Causal inference and control

Week 8:

Time series

Søren Wengel Mogensen
Department of Automatic Control
soren.wengel_mogensen@control.lth.se

Assignment 3

Graphical marginalization

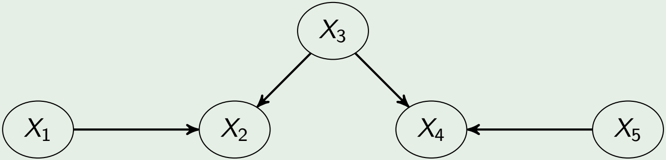
In causal modeling, the idea of *hidden* variables is central. In Week 4, we looked at identification methods that do not require full observation.

In causal discovery, we may be interested in learning a ‘marginal’ of the causal graph when there are hidden variables, $\boldsymbol{H}$, as well as observed variables, $\boldsymbol{O}$. One natural requirement is that the implied conditional independences are the same when restricting to $\boldsymbol{O}$.

Graphical marginalization

Example (DAGs are not closed under marginalization, Richardson and Spirtes [2002])

Assume X_3 is unobserved. There is no DAG on nodes $\{X_1, X_2, X_4, X_5\}$ that encodes the same conditional independences (using d-separation).



Acyclic directed mixed graphs

We say that a graph is a *directed acyclic mixed graph* (ADMG) if every edge is either *directed*, $\rightarrow$, or *bidirected*, $\leftrightarrow$.

The extension of *d*-separation to ADMGs is known as *m-separation*.

Latent projection

Let $\mathcal{G} = (\boldsymbol{V}, E)$ be an ADMG, $\boldsymbol{V} = \boldsymbol{O} \cup \boldsymbol{H}$. We define the following transformation.

Definition (Latent projection)

We define $m(\mathcal{G}, \boldsymbol{O})$ as the graph such that for $X, Y \in \boldsymbol{O}$

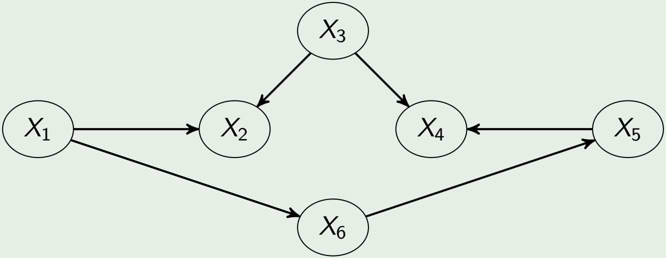
- $X \rightarrow Y$ in $m(\mathcal{G}, \boldsymbol{O})$ if there is a directed path $X \rightarrow \ldots \rightarrow Y$ in $\mathcal{G}$ such that every non-endpoint node is in $\boldsymbol{H}$,
- $X \leftrightarrow Y$ in $m(\mathcal{G}, \boldsymbol{O})$ if there is a path between X and Y such that all non-endpoint nodes are in $\boldsymbol{H}$, all non-endpoint nodes are non-colliders, and there are arrowheads at both X and Y.

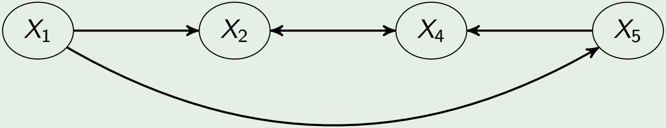
The latent projection is also an ADMG!

Latent projection

Example

Let $\boldsymbol{V} = \{X_1, X_2, X_3, X_4, X_5, X_6\}$ and $\boldsymbol{O} = \{X_1, X_2, X_4, X_5\}$





Latent projection as a marginal

Let $\mathcal{G} = (\boldsymbol{O} \cup \boldsymbol{H}, E)$ be a DAG and let $\mathcal{M} = m(\mathcal{G}, \boldsymbol{O})$.

Proposition

Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \subseteq \boldsymbol{O}$. We have that $\boldsymbol{A}$ and $\boldsymbol{B}$ are d-separated by $\boldsymbol{C}$ in $\mathcal{G}$ if and only if $\boldsymbol{A}$ and $\boldsymbol{B}$ are m-separated by $\boldsymbol{C}$ in $m(\mathcal{G}, \boldsymbol{O})$.

Latent projection as a marginal

Let $\mathcal{G} = (\boldsymbol{O} \cup \boldsymbol{H}, E)$ be a DAG and let $\mathcal{M} = m(\mathcal{G}, \boldsymbol{O})$.

