Identification in time series summary causal graphs

A. Meynaoui*, C. K. Assaad**, E. Devijver†, E. Gaussier†, and G. Gössler‡

** EasyVista
* Univ. Rennes 2, IRMAR
† Univ. Grenoble Alpes, CNRS, INP, LIG
‡ Univ. Grenoble Alpes, CNRS, INRIA, Grenoble INP, LIG

Identification in non-temporal causal graphs

Notations:

- $\mathscr{G}$: directed acyclic graph (DAG)
- V: set of nodes in $\mathscr{G}$
- E: set of directed edges in $\mathscr{G}$
- p: number of nodes in $\mathscr{G}$

Action: Assign fixed values to a set of variables (*do* operator)

Causal query: Causal effect under the action

Estimand: Distribution of the target without the operator do

Causal query → **Estimand** → Estimation

Example: Diamond

Question: $P(V^4|do(V^2 = v^2))$?

Answer: $\sum_{v^1} P(V^1 = v^1)P(V^4|V^2 = v^2, V^1 = v^1)$

J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009.

Identification in non-temporal causal graphs

Notations:

- $\mathscr{G}$: directed acyclic graph (DAG)
- V: set of nodes in $\mathscr{G}$
- E: set of directed edges in $\mathscr{G}$
- p: number of nodes in $\mathscr{G}$

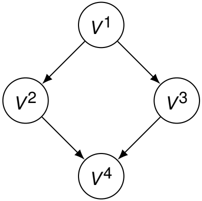
Action: Assign fixed values to a set of variables (*do* operator)

Causal query: Causal effect under the action

Estimand: Distribution of the target without the operator do

Causal query → **Estimand** → Estimation

Example: Diamond



Question: $P(V^4|do(V^2 = v^2))$?

Answer: $\sum_{v^1} P(V^1 = v^1)P(V^4|V^2 = v^2, V^1 = v^1)$

J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009.

Identification in non-temporal causal graphs

Notations:

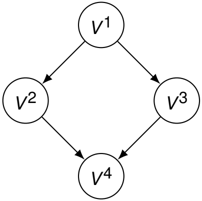
- $\mathcal{G}$: directed acyclic graph (DAG)
- V: set of nodes in $\mathcal{G}$
- E: set of directed edges in $\mathcal{G}$
- p: number of nodes in $\mathcal{G}$

Action: Assign fixed values to a set of variables (*do* operator)

Causal query: Causal effect under the action

Estimand: Distribution of the target without the operator do

Example: Diamond



Question: $P(V^4|do(V^2 = v^2))$?

Answer: $\sum_{v^1} P(V^1 = v^1)P(V^4|V^2 = v^2, V^1 = v^1)$

J. Pearl. *Causality: Models, Reasoning and Inference.* Cambridge University Press, 2009.

Identification in non-temporal causal graphs

Notations:

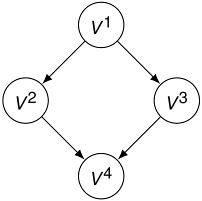
- $\mathscr{G}$: directed acyclic graph (DAG)
- V: set of nodes in $\mathscr{G}$
- E: set of directed edges in $\mathscr{G}$
- p: number of nodes in $\mathscr{G}$

Action: Assign fixed values to a set of variables (*do* operator)

Causal query: Causal effect under the action

Estimand: Distribution of the target without the operator do

Example: Diamond



Question: $P(V^4 | do(V^2 = v^2))$?

Answer: $\sum_{v^1} P(V^1 = v^1) P(V^4 | V^2 = v^2, V^1 = v^1)$

J. Pearl. *Causality: Models, Reasoning and Inference.* Cambridge University Press, 2009.

Identification in non-temporal causal graphs

Notations:

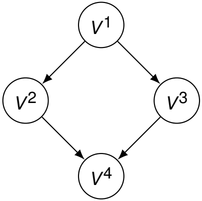
- $\mathscr{G}$: directed acyclic graph (DAG)
- V: set of nodes in $\mathscr{G}$
- E: set of directed edges in $\mathscr{G}$
- p: number of nodes in $\mathscr{G}$

Action: Assign fixed values to a set of variables (*do* operator)

Causal query: Causal effect under the action

Estimand: Distribution of the target without the operator do

Example: Diamond



Question: $P(V^4|do(V^2 = v^2))$?

Answer: $\sum_{v^1} P(V^1 = v^1)P(V^4|V^2 = v^2, V^1 = v^1)$

J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009.

Identification in non-temporal causal graphs

Notations:

- $\mathcal{G}$: directed acyclic graph (DAG)
- V: set of nodes in $\mathcal{G}$
- E: set of directed edges in $\mathcal{G}$
- p: number of nodes in $\mathcal{G}$

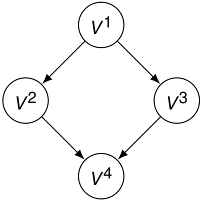
Action: Assign fixed values to a set of variables (*do* operator)

Causal query: Causal effect under the action

Estimand: Distribution of the target without the operator do

Causal query → **Estimand** → Estimation

Example: Diamond



Question: $P(V^4|do(V^2 = v^2))$?

