Causal discovery: constraint-based methods

Charles K. Assaad, Emilie Devijver, Eric Gaussier

charles.assaad@ens-lyon.fr

Table of content

Preliminaries

Causal discovery with causal sufficiency

Tests

Conclusion

Table of content

Preliminaries

Causal discovery with causal sufficiency

Tests

Conclusion

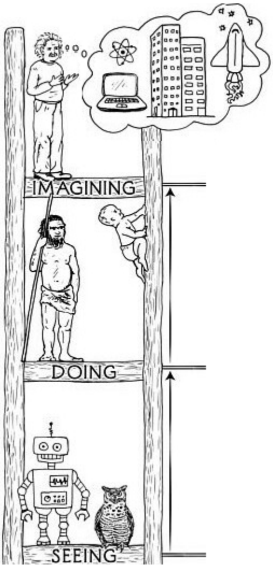
Where does a causal graph come from?

Where does a causal graph come from?

- Theory
  - Sometimes infeasible

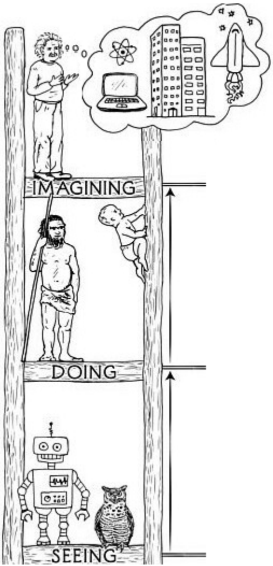
Where does a causal graph come from?

- Theory
  - Sometimes infeasible
- Experts
  - Sometimes infeasible



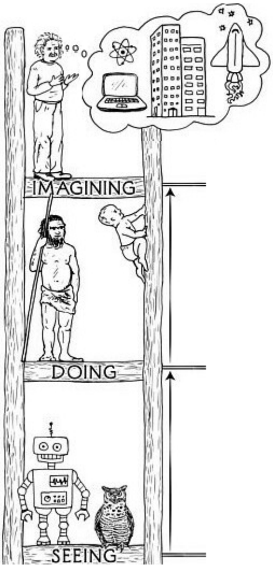
Where does a causal graph come from?

- Theory
  - Sometimes infeasible
- Experts
  - Sometimes infeasible
- Experimentations
  - Sometimes infeasible
  - Sometimes unethical
  - Costly

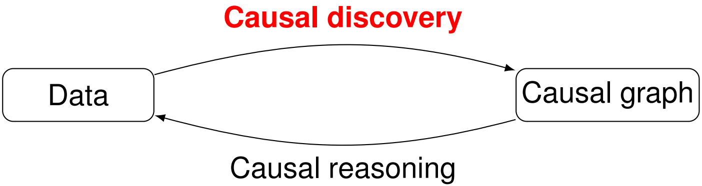


Where does a causal graph come from?

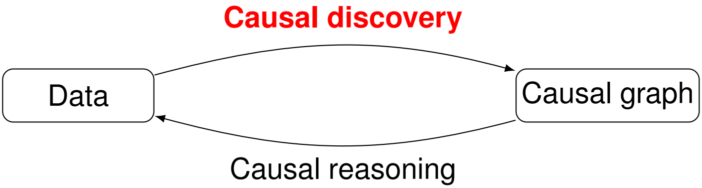
- Theory
  - Sometimes infeasible
- Experts
  - Sometimes infeasible
- Experimentations
  - Sometimes infeasible
  - Sometimes unethical
  - Costly
- Observations
  - Correlation does not imply causation!



Causal discovery (1/2)

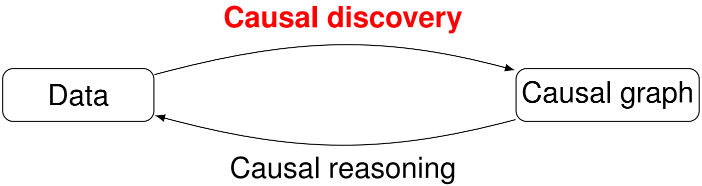


Causal discovery (1/2)



In general, causal discovery from observational data is not possible.

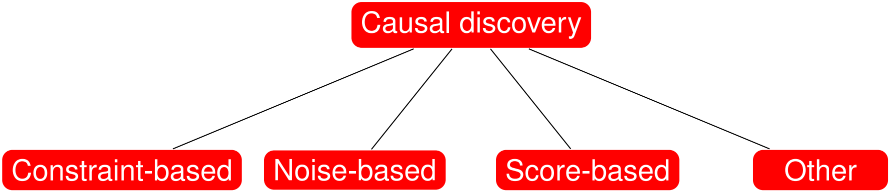
Causal discovery (1/2)



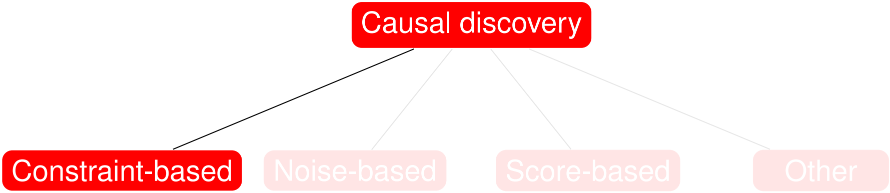
In general, causal discovery from observational data is not possible.

But it is possible under **additional assumptions**.