Proposition

Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \subseteq \boldsymbol{O}$. We have that $\boldsymbol{A}$ and $\boldsymbol{B}$ are d-separated by $\boldsymbol{C}$ in $\mathcal{G}$ if and only if $\boldsymbol{A}$ and $\boldsymbol{B}$ are m-separated by $\boldsymbol{C}$ in $m(\mathcal{G}, \boldsymbol{O})$.

We should just prove that if there is a d-connecting walk between $i \in \boldsymbol{A}$ and $j \in \boldsymbol{B}$ given $\boldsymbol{C}$ in $\mathcal{G}$, then we can also find an m-connecting walk between $i \in \boldsymbol{A}$ and $j \in \boldsymbol{B}$ given $\boldsymbol{C}$ in $m(\mathcal{G}, \boldsymbol{O})$, and vice versa.

Latent projection as a marginal

Sketch of proof.

Latent projection as a marginal

Sketch of proof.

We consider a d-connecting walk between $i = i_0 \in \boldsymbol{A}$ and $j = i_m \in \boldsymbol{B}$ given $\boldsymbol{C}$ in $\mathcal{G}$,

$$i_0 \sim i_1 \sim i_2 \sim i_3 \sim \ldots \sim i_{m-1} \sim i_m.$$

The symbol $\sim$ denotes an edge. The endpoints are in $\boldsymbol{O}$ so every node on the path which is in $\boldsymbol{H}$ is on a subpath

$$i_k \sim i_{k+1} \sim \ldots \sim i_{k+l}$$

where $i_k, i_k \in \boldsymbol{O}$ and every other node is in $\mathcal{H}$. If there are colliders on this subpath, they must be ancestors of nodes in $\boldsymbol{C} \subseteq \boldsymbol{O}$. When there are no colliders, the endpoint identical edge is in $m(\mathcal{G}, \boldsymbol{O})$.

Latent projection as a marginal

We consider a d-connecting walk between $i = i_0 \in \boldsymbol{A}$ and $j = i_m \in \boldsymbol{B}$ given $\boldsymbol{C}$ in $\mathcal{G}$,

$$\underbrace{i_0 \sim i_1}_{O} \sim \underbrace{i_2 \sim i_3}_{H} \sim \underbrace{i_4 \sim i_5}_{O} \dots \sim \underbrace{i_{m-1} \sim i_m}_{O}.$$

For instance, the subpath between i_1 and i_4 is collapsed into an edge in $m(\mathcal{G}, \boldsymbol{O})$. This edge is endpoint-identical with the subpath between i_1 and i_4.

$$\underbrace{i_0 \sim i_1}_{O} \leftarrow \underbrace{i_2 \rightarrow i_3}_{H} \rightarrow \underbrace{i_4 \sim i_5}_{O} \dots \sim \underbrace{i_{m-1} \sim i_m}_{O}.$$

Latent projection as a marginal

We consider a d-connecting walk between $i = i_0 \in \boldsymbol{A}$ and $j = i_m \in \boldsymbol{B}$ given $\boldsymbol{C}$ in $\mathcal{G}$,

$$\underbrace{i_0 \sim i_1}_{O} \sim \underbrace{i_2 \sim i_3}_{H} \sim \underbrace{i_4 \sim i_5}_{O} \ldots \sim \underbrace{i_{m-1} \sim i_m}_{O}.$$

For instance, the subpath between i_1 and i_4 is collapsed into an edge in $m(\mathcal{G}, \boldsymbol{O})$. This edge is endpoint-identical with the subpath between i_1 and i_4.

$$\underbrace{i_0 \sim i_1}_{O} \leftarrow \underbrace{i_2 \rightarrow i_3}_{H} \rightarrow \underbrace{i_4 \sim i_5}_{O} \ldots \sim \underbrace{i_{m-1} \sim i_m}_{O}.$$

$$\underbrace{i_0 \sim i_1}_{O} \longleftarrow\!\!\!\longrightarrow \underbrace{i_4 \sim i_5}_{O} \ldots \sim \underbrace{i_{m-1} \sim i_m}_{O}.$$

Latent projection as a marginal

The reverse implication just ‘expands’ paths to construct a d-connecting walk in $\mathcal{G}$ from a d-connecting walk in $m(\mathcal{G}, \boldsymbol{O})$.

