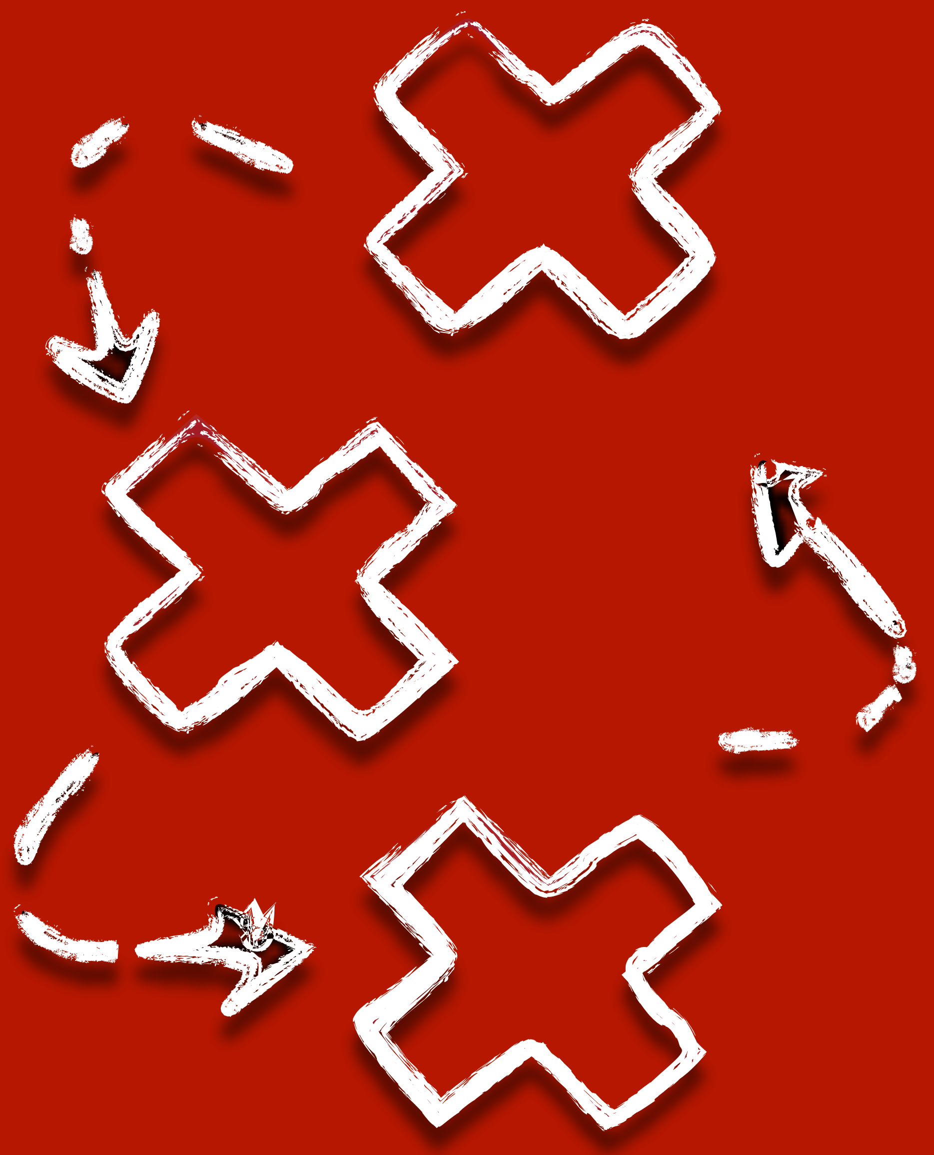


Causal Data Science

Lecture 11.2: Invariant causal prediction

Lecturer: Sara Magliacane

UvA - Spring 2023

Causal discovery overview

Constraint-based causal discovery

- Conditional independence tests
- Observational data
- Output: MEC
- SGS, PC

Score-based causal discovery

- Penalised likelihood
- Observational data
- Output: MEC
- GES

Restricted models

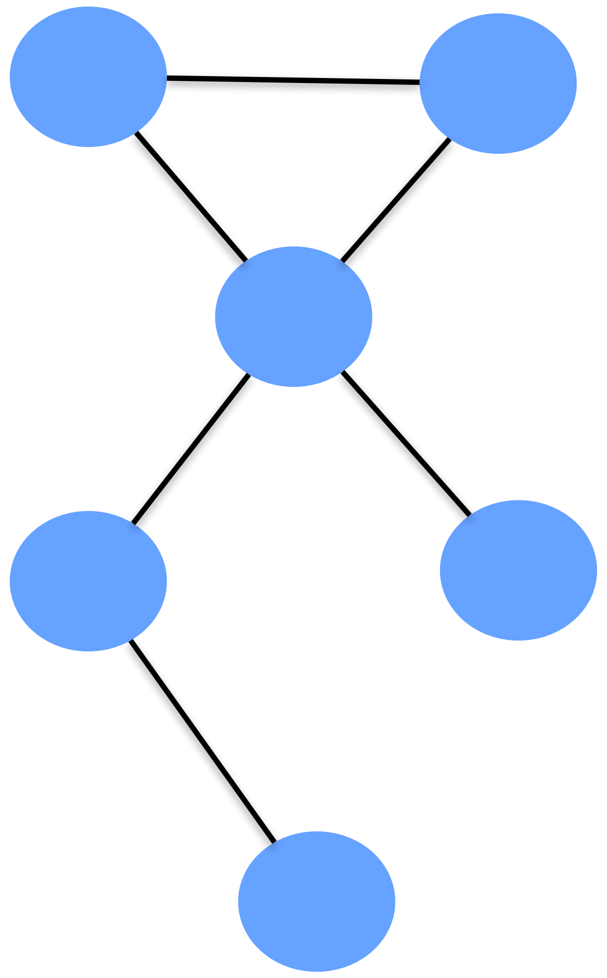
