Generalizing the Effect of Soft Interventions

Juan D. Correa <jcorrea@autonoma.edu.co>

Online Causal Inference Seminar

October 19, 2021

UNIVERSIDAD AUTÓNOMA DE MANIZALES
UAM®
VIGILADA MINEDUCACIÓN

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

CS@CU COMPUTER SCIENCE

Outline

- Motivation
- Soft-Interventions and Structural Causal Models
- Transportability
- Symbolic solution
- Algorithmic solution
- Conclusions

This presentation is based on the paper:

General Transportability of Soft Interventions: Completeness Results,
which is joint work with **Prof. Elias Bareinboim** at Columbia University.

Motivation

- Establishing the effect of new interventions/policies from data is a pervasive task across the empirical sciences.

- Controlled experimentation is considered the gold standard to learn such causal effects in many settings. However, experiments rarely generalize to domains outside where they were originally performed.
  Many significant problems in the empirical sciences are instances of this task (Banerjee et al. 07, Duflo et al. 07, Bertrand et al. 10).

- This problem has been studied in the causal inference literature under the rubric of transportability theory (Bareinboim-Pearl 2011, 2012, 2013, Bareinboim et al. 2013). There are sufficient and necessary conditions that solve transportability of atomic interventions.

- In this talk, we discuss the task of generalizing policies (or soft interventions) from a collection of heterogenous data, including observations and experiments.

Example tasks/applications

- Predict the impact of public policies for a country using data from other, similar but different, countries.
- Adapt a classifier to a new domain while using minimal data in the new domain.
- Recover the results from an experimental study carried out on a population that is known to misrepresent the general target population.
- Combine the results of A/B (or Multivariable) experiments in advertisement to predict the effect of a new (not-tested) strategy and over a new niche.

Soft Interventions

Soft Interventions (Motivation)

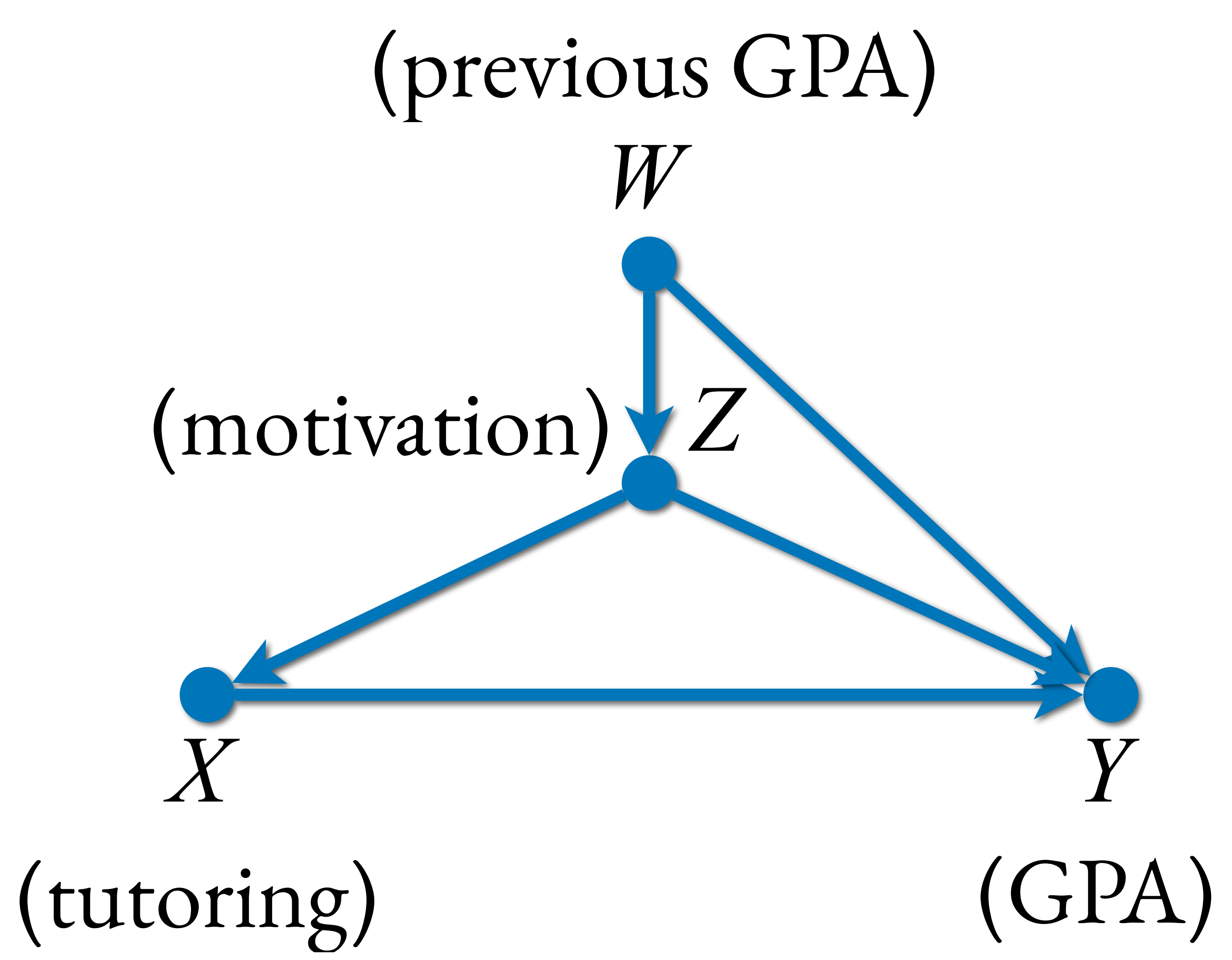
In decision making scenarios, even if the effect of a do() intervention is identifiable …

- Available resources may be insufficient to implement the corresponding policy.
  - There are not enough teachers to cover all the hours of tutoring needed for every single student in a school.

- Effectiveness of the intervention cannot be guaranteed:
  - Patients assigned treatment may not follow it.

Example - Tutoring Program

- For a group of students we observe their GPA at the beginning of the term, their motivation level (low, high), whether they get tutoring or not, and their GPA at the end of the semester.



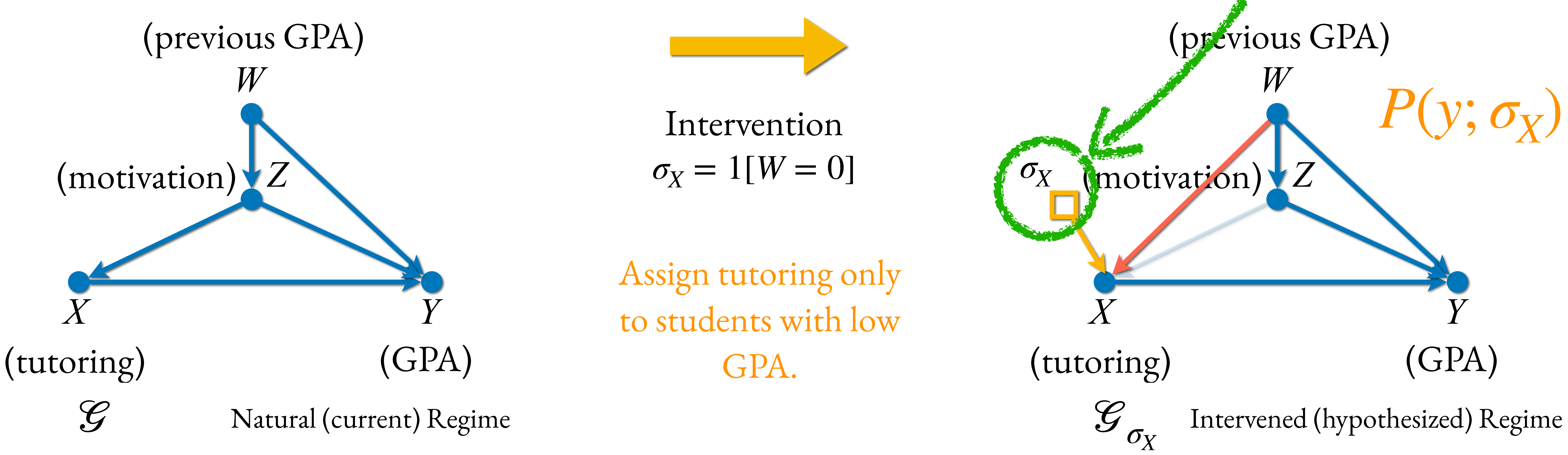
- Using machine learning, and with enough data, a student's GPA can be **predicted** with small error given other features i.e., $P(y \mid w, z, x)$.

