Marginalization and Reduction of Structural Causal Models

A Functional Approach

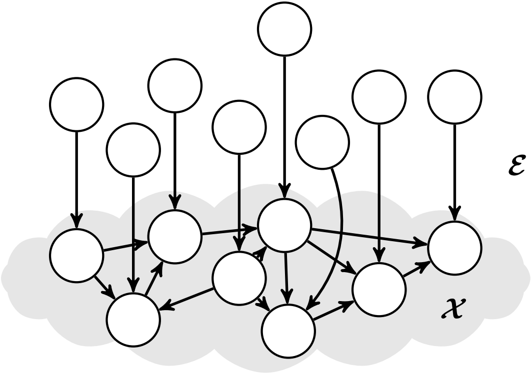
Stephan Bongers

Joris Mooij, Jonas Peters, Bernhard Schölkopf

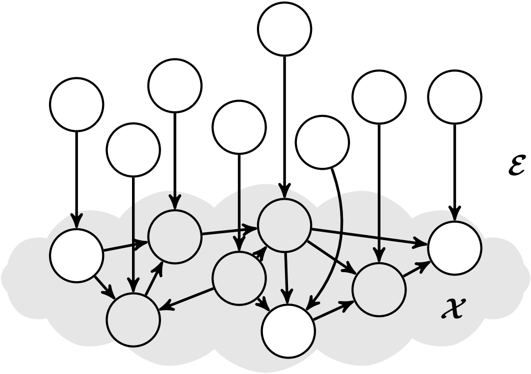
University of Amsterdam

May 10, 2016

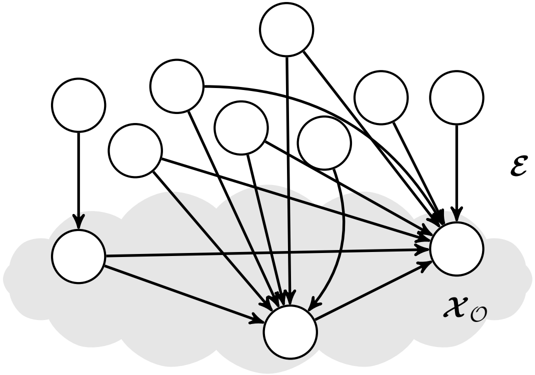
Motivation



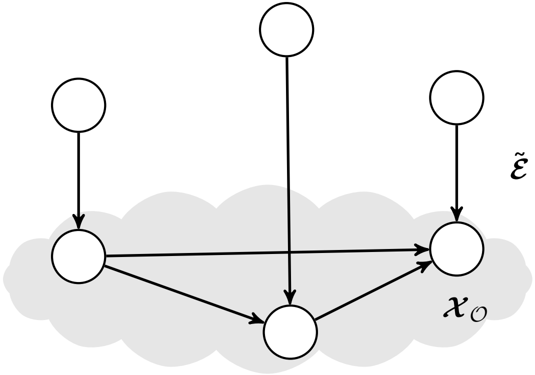
Motivation



Motivation



Motivation



Outline

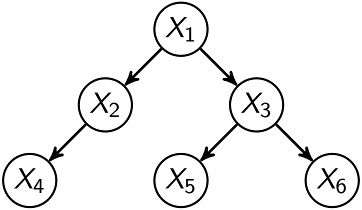
Introduction: Causal Modelling

Introduction: Causal Modelling

Causal Network

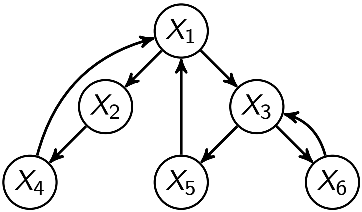
Introduction: Causal Modelling

Causal Bayesian Network



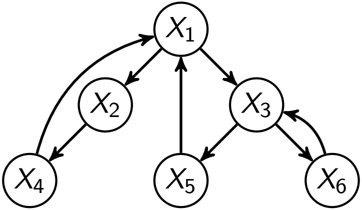
Introduction: Causal Modelling

Causal Bayesian Network



Introduction: Causal Modelling

Causal ~~Bayesian~~ Network

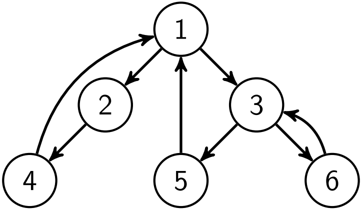


We forget

- Bayesian Network

Introduction: Causal Modelling

Causal ~~Bayesian~~ Network

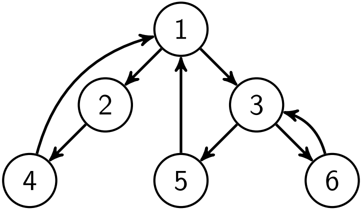


We forget

- Bayesian Network
- Random Variables

Introduction: Causal Modelling

Causal ~~Bayesian~~ Network



We forget

- Bayesian Network
- Random Variables
- Underlying graph

Introduction: Causal Modelling

Causal ~~Bayesian~~ Network

We forget

- Bayesian Network
- Random Variables
- Underlying graph

This make everything much clearer!

Introduction: Causal Modelling

Causal ~~Bayesian~~ Network

We forget

- Bayesian Network
- Random Variables
- Underlying graph

This make everything much clearer!

So, what do we get . . .