- Nonlinear additive noise, Linear Non-Gaussianity
- Observational data
- Output: DAG
- RESIT, LINGAM

Interventional causal discovery / causal invariance

- Observational and Interventional data
- Output: parents of Y
- ICP

Intervention design/Experiment selection

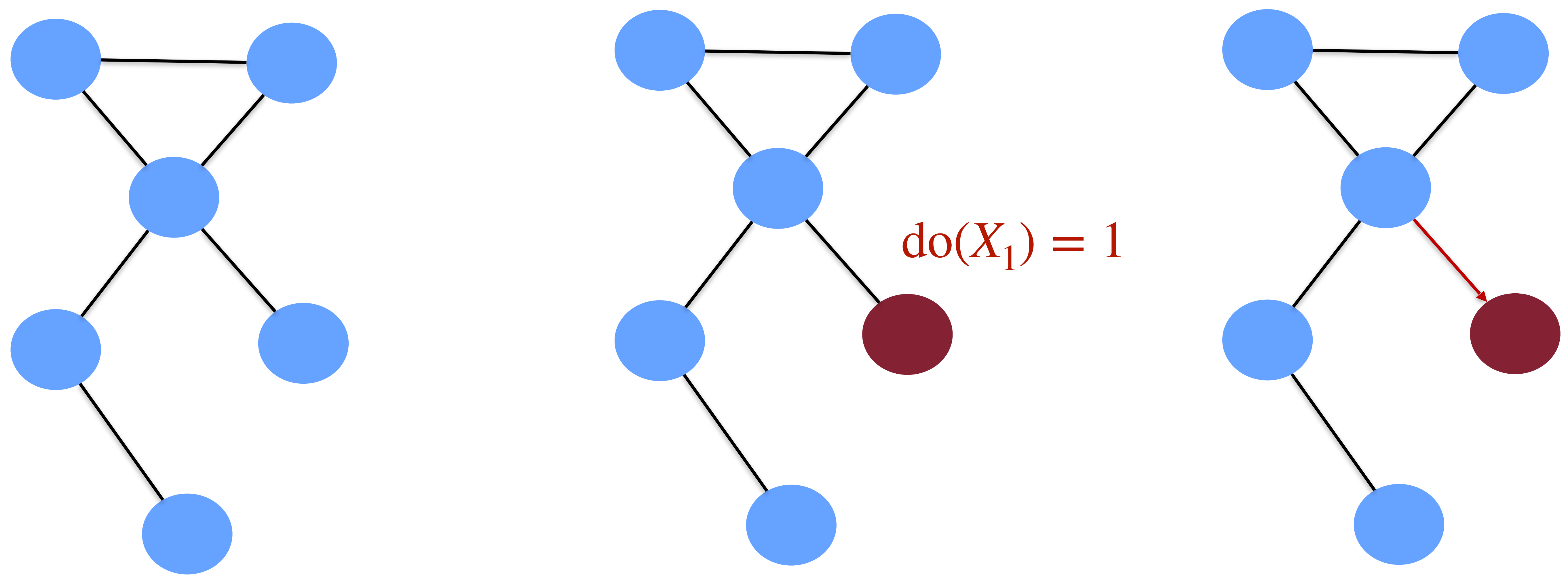
Design a set of interventions, so that we can **accurately** reconstruct **as much as possible** the causal graph **with the least samples,** also when **noisy**