- This data reflects the **current/natural** regime, yet we aim to assess the impact of a new unobserved policy (intervention) on the students GPA.

Consider a Soft Intervention

- Resources are limited so we want to focus on students that need tutoring the most.
- From now on, students with low GPA have to get t… be available to them. That is: $P^*(X = 1 \mid W = 0)$

Regime node used to encode the fact that X has been intervened on.



Some Canonical types of Interventions

[Dawid 02, Eberhardt&Scheines 07, Tian 08]

- **Hard/atomic:** $\sigma_X = do(X = x)$ set variable X to a constant value x.
  (Do-calculus original treatment considered mostly this type of intervention.)
  - Every student gets tutoring.
- **Conditional:** $\sigma_X = g(w)$ sets the variable X to output the result of a function g that depends on a set of observable variables W.
  - Students get tutoring if and only if they have a low GPA.
- **Stochastic:** $\sigma_X = P^*(x \mid w)$ sets the variable X to follow a given probability distribution conditional on a set of variables W.
  - Students with low GPA enter a raffle for 80% of the spots, other interested students enter for the remaining 20%.

Transportability

Observation: All data is not created equal…

- Heterogenous datasets are pervasive. They could …

(1) have different **experimental conditions**, → Surrogate Experiments

(2) come from different underlying **populations**, → Transportability

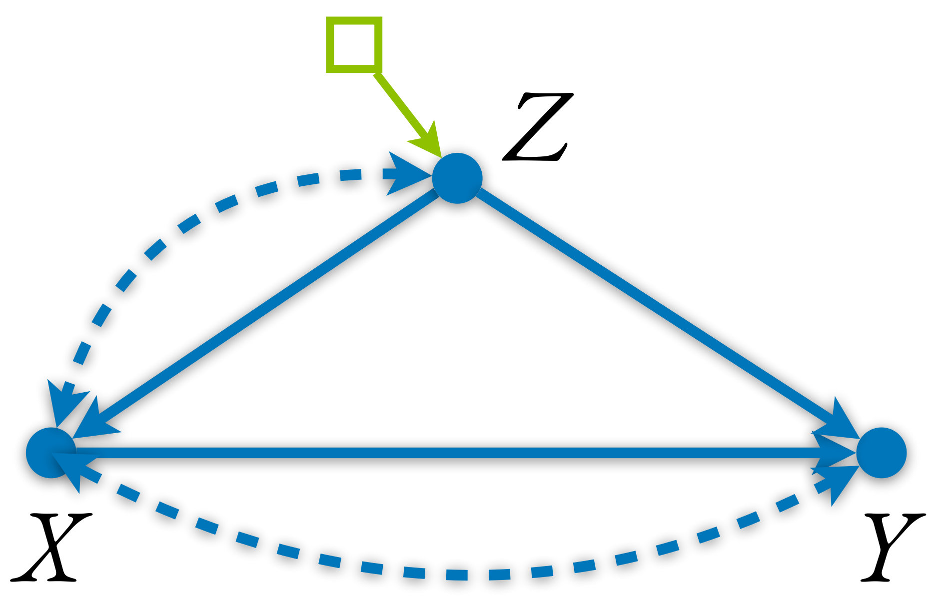
(3) suffer from non-random **sampling mechanisms**, → Sample Selection Bias

(4) **measure** different sets of variables. → Partial Observability

Transportability (TR)

[Bareinboim&Pearl 14]

- Suppose data comes from different domains $\pi^*, \pi^1, \pi^2, \ldots,$ where π^* represents the target domain in which the causal effect is to be identified.
  - Use experiments on mice to assess the effect of a treatment on humans
  - Use data from a study carried out in Los Angeles to estimate the impact of a new policy in New York City



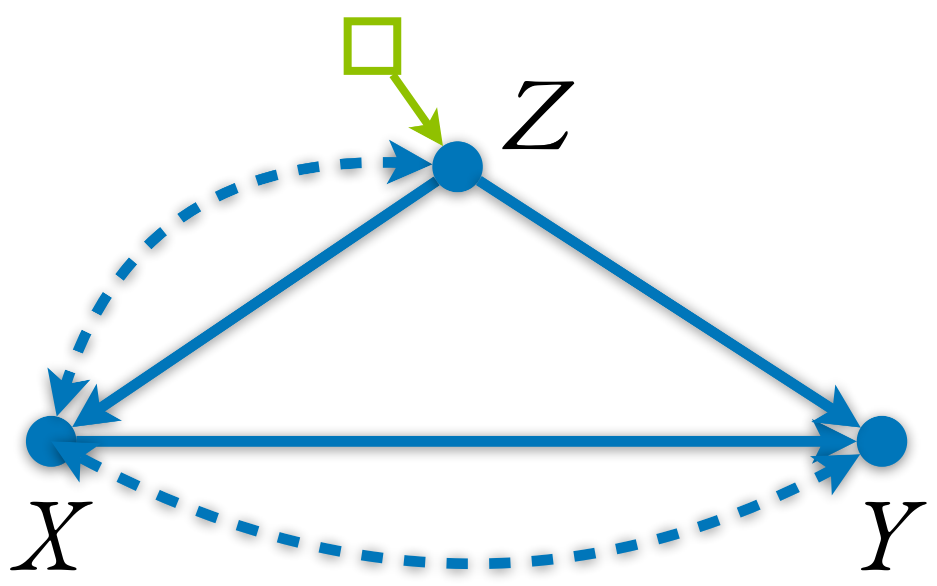
	π	π^*
domain	LA →	NYC
data	$P(\mathbf{v})$ $P(\mathbf{v}; \sigma_X = do(x))$	$P^*(\mathbf{v})$
query		$P^*(y; \sigma_X = g(z))$

square node (□) indicates that the mechanism behind Z differs between domains π and π^*

Selection Diagram

[Bareinboim&Pearl 11]

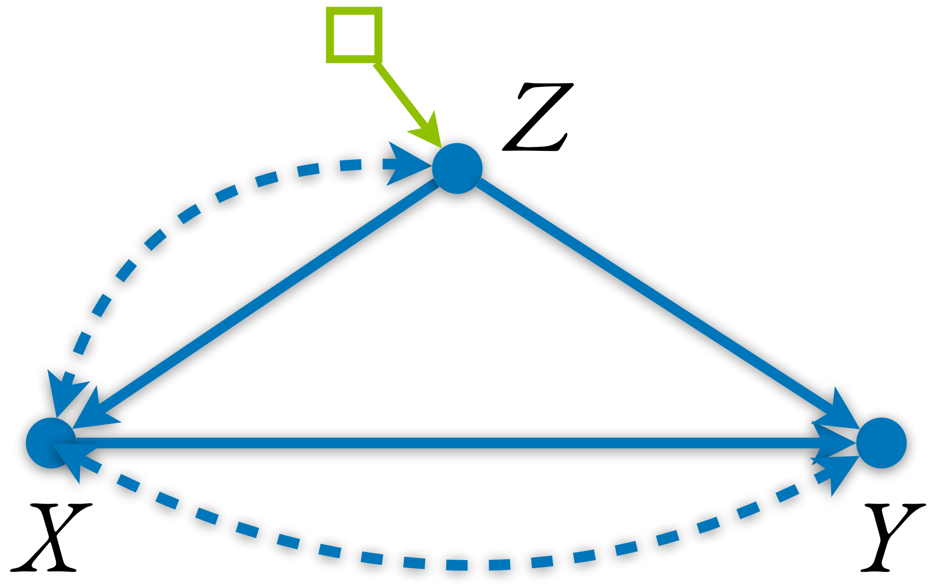
- Each domain is assumed to have an underlying SCM which structure is summarized with a causal diagram annotated with square nodes to show the difference between domains. This is called selection diagram.