Structural Causal Models

Structural Causal Models

$$\boldsymbol{\mathcal{X}} = \prod_{i \in \mathcal{I}} \boldsymbol{\mathcal{X}}_i$$

'Endogenous variables'

Structural Causal Models

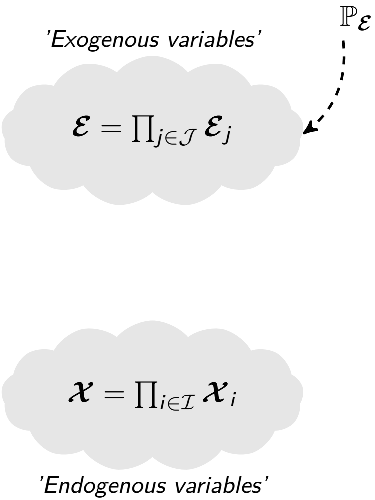
'Exogenous variables'

$$\boldsymbol{\mathcal{E}} = \prod_{j \in \mathcal{J}} \boldsymbol{\mathcal{E}}_j$$

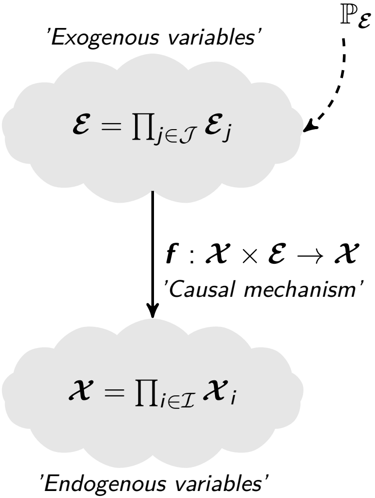
$$\boldsymbol{\mathcal{X}} = \prod_{i \in \mathcal{I}} \boldsymbol{\mathcal{X}}_i$$

'Endogenous variables'

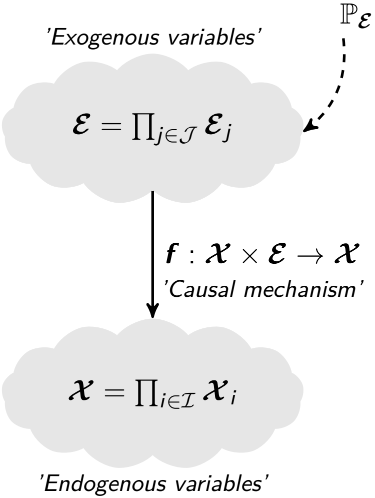
Structural Causal Models



Structural Causal Models



Structural Causal Models



Define a SCM as:

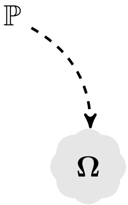
$$\mathcal{M} := \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$$

"f_i may depend on x_i"
(i.e. we allow self-loops)

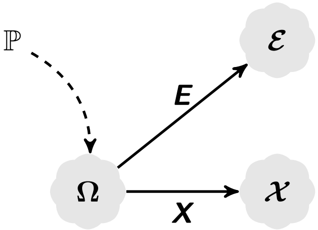
Structural Causal Models: Solutions

Structural Causal Models: Solutions

Structural Causal Models: Solutions

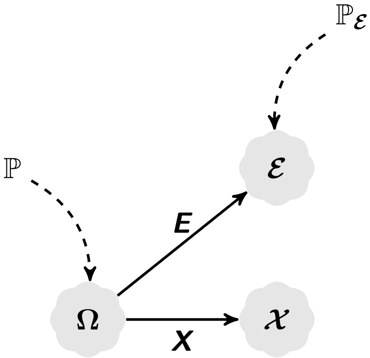


Structural Causal Models: Solutions



Random variables $(\boldsymbol{E}, \boldsymbol{X})$

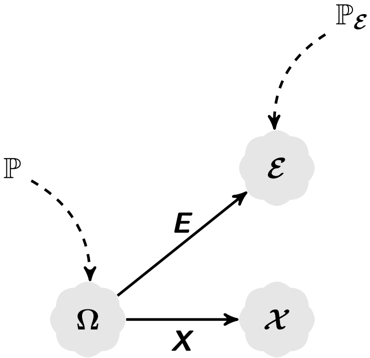
Structural Causal Models: Solutions



Random variables $(\boldsymbol{E}, \boldsymbol{X})$ are a *solution* of $\mathcal{M}$ if

- $\mathbb{P}^{\boldsymbol{E}} = \mathbb{P}_{\mathcal{E}}$
- Structural Equation: $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{E})$ a.s.

Structural Causal Models: Solutions



Random variables $(\boldsymbol{E}, \boldsymbol{X})$ are a *solution* of $\mathcal{M}$ if

- $\mathbb{P}^{\boldsymbol{E}} = \mathbb{P}_{\mathcal{E}}$
- Structural Equation: $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{E})$ a.s.

Remark: Two different SCMs can have the same solution space.

Structural Causal Models: Graphical Representation

Structural Causal Models: Graphical Representation

$\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$

where $(\boldsymbol{E}, \boldsymbol{X})$ is a solution such that $\boldsymbol{E}$ is jointly independent.

Structural Causal Models: Graphical Representation

$\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\mathcal{E}} \rangle$

where $(\boldsymbol{E}, \boldsymbol{X})$ is a solution such that $\boldsymbol{E}$ is jointly independent.

$\Longrightarrow$

Augmented directed graph $\mathcal{G}^a_{\mathcal{M}}$:

- Nodes $\mathcal{I} \cup \mathcal{J}$
- $i \to j$ if i is a direct cause of j (i.e. f_j depends on x_i/e_i).

Structural Causal Models: Graphical Representation

$\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$

where $(\boldsymbol{E}, \boldsymbol{X})$ is a solution such that $\boldsymbol{E}$ is jointly independent.