https://papers.nips.cc/paper/2019/hash/5ee5605917626676f6a285fa4c10f7b0-Abstract.html https://arxiv.org/abs/2011.00641

Intervention design/Experiment selection

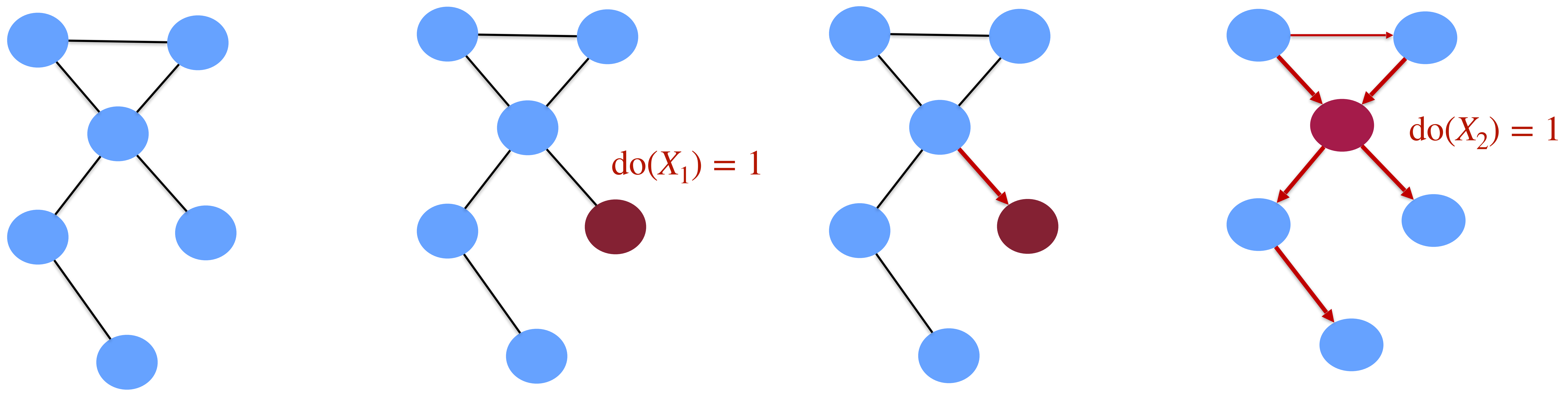
Design a set of interventions, so that we can **accurately** reconstruct **as much as possible** the causal graph **with the least samples,** also when **noisy**



https://papers.nips.cc/paper/2019/hash/5ee5605917626676f6a285fa4c10f7b0-Abstract.html https://arxiv.org/abs/2011.00641