Answer: $\sum_{v^1} P(V^1 = v^1)P(V^4|V^2 = v^2, V^1 = v^1)$

J. Pearl. *Causality: Models, Reasoning and Inference.* Cambridge University Press, 2009.

Identification in non-temporal causal graphs

Notations:

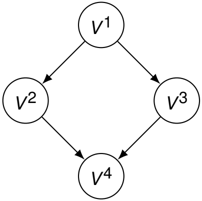
- $\mathcal{G}$: directed acyclic graph (DAG)
- V: set of nodes in $\mathcal{G}$
- E: set of directed edges in $\mathcal{G}$
- p: number of nodes in $\mathcal{G}$

Action: Assign fixed values to a set of variables (*do* operator)

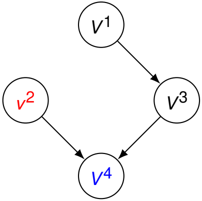
Causal query: Causal effect under the action

Estimand: Distribution of the target without the operator do

Example: Diamond



Question: $P(V^4|do(V^2=v^2))$?



Answer: $\sum_{v^1} P(V^1=v^1)P(V^4|V^2=v^2, V^1=v^1)$

J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009.

Identification in non-temporal causal graphs

Notations:

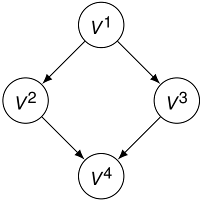
- $\mathscr{G}$: directed acyclic graph (DAG)
- V: set of nodes in $\mathscr{G}$
- E: set of directed edges in $\mathscr{G}$
- p: number of nodes in $\mathscr{G}$

Action: Assign fixed values to a set of variables (*do* operator)

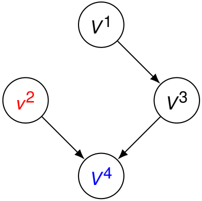
Causal query: Causal effect under the action

Estimand: Distribution of the target without the operator do

Example: Diamond



Question: $P(V^4 | do(V^2 = v^2))$?



Answer: $\sum_{v^1} P(V^1 = v^1) P(V^4 | V^2 = v^2, V^1 = v^1)$

J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009.

Identification in non-temporal causal graphs

$\mathbf{X} = (X^1, X^2, X^3)$: set of time series

Full-time causal graph

Assumption 1: consistency throughout time
Assumption 2: causal sufficiency
λ_{max}: maximal lag between cause and effect

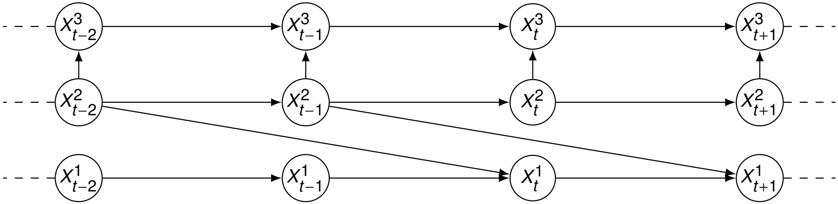
Window causal graph

C. K. Assaad, E. Devijver, E. Gaussier. Survey and Evaluation of Causal Discovery Methods for Time Series. JAIR, 2022.

Identification in non-temporal causal graphs

$\mathbf{X} = (X^1, X^2, X^3)$: set of time series

Full-time causal graph



Assumption 1: consistency throughout time
Assumption 2: causal sufficiency
$\lambda_{\max}$: maximal lag between cause and effect

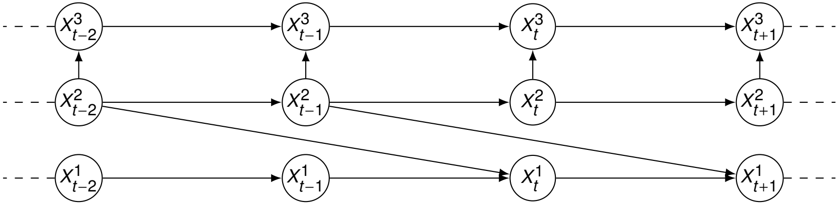
Window causal graph

C. K. Assaad, E. Devijver, E. Gaussier. Survey and Evaluation of Causal Discovery Methods for Time Series. JAIR, 2022.

Identification in non-temporal causal graphs

$\mathbf{X} = (X^1, X^2, X^3)$: set of time series

Full-time causal graph



Assumption 1: consistency throughout time
Assumption 2: causal sufficiency
$\lambda_{\max}$: maximal lag between cause and effect

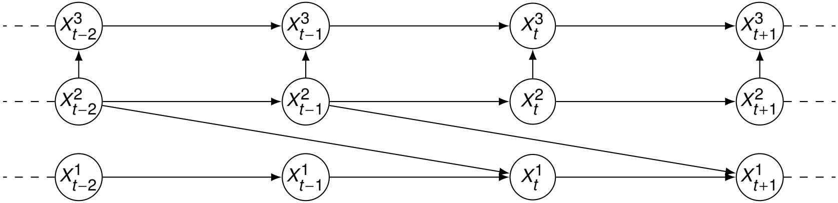
Window causal graph

C. K. Assaad, E. Devijver, E. Gaussier. Survey and Evaluation of Causal Discovery Methods for Time Series. JAIR, 2022.

Identification in non-temporal causal graphs

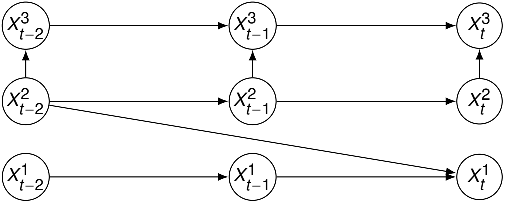
$\mathbf{X} = (X^1, X^2, X^3)$: set of time series