Latent projection as a marginal

In order to actually compute a latent projection, we used the following approach. We say that (k, l, m) is an *unshielded collider* if $l \in \boldsymbol{H}$ and (1), (2), or (3) holds.

(1) $k \to l \to m$ is in $\mathcal{G}$ and $k \to m$ is not in $\mathcal{G}$.

(2) $k \leftarrow l \to m$ is in $\mathcal{G}$ and $k \leftrightarrow m$ is not in $\mathcal{G}$.

(3) $k \leftarrow l \leftrightarrow m$ is in $\mathcal{G}$ and $k \leftrightarrow m$ is not in $\mathcal{G}$.

That is, (k, l, m), $k \sim l \sim m$, is unshielded if the middle node is unobserved and the corresponding endpoint-identical edge is not in the graph.

Latent projection as a marginal

In order to actually compute a latent projection, we used the following approach. We say that (k, l, m) is an *unshielded collider* if $l \in \boldsymbol{H}$ and (1), (2), or (3) holds.

(1) $k \to l \to m$ is in $\mathcal{G}$ and $k \to m$ is not in $\mathcal{G}$.

(2) $k \leftarrow l \to m$ is in $\mathcal{G}$ and $k \leftrightarrow m$ is not in $\mathcal{G}$.

(3) $k \leftarrow l \leftrightarrow m$ is in $\mathcal{G}$ and $k \leftrightarrow m$ is not in $\mathcal{G}$.

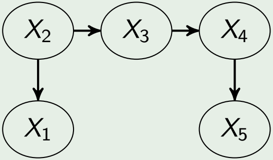
That is, (k, l, m), $k \sim l \sim m$, is unshielded if the middle node is unobserved and the corresponding endpoint-identical edge is not in the graph.

Given $\mathcal{G} = (V, E)$ and $\boldsymbol{O}$, we can compute $m(\mathcal{G}, \boldsymbol{O})$ by simply adding the 'shielding' edges until there are no more unshielded colliders.

Latent projection as a marginal

Example

Let $\boldsymbol{V} = \{X_1, X_2, X_3, X_4, X_5\}$ and $\boldsymbol{O} = \{X_1, X_5\}$

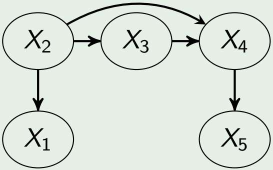


$X_2 \to X_3 \to X_4$ is unshielded, so $X_2 \to X_4$ is added.

Latent projection as a marginal

Example

Let $\boldsymbol{V} = \{X_1, X_2, X_3, X_4, X_5\}$ and $\boldsymbol{O} = \{X_1, X_5\}$

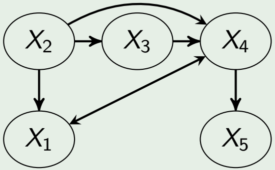


$X_1 \leftarrow X_2 \rightarrow X_4$ is unshielded, so $X_1 \leftrightarrow X_4$ is added.

Latent projection as a marginal

Example

Let $\boldsymbol{V} = \{X_1, X_2, X_3, X_4, X_5\}$ and $\boldsymbol{O} = \{X_1, X_5\}$



This is repeated and at the end we output the graph restricted to $\boldsymbol{O}$. In this example, the output is $X_1 \leftrightarrow X_5$.

Time series and structural causal models

Time series

We assume now that we are modeling a *time series*

$$X_0, X_1, X_2, \ldots$$

where $X_t = (X_t^1, \ldots, X_t^n)^T$ is an n-dimensional random vector.

Time series

We assume now that we are modeling a *time series*

$$X_0, X_1, X_2, \ldots$$

where $X_t = (X_t^1, \ldots, X_t^n)^T$ is an n-dimensional random vector.

For simplicity, we assume that the process ‘starts’ at time 0. One could also consider a doubly infinite process $\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$.

Time series

We will assume that X is generated as an SCM, that is, for each $t = 0, 1, 2, \dots,$ and $i \in V = \{1, 2, \dots, n\}$,

$$X_t^i = f_t^i(\boldsymbol{PA}_{X_t^i}, \varepsilon_t^i)$$

and $\boldsymbol{PA}_{X_t^i} \subseteq X_{p(t)}^V$ where $p(t) = \{0, 1, 2, \dots, t-1\}$ such that $X_{p(t)}^V = \{X_s^i : i \in V, s < t\}$. That is, f_t^i only depends on the past. One can also impose more 'temporal regularity'.