Intervention design/Experiment selection

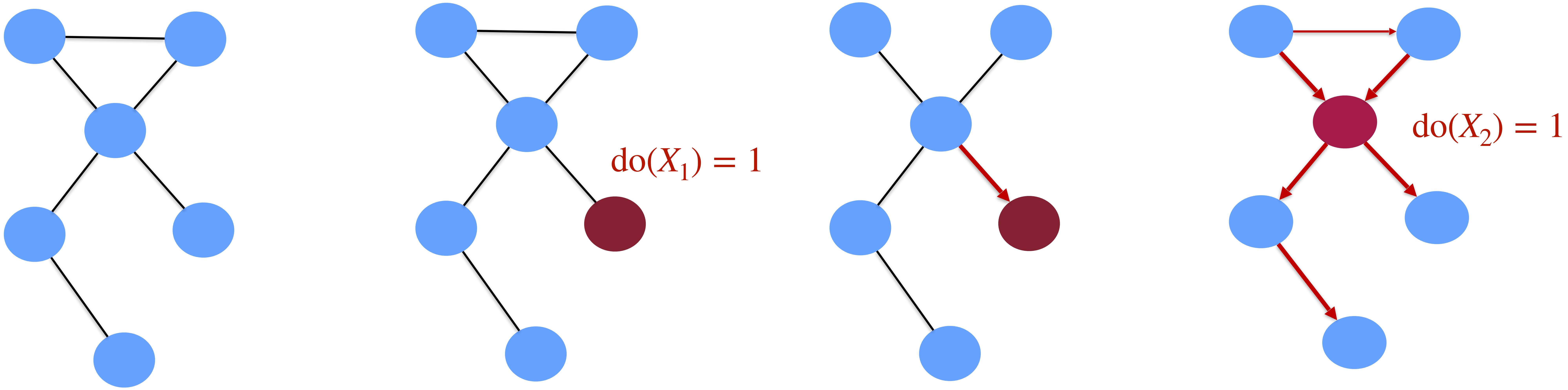
Design a set of interventions, so that we can **accurately** reconstruct **as much as possible** the causal graph **with the least samples,** also when **noisy**



https://papers.nips.cc/paper/2019/hash/5ee5605917626676f6a285fa4c10f7b0-Abstract.html https://arxiv.org/abs/2011.00641

Intervention design/Experiment selection

Design a set of interventions, so that we can **accurately** reconstruct **as much as possible** the causal graph **with the least samples,** also when **noisy**



We formalised an algorithm/policy based on the concept of **central node** for forests with noisy interventions and for DAGs with noiseless interventions.

https://papers.nips.cc/paper/2019/hash/5ee5605917626676f6a285fa4c10f7b0-Abstract.html https://arxiv.org/abs/2011.00641

Learning from multiple contexts

- Now we cannot decide which intervention to perform (**intervention design**)
  - We then also have **known intervention targets**, e.g. $\mathrm{do}(S = 1)$
- Instead, somebody gives us a **set of data from multiple contexts**
  - Possibly **unknown intervention targets**
  - Possibly **soft interventions** instead of **perfect interventions**

Invariant Causal Prediction example

$$\begin{cases} X_1 = \varepsilon_1 \\ Y = X_1 + \varepsilon_Y \\ X_2 = Y + \varepsilon_{X_2} \end{cases}$$

$X_1 \to Y \to X_2$

$\varepsilon_1, \varepsilon_Y \sim N(0,1)$, $\varepsilon_{X_2} \sim N(0, 0.01)$

M1: $Y \sim X_1$

M2: $Y \sim X_2$

M2 has smaller error

Invariant Causal Prediction example

$$\begin{cases} X_1 = \varepsilon_1 \\ Y = X_1 + \varepsilon_Y \\ X_2 = Y + \varepsilon_{X_2} \end{cases}$$