A square node pointing to variable Z encodes the assumption that:

- $f_Z \neq f_Z^*$, or
- $P(U_z) \neq P^*(U_z)$

Transportability Formula



	π		π^*
domain	LA	→	NYC
data	$P(\mathbf{v})$ $P(\mathbf{v}; \sigma_X = do(x))$		$P^*(\mathbf{v})$
query			$P^*(y; \sigma_X = g(z))$

When transportable, the effect of interest can be written in terms of available distributions. For instance:

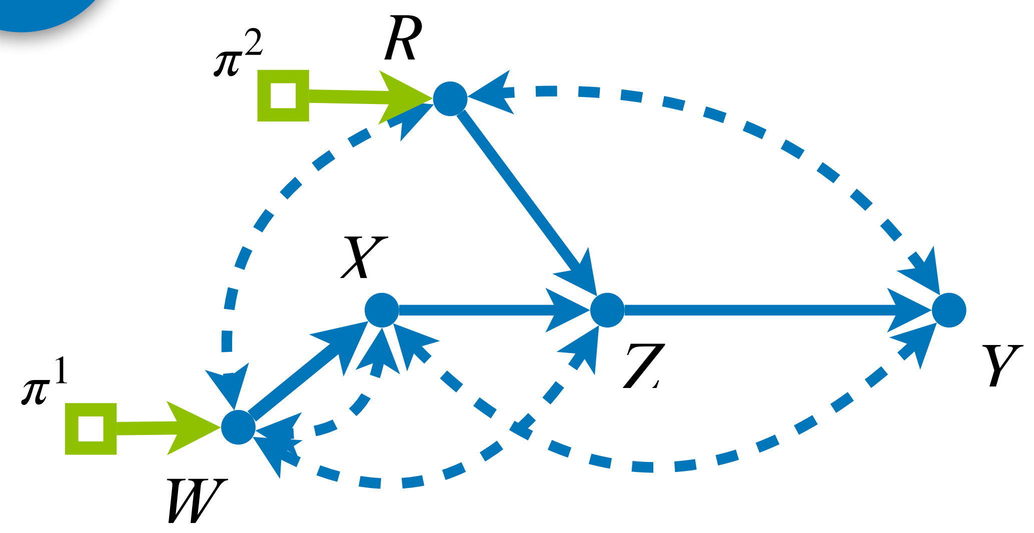
$$P^*(y; \sigma_X = g(z)) = \sum_{z,x} P(y \mid do(x), z) P(x \mid z; \sigma_X = g(z)) P^*(z)$$

The Soft Transportability Task

1 **Target Query**

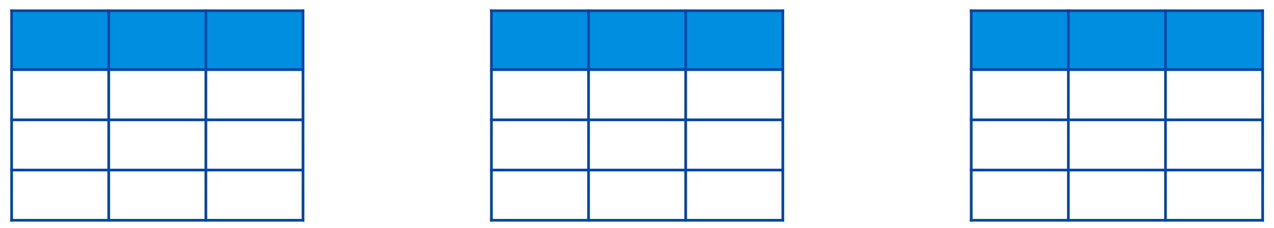
e.g., $Q = P(\mathbf{y} \mid \mathbf{z}; \sigma_{\mathbf{x}})$

2 **Selection Diagram**



3 **Data**

π^*	π^1	π^2



$P^*(\mathbf{V})$	$P^1(\mathbf{V}; \sigma_Z^1)$	$P^1(\mathbf{V}; \sigma_W^2)$

Inference Engine

No

Yes

$$Q = f(P_1, \ldots, P_k)$$

Another Example

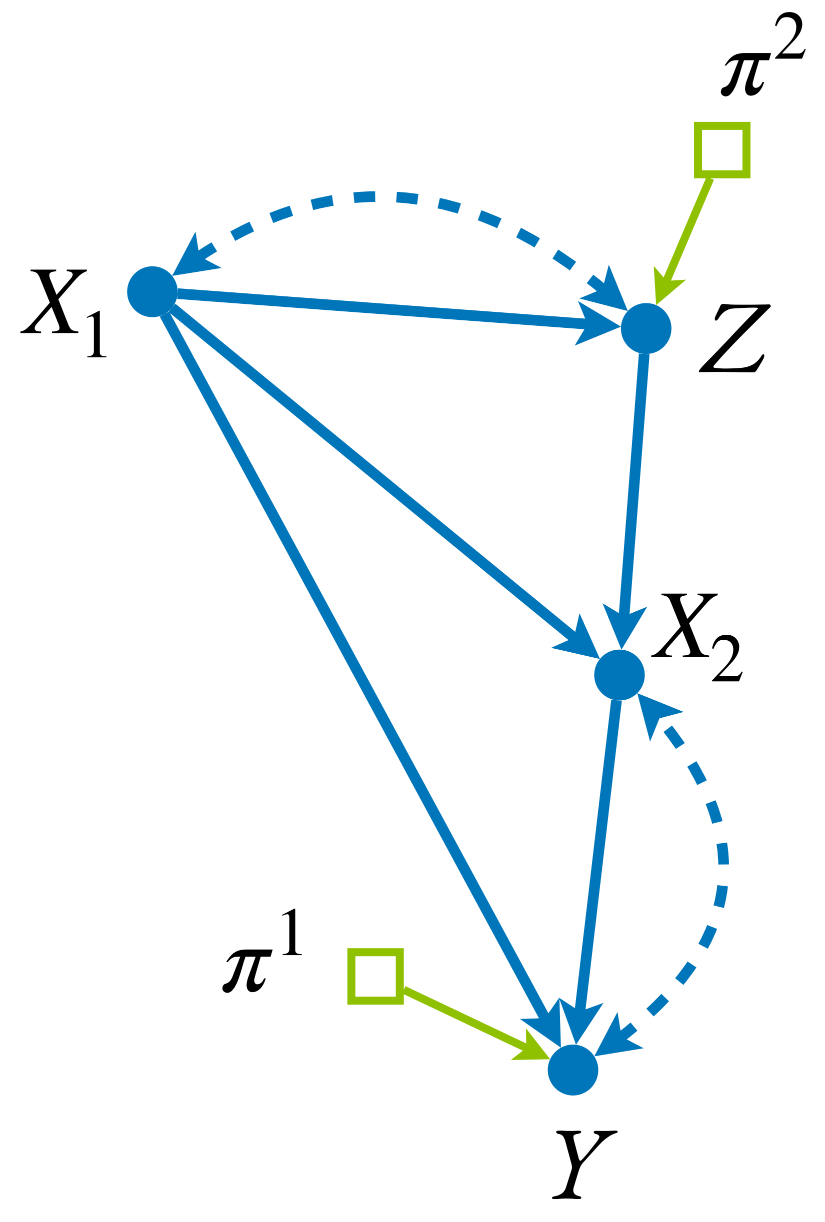
- Query: The effect of an stochastic policy $\sigma^* = \{\sigma_{X_1} = \hat{P}(X_1), \sigma_{X_2} = \hat{P}(X_2 \mid X_1, Z)\}$ on Y in a target domain π^*, namely, $P^*(y; \sigma^*)$

- Available data:

  - Controlled trial in domain π^1:
    $$P^1(\mathbf{V}; \sigma_{X_1,X_2} = do(x_1, x_2))$$

  - Conditional experiment in π^2:
    $$P^2(\mathbf{V}; \sigma_{X_2} = do(x_2 = g(X_1, Z)))$$