Time series

We will assume that X is generated as an SCM, that is, for each $t = 0, 1, 2, \ldots,$ and $i \in V = \{1, 2, \ldots, n\}$,

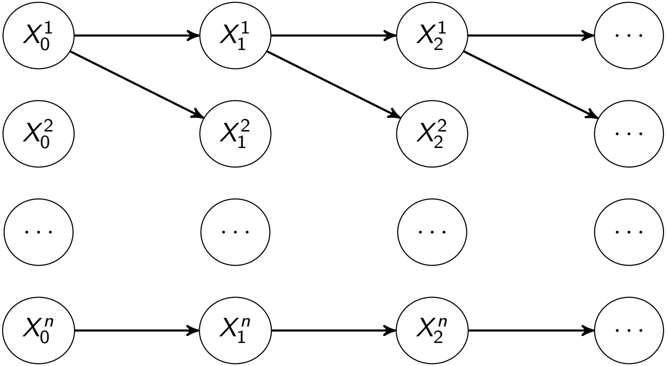
$$X_t^i = f_t^i(\boldsymbol{PA}_{X_t^i}, \varepsilon_t^i)$$

and $\boldsymbol{PA}_{X_t^i} \subseteq X_{p(t)}^V$ where $p(t) = \{0, 1, 2, \ldots, t-1\}$ such that $X_{p(t)}^V = \{X_s^i : i \in V, s < t\}$. That is, f_t^i only depends on the past. One can also impose more 'temporal regularity'.

One could generalize this to allow *instantaneous effects* in which case f_t^i may also depend on X_t^{-i}.

Full time graph

The *full time graph* is simply the (infinite) graph defined by the SCM as before.
An example,



Identifiability of graph

Let us fix t and consider the finite SCM from time points 0 to t and its graph. Let $\mathcal{D}$ be the full time graph. We say that $\mathcal{D}$ restricted to nodes $X^V_{p(t+1)}$ is a *finite causal graph*.

Theorem (ECI, Theorem 10.1)

If two full time graphs are Markov equivalent, then they are equal.

Identifiability of graph

Let us fix t and consider the finite SCM from time points 0 to t and its graph. Let $\mathcal{D}$ be the full time graph. We say that $\mathcal{D}$ restricted to nodes $X^V_{p(t+1)}$ is a *finite causal graph*.

Theorem (ECI, Theorem 10.1)

If two full time graphs are Markov equivalent, then they are equal.

(the assumption that there are no instantaneous effects is necessary for the above result)

Subsampling

A word of caution.

Summary graph

We consider a time series, X, corresponding to an SCM with causal DAG, $\mathcal{D} = (W, E)$. We have $W = \{X_s^i : s \in \mathbb{N}_0, i \in V\}$, $V = \{1, 2, \ldots, n\}$, that is, every node in $\mathcal{D}$ corresponds to a random variable in the SCM. We now define a graph in which each node represents an entire *coordinate process*, $X_0^i, X_1^i, X_2^i, \ldots$.

The *summary graph* of X, $\mathcal{S} = (V, F)$, is the directed graph on nodes $X^1, \ldots, X^n$, or equivalently on nodes $V = \{1, 2, \ldots, n\}$, such that for $i \neq j$, $i \to j$ in $\mathcal{S}$ if there exists s, t such that $X_s^i \to X_t^j$ in $\mathcal{D}$.

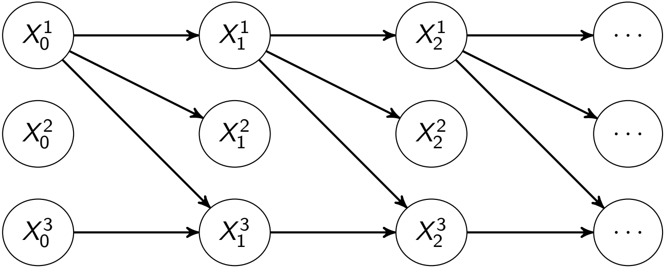
Summary graph

We consider a time series, X, corresponding to an SCM with causal DAG, $\mathcal{D} = (W, E)$. We have $W = \{X^i_s : s \in \mathbb{N}_0, i \in V\}$, $V = \{1, 2, \ldots, n\}$, that is, every node in $\mathcal{D}$ corresponds to a random variable in the SCM. We now define a graph in which each node represents an entire *coordinate process*, $X^i_0, X^i_1, X^i_2, \ldots$.