$\varepsilon_1, \varepsilon_Y \sim N(0,1)$, $\varepsilon_{X_2} \sim N(0, 0.01)$

$X_1 \to Y \to X_2$

$X_1 \to Y \quad X_2 \quad do(X_2)$

M1: $Y \sim X_1$

M2: $Y \sim X_2$

M2 has smaller error but it fails in $do(X_2)$

Invariant Causal Prediction example

$$\begin{cases} X_1 = \varepsilon_1 \\ Y = X_1 + \varepsilon_Y \\ X_2 = Y + \varepsilon_{X_2} \end{cases}$$

$X_1 \to Y \to X_2$

$\varepsilon_1, \varepsilon_Y \sim N(0,1)$, $\varepsilon_{X_2} \sim N(0, 0.01)$

M1: $Y \sim X_1$ M2: $Y \sim X_2$

Using causal parents of Y helps with DISTRIBUTION SHIFTS

Invariant Causal Prediction (ICP) [Peters et al 2016]

- Given a target variable Y and features $(X_1, \ldots, X_p)$, we want to **find the causal parents of Y, i.e. $\mathrm{Pa}(Y)$**

Invariant Causal Prediction (ICP) [Peters et al 2016]

- Given a target variable Y and features $(X_1, \ldots, X_p)$, we want to **find the causal parents of Y, i.e. $\mathrm{Pa}(Y)$**
- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$

Invariant Causal Prediction (ICP) [Peters et al 2016]

- Given a target variable Y and features $(X_1, \ldots, X_p)$, we want to **find the causal parents of Y, i.e. $\mathrm{Pa}(Y)$**
- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments Y is intervened upon**
- We can then show that $e, f \in E : P^e(Y^e \,|\, \mathrm{Pa}(Y^e)) = P^f(Y^f \,|\, \mathrm{Pa}(Y^f))$

Invariant Causal Prediction example

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

Invariant Causal Prediction example

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 1 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_Y \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

Invariant Causal Prediction example

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \qquad E = 0$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 1 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \qquad E = 1$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_Y \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \qquad E = 2$$

Invariant Causal Prediction example

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \quad E = 0$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 1 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \quad E = 1$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_Y \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \quad E = 2$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = \begin{cases} -2Y + \epsilon_2 & \text{if } E = 0 \\ 1 & \text{if } E = 1 \\ 10Y + \epsilon_Y & \text{if } E = 2 \end{cases} \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