$\Longrightarrow$

Augmented directed graph $\mathcal{G}^a_{\mathcal{M}}$:

- Nodes $\mathcal{I} \cup \mathcal{J}$
- $i \to j$ if i is a direct cause of j (i.e. f_j depends on x_i/e_i).

Example (Acyclic SCM)

$$f_1(E_1) = E_1$$
$$f_2(X_1, E_2) = X_1 + E_2$$
$$f_3(X_1, E_3) = X_1 + E_3$$

$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

Structural Causal Models: Graphical Representation

$$\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$$

where $(\boldsymbol{E}, \boldsymbol{X})$ is a solution such that $\boldsymbol{E}$ is jointly independent.

$\Longrightarrow$

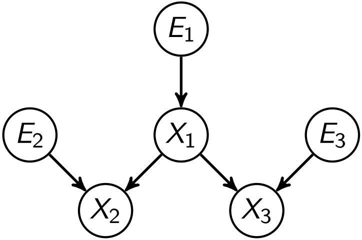
Augmented directed graph $\mathcal{G}^a_{\mathcal{M}}$:

- Nodes $\mathcal{I} \cup \mathcal{J}$
- $i \to j$ if i is a direct cause of j (i.e. f_j depends on x_i/e_i).

Example (Acyclic SCM)

$$f_1(E_1) = E_1$$
$$f_2(X_1, E_2) = X_1 + E_2$$
$$f_3(X_1, E_3) = X_1 + E_3$$

$\Longrightarrow$



$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

Structural Causal Models: Graphical Representation

$\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$

where $(\boldsymbol{E}, \boldsymbol{X})$ is a solution such that $\boldsymbol{E}$ is jointly independent.

$\Longrightarrow$

Augmented directed graph $\mathcal{G}^a_{\mathcal{M}}$:

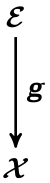
- Nodes $\mathcal{I} \cup \mathcal{J}$
- $i \to j$ if i is a direct cause of j (i.e. f_j depends on x_i/e_i).

Example (Acyclic SCM)

$$f_1(E_1) = E_1$$
$$f_2(X_1, E_2) = X_1 + E_2$$
$$f_3(X_1, E_3) = X_1 + E_3$$

$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

$\Longrightarrow$



Structural Causal Models: Graphical Representation

$\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$

where $(\boldsymbol{E}, \boldsymbol{X})$ is a solution such that $\boldsymbol{E}$ is jointly independent.

$\Longrightarrow$

Augmented directed graph $\mathcal{G}^a_{\mathcal{M}}$:

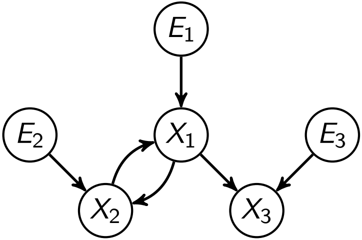
- Nodes $\mathcal{I} \cup \mathcal{J}$
- $i \to j$ if i is a direct cause of j (i.e. f_j depends on x_i/e_i).

Example (Cyclic SCM)

$$f_1(E_1) = X_2 + E_1$$
$$f_2(X_1, E_2) = X_1 + E_2$$
$$f_3(X_1, E_3) = X_1 + E_3$$

$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

$\Longrightarrow$



Structural Causal Models: Graphical Representation

$\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$

where $(\boldsymbol{E}, \boldsymbol{X})$ is a solution such that $\boldsymbol{E}$ is jointly independent.

$\Longrightarrow$

Augmented directed graph $\mathcal{G}^a_{\mathcal{M}}$:

- Nodes $\mathcal{I} \cup \mathcal{J}$
- $i \to j$ if i is a direct cause of j (i.e. f_j depends on x_i / e_i).

Example (Cyclic SCM)

$$f_1(E_1) = X_2 + E_1$$
$$f_2(X_1, E_2) = X_1 + E_2$$
$$f_3(X_1, E_3) = X_1 + E_3$$

$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

$\Longrightarrow$

No such mapping

$\boldsymbol{\mathcal{E}}$
$\downarrow \boldsymbol{g}$
$\boldsymbol{\mathcal{X}}$

Structural Causal Models: Solvability

Structural Causal Models: Solvability

We call $\mathcal{M}$ *solvable* if for $\mathbb{P}_{\mathcal{E}}$-almost every $\boldsymbol{e}$

$$\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e})$$

has a unique solution $\boldsymbol{x}$.

Structural Causal Models: Solvability

We call $\mathcal{M}$ *solvable* if for $\mathbb{P}_{\mathcal{E}}$-almost every $\boldsymbol{e}$

$$\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e})$$

has a unique solution $\boldsymbol{x}$.

Proposition

A solvable SCM $\mathcal{M}$ induces a (unique) mapping $\boldsymbol{g} : \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}$ such that for $\mathbb{P}_{\mathcal{E}}$-almost every $\boldsymbol{e}$

$$\boldsymbol{g}(\boldsymbol{e}) = \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{e}), \boldsymbol{e}),$$

which induces a distribution on $\boldsymbol{\mathcal{X}}$

Structural Causal Models: Solvability

We call $\mathcal{M}$ *solvable* if for $\mathbb{P}_{\mathcal{E}}$-almost every $\boldsymbol{e}$

$$\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e})$$

has a unique solution $\boldsymbol{x}$.