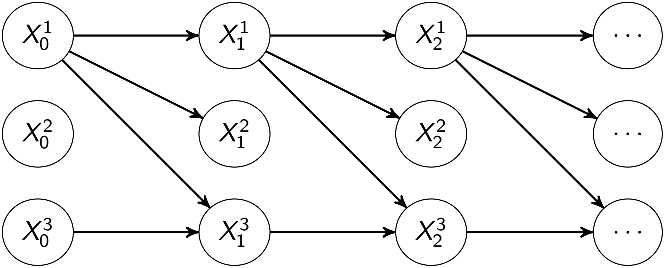
The *summary graph* of X, $\mathcal{S} = (V, F)$, is the directed graph on nodes $X^1, \ldots, X^n$, or equivalently on nodes $V = \{1, 2, \ldots, n\}$, such that for $i \neq j$, $i \to j$ in $\mathcal{S}$ if there exists s, t such that $X^i_s \to X^j_t$ in $\mathcal{D}$.

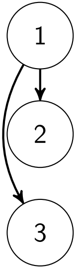
A summary graph need not be acyclic (the underlying full time graph is of course acyclic as it corresponds to an acyclic SCM).

Summary graph, example



Summary graph, example





Summary graph of the above full time graph (assuming that the edges of the full time repeat).

Bivariate Granger causality

Granger causality essentially describes whether the past of a process is predictive of the present of another process [Granger, 1969].

Let $X_t = (X_1, X_2)^T$ be a time series.

Definition (Bivariate Granger causality)

We say that X^1 is *Granger noncausal* for X^2 if for all t

$$X_t^2 \perp\!\!\!\perp X_{p(t)}^1 \mid X_{p(t)}^2.$$

Bivariate Granger causality

Granger causality essentially describes whether the past of a process is predictive of the present of another process [Granger, 1969].

Let $X_t = (X_1, X_2)^T$ be a time series.

Definition (Bivariate Granger causality)

We say that X^1 is *Granger noncausal* for X^2 if for all t

$$X_t^2 \perp\!\!\!\perp X_{p(t)}^1 \mid X_{p(t)}^2.$$

Note that we define *Granger noncausality* because this leads to an independence relation, not unlike conditional independence.

Multivariate Granger causality

Let $X_t = (X_1, \ldots, X_n)^T$ be a time series.

Definition (Multivariate Granger causality)

We say that X^i is *Granger noncausal* for X^j if for all t

$$X_t^j \perp\!\!\!\perp X_{p(t)}^i \mid X_{p(t)}^{-i}.$$

where $-i$ denotes $V \setminus \{i\}$.

Multivariate Granger causality

Let $X_t = (X_1, \ldots, X_n)^T$ be a time series.

Definition (Multivariate Granger causality)

We say that X^i is *Granger noncausal* for X^j if for all t

$$X_t^j \perp\!\!\!\perp X_{p(t)}^i \mid X_{p(t)}^{-i}.$$

where $-i$ denotes $V \setminus \{i\}$.

Note, again, that we define *Granger noncausality* because this leads to an independence relation, not unlike conditional independence.

Granger causality and summary graph

If all coordinate processes are observed, one can recover the summary graph from tests of Granger causality (we are still assuming no instantaneous effects).

Theorem (ECI Theorem 10.3)

Assume faithfulness of X and its full time graph. The summary graph of X has an edge $i \to j$ if and only if there exists t such that

$$X_t^j \not\perp\!\!\!\perp X_{p(t)}^i \mid X_{p(t)}^{-i} \tag{1}$$

where $-i = V \setminus \{i\}$.

Granger causality and summary graph

If all coordinate processes are observed, one can recover the summary graph from tests of Granger causality (we are still assuming no instantaneous effects).