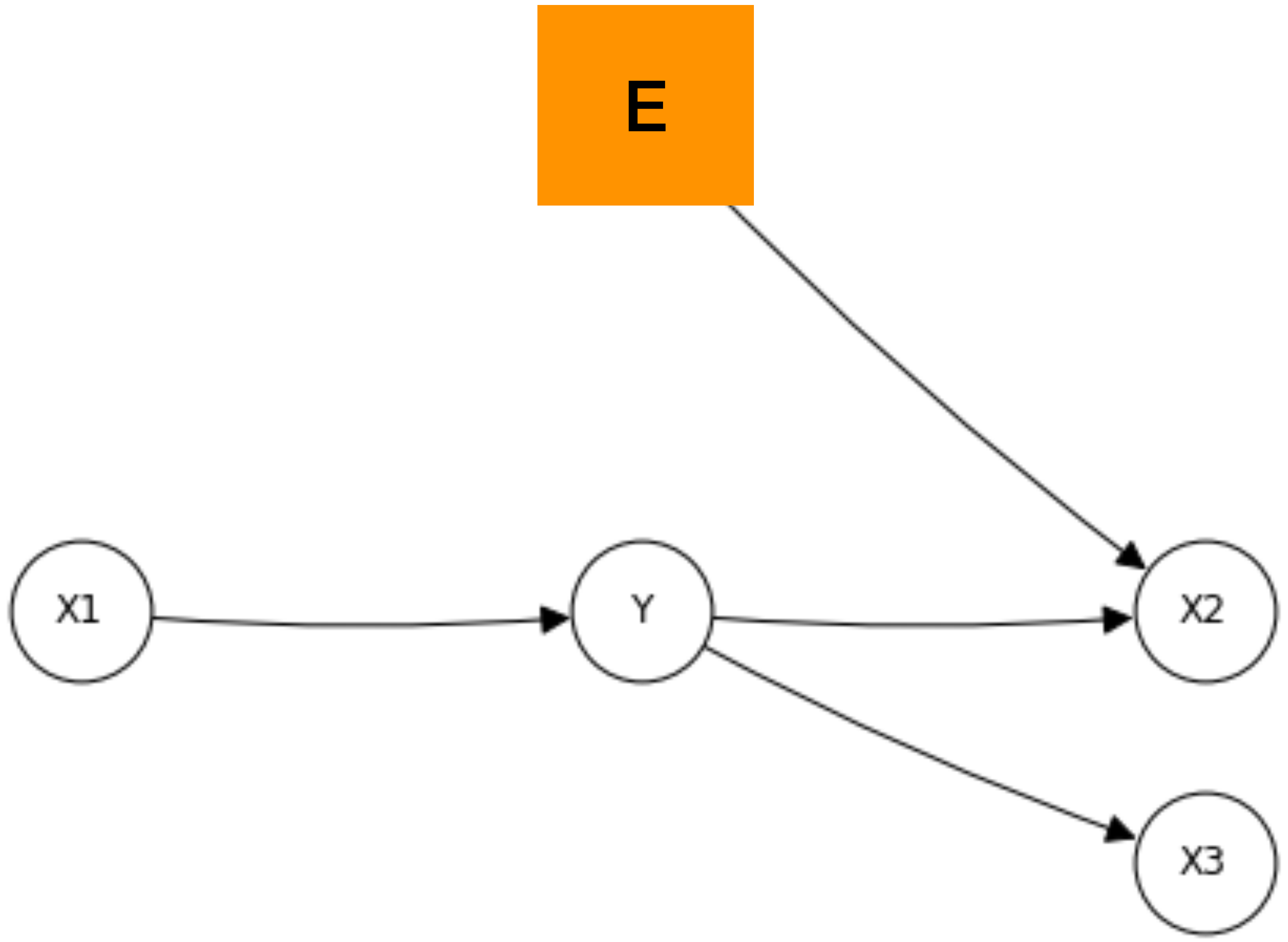
Invariant Causal Prediction example

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \qquad E = 0$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 1 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \qquad E = 1$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_Y \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \qquad E = 2$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = \begin{cases} -2Y + \epsilon_2 & \text{if } E = 0 \\ 1 & \text{if } E = 1 \\ 10Y + \epsilon_Y & \text{if } E = 2 \end{cases} \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$



Invariant Causal Prediction example 2

$$\begin{cases} X_1 = \varepsilon_1 & \longrightarrow \ do(X_1 = 1) \\ Y = X_1 + \varepsilon_Y & \longrightarrow \ \text{unchanged} \\ X_2 = Y + \varepsilon_{X_2} & \longrightarrow \ do(X_2 = 1) \end{cases}$$

$\varepsilon_1, \varepsilon_Y \sim N(0,1)$, $\varepsilon_{X_2} \sim N(0, 0.01)$

$$P^e(Y^e \,|\, \mathrm{Pa}(Y^e)) = P^f(Y^f \,|\, \mathrm{Pa}(Y^f))$$

Invariant Causal Prediction example 2

env e_1		env e_2
$X_1 = \varepsilon_1$	$\longrightarrow$	$do(X_1 = 1)$
$Y = X_1 + \varepsilon_Y$	$\longrightarrow$	unchanged
$X_2 = Y + \varepsilon_{X_2}$	$\longrightarrow$	$do(X_2 = 1)$

$\varepsilon_1, \varepsilon_Y \sim N(0,1)$, $\varepsilon_{X_2} \sim N(0, 0.01)$

$X_1 \to Y \to X_2$ env e_1

$X_1 \to Y \quad X_2$ env e_2

$$\begin{tikzcd} & E \arrow[dl] \arrow[drr] & & \\ X_1 \arrow[r] & Y \arrow[rr] & & X_2 \end{tikzcd}$$