Proposition

A solvable SCM $\mathcal{M}$ induces a (unique) mapping $\boldsymbol{g} : \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}$ such that for $\mathbb{P}_{\mathcal{E}}$-almost every $\boldsymbol{e}$

$$\boldsymbol{g}(\boldsymbol{e}) = \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{e}), \boldsymbol{e}),$$

which induces a distribution on $\boldsymbol{\mathcal{X}}$

Every solvable SCM has a solution and every solvable SCM induces a unique distribution.

Interventions

Interventions

Perfect intervention on $I \subseteq \mathcal{I}$

$$\mathcal{M} \xrightarrow{\mathrm{do}(I, \boldsymbol{\xi}_I)} \mathcal{M}_{\mathrm{do}(I, \boldsymbol{\xi}_I)} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \tilde{\boldsymbol{f}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$$

where

$$\tilde{f}_i(\boldsymbol{x}, \boldsymbol{e}) := \begin{cases} \xi_i & i \in I \\ f_i(\boldsymbol{x}, \boldsymbol{e}) & i \in \mathcal{I} \setminus I. \end{cases}$$

Interventions

Perfect intervention on $I \subseteq \mathcal{I}$

$$\mathcal{M} \xrightarrow{\mathrm{do}(I, \boldsymbol{\xi}_I)} \mathcal{M}_{\mathrm{do}(I, \boldsymbol{\xi}_I)} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \tilde{\boldsymbol{f}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$$

where

$$\tilde{f}_i(\boldsymbol{x}, \boldsymbol{e}) := \begin{cases} \xi_i & i \in I \\ f_i(\boldsymbol{x}, \boldsymbol{e}) & i \in \mathcal{I} \setminus I. \end{cases}$$

Proposition

1. *Disjoint perfect interventions commute.*
2. *Acyclicity is preserved under perfect intervention.*
3. *Solvability is, in general, not preserved under perfect intervention.*

Equivalence relations between SCMs

Equivalence relations between SCMs

We are mainly interested in the induced distribution on the endogenous variables!

Equivalence relations between SCMs

We are mainly interested in the induced distribution on the endogenous variables!

Two SCMs $\mathcal{M}$ and $\tilde{\mathcal{M}}$ are called

1. *Observationally equivalent*: if $\mathcal{M}$ and $\tilde{\mathcal{M}}$ are
   - both not solvable, or
   - both solvable and induce the same (observational) distribution on $\boldsymbol{\mathcal{X}}_{\mathcal{O}}$.

w.r.t $\mathcal{O} \subseteq \mathcal{I} \cap \tilde{\mathcal{I}}$.

Equivalence relations between SCMs

We are mainly interested in the induced distribution on the endogenous variables!

Two SCMs $\mathcal{M}$ and $\tilde{\mathcal{M}}$ are called

1. *Observationally equivalent*: if $\mathcal{M}$ and $\tilde{\mathcal{M}}$ are
   - both not solvable, or
   - both solvable and induce the same (observational) distribution on $\boldsymbol{\mathcal{X}}_{\mathcal{O}}$.
2. *Interventionally equivalent*: if for any $I \subseteq \mathcal{O}$ and value $\boldsymbol{\xi}_I$, $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ and $\tilde{\mathcal{M}}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ are observationally equivalent w.r.t. $\mathcal{O}$.

w.r.t $\mathcal{O} \subseteq \mathcal{I} \cap \tilde{\mathcal{I}}$.

Marginalization: acyclic SCMs - Substitution interpretation

Marginalization: acyclic SCMs - Substitution interpretation

For an acyclic SCM $\mathcal{M}$ we can define the marginalization as:

$$\mathcal{M} \xrightarrow{\text{marg}(\ell)} \mathcal{M}_{\text{marg}(\ell)} = \langle \mathcal{O}, \mathcal{J}, \boldsymbol{\mathcal{X}}_{\mathcal{I}\setminus\{\ell\}}, \boldsymbol{\mathcal{E}}, \tilde{\boldsymbol{f}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$$

where $\tilde{\boldsymbol{f}} : \boldsymbol{\mathcal{X}}_{\mathcal{I}\setminus\{\ell\}} \times \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}_{\mathcal{I}\setminus\{\ell\}}$ is defined by

$$\tilde{\boldsymbol{f}}(\boldsymbol{x}_{\mathcal{I}\setminus\{\ell\}}, \boldsymbol{e}) := \boldsymbol{f}_{\mathcal{I}\setminus\{\ell\}}(\boldsymbol{x}_{\mathcal{I}\setminus\{\ell\}}, \boldsymbol{f}_{\ell}(\boldsymbol{x}_{\mathcal{I}\setminus\{\ell\}}, \boldsymbol{e}), \boldsymbol{e})$$

Marginalization: acyclic SCMs - Substitution interpretation

For an acyclic SCM $\mathcal{M}$ we can define the marginalization as:

$$\mathcal{M} \xrightarrow{\text{marg}(\ell)} \mathcal{M}_{\text{marg}(\ell)} = \langle \mathcal{O}, \mathcal{J}, \boldsymbol{\mathcal{X}}_{\mathcal{I}\setminus\{\ell\}}, \boldsymbol{\mathcal{E}}, \tilde{\boldsymbol{f}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$$

where $\tilde{\boldsymbol{f}} : \boldsymbol{\mathcal{X}}_{\mathcal{I}\setminus\{\ell\}} \times \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}_{\mathcal{I}\setminus\{\ell\}}$ is defined by