Theorem (ECI Theorem 10.3)

Assume faithfulness of X and its full time graph. The summary graph of X has an edge $i \to j$ if and only if there exists t such that

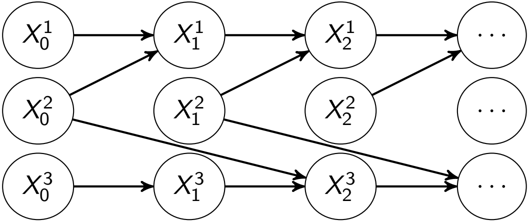
$$X_t^j \not\perp\!\!\!\perp X_{p(t)}^i \mid X_{p(t)}^{-i} \tag{1}$$

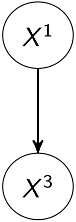
where $-i = V \setminus \{i\}$.

This is a causal interpretation of Granger causality. However, it is important that all coordinate processes are observed.

Granger causality and summary graph

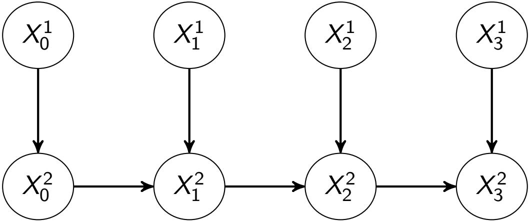
Assume we for all $i, j \in V$ do the test in (1) and construct the corresponding graph. If we do not have full observation (some coordinate processes are unobserved), this graph may still be a useful summary of 'dynamic dependence'. However, there is no reason to think that it is *causal*.





Granger causality, examples

Example with instantaneous effects where X^1 is Granger noncausal for X^2.



Conditional Granger causality

Let $X_t = (X_1, \ldots, X_n)^T$ be a time series.

Definition (Conditional Granger causality)

Let $A, B, C \subseteq V$ be disjoint. We say that X^A is *Granger noncausal* for X^B given X^C if for all t

$$X_t^B \perp\!\!\!\perp X_{p(t)}^A \mid X_{p(t)}^{B \cup C}.$$

Conditional Granger causality

Let $X_t = (X_1, \ldots, X_n)^T$ be a time series.

Definition (Conditional Granger causality)

Let $A, B, C \subseteq V$ be disjoint. We say that X^A is *Granger noncausal* for X^B given X^C if for all t

$$X_t^B \perp\!\!\!\perp X_{p(t)}^A \mid X_{p(t)}^{B \cup C}.$$

Compare with

Definition (Multivariate Granger causality)

We say that X^i is *Granger noncausal* for X^j if for all t

$$X_t^j \perp\!\!\!\perp X_{p(t)}^i \mid X_{p(t)}^{-i}.$$

δ-separation

We know that d-separation in the causal DAG implies conditional independence. Analogously, we can read off conditional Granger noncausality from the summary graph, $\mathcal{S}$, of the causal DAG.

Definition (δ-separation, Eichler [2007], Didelez [2008])

Let $B \subseteq V$. We let $\mathcal{S}^B$ denote the graph obtained from $\mathcal{S}$ by removing all edges out of B, i.e., all edges $k \to l$ such that $k \in B$. We say that B is *δ-separated* from A given C in $\mathcal{S}$ if there are no d-connecting paths between any $i \in A$ and $j \in B$ given C in $\mathcal{S}^B$.

δ-separation

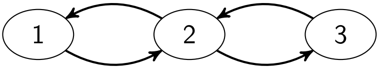
We know that d-separation in the causal DAG implies conditional independence. Analogously, we can read off conditional Granger noncausality from the summary graph, $\mathcal{S}$, of the causal DAG.

Definition (δ-separation, Eichler [2007], Didelez [2008])

Let $B \subseteq V$. We let $\mathcal{S}^B$ denote the graph obtained from $\mathcal{S}$ by removing all edges out of B, i.e., all edges $k \to l$ such that $k \in B$. We say that B is *δ-separated* from A given C in $\mathcal{S}$ if there are no d-connecting paths between any $i \in A$ and $j \in B$ given C in $\mathcal{S}^B$.

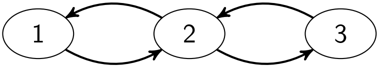
δ-separation is *asymmetric*, that is, if B is δ-separated from A given C, this does not imply that A is δ-separated from B given C (Granger causality is asymmetric too).

δ-separation

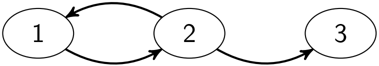


Summary graph, $\mathcal{S}$. Let $A = \{1\}, B = \{3\}, C = \{2\}$.