$e_1 + e_2$

$$Y \perp\!\!\!\perp E \mid X_1 \Rightarrow X_1 \in Pa(Y)$$

Invariant Causal Prediction (ICP)

- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments Y is intervened upon**

Invariant Causal Prediction (ICP)

- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments Y is intervened upon**
- We represent the environment index with E
- If there are no latent confounders, one can prove that:

$$\bigcap_{\mathbf{S} \subseteq \{1,\ldots,p\} \text{ s.t. } Y \perp\!\!\!\perp E | \mathbf{S}} \mathbf{S} \subseteq \mathrm{Pa}(Y)$$

Invariant Causal Prediction example 3

env e_1

$$\begin{cases} X_1 = \varepsilon_1 \\ Y = X_1 + \varepsilon_Y \\ X_2 = Y + \varepsilon_{X_2} \end{cases}$$

env e_2

$\to do(X_1 = 1)$

$\varepsilon_1, \varepsilon_Y \sim N(0,1)$, $\varepsilon_{X_2} \sim N(0, 0.01)$

$X_1 \to Y \to X_2$ env e_1

$X_1 \to Y \to X_2$ env e_2

$$\begin{tikzcd} & E \arrow[dl] & \\ X_1 \arrow[r] & Y \arrow[r] & X_2 \end{tikzcd}$$

$Y \perp\!\!\!\perp E \mid X_1$

$Y \perp\!\!\!\perp E \mid X_1, X_2$

$\{X_1\} \cap \{X_1, X_2\} = \{X_1\}$

$\Rightarrow X_1 \in Pa(Y)$

Invariant Causal Prediction example 4

env e_1

$$\begin{cases} X_1 = \varepsilon_1 \\ Y = X_1 + \varepsilon_Y \\ X_2 = Y + \varepsilon_{X_2} \end{cases}$$

$\varepsilon_1, \varepsilon_Y \sim N(0,1)$, $\varepsilon_{X_2} \sim N(0, 0.01)$

env e_2

$\rightarrow do(X_1 = 1)$

$X_1 \rightarrow Y \rightarrow X_2$ env e_1

$X_1 \rightarrow Y \rightarrow X_2$ env e_2

E

$X_1 \rightarrow Y \rightarrow X_2$

$pa(X_2) = ?$

$X_2 \perp\!\!\!\perp E \mid X_1$

$X_2 \perp\!\!\!\perp E \mid X_1, Y$

$X_2 \perp\!\!\!\perp E \mid Y$

$\{X_1\} \cap \{X_1, Y\} \cap \{Y\} = \emptyset$

$\emptyset \subseteq pa(X_2)$

Invariant Causal Prediction (ICP)

- We assume we have access to **a set of different environments** E (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments** Y **is intervened upon**
- Approximate test on residuals for each $\mathbf{S} \subseteq \{1, \ldots, p\}$
  - Fit linear regression with $\mathbf{S}$ and let $R = Y - \hat{f}(X_{\mathbf{S}})$
  - Test null hypothesis that mean and variance of R are the same across E
  - Combine the two p-values and reject $\mathbf{S}$ if the combined p-value $\leq \alpha$

ICP improves with more interventions

$E \rightarrow X_1 \rightarrow X_2 \rightarrow Y$

$$\left.\begin{array}{l} E \perp\!\!\!\perp Y \mid X_1 \\ E \perp\!\!\!\perp Y \mid X_2 \\ E \perp\!\!\!\perp Y \mid X_1, X_2 \end{array}\right\} \cap \quad \phi$$

+ new environment e3

$X_2 = g^{new}(X_1, \varepsilon_2)$

$E \rightarrow X_1 \rightarrow X_2 \rightarrow Y$, $E \rightarrow X_2$

$$\begin{array}{l} E \perp\!\!\!\perp Y \mid X_2 \\ E \perp\!\!\!\perp Y \mid X_2, X_1 \end{array} \cap \quad X_2 \in Pa(Y)$$