Full-time causal graph



Assumption 1: consistency throughout time
Assumption 2: causal sufficiency
$\lambda_{\max}$: maximal lag between cause and effect

Window causal graph

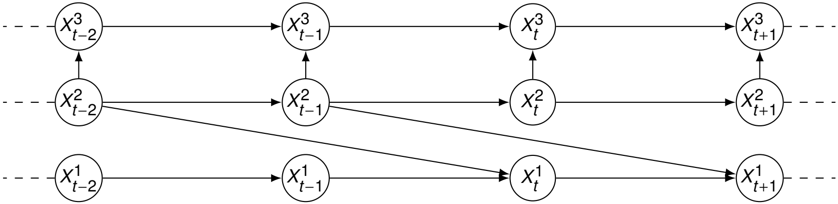


C. K. Assaad, E. Devijver, E. Gaussier. Survey and Evaluation of Causal Discovery Methods for Time Series. JAIR, 2022.

Identification in non-temporal causal graphs

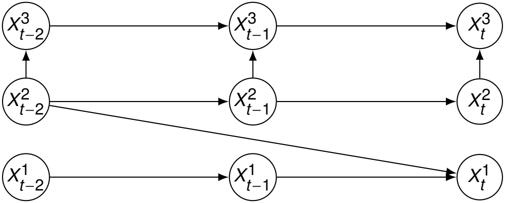
$\mathbf{X} = (X^1, X^2, X^3)$: set of time series

Full-time causal graph

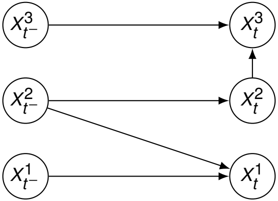


Assumption 1: consistency throughout time
Assumption 2: causal sufficiency
$\lambda_{\max}$: maximal lag between cause and effect

Window causal graph



Extended summary causal graph

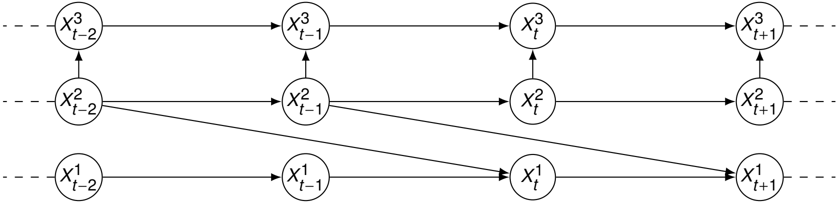


C. K. Assaad, E. Devijver, E. Gaussier. Survey and Evaluation of Causal Discovery Methods for Time Series. JAIR, 2022.

Identification in non-temporal causal graphs

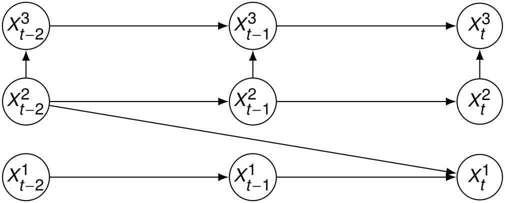
$\mathbf{X} = (X^1, X^2, X^3)$: set of time series

Full-time causal graph

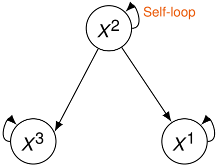


Assumption 1: consistency throughout time
Assumption 2: causal sufficiency
$\lambda_{\max}$: maximal lag between cause and effect

Window causal graph



Summary causal graph



C. K. Assaad, E. Devijver, E. Gaussier. Survey and Evaluation of Causal Discovery Methods for Time Series. JAIR, 2022.

Problem setting

Problem

Given a(n) (extended) summary graph $G_{(e)s}$, a maximal lag λ_{max} and a causal query Q, find the estimand of the causal query.

⚠ *Several estimands of the same causal query!*

Example:

- $Q = P\left(X_t^3 = x_t^3 \,|\, do(X_{t-1}^2 = x_{t-1}^2)\right)$
- $\lambda_{max} = 2$

Summary causal graph

Window causal graph 1

Window causal graph 2

$$Q = \sum_{x_{t-2}^2} P(X_{t-2}^2 = x_{t-2}^2) P\left(X_t^3 = x_t^3 \,|\, X_{t-1}^2 = x_{t-1}^2, x_{t-2}^2\right)$$

$$Q = P(X_t^3 = x_t^3)$$

Problem setting

Problem

Given a(n) (extended) summary graph $G_{(e)s}$, a maximal lag λ_{max} and a causal query Q, find the estimand of the causal query.

