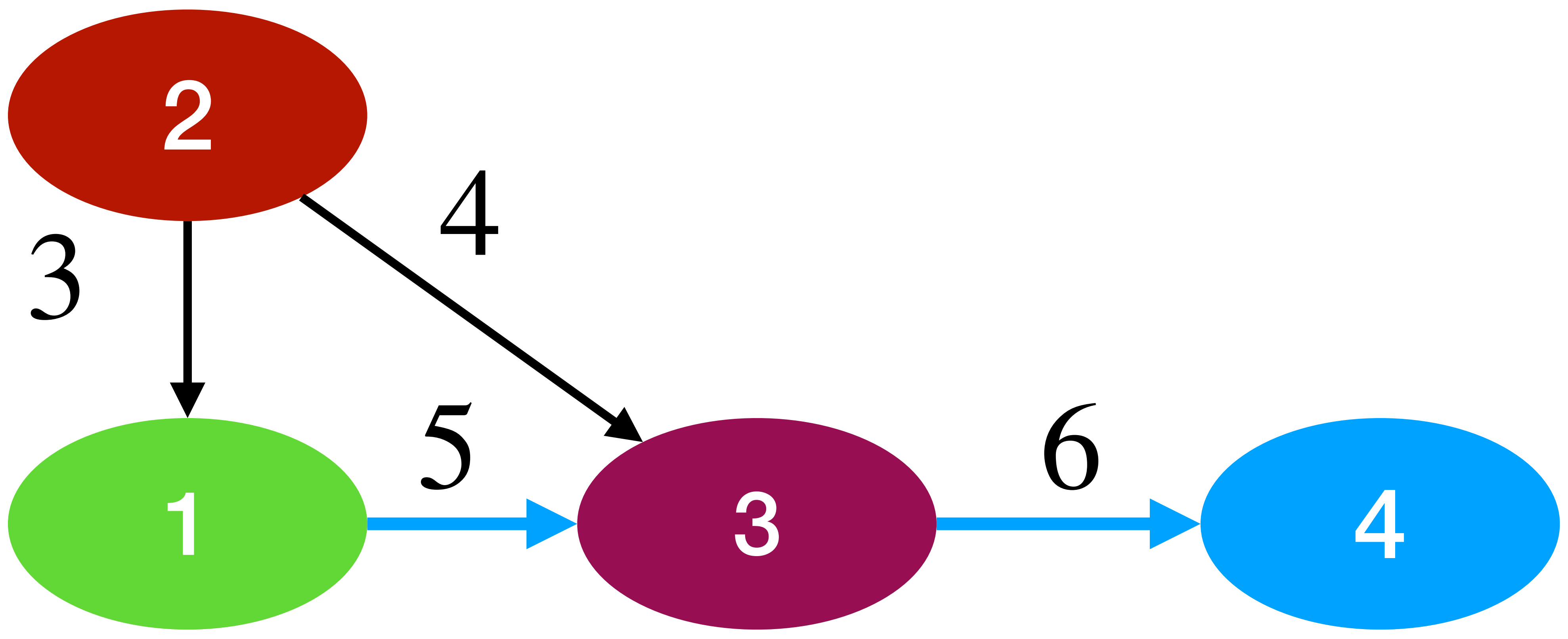
```
array([29.87150906])
```

- Let's regress $lm(X_4 \sim X_1, X_2, X_3)$

```
linear_regressorX123 = LinearRegression()
linear_regressorX123.fit(X, Y)
linear_regressorX123.coef_[:,1]
```

```
array([0.15806091])
```

Example in Jupyter notebook Linear SCM Example



$E[X_4 \mid do(X_1 = 1)] - E[X_4 \mid do(X_1 = 0)] = 30$

- Let's regress $lm(X_4 \sim X_1)$

```
linear_regressor = LinearRegression()
linear_regressor.fit(X1, Y)
linear_regressor.coef_
```

```
array([[37.15893506]])
```

- Let's regress $lm(X_4 \sim X_1, X_2)$

```
linear_regressorX12 = LinearRegression()
linear_regressorX12.fit(X21, Y)
linear_regressorX12.coef_[:,1]
```

```
array([29.87150906])
```