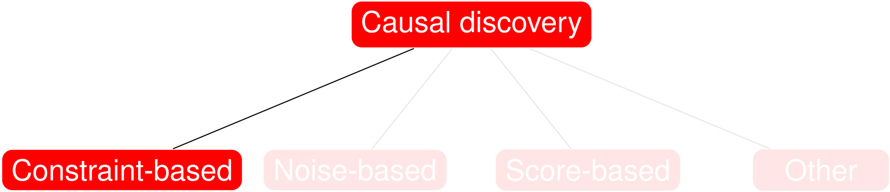
Causal discovery (2/2)



Causal discovery (2/2)



Causal discovery (2/2)



Constraint-based: run local tests of independence to create constraints on space of possible graphs.

Recap about causal graphical models (1/2)

Parental Markov Condition Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,

$$\forall X \in \mathcal{V}, X \perp\!\!\!\perp_P \mathcal{V} \backslash \{Parents(X), Descendants(X)\} \mid Parents(X).$$

Causal sufficiency

$$\forall X \leftarrow Z \rightarrow Y, \text{ if } X, Y \in \mathcal{V} \text{ then } Z \in \mathcal{V}.$$

Skeleton the skeleton of a DAG $\mathcal{G}$ is an undirected graph with same adjacencies as $\mathcal{G}$.

Collider $X \rightarrow Z \leftarrow Y$.

V-structure (or unsheilded colliders, or immorality) If the two parent vertices are not adjacent, the collider is a v-structure.

Recap about causal graphical models (1/2)

Parental Markov Condition Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,

$\forall X \in \mathcal{V}, X \perp\!\!\!\perp_P \mathcal{V} \backslash \{Parents(X), Descendants(X)\} \mid Parents(X)$.

Causal sufficiency

$$\forall X \leftarrow Z \rightarrow Y, \text{ if } X, Y \in \mathcal{V} \text{ then } Z \in \mathcal{V}.$$

Skeleton the skeleton of a DAG $\mathcal{G}$ is an undirected graph with same adjacencies as $\mathcal{G}$.

Collider $X \rightarrow Z \leftarrow Y$.

V-structure (or unsheilded colliders, or immorality) If the two parent vertices are not adjacent, the collider is a v-structure.

Recap about causal graphical models (1/2)

Parental Markov Condition Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,

$$\forall X \in \mathcal{V}, X \perp\!\!\!\perp_P \mathcal{V} \backslash \{Parents(X), Descendants(X)\} \mid Parents(X).$$

Causal sufficiency

$$\forall X \leftarrow Z \rightarrow Y, \text{ if } X, Y \in \mathcal{V} \text{ then } Z \in \mathcal{V}.$$

Skeleton the skeleton of a DAG $\mathcal{G}$ is an undirected graph with same adjacencies as $\mathcal{G}$.

Collider $X \rightarrow Z \leftarrow Y$.

V-structure (or unsheilded colliders, or immorality) If the two parent vertices are not adjacent, the collider is a v-structure.

Recap about causal graphical models (1/2)

Parental Markov Condition Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,

$\forall X \in \mathcal{V}, X \perp\!\!\!\perp_P \mathcal{V} \backslash \{Parents(X), Descendants(X)\} \mid Parents(X)$.

Causal sufficiency

$$\forall X \leftarrow Z \rightarrow Y, \text{ if } X, Y \in \mathcal{V} \text{ then } Z \in \mathcal{V}.$$

Skeleton the skeleton of a DAG $\mathcal{G}$ is an undirected graph with same adjacencies as $\mathcal{G}$.

Collider $X \rightarrow Z \leftarrow Y$.

V-structure (or unsheilded colliders, or immorality) If the two parent vertices are not adjacent, the collider is a v-structure.

Recap about causal graphical models (1/2)

Parental Markov Condition Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,

$$\forall X \in \mathcal{V}, X \perp\!\!\!\perp_P \mathcal{V} \backslash \{Parents(X), Descendants(X)\} \mid Parents(X).$$

Causal sufficiency

$$\forall X \leftarrow Z \rightarrow Y, \text{ if } X, Y \in \mathcal{V} \text{ then } Z \in \mathcal{V}.$$

Skeleton the skeleton of a DAG $\mathcal{G}$ is an undirected graph with same adjacencies as $\mathcal{G}$.

Collider $X \rightarrow Z \leftarrow Y$.

V-structure (or unsheilded colliders, or immorality) If the two parent vertices are not adjacent, the collider is a v-structure.

Recap about causal graphical models (2/2)

Theorem (probabilistic implications of d-separation) Given a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a distribution $P(\mathcal{V})$ compatible with $\mathcal{G}$ and disjoint sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subset \mathcal{V}$:

(i) $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z} \Rightarrow \mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$ in every distribution P compatible with $\mathcal{G}$ (Also known as the global Markov property);

(ii) If $\mathcal{X} \not\perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$, then there exists a distribution P compatible with $\mathcal{G}$ such that $\mathcal{X} \not\perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$;

(ii) If $\mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$ holds in all distributions compatible with $\mathcal{G}$, then $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$.

Theorem (Markov equivalence for DAGs) Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent (have the same d-separations) *iff* they have the same skeleton and the same v-structures.

Recap about causal graphical models (2/2)

Theorem (probabilistic implications of d-separation) Given a DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a distribution $P(\mathcal{V})$ compatible with $\mathcal{G}$ and disjoint sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subset \