⚠ *Several estimands of the same causal query!*

Example:

- $Q = P\left(X_t^3 = x_t^3 | do(X_{t-1}^2 = x_{t-1}^2)\right)$
- $\lambda_{max} = 2$

Summary causal graph

Window causal graph 1

Window causal graph 2

$$Q = \sum_{x_{t-2}^2} P(X_{t-2}^2 = x_{t-2}^2) P\left(X_t^3 = x_t^3 | X_{t-1}^2 = x_{t-1}^2, x_{t-2}^2\right)$$

$$Q = P(X_t^3 = x_t^3)$$

Problem setting

Problem

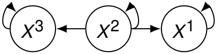
Given a(n) (extended) summary graph $G_{(e)s}$, a maximal lag λ_{max} and a causal query Q, find the estimand of the causal query.

⚠ *Several estimands of the same causal query!*

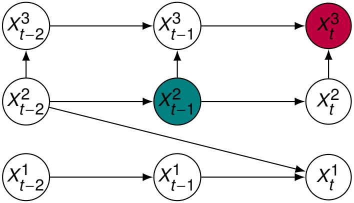
Example:

- $Q = P\left(X_t^3 = x_t^3 | do(X_{t-1}^2 = x_{t-1}^2)\right)$
- $\lambda_{max} = 2$

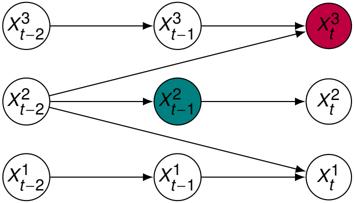
Summary causal graph



Window causal graph 1



Window causal graph 2



$$Q = \sum_{x_{t-2}^2} P(X_{t-2}^2 = x_{t-2}^2) P\left(X_t^3 = x_t^3 | X_{t-1}^2 = x_{t-1}^2, x_{t-2}^2\right)$$

$$Q = P(X_t^3 = x_t^3)$$

Equivalence classes

Objective

Given a(n) (extended) summary graph G_s (or G_{es}), a maximal lag λ_{max} and a causal query

$$Q = P\left(X_t^2 = x_t^2 | do(X_{t-\mu}^1 = x_{t-\mu}^1)\right),$$

with $\mu \geq 0$, we aim to grouping together all the window graphs giving the same estimand.

Abuse of notation: $P(x_t^2 | do(x_{t-\mu}^1)) := Q$.

Definition (Equivalence classes)

Two window graphs G_{w_1} and G_{w_2} compatible with G_s (or G_{es}) are equivalent if the estimands of Q in G_{w_1} and G_{w_2} are the sames.

Technical problem: The estimand is not uniquely defined (many back-door sets).

Equivalence classes

Objective

Given a(n) (extended) summary graph G_s (or G_{es}), a maximal lag λ_{max} and a causal query

$$Q = P\left(X_t^2 = x_t^2 | do(X_{t-\mu}^1 = x_{t-\mu}^1)\right),$$

with $\mu \geq 0$, we aim to grouping together all the window graphs giving the same estimand.

Abuse of notation: $P(x_t^2 | do(x_{t-\mu}^1)) := Q$.

Definition (Equivalence classes)

Two window graphs G_{w_1} and G_{w_2} compatible with G_s (or G_{es}) are equivalent if the estimands of Q in G_{w_1} and G_{w_2} are the sames.

Technical problem: The estimand is not uniquely defined (many back-door sets).

Equivalence classes

Objective

Given a(n) (extended) summary graph G_s (or G_{es}), a maximal lag λ_{max} and a causal query

$$Q = P\left(X_t^2 = x_t^2 | do(X_{t-\mu}^1 = x_{t-\mu}^1)\right),$$

with $\mu \geq 0$, we aim to grouping together all the window graphs giving the same estimand.

Abuse of notation: $P(x_t^2 | do(x_{t-\mu}^1)) := Q$.

Definition (Equivalence classes)

Two window graphs G_{w_1} and G_{w_2} compatible with G_s (or G_{es}) are equivalent if the estimands of Q in G_{w_1} and G_{w_2} are the sames.

Technical problem: The estimand is not uniquely defined (many back-door sets).

Equivalence classes

Idea

Each window graph G_w is characterized by the set Pa_w of X_t^2 parents in G_w.

Characterization: Each class $\mathscr{C}$ is defined by a set of parents $Pa_{\mathscr{C}}$. The estimand Q_e of Q associated with $\mathscr{C}$ is

$$Q_e = \sum_{y_{\mathscr{C}}} P(x_t^2 | x_{t-\mu}^1, y_{\mathscr{C}}) \times P(y_{\mathscr{C}}),$$

where the sum covers all potential values of $y_{\mathscr{C}}$ in $Pa_{\mathscr{C}}$.

Open question: What is the "optimal" characterization back-door set ?

Equivalence classes

Idea

Each window graph G_w is characterized by the set Pa_w of X_t^2 parents in G_w.

Characterization: Each class $\mathscr{C}$ is defined by a set of parents $Pa_{\mathscr{C}}$. The estimand Q_e of Q associated with $\mathscr{C}$ is

$$Q_e = \sum_{y_{\mathscr{C}}} P(x_t^2 | x_{t-\mu}^1, y_{\mathscr{C}}) \times P(y_{\mathscr{C}}),$$

where the sum covers all potential values of $y_{\mathscr{C}}$ in $Pa_{\mathscr{C}}$.