δ-separation

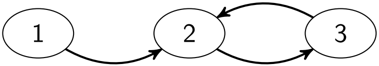


Summary graph, $\mathcal{S}$. Let $A = \{1\}, B = \{3\}, C = \{2\}$.



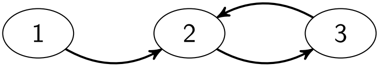
Graph, $\mathcal{S}^B$.

δ-separation

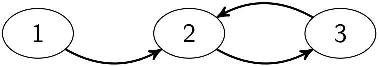


Summary graph, $\mathcal{S}$. Let $A = \{1\}, B = \{3\}, C = \emptyset$. B is not δ-separated from A given C.

δ-separation



Summary graph, $\mathcal{S}$. Let $A = \{1\}, B = \{3\}, C = \emptyset$. B is not δ-separated from A given C.



Summary graph, $\mathcal{S}$. Let $A = \{3\}, B = \{1\}, C = \emptyset$. B is δ-separated from A given C.

Global Markov property in summary graphs

Theorem

Let $A, B, C \subseteq V$ be disjoint and let $\mathcal{S}$ be the summary graph of the causal DAG of X. If B is δ-separated from A given C in $\mathcal{S}$, then A is Granger noncausal for B given C,

$$X_t^B \perp\!\!\!\perp X_{p(t)}^A \mid X_{p(t)}^{B \cup C}.$$

Granger causality as an independence relation

	Causal DAG	Summary graph
nodes	variables	coordinate processes
separation	d	δ
independence relation	conditional indep	Granger (non)causality
margins	ADMGs (m-sep)	DMGs (μ-sep)

The two frameworks are completely analogous. Granger (non)causality can be thought of as an independence relation on the set of coordinate processes. Under full observation, the Markov equivalence classes are singletons (which is different from the DAGs!).

Markov equivalence

Let $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ be DMGs. We say that $\mathcal{G}_1$ and $\mathcal{G}_2$ are *Markov equivalent* if for all $A, B, C \subseteq V$,

$$A \perp_\mu B \mid C \; [\mathcal{G}_1] \Leftrightarrow A \perp_\mu B \mid C \; [\mathcal{G}_2].$$

We use $[\mathcal{G}_1]$ to denote the Markov equivalence class of $\mathcal{G}_1$. What can we say about the Markov equivalence classes?

Markov equivalence

Theorem (Mogensen and Hansen [2020])

Let $\mathcal{G} = (V, E)$ be a DMG. The equivalence class of $\mathcal{G}$ has a greatest element $\mathcal{N} = (V, F)$. That is, if $\tilde{\mathcal{G}} = (V, \tilde{E})$ is Markov equivalent with $\mathcal{G}$, then $\tilde{E} \subseteq F$.

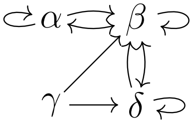
We use the theorem to define the following graph which represents the entire Markov equivalence class.

Definition

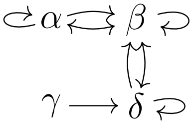
Let $\mathcal{N} = (V, F)$ be a maximal DMG. The corresponding *directed mixed equivalence graph* (DMEG) is the triple $(V, \bar{F}, \tilde{F})$ such that $F = \bar{F} \dot{\cup} \tilde{F}$ where for each $e \in F$ it holds that $e \in \bar{F}$ if and only if $e \in \mathcal{G}$ for every $\mathcal{G} \in [\mathcal{N}]$.

An equivalence class and its DMEG

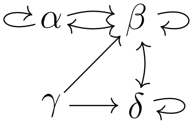
①



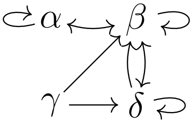
②



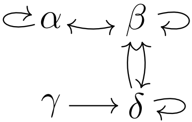
③



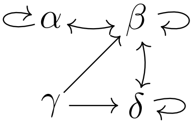
④



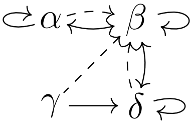
⑤



⑥



⑦



References I

Vanessa Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):245–264, 2008.

Michael Eichler. Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, 137:334–353, 2007.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

Søren Wengel Mogensen and Niels Richard Hansen. Markov equivalence of marginalized local independence graphs. *The Annals of Statistics*, 48(1): 539–559, 2020.

Thomas S. Richardson and Peter Spirtes. Ancestral graph markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.