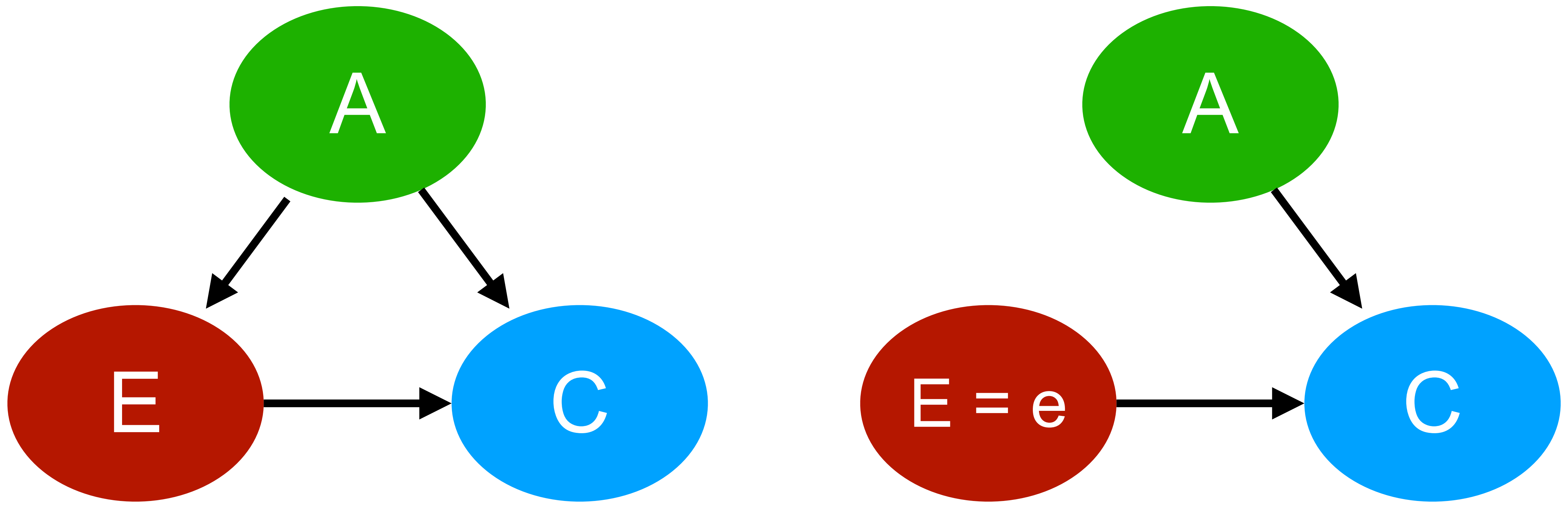
- Let's regress $lm(X_4 \sim X_1, X_2, X_3)$