Open question: What is the "optimal" characterization back-door set ?

Equivalence classes

Idea

Each window graph G_w is characterized by the set Pa_w of X_t^2 parents in G_w.

Characterization: Each class $\mathscr{C}$ is defined by a set of parents $Pa_{\mathscr{C}}$. The estimand Q_e of Q associated with $\mathscr{C}$ is

$$Q_e = \sum_{y_{\mathscr{C}}} P(x_t^2 | x_{t-\mu}^1, y_{\mathscr{C}}) \times P(y_{\mathscr{C}}),$$

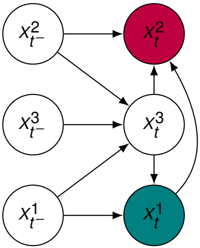
where the sum covers all potential values of $y_{\mathscr{C}}$ in $Pa_{\mathscr{C}}$.

Open question: What is the "optimal" characterization back-door set ?

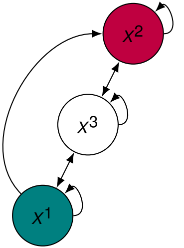
Equivalence classes

Example: Consider $\lambda_{\max} = 2$, $\mu = 0$ and

Extended summary graph



Summary graph



3 classes.

Classes	Parent set
$\mathscr{C}_1$	$\{X^1_{t-1}, X^3_t\}$
$\mathscr{C}_2$	$\{X^1_{t-2}, X^3_t\}$
$\mathscr{C}_3$	$\{X^1_{t-2}, X^1_{t-1}, X^3_t\}$

Number of window graphs: **243**

21 classes.

Number of window graphs: **453789**

Largest class: $\{X^1_{t-2}, X^1_{t-1}, X^3_{t-2}, X^3_{t-1}, X^3_t\}$

Equivalence classes

Back to the general case: let $\mathscr{C}_i$ and $\mathscr{C}_j$ be two different classes. Under which condition, we can have a common estimand in both classes ?

Answer: If for all window graphs G_w in $\mathscr{C}_i \cup \mathscr{C}_j$, we have

$$\left[Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j}\right] \cap De(X^1_{t-\mu}, X^2_t) = \varnothing,$$

where $De(\cdot,\cdot)$ is the set of common descendants.

Proposition

Let G_{es} be an extended summary graph and $\mathscr{C}_i$, $\mathscr{C}_j$ two different classes. Then, for all G_w in $\mathscr{C}_i \cup \mathscr{C}_j$, we have

$$Q_e = \sum_{y_{\mathscr{C}_{ij}}} P(x^2_t | x^1_{t-\mu}, y_{\mathscr{C}_{ij}}) \times P(y_{\mathscr{C}_{ij}}),$$

where the sum covers all potential values of $y_{\mathscr{C}_{ij}}$ in $Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j}$.

Idea of the proof: a potential parent of $X^1_{t-\mu}$ is a potential descendant of X^2_t only if $\mu = 0$. This is in conflict with the DAG assumption.

Equivalence classes

Back to the general case: let $\mathscr{C}_i$ and $\mathscr{C}_j$ be two different classes. Under which condition, we can have a common estimand in both classes ?

Answer: If for all window graphs G_w in $\mathscr{C}_i \cup \mathscr{C}_j$, we have

$$\left[Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j}\right] \cap De(X^1_{t-\mu}, X^2_t) = \varnothing,$$

where $De(\cdot,\cdot)$ is the set of common descendants.

Proposition

Let G_{es} be an extended summary graph and $\mathscr{C}_i$, $\mathscr{C}_j$ two different classes. Then, for all G_w in $\mathscr{C}_i \cup \mathscr{C}_j$, we have

$$Q_e = \sum_{y_{\mathscr{C}_{ij}}} P(x^2_t | x^1_{t-\mu}, y_{\mathscr{C}_{ij}}) \times P(y_{\mathscr{C}_{ij}}),$$

where the sum covers all potential values of $y_{\mathscr{C}_{ij}}$ in $Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j}$.

Idea of the proof: a potential parent of $X^1_{t-\mu}$ is a potential descendant of X^2_t only if $\mu = 0$. This is in conflict with the DAG assumption.

Equivalence classes

Back to the general case: let $\mathscr{C}_i$ and $\mathscr{C}_j$ be two different classes. Under which condition, we can have a common estimand in both classes ?

Answer: If for all window graphs G_w in $\mathscr{C}_i \cup \mathscr{C}_j$, we have

$$\left[Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j} \right] \cap De(X^1_{t-\mu}, X^2_t) = \emptyset,$$

where $De(\cdot,\cdot)$ is the set of common descendants.

Proposition

Let G_{es} be an extended summary graph and $\mathscr{C}_i$, $\mathscr{C}_j$ two different classes. Then, for all G_w in $\mathscr{C}_i \cup \mathscr{C}_j$, we have

$$Q_e = \sum_{y_{\mathscr{C}_{ij}}} P(x^2_t | x^1_{t-\mu}, y_{\mathscr{C}_{ij}}) \times P(y_{\mathscr{C}_{ij}}),$$

where the sum covers all potential values of $y_{\mathscr{C}_{ij}}$ in $Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j}$.