- Selection diagram



Different Diagrams for Different Distributions

$P^*(y; \sigma^*)$

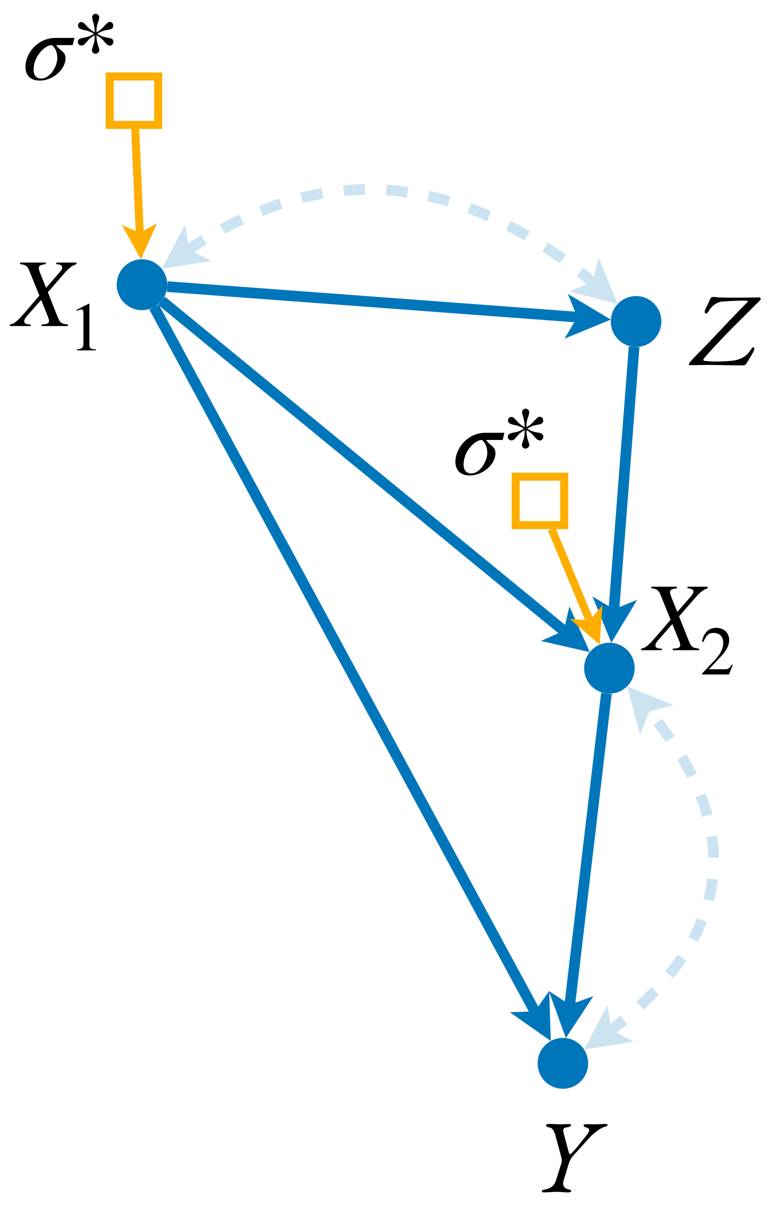
(π^*)

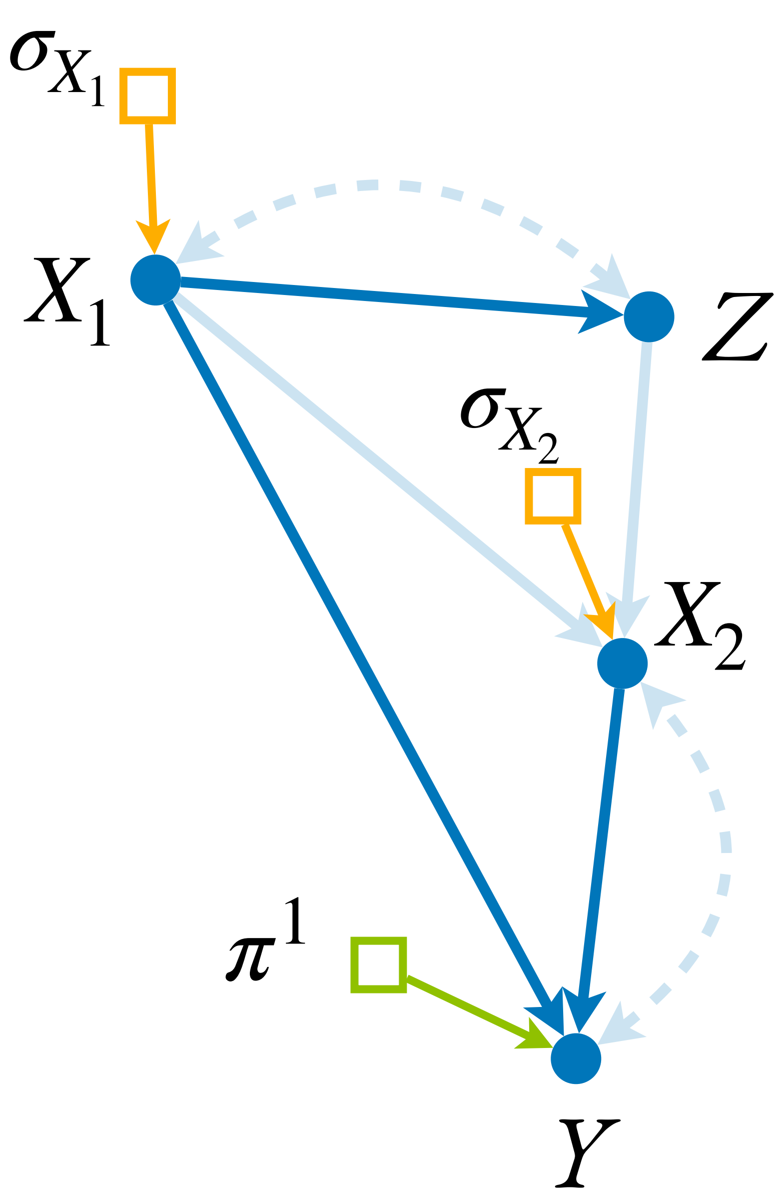
$P^1(\mathbf{V}; \sigma_{X_1,X_2} = do(x_1, x_2))$

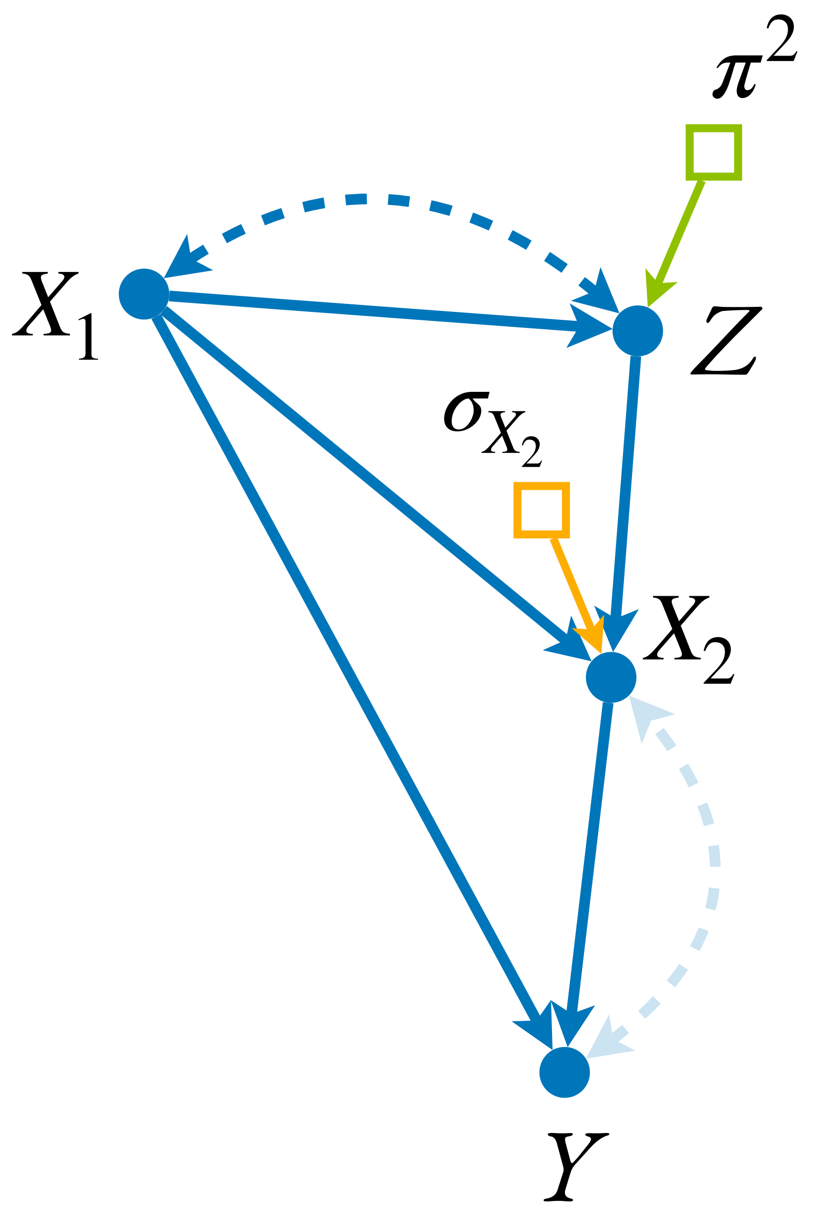
(π^1)