ICP improves with more interventions

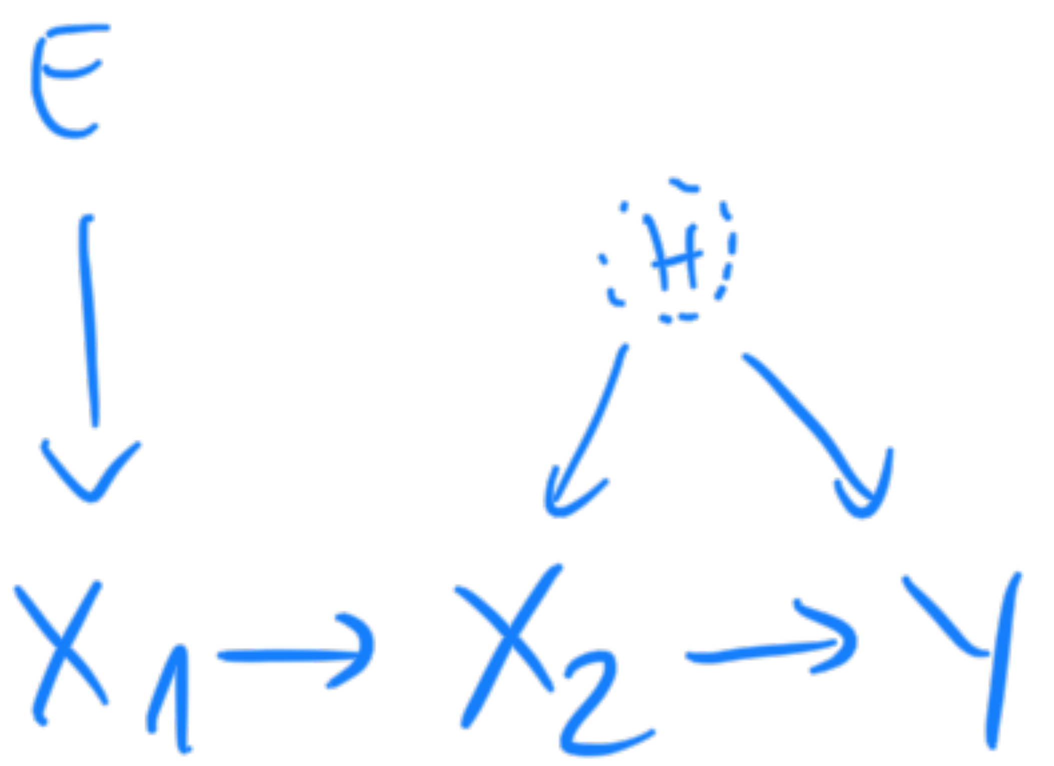
- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments Y is intervened upon**
- We represent the environment index with E, no latent confounders
- If the **environment variable causes all $(X_1, \ldots, X_p)$**

$$\mathrm{Pa}(Y) = \bigcap_{\mathbf{S} \subseteq \{1,\ldots,p\} \text{ s.t. } Y \perp\!\!\!\perp E | \mathbf{S}} \mathbf{S}$$

Invariant Causal Prediction - latent confounders

- If there are latent confounders, one can prove that:

$$\bigcap_{\mathbf{S}\subseteq\{1,\ldots,p\} \text{ s.t. } Y \perp\!\!\!\perp E \mid \mathbf{S}} \mathbf{S} \subseteq \mathrm{Anc}(Y)$$



$Y \not\perp\!\!\!\perp E \mid X_2$

$Y \perp\!\!\!\perp E \mid X_1$ $\Rightarrow X_1 \in Anc(Y)$

$Y \perp\!\!\!\perp E \mid X_1, X_2$

Learning from multiple contexts

- Now we cannot decide which intervention to perform (**intervention design**)
  - We then also have **known intervention targets**, e.g. $\mathrm{do}(S = 1)$
- Instead, somebody gives us a **set of data from multiple contexts**
  - Possibly **unknown intervention targets**
  - Possibly **soft interventions** instead of **perfect interventions**

ICP finds subsets of parents, what about finding (an equivalence class of) the causal graph?