Idea of the proof: a potential parent of $X^1_{t-\mu}$ is a potential descendant of X^2_t only if $\mu = 0$. This is in conflict with the DAG assumption.

Equivalence classes

Back to the general case: let $\mathscr{C}_i$ and $\mathscr{C}_j$ be two different classes. Under which condition, we can have a common estimand in both classes ?

Answer: If for all window graphs G_w in $\mathscr{C}_i \cup \mathscr{C}_j$, we have

$$\left[Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j}\right] \cap De(X^1_{t-\mu}, X^2_t) = \varnothing,$$

where $De(\cdot,\cdot)$ is the set of common descendants.

Proposition

Let G_{es} be an extended summary graph and $\mathscr{C}_i$, $\mathscr{C}_j$ two different classes. Then, for all G_w in $\mathscr{C}_i \cup \mathscr{C}_j$, we have

$$Q_e = \sum_{y_{\mathscr{C}_{ij}}} P(x^2_t | x^1_{t-\mu}, y_{\mathscr{C}_{ij}}) \times P(y_{\mathscr{C}_{ij}}),$$

where the sum covers all potential values of $y_{\mathscr{C}_{ij}}$ in $Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j}$.

Idea of the proof: a potential parent of $X^1_{t-\mu}$ is a potential descendant of X^2_t only if $\mu = 0$. This is in conflict with the DAG assumption.

Equivalence classes

Proposition

Let G_{es} be an extended summary graph with a maximal lag $\lambda_{\max}$. Let $\mathscr{C}_{\max}$ be the largest class (with the largest cardinal). Then, for all window graph G_w, we have

$$Q_e = \sum_{y_{\mathscr{C}max}} P(x_t^2 | x_{t-\mu}^1, y_{\mathscr{C}_{\max}}) \times P(y_{\mathscr{C}_{\max}}),$$

where the sum covers all potential values of $y_{\mathscr{C}_{\max}}$ in $Pa_{\mathscr{C}_{\max}}$.

But...

The same result does not hold for summary graphs!

Equivalence classes

Proposition

Let G_{es} be an extended summary graph with a maximal lag $\lambda_{\max}$. Let $\mathscr{C}_{\max}$ be the largest class (with the largest cardinal). Then, for all window graph G_w, we have

$$Q_e = \sum_{y_{\mathscr{C}max}} P(x_t^2 | x_{t-\mu}^1, y_{\mathscr{C}_{\max}}) \times P(y_{\mathscr{C}_{\max}}),$$

where the sum covers all potential values of $y_{\mathscr{C}_{\max}}$ in $Pa_{\mathscr{C}_{\max}}$.

But...

The same result does not hold for summary graphs!

Equivalence classes

Proposition

Let G_{es} be an extended summary graph with a maximal lag $\lambda_{\max}$. Let $\mathscr{C}_{\max}$ be the largest class (with the largest cardinal). Then, for all window graph G_w, we have

$$Q_e = \sum_{y_{\mathscr{C}max}} P(x_t^2 | x_{t-\mu}^1, y_{\mathscr{C}_{\max}}) \times P(y_{\mathscr{C}_{\max}}),$$

where the sum covers all potential values of $y_{\mathscr{C}_{\max}}$ in $Pa_{\mathscr{C}_{\max}}$.

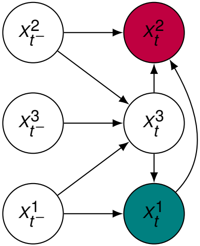
But...

The same result does not hold for summary graphs!

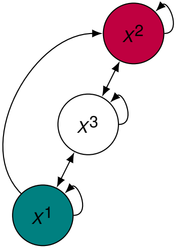
Equivalence classes

Example: Consider $\lambda_{\max} = 2$, $\mu = 0$ and

Extended summary graph



Summary graph



Universal estimand:

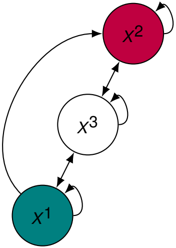
$$Q_e = \sum_{y_{\mathscr{C}max}} P(x^2_t | x^1_{t-\mu}, y_{\mathscr{C}_{\max}}) \times P(y_{\mathscr{C}_{\max}}),$$

where the sum covers all potential values of $y_{\mathscr{C}_{\max}}$ in $\left\{X^1_{t-2}, X^1_{t-1}, X^3_t\right\}$.

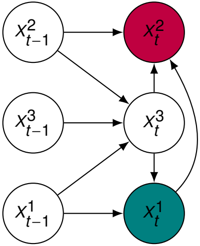
No universal estimand ...

Equivalence classes

No universal estimand for a summary graph.

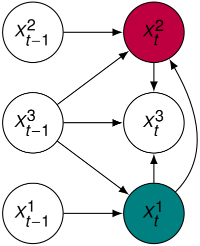


Summary graph



Window graph G_{w_1} in $\mathscr{C}_i$

$Pa_{\mathscr{C}_i} = \left\{X^1_{t-1}, X^3_t\right\}$



Window graph G_{w_2} in $\mathscr{C}_j$

$Pa_{\mathscr{C}_i} = \left\{X^1_{t-1}, X^3_{t-1}\right\}$

The set $Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j} = \left\{X^1_{t-1}, X^3_{t-1}, X^3_t\right\}$ is not a back-door set in G_{w_2}.