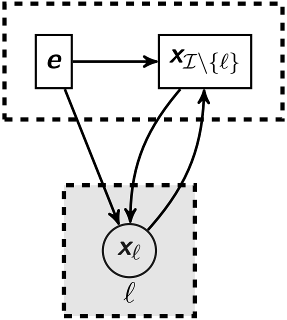
$$\tilde{\boldsymbol{f}}(\boldsymbol{x}_{\mathcal{I}\setminus\{\ell\}}, \boldsymbol{e}) := \boldsymbol{f}_{\mathcal{I}\setminus\{\ell\}}(\boldsymbol{x}_{\mathcal{I}\setminus\{\ell\}}, \boldsymbol{f}_\ell(\boldsymbol{x}_{\mathcal{I}\setminus\{\ell\}}, \boldsymbol{e}), \boldsymbol{e})$$

Here we used:

$$\boldsymbol{x}_\ell = \boldsymbol{f}_\ell(\boldsymbol{x}_{\mathcal{I}\setminus\{\ell\}}, \boldsymbol{e})$$

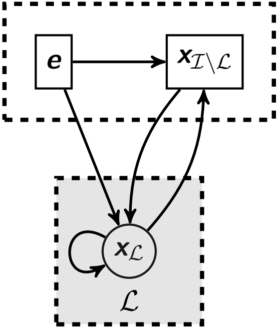
Marginalization: Subsystem interpretation

Marginalization: Subsystem interpretation



Acyclic Case

Marginalization: Subsystem interpretation



General Case: for $\mathcal{L} \subseteq \mathcal{I}$

Marginalization: Partial Solvability

Marginalization: Partial Solvability

We call $\mathcal{M}$ *partially solvable w.r.t.* $\mathcal{L} \subseteq \mathcal{I}$ if for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e}$ for all $\boldsymbol{\xi}_{\mathcal{O}} \in \boldsymbol{\mathcal{X}}_{\mathcal{O}}$

$$\boldsymbol{x}_{\mathcal{L}} = \boldsymbol{f}_{\mathcal{L}}(\boldsymbol{\xi}_{\mathcal{O}}, \boldsymbol{x}_{\mathcal{L}}, \boldsymbol{e})$$

has a unique solution $\boldsymbol{x}_{\mathcal{L}} \in \boldsymbol{\mathcal{X}}_{\mathcal{L}}$.

Marginalization: Partial Solvability

We call $\mathcal{M}$ *partially solvable w.r.t.* $\mathcal{L} \subseteq \mathcal{I}$ if for $\mathbb{P}_{\mathcal{E}}$-almost every $\boldsymbol{e}$ for all $\boldsymbol{\xi}_{\mathcal{O}} \in \boldsymbol{\mathcal{X}}_{\mathcal{O}}$

$$\boldsymbol{x}_{\mathcal{L}} = \boldsymbol{f}_{\mathcal{L}}(\boldsymbol{\xi}_{\mathcal{O}}, \boldsymbol{x}_{\mathcal{L}}, \boldsymbol{e})$$

has a unique solution $\boldsymbol{x}_{\mathcal{L}} \in \boldsymbol{\mathcal{X}}_{\mathcal{L}}$.

Proposition

If $\mathcal{M}$ is partially solvable w.r.t. $\mathcal{L}$, then this induces a (unique) mapping $\boldsymbol{g}_{\mathcal{L}} : \boldsymbol{\mathcal{X}}_{\mathcal{O}} \times \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}_{\mathcal{L}}$ such that for $\mathbb{P}_{\mathcal{E}}$-almost every $\boldsymbol{e}$ we have

$$\forall_{\boldsymbol{x}_{\mathcal{O}} \in \boldsymbol{\mathcal{X}}_{\mathcal{O}}} : \quad \boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}) = \boldsymbol{f}_{\mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}), \boldsymbol{e}).$$

Marginalization: Marginal SCMs

Marginalization: Marginal SCMs

For a $\mathcal{M}$ that is partially solvable w.r.t. $\mathcal{L}$ we can define the marginalization as:

$$\mathcal{M} \xrightarrow{\operatorname{marg}(\mathcal{L})} \mathcal{M}_{\operatorname{marg}(\mathcal{L})} = \langle \mathcal{O}, \mathcal{J}, \boldsymbol{\mathcal{X}}_{\mathcal{O}}, \boldsymbol{\mathcal{E}}, \tilde{\boldsymbol{f}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$$

where $\tilde{\boldsymbol{f}} : \boldsymbol{\mathcal{X}}_{\mathcal{O}} \times \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ is defined by

$$\tilde{\boldsymbol{f}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}) := \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}), \boldsymbol{e}).$$

where $\boldsymbol{g}_{\mathcal{L}} : \boldsymbol{\mathcal{X}}_{\mathcal{O}} \times \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}_{\mathcal{L}}$ is the mappings defined by

$$\boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}) = \boldsymbol{f}_{\mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}), \boldsymbol{e})$$

for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e}$.

Marginalization

Marginalization

Proposition

1. *Marginalizing separately over two disjoint subsets of endogenous variables commute, whenever defined.*
2. *Marginalizing over endogenous variables* $\mathcal{L}$ *commutes with a perfect intervention on* $I \subseteq \mathcal{I} \setminus \mathcal{L}$.
3. $\mathcal{M}$ *and* $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ *(if defined) are observationally equivalent w.r.t.* $\mathcal{I} \setminus \mathcal{L}$.