$P^2(\mathbf{V}; \sigma_{X_2} = do(x_2 = g(X_1, Z)))$

(π^2)







Symbolic Solution

σ-calculus

[Correa&Bareinboim, 20]

Theorem. Let $\mathcal{G}$ be a causal diagram, with endogenous variables $\mathbf{V}$. For any disjoint subsets $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, two disjoint subsets $\mathbf{T}, \mathbf{W} \subseteq \mathbf{V} \backslash (\mathbf{Z} \cup \mathbf{Y})$ (i.e., possibly including elements in $\mathbf{X}$), the following rules are valid for any intervention strategies $\sigma_{\mathbf{X}}$, $\sigma_{\mathbf{Z}}$:

Rule 1 (Insertion/Deletion of observations):

$$P(\mathbf{y} \mid \mathbf{w}, \mathbf{t}; \sigma_{\mathbf{X}}) = P(\mathbf{y} \mid \mathbf{w}; \sigma_{\mathbf{X}}) \quad \text{if } (\mathbf{Y} \perp \mathbf{T} \mid \mathbf{W}) \text{ in } \mathcal{G}_{\sigma_{\mathbf{X}}}$$

Rule 2 (Change of regimes under observation):

$$P(\mathbf{y} \mid \mathbf{x}, \mathbf{w}; \sigma_{\mathbf{X}}) = P(\mathbf{y} \mid \mathbf{x}, \mathbf{w}) \quad \text{if } (\mathbf{Y} \perp \mathbf{X} \mid \mathbf{W}) \text{ in } \mathcal{G}_{\sigma_{\mathbf{X}}\underline{\mathbf{X}}} \text{ and } \mathcal{G}_{\underline{\mathbf{X}}}$$

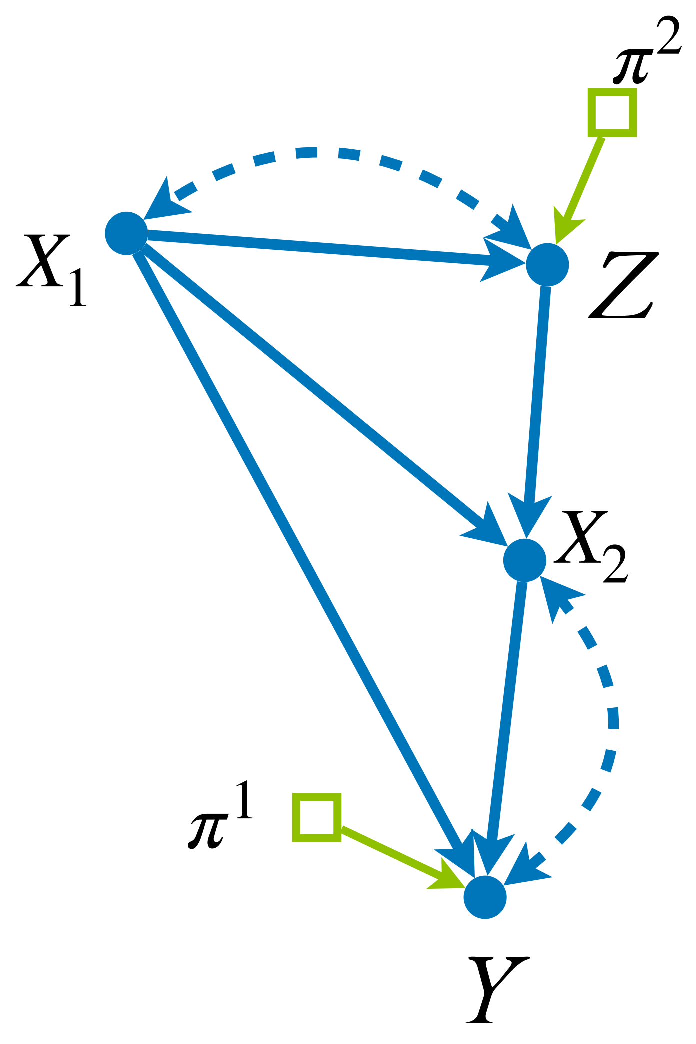
Rule 3 (Change of regimes without observation):

$$P(\mathbf{y} \mid \mathbf{w}; \sigma_{\mathbf{X}}) = P(\mathbf{y} \mid \mathbf{w}) \quad \text{if } (\mathbf{Y} \perp \mathbf{X} \mid \mathbf{W}) \text{ in } \mathcal{G}_{\sigma_{\mathbf{X}}\overline{\mathbf{X}(\mathbf{W})}} \text{ and } \mathcal{G}_{\overline{\mathbf{X}(\mathbf{W})}}$$

Solving the instance with σ-calculus

$$P^*(y;\sigma^*) = \sum_{x_1,z,x_2} P^*(y \mid x_1, z, x_2;\sigma^*) P^*(x_2 \mid z, x_1;\sigma^*) P^*(z \mid x_1;\sigma^*) P^*(x_1;\sigma^*)$$

$$P^*(x_2 \mid z, x_1;\sigma^*) P^*(x_1;\sigma^*) = \hat{P}(x_2 \mid z, x_1)\hat{P}(x_1) \qquad \text{by def. of } \sigma^*$$



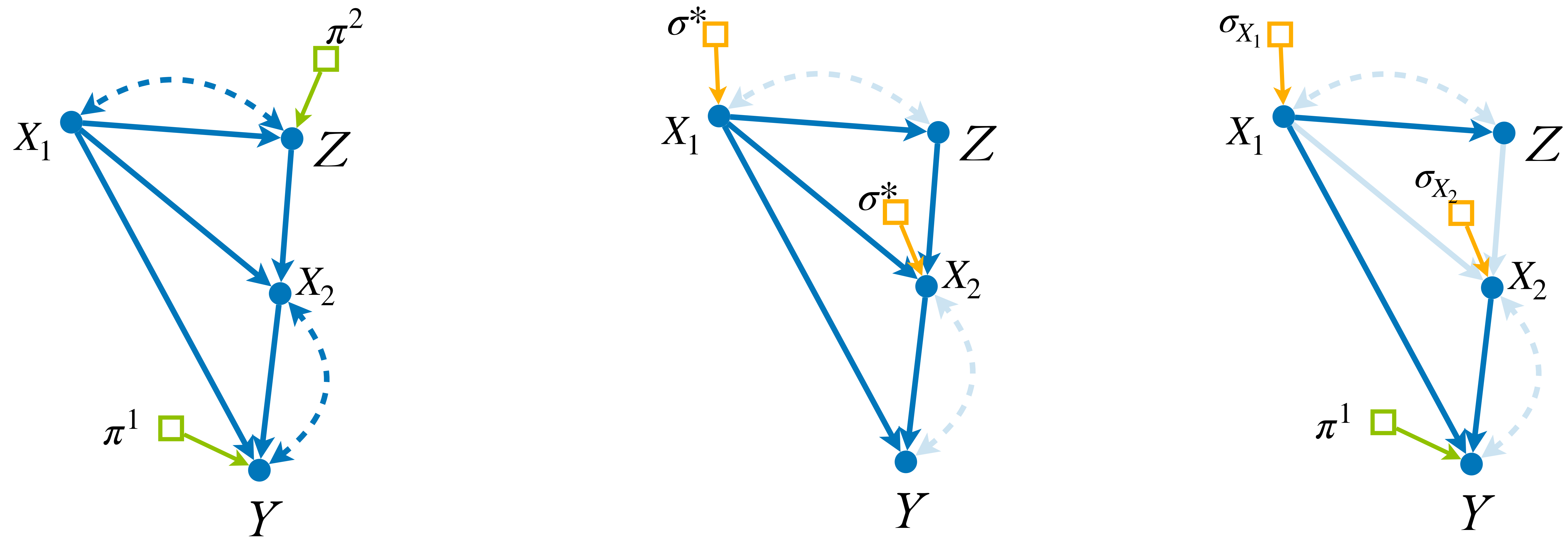
Solving the instance with σ-calculus

$$P^*(y;\sigma^*) = \sum_{x_1,z,x_2} P^*(y \mid x_1, z, x_2; \sigma^*) P^*(x_2 \mid z, x_1; \sigma^*) P^*(z \mid x_1; \sigma^*) P^*(x_1; \sigma^*)$$