The classes $\mathscr{C}_i$ and $\mathscr{C}_j$ are not super equivalent.

Equivalence classes

Definition (Super equivalence)

Let G_s be a summary graph, two classes $\mathscr{C}_i$ and $\mathscr{C}_j$ are super equivalent if $Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j}$ is a valid back-door set of the query Q over $\mathscr{C}_i$ and $\mathscr{C}_j$.

Definition (Super class)

Let G_s be a summary graph, a super class is the maximal set of equivalence classes that are super equivalent for the causal query Q.

General result. If G_{es} is an extended summary graph, then there exists a unique super class. The universal estimand is

$$Q_e = \sum_{y_{\mathscr{C}}} P(x_t^2 | x_{t-\mu}^1, y_{\mathscr{C}}) \times P(y_{\mathscr{C}}),$$

where the sum covers all potential values of $y_{\mathscr{C}_{\max}}$ in the largest class.

Equivalence classes

Definition (Super equivalence)

Let G_s be a summary graph, two classes $\mathscr{C}_i$ and $\mathscr{C}_j$ are super equivalent if $Pa_{\mathscr{C}_i} \cup Pa_{\mathscr{C}_j}$ is a valid back-door set of the query Q over $\mathscr{C}_i$ and $\mathscr{C}_j$.

Definition (Super class)

Let G_s be a summary graph, a super class is the maximal set of equivalence classes that are super equivalent for the causal query Q.

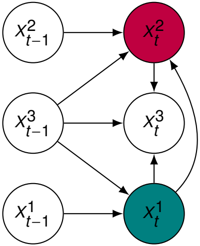
General result. If G_{es} is an extended summary graph, then there exists a unique super class. The universal estimand is

$$Q_e = \sum_{y_{\mathscr{C}}} P(x_t^2 | x_{t-\mu}^1, y_{\mathscr{C}}) \times P(y_{\mathscr{C}}),$$

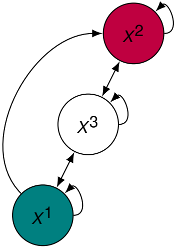
where the sum covers all potential values of $y_{\mathscr{C}_{\max}}$ in the largest class.

Numerical simulations

Consider $\lambda_{\max} = 2$, $\mu = 0$ and



Extended summary graph



Summary graph

Generative model. Given a window causal graph G_w, the time series X^i for $i \in \{1, 2, 3\}$ is generated using the equations:

$$X^i_t = \sum_{X^j_{t'} \in Pa_{G_w}(X^i_t)} \alpha_{ij} X^j_{t'} + 0.1 \xi^i_t,$$

where $\xi^i_t \sim N(0, 1)$ and α_{ij} in $[0.2, 1]$ (randomly chosen).

We aim to estimate the expectation $\mathbb{E}[X^2_t | do(X^1_t = x^1_t)]$ for a given window graph G_w.

Numerical simulations

We show that

$$\mathbb{E}[X_t^2 | do(X_t^1 = x_t^1)] = \alpha_{21} x_t^1.$$

We estimate α_{21} using a linear regression.

Extended summary graph. Consider the classes $\mathscr{C}_1$, $\mathscr{C}_4$ and $\mathscr{C}_9$ with

$$Pa_{\mathscr{C}_1} = \left\{X_t^1, X_{t-1}^1, X_{t-1}^3\right\}, Pa_{\mathscr{C}_4} = \left\{X_t^1, X_{t-2}^1, X_{t-2}^3\right\}$$
$$Pa_{\mathscr{C}_9} = \left\{X_t^1, X_{t-1}^1, X_{t-1}^3, X_{t-2}^1, X_{t-2}^3\right\}$$

The estimation is performed by generating a 100 points for each time series.

Numerical simulations

We show that

$$\mathbb{E}[X_t^2 | do(X_t^1 = x_t^1)] = \alpha_{21} x_t^1.$$

We estimate α_{21} using a linear regression.

Extended summary graph. Consider the classes $\mathscr{C}_1$, $\mathscr{C}_4$ and $\mathscr{C}_9$ with

$$Pa_{\mathscr{C}_1} = \left\{X_t^1, X_{t-1}^1, X_{t-1}^3\right\}, Pa_{\mathscr{C}_4} = \left\{X_t^1, X_{t-2}^1, X_{t-2}^3\right\}$$
$$Pa_{\mathscr{C}_9} = \left\{X_t^1, X_{t-1}^1, X_{t-1}^3, X_{t-2}^1, X_{t-2}^3\right\}$$

The estimation is performed by generating a 100 points for each time series.

Numerical simulations

We show that

$$\mathbb{E}[X_t^2 | do(X_t^1 = x_t^1)] = \alpha_{21} x_t^1.$$

We estimate α_{21} using a linear regression.

Extended summary graph. Consider the classes $\mathscr{C}_1$, $\mathscr{C}_4$ and $\mathscr{C}_9$ with

$$Pa_{\mathscr{C}_1} = \left\{X_t^1, X_{t-1}^1, X_{t-1}^3\right\}, Pa_{\mathscr{C}_4} = \left\{X_t^1, X_{t-2}^1, X_{t-2}^3\right\}$$
$$Pa_{\mathscr{C}_9} = \left\{X_t^1, X_{t-1}^1, X_{t-1}^3, X_{t-2}^1, X_{t-2}^3\right\}$$

The estimation is performed by generating a 100 points for each time series.