Marginalization

Proposition

1. *Marginalizing separately over two disjoint subsets of endogenous variables commute, whenever defined.*
2. *Marginalizing over endogenous variables $\mathcal{L}$ commutes with a perfect intervention on $I \subseteq \mathcal{I} \setminus \mathcal{L}$.*
3. *$\mathcal{M}$ and $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ (if defined) are observationally equivalent w.r.t. $\mathcal{I} \setminus \mathcal{L}$.*

Theorem

$\mathcal{M}$ and $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ (if defined) are interventionally equivalent w.r.t. $\mathcal{I} \setminus \mathcal{L}$.

Example: Cyclic Linear Gaussian Case (Hyttinen et al.)

Example: Cyclic Linear Gaussian Case (Hyttinen et al.)

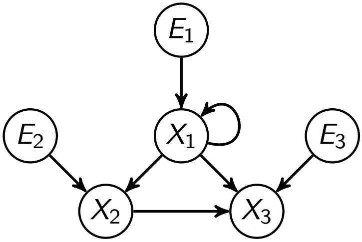
$\mathcal{M} = \langle \{3\}, \{3\}, \mathbb{R}^3, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3} \rangle$

$$
\begin{aligned}
f_1(X_1, E_1) &= \frac{1}{2} X_1 + \frac{1}{2} E_1 \\
f_2(X_1, E_2) &= X_1 + E_2 \\
f_3(X_1, X_2, E_3) &= X_1 + X_2 + E_3 \\
\boldsymbol{E} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})
\end{aligned}
$$

Example: Cyclic Linear Gaussian Case (Hyttinen et al.)

$\mathcal{M} = \langle \{3\}, \{3\}, \mathbb{R}^3, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3} \rangle$

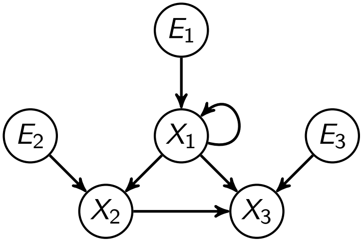
$$f_1(X_1, E_1) = \frac{1}{2}X_1 + \frac{1}{2}E_1$$
$$f_2(X_1, E_2) = X_1 + E_2$$
$$f_3(X_1, X_2, E_3) = X_1 + X_2 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$



Example: Cyclic Linear Gaussian Case (Hyttinen et al.)

$$\mathcal{M} = \langle \{3\}, \{3\}, \mathbb{R}^3, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3} \rangle$$

$$f_1(X_1, E_1) = \frac{1}{2} X_1 + \frac{1}{2} E_1$$
$$f_2(X_1, E_2) = X_1 + E_2$$
$$f_3(X_1, X_2, E_3) = X_1 + X_2 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

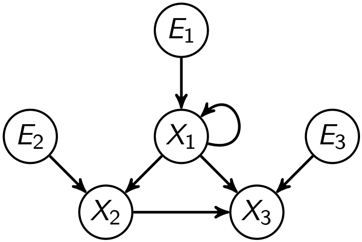


$$\mathcal{M}_{\mathrm{marg}(1)} = \langle \{2\}, \{3\}, \mathbb{R}^2, \mathbb{R}^3, \tilde{\boldsymbol{f}}, \mathbb{P}_{\mathbb{R}^3} \rangle$$

$$\tilde{f}_2(E_1, E_2) = E_1 + E_2$$
$$\tilde{f}_3(X_2, E_1, E_3) = X_2 + E_1 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

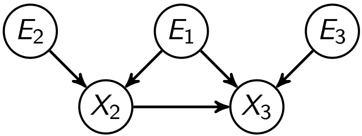
Example: Cyclic Linear Gaussian Case (Hyttinen et al.)

$$\mathcal{M} = \langle \{3\}, \{3\}, \mathbb{R}^3, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3} \rangle$$



$$f_1(X_1, E_1) = \frac{1}{2} X_1 + \frac{1}{2} E_1$$
$$f_2(X_1, E_2) = X_1 + E_2$$
$$f_3(X_1, X_2, E_3) = X_1 + X_2 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

$$\mathcal{M}_{\mathrm{marg}(1)} = \langle \{2\}, \{3\}, \mathbb{R}^2, \mathbb{R}^3, \tilde{\boldsymbol{f}}, \mathbb{P}_{\mathbb{R}^3} \rangle$$

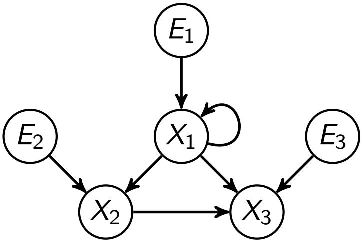


$$\tilde{f}_2(E_1, E_2) = E_1 + E_2$$
$$\tilde{f}_3(X_2, E_1, E_3) = X_2 + E_1 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

Example: Cyclic Linear Gaussian Case (Hyttinen et al.)

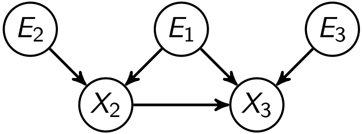
$\mathcal{M} = \langle\{3\}, \{3\}, \mathbb{R}^3, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3}\rangle$

$$f_1(X_1, E_1) = \frac{1}{2}X_1 + \frac{1}{2}E_1$$
$$f_2(X_1, E_2) = X_1 + E_2$$
$$f_3(X_1, X_2, E_3) = X_1 + X_2 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$



$\mathcal{M}_{\mathrm{marg}(1)} = \langle\{2\}, \{3\}, \mathbb{R}^2, \mathbb{R}^3, \tilde{\boldsymbol{f}}, \mathbb{P}_{\mathbb{R}^3}\rangle$

$$\tilde{f}_2(E_1, E_2) = E_1 + E_2$$
$$\tilde{f}_3(X_2, E_1, E_3) = X_2 + E_1 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$



Can we somehow reduce the space of exogenous variables?