$P^*(z \mid x_1; \sigma^*) = P^*(z \mid x_1; \sigma^*_{x_1})$ by rule 3 remove intervention σ^*_2

$= P^*(z \mid x_1; \sigma_{x_1})$ by rule 2 change to intervention σ_1

$= P^1(z \mid x_1; \sigma_{x_1})$ $(Z \perp \square_{\pi_1} \mid X_1)$ change domain



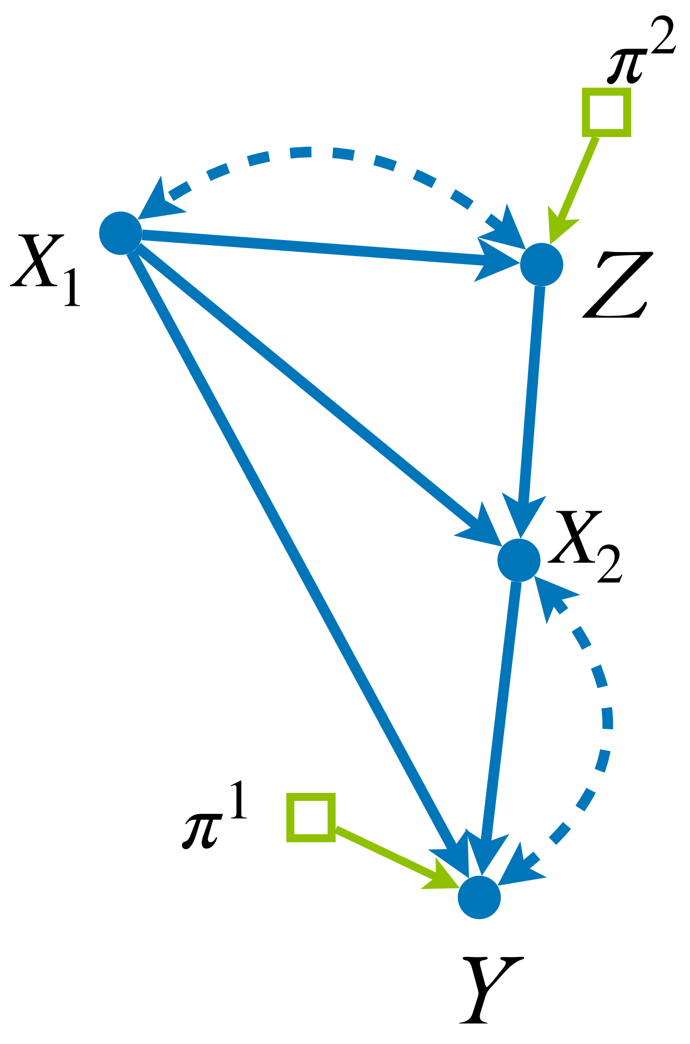
Solving the instance with σ-calculus

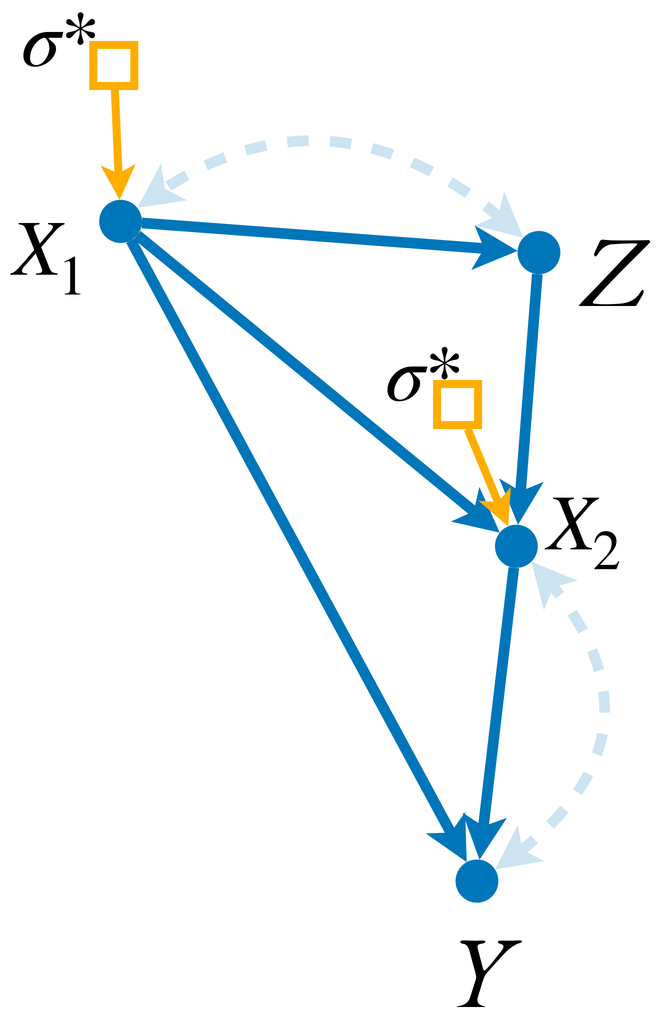
$$P^*(y;\sigma^*) = \sum_{x_1,z,x_2} P^*(y \mid x_1, z, x_2; \sigma^*) P^*(x_2 \mid z, x_1; \sigma^*) P^*(z \mid x_1; \sigma^*) P^*(x_1; \sigma^*)$$

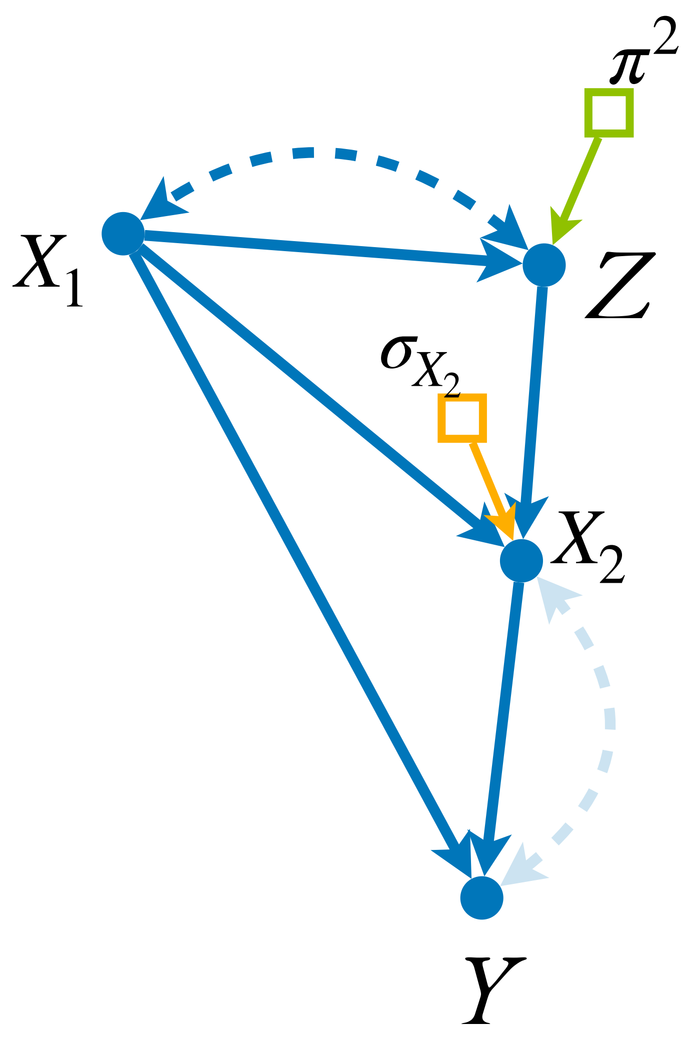
$P^*(y \mid x_1, z, x_2; \sigma^*) = P^*(y \mid x_1, x_2; \sigma^*)$ by rule 1 remove observation on Z