Numerical simulations

MAE: Mean Absolute Error.

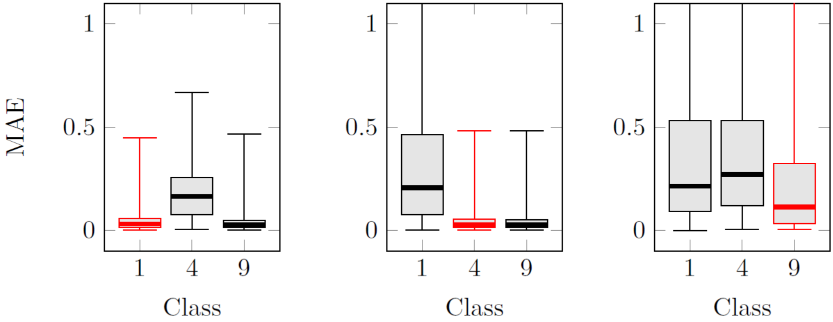
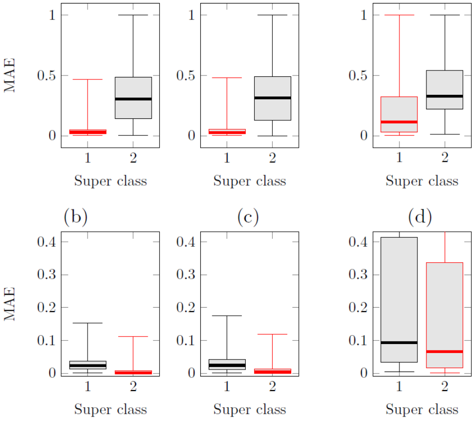


Figure: Estimation boxplots over each class.

Numerical simulations

Summary graph. Two super classes S_1 and S_2.

$$S_1 : \left\{X_t^1, X_{t-1}^1, X_{t-1}^3, X_{t-2}^1, X_{t-2}^3\right\}$$
$$S_2 : \left\{X_t^1, X_t^3, X_{t-1}^1, X_{t-1}^3, X_{t-2}^1, X_{t-2}^3\right\}$$



(b) (c) (d)

(e) (f) (g)

Practical use of the algorithm output

Algorithm output:

- Super classes $S_1, \ldots, S_p$,
- Associated estimands $Q_e^1, \ldots, Q_e^p$,
- Proportions $w_1, \ldots, w_p$.

Practical use:

- Most likely estimand:

$$MLE = \mathcal{E}_{argmax(w_1,\ldots,w_p)},$$

- Weighted mean estimand:

$$WME = \sum_{i=1}^{p} w_i Q_e^i.$$

Practical use of the algorithm output

Algorithm output:

- Super classes $S_1, \dots, S_p$,
- Associated estimands $Q_e^1, \dots, Q_e^p$,
- Proportions $w_1, \dots, w_p$.

Practical use:

- Most likely estimand:

$$MLE = \mathscr{E}_{argmax(w_1,\dots,w_p)},$$

- Weighted mean estimand:

$$WME = \sum_{i=1}^{p} w_i Q_e^i.$$

Practical use of the algorithm output

Algorithm output:

- Super classes $S_1, \ldots, S_p$,
- Associated estimands $Q_e^1, \ldots, Q_e^p$,
- Proportions $w_1, \ldots, w_p$.

Practical use:

- Most likely estimand:

$$MLE = \mathscr{E}_{argmax(w_1,\ldots,w_p)},$$

- Weighted mean estimand:

$$WME = \sum_{i=1}^{p} w_i Q_e^i.$$

Conclusion and Prospect

- Searching for equivalence classes for a causal query from an extended summary or a summary graph.
- Characterizing each class of window graphs by the set of parents.
- Introducing the notion of super class.
- Performing some numerical simulations.

Conclusion and Prospect

- Searching for equivalence classes for a causal query from an extended summary or a summary graph.
- Characterizing each class of window graphs by the set of parents.
- Introducing the notion of super class.
- Performing some numerical simulations.

Conclusion and Prospect

- Searching for equivalence classes for a causal query from an extended summary or a summary graph.
- Characterizing each class of window graphs by the set of parents.
- Introducing the notion of super class.
- Performing some numerical simulations.

Conclusion and Prospect

- Searching for equivalence classes for a causal query from an extended summary or a summary graph.
- Characterizing each class of window graphs by the set of parents.
- Introducing the notion of super class.
- Performing some numerical simulations.

Conclusion and Prospect

- Extension to hidden common causes ?
- General Estimation of the equivalence class estimands (non i.i.d data).
- Searching for the Optimal characterization of the equivalence classes ?

Conclusion and Prospect

- Extension to hidden common causes ?
- General Estimation of the equivalence class estimands (non i.i.d data).
- Searching for the Optimal characterization of the equivalence classes ?

Conclusion and Prospect

- Extension to hidden common causes **?**
- General Estimation of the equivalence class estimands (non i.i.d data).
- Searching for the Optimal characterization of the equivalence classes **?**