Example: Cyclic Linear Gaussian Case

Answer: Yes!

Example: Cyclic Linear Gaussian Case

Answer: Yes!

$$\tilde{\mathcal{M}} = \langle \{2\}, \{3\}, \mathbb{R}^2, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3} \rangle$$

$$\begin{aligned} f_2(E_1, E_2) &= E_1 + E_2 \\ f_3(X_2, E_1, E_3) &= X_2 + E_1 + E_3 \\ \boldsymbol{E} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) \end{aligned}$$

Example: Cyclic Linear Gaussian Case

Answer: Yes!

$$\tilde{\mathcal{M}} = \langle \{2\}, \{3\}, \mathbb{R}^2, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3} \rangle$$

$$f_2(E_1, E_2) = E_1 + E_2$$
$$f_3(X_2, E_1, E_3) = X_2 + E_1 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

$$\implies$$

$$\tilde{\mathcal{M}}_{\mathrm{red}(\phi)} = \langle \{2\}, \{2\}, \mathbb{R}^2, \mathbb{R}^2, \tilde{\boldsymbol{f}}, \mathbb{P}_{\mathbb{R}^2} \rangle$$

$$\tilde{f}_2(\tilde{E}_1) = \tilde{E}_1$$
$$\tilde{f}_3(X_2, \tilde{E}_2) = X_2 + \tilde{E}_2$$

Example: Cyclic Linear Gaussian Case

Answer: Yes!

$$\tilde{\mathcal{M}} = \langle \{2\}, \{3\}, \mathbb{R}^2, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3} \rangle$$

$$f_2(E_1, E_2) = E_1 + E_2$$
$$f_3(X_2, E_1, E_3) = X_2 + E_1 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$$

$$\Longrightarrow$$

$$\tilde{\mathcal{M}}_{\mathrm{red}(\phi)} = \langle \{2\}, \{2\}, \mathbb{R}^2, \mathbb{R}^2, \tilde{\boldsymbol{f}}, \mathbb{P}_{\mathbb{R}^2} \rangle$$

$$\tilde{f}_2(\tilde{E}_1) = \tilde{E}_1$$
$$\tilde{f}_3(X_2, \tilde{E}_2) = X_2 + \tilde{E}_2$$

Here:

$$\phi : \mathbb{R}^3 \to \mathbb{R}^2$$
$$(E_1, E_2, E_3) \mapsto (E_1 + E_2, E_1 + E_3)$$

such that

$$\boldsymbol{f} = \tilde{\boldsymbol{f}} \circ (\mathbf{Id}_{\mathcal{X}}, \phi)$$

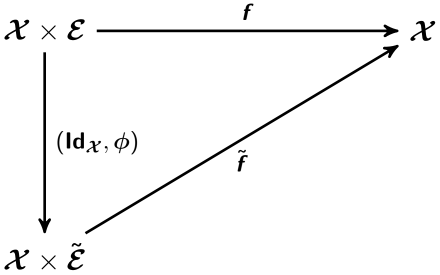
Reduction

Reduction

Reducing the number of exogenous variables by a transformation $\phi : \boldsymbol{\mathcal{E}} \to \tilde{\boldsymbol{\mathcal{E}}}$, called a *reduction*,

$$\mathcal{M} \xrightarrow{\mathsf{red}(\phi)} \mathcal{M}_{\mathsf{red}(\phi)} = \langle \mathcal{I}, \tilde{\mathcal{J}}, \boldsymbol{\mathcal{X}}, \tilde{\boldsymbol{\mathcal{E}}}, \tilde{\boldsymbol{f}}, \mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}} \rangle$$

holds if there exists for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e}$ we have the commuting diagram



Reduction

Reduction

Consider a reduction

$$\mathcal{M} \xrightarrow{\mathsf{red}(\phi)} \mathcal{M}_{\mathsf{red}(\phi)}$$

Reduction

Consider a reduction

$$\mathcal{M} \xrightarrow{\text{red}(\phi)} \mathcal{M}_{\text{red}(\phi)}$$

Proposition

If $\mathcal{M}$ is solvable, then

- $\mathcal{M}_{red(\phi)}$ *is also solvable, and*
- $\mathcal{M}$ *and* $\mathcal{M}_{red(\phi)}$ *are observationally equivalent,*
- *Reduction and perfect intervention of a SCM commute.*

Reduction

Consider a reduction

$$\mathcal{M} \xrightarrow{\text{red}(\phi)} \mathcal{M}_{\text{red}(\phi)}$$

Proposition

If $\mathcal{M}$ is solvable, then

- *$\mathcal{M}_{red(\phi)}$ is also solvable, and*
- *$\mathcal{M}$ and $\mathcal{M}_{red(\phi)}$ are observationally equivalent,*
- *Reduction and perfect intervention of a SCM commute.*

Theorem

$\mathcal{M}$ and $\mathcal{M}_{red(\phi)}$ are interventionally equivalent.