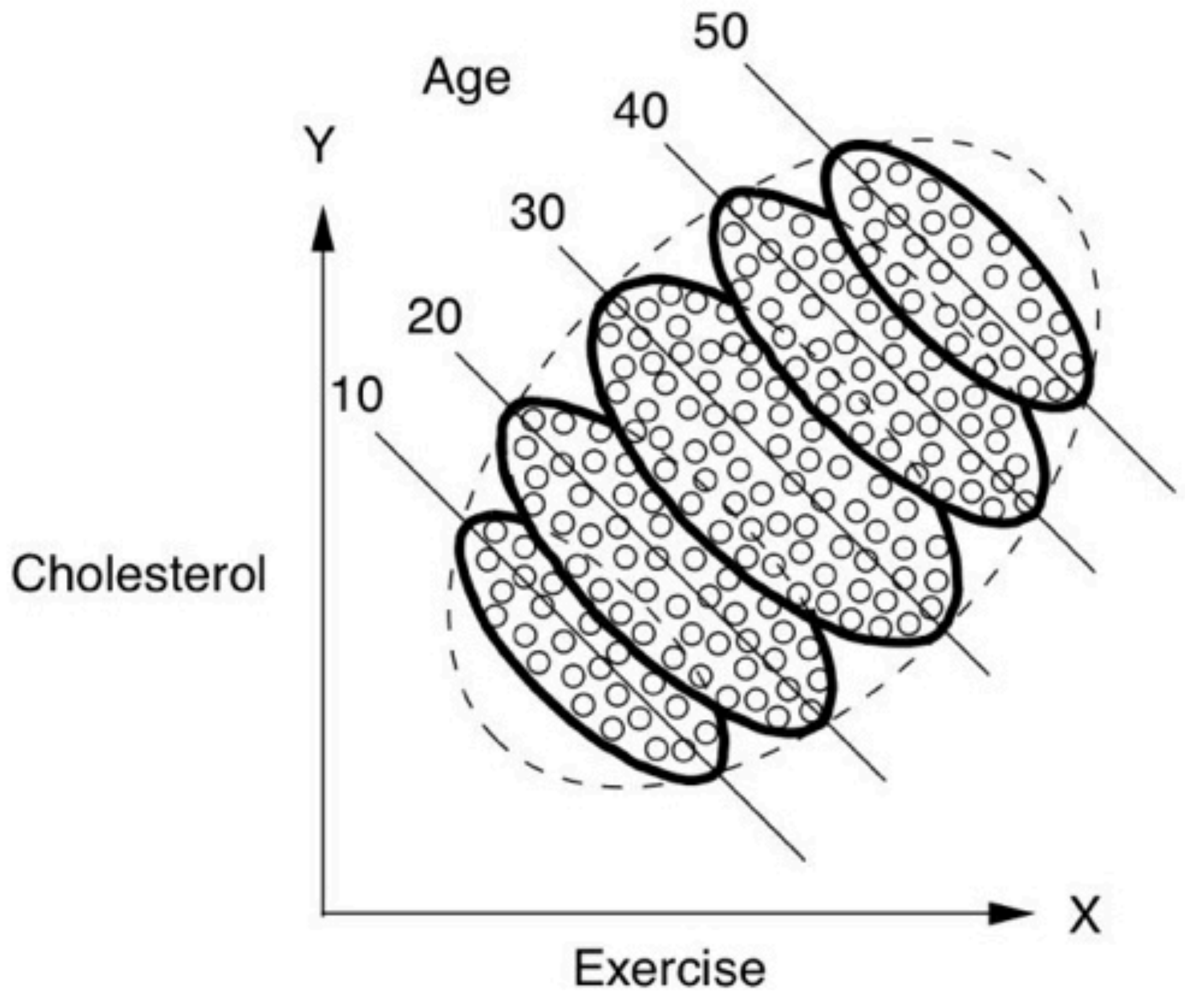
```
linear_regressorX123 = LinearRegression()
linear_regressorX123.fit(X, Y)
linear_regressorX123.coef_[:,1]
```

```
array([0.15806091])
```

- How do we know which set to adjust for?

Simpson paradox examples





From the Book of Why [Pearl 2018]

$$P(C \mid do(E=e)) = \sum_a P(A=a) \cdot P(C \mid A=a, E=e)$$

adjusting for A

- Can we determine **the adjustment sets** from the graph?

More formally: adjustment sets

- Given a causal Bayesian network (G, p) with DAG $G = (\mathbf{V}, \mathbf{E})$

More formally: adjustment sets

- Given a causal Bayesian network (G, p) with DAG $G = (\mathbf{V}, \mathbf{E})$
- We call **(valid) adjustment sets** for the causal effect of X_i on X_j with $i \neq j$, the sets $\mathbf{Z} \subseteq \mathbf{V}$ such that:

TOTAL CAUSAL EFFECT

More formally: adjustment sets

- Given a causal Bayesian network (G, p) with DAG $G = (\mathbf{V}, \mathbf{E})$
- We call **(valid) adjustment sets** for the causal effect of X_i on X_j with $i \neq j$, the sets $\mathbf{Z} \subseteq \mathbf{V}$ such that:

TOTAL CAUSAL EFFECT

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

(ADJUSTMENT FORMULA)

More formally: adjustment sets

- Given a causal Bayesian network (G, p) with DAG $G = (\mathbf{V}, \mathbf{E})$
- We call **(valid) adjustment sets** for the causal effect of X_i on X_j with $i \neq j$, the sets $\mathbf{Z} \subseteq \mathbf{V}$ such that:

TOTAL CAUSAL EFFECT

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

(ADJUSTMENT FORMULA)

Adjustment sets allow us to estimate the post-interventional distribution from a combination of observational ones

More formally: adjustment sets

- Given a causal Bayesian network (G, p) with DAG $G = (\mathbf{V}, \mathbf{E})$
- We call **(valid) adjustment sets** for the causal effect of X_i on X_j with $i \neq j$, the sets $\mathbf{Z} \subseteq \mathbf{V}$ such that:

TOTAL CAUSAL EFFECT

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

(ADJUSTMENT FORMULA)

- Can we use the graphical structure to find these valid adjustment sets?