$= P^*(y \mid x_1, x_2; \sigma_{X_2})$ by rule 2 change to intervention σ_2

$= P^2(y \mid x_1, x_2; \sigma_{X_2})$ $(Z \perp \square_{\pi_2} \mid X_1, X_2)$ in $\mathcal{G}_{\sigma_{X_2}}$ change domain







Solving the instance with σ-calculus

$$P^*(y;\sigma^*) = \sum_{x_1,z,x_2} P^*(y \mid x_1, z, x_2; \sigma^*) P^*(x_2 \mid z, x_1; \sigma^*) P^*(z \mid x_1; \sigma^*) P^*(x_1; \sigma^*)$$

$$= \sum_{x_1,z,x_2} P^2(y \mid x_1, x_2; \sigma_2) \hat{P}(x_2 \mid z, x_1) P^1(z \mid x_1; \sigma_{X_1}) \hat{P}(x_1)$$

Each term is estimable from one of the domains π^1, π^2 or given by the target intervention σ^*.

Theorem: σ-calculus and basic probability axioms are sound and complete for the σ-TR task.

Algorithmic Solution

Factorization in the Presence of Latents: Confounded Factors

- In the absence of bidirected edges, a query or input distribution always decomposes into factors of the for $P(v_i \mid pa_i)$.
- With latent confounding, we can define coarser factors taking the hidden variables into account, called **c-factors** (confounded factors, Tian&Pearl01).
- Let $\mathbf{C} \subseteq \mathbf{V}$, then the c-factor associated with $\mathbf{C}$ is given by the following function:

$$Q[\mathbf{C}](\mathbf{c}, pa_{\mathbf{c}}) = \sum_{u(\mathbf{C})} P(u(\mathbf{C})) \prod_{V_i \in \mathbf{C}} P(v_i \mid pa_i, u_i), \quad \text{where } U(\mathbf{C}) = \bigcup_{V_i \in \mathbf{C}} \mathbf{U}_i.$$

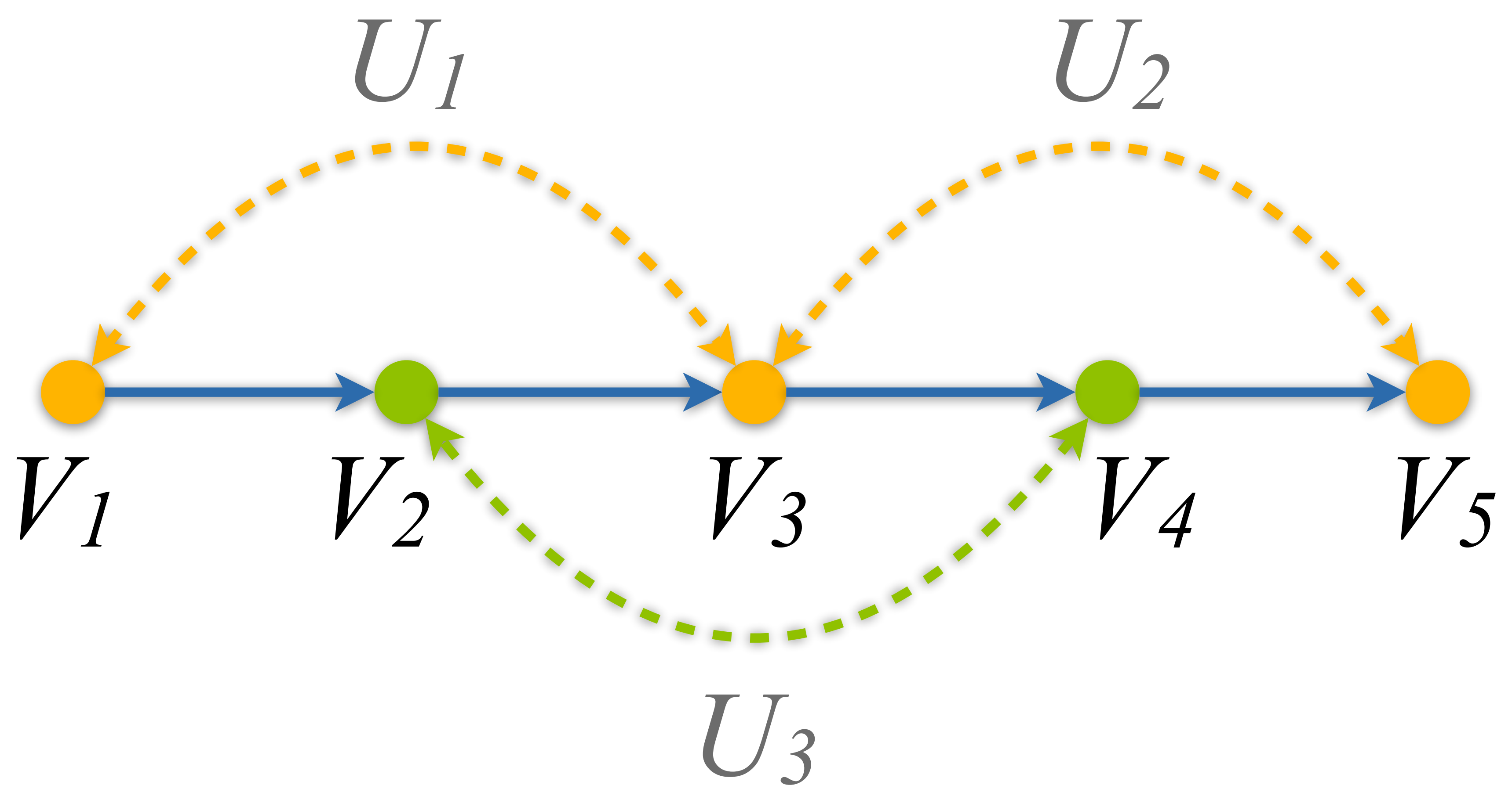
For simplicity, $Q[\mathbf{C}](\mathbf{c}, \mathbf{pa_c})$ is often written as $Q[\mathbf{C}]$.

Also, notice that $Q[\mathbf{V}] = P(\mathbf{v})$.

Confounded Components (C-Components)

[Tian&Pearl02]

- Definition (c-component). Two variables are in the same c-component if and only if they are connected by a bidirected path, a path composed entirely of bidirected edges.



- V_1, V_3 and V_5 are in the same c-component.
- V_2 and V_4 are in the same c-component.