Reduction: Sufficient Condition

Reduction: Sufficient Condition

Given a SCM
$\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$

If $\phi : \boldsymbol{\mathcal{E}} \to \tilde{\boldsymbol{\mathcal{E}}}$ has a measurable right-sided inverse $\boldsymbol{\pi}$ (i.e., $\phi \circ \boldsymbol{\pi} = \mathbf{Id}_{\tilde{\boldsymbol{\mathcal{E}}}}$) such that

$$\boldsymbol{f} = \boldsymbol{f} \circ (\mathrm{Id}_{\boldsymbol{\mathcal{X}}}, \boldsymbol{\pi} \circ \phi)$$

Reduction: Sufficient Condition

Given a SCM
$\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$

If $\phi : \boldsymbol{\mathcal{E}} \to \tilde{\boldsymbol{\mathcal{E}}}$ has a measurable right-sided inverse $\boldsymbol{\pi}$ (i.e., $\phi \circ \boldsymbol{\pi} = \mathbf{Id}_{\tilde{\boldsymbol{\mathcal{E}}}}$) such that

$$\boldsymbol{f} = \boldsymbol{f} \circ (\mathrm{Id}_{\boldsymbol{\mathcal{X}}}, \boldsymbol{\pi} \circ \phi)$$

$\implies$ $\boldsymbol{f}_{\mathrm{red}(\phi)} := \boldsymbol{f} \circ (\mathrm{Id}_{\boldsymbol{\mathcal{X}}}, \boldsymbol{\pi})$ defines a reduced causal mechanism s.t.

$$\mathcal{M}_{\mathsf{red}(\phi)} = \langle \mathcal{I}, \tilde{\mathcal{J}}, \boldsymbol{\mathcal{X}}, \tilde{\boldsymbol{\mathcal{E}}}, \tilde{\boldsymbol{f}}, \mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}} \rangle$$

Reduction: Suprising Measure-Theoretic Fact

Reduction: Suprising Measure-Theoretic Fact

Theorem
Two standard measurable spaces are Borel-isomorphic if and only if they have the same cardinality.

Reduction: Suprising Measure-Theoretic Fact

Theorem

Two standard measurable spaces are Borel-isomorphic if and only if they have the same cardinality.

This implies: there exist for every $n \in \mathbb{N}$ a measurable isomorphism $\phi : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathcal{M} \xrightarrow{\mathrm{red}(\phi)} \mathcal{M}_{\mathrm{red}(\phi)} = \langle \mathcal{I}, \{1\}, \mathbb{R}^{\mathcal{I}}, \mathbb{R}, \tilde{\boldsymbol{f}}, \tilde{\mathbb{P}}_{\boldsymbol{\mathcal{E}}} \rangle$$

Reduction: Suprising Measure-Theoretic Fact

Theorem

Two standard measurable spaces are Borel-isomorphic if and only if they have the same cardinality.

This implies: there exist for every $n \in \mathbb{N}$ a measurable isomorphism $\phi : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathcal{M} \xrightarrow{\text{red}(\phi)} \mathcal{M}_{\text{red}(\phi)} = \langle \mathcal{I}, \{1\}, \mathbb{R}^{\mathcal{I}}, \mathbb{R}, \tilde{\boldsymbol{f}}, \tilde{\mathbb{P}}_{\boldsymbol{\mathcal{E}}} \rangle$$

This mapping is in general not so nice! What other mathematical structure do we need to make it suitable for statistical learning?

Reduction: Strictly Monotonic Reduction

For a SCM $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$, with $\boldsymbol{\mathcal{X}} = \mathbb{R}^{\mathcal{I}}$ and $\boldsymbol{\mathcal{E}} = \mathbb{R}^{\mathcal{J}}$, we call a reduction $\phi : \mathbb{R}^{\mathcal{J}} \to \mathbb{R}^{\mathcal{I}}$ *strictly monotonic* if for each $i \in \mathcal{I}$:

(i) for all $\boldsymbol{x} \in \mathbb{R}^{\mathcal{I}}$, the function $\tilde{e}_i \mapsto \tilde{f}_i(\boldsymbol{x}, \tilde{e}_i)$ is strictly monotonic

(ii) the cumulative density function of the i'th component, that is

$$F_i^{\tilde{\mathcal{M}}} : \tilde{e}_i \mapsto \mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}[A_1 \times \ldots \times A_{\tilde{\mathcal{J}}}]$$

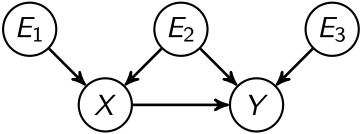
with $A_i = (-\infty, \tilde{e}_i)$ and $A_j = (-\infty, \infty)$ for all $j \neq i$, is strictly monotonic.

Reduction: Strictly Monotonic Reduction Example

Consider the following SCM

$$\mathcal{M} = \langle \{1, 2\}, \{1, 2, 3\}, \mathbb{R}^2, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3} \rangle$$

$$f_X(E_1, E_2) = E_1 + E_2$$
$$f_Y(X, E_2, E_3) = XE_2 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}).$$

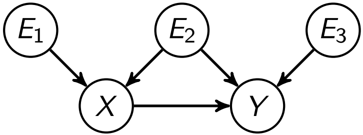


Reduction: Strictly Monotonic Reduction Example

Consider the following SCM

$$\mathcal{M} = \langle \{1,2\}, \{1,2,3\}, \mathbb{R}^2, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3} \rangle$$

$$f_X(E_1, E_2) = E_1 + E_2$$
$$f_Y(X, E_2, E_3) = XE_2 + E_3$$
$$\boldsymbol{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}).$$



Assuming the existence of a "strictly monotonic" reduction to $\mathbb{R}^2$ on $\mathcal{M}$ leads to a contradiction.

Outlook

Open questions:

1. What suitable mathematical structure do we need to impose besides measurability in order to make the reduction suitable for statistical learning?
2. What is the “right” graphical representation whenever the exogenous variables of a solution of a SCM are not jointly independent?
3. How does D-separation extends to the cyclic case?
4. . . .

Questions?

...