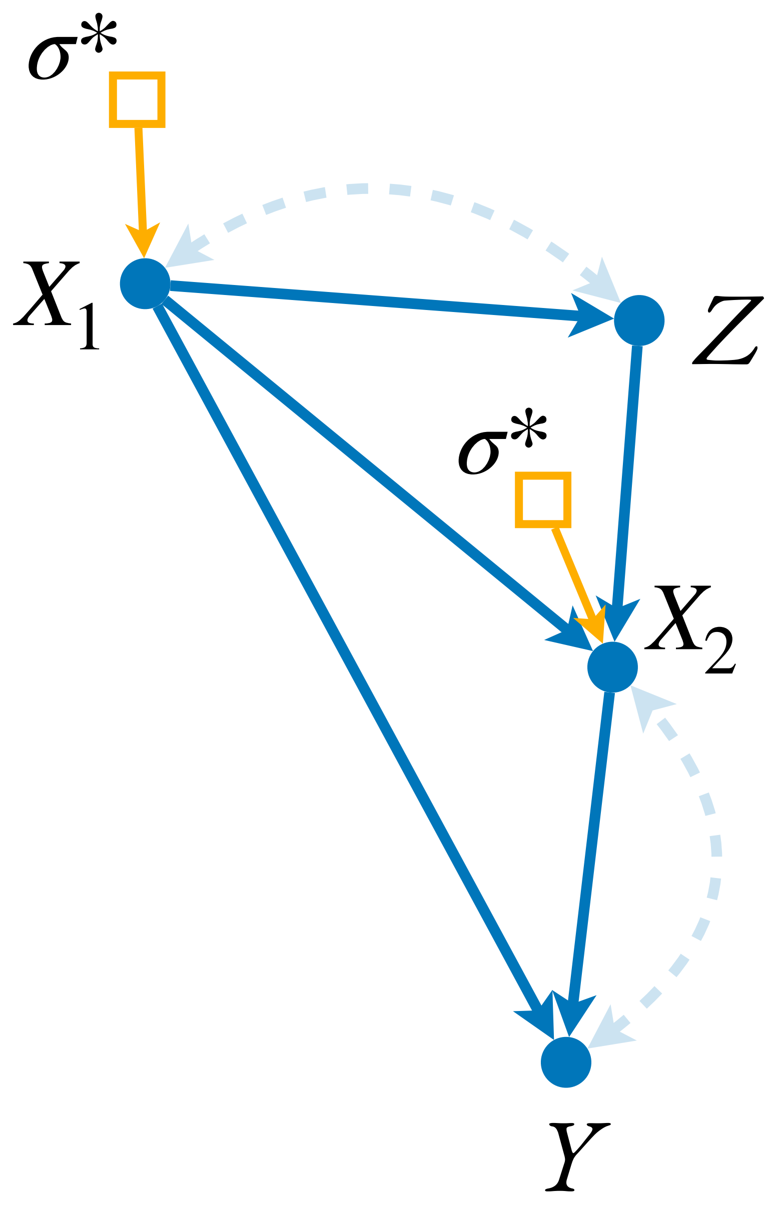
Solving the instance with σ-TR (algorithm)

- Target c-factors:

$$P^*(y;\sigma^*) = \sum_{x_1,z,x_2} P^*(\mathbf{v};\sigma^*)$$

$$= \sum_{x_1,z,x_2} Q^*[\mathbf{V};\sigma^*]$$

$$= \sum_{x_1,z,x_2} Q^*[Y;\sigma^*]Q^*[Z;\sigma^*]Q^*[X_2;\sigma^*]Q^*[X_1;\sigma^*]$$

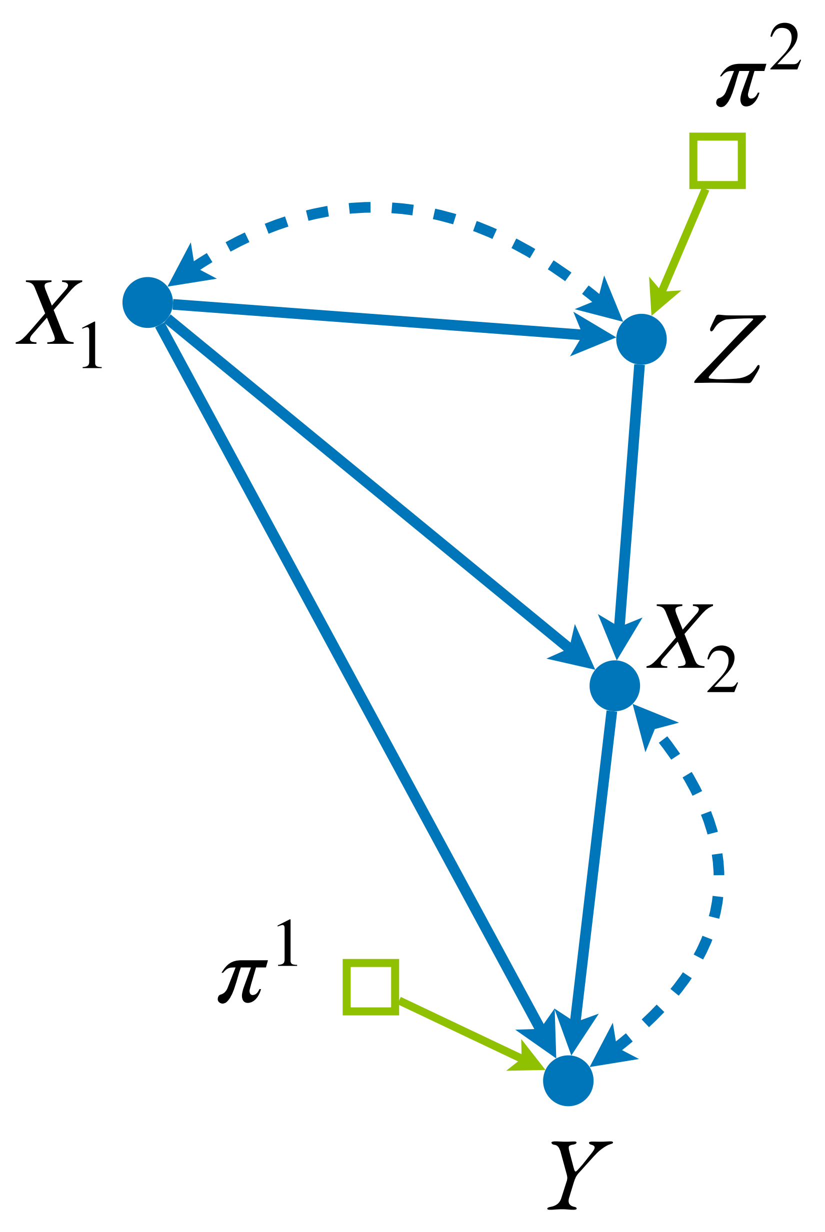


We try to get these c-factors from the available data

Given by the intervention σ^* as $\hat{P}(x_2 \mid z, x_1)\hat{P}(x_1)$

Transportability of c-factor

- A c-factor $Q^a[\mathbf{C}]$ is transportable from domain π^a to domain π^* if no variable in $\mathbf{C}$ is pointed by a square node corresponding to domain π^a.
- There is no need to use d-separation to determine transportability due to the canonical form of the c-factor.



In this example

- $Q^*[Y; \sigma^*]$ is transportable from π^2
- $Q^*[Z; \sigma^*]$ is transportable from π^1

Still, we need to identify the corresponding $Q^2[Y; \sigma^*]$ and $Q^1[Z; \sigma^*]$ from the distributions available in those domains.

Identifying the c-factors in each domain

For this examples we can show

> **Theorem:** An effect $P(\mathbf{y} \mid \mathbf{w}; \sigma_{\mathbf{X}})$ is transportable from a combination of observations an experiments $\mathbb{Z}$ and a selection diagram $\mathcal{G}^{\Delta}$ if and only if σ-TR outputs an estimand for it. The algorithm takes $O(n^2(n+m)p)$ time to output an expression or fail, where $n = |\mathbf{V}|$, m is the number of edges of $\mathcal{G}$ and $p = |\mathbb{Z}|$

- $Q^*[Z; \sigma^*] = Q^1[Z; \sigma^*]$
- $Q^*[Y; \sigma^*] = Q^2[Y; \sigma^*]$

Then, the query can be expressed as

$$P^*(y; \sigma^*) = \sum_{x_1, z, x_2} Q^*[Y; \sigma^*] Q^*[Z; \sigma^*] Q^*[X_2; \sigma^*] Q^*[X_1; \sigma^*]$$

$$= \sum_{x_1, z, x_2} P^2(y \mid x_1, x_2, z; \sigma_{X_2}) P^1(z \mid x_1; \sigma_{X_1, X_2}) \hat{P}(x_2 \mid z, x_1) \hat{P}(x_1)$$

Conclusions

- The problem of assessing the effect of a policy in a target domain, using a combination of observational and experimental data from multiple domains, can be solved non-parametrically with the help of a selection diagram encoding the assumptions about the causal mechanism and the differences between the domains
- We provide a necessary and sufficient graphical condition that characterizes the existence of an unbiased estimator for the effect of a target policy (possibly stochastic) given assumption in the form of a diagram and heterogeneous datasets.
- We develop a sound and complete algorithm (σ-TR) to efficiently determine whether the transport formula exists, and output an unbiased estimator of the corresponding transport formula (whenever it exists).
- We prove that σ-calculus, are complete for this task.

Thank you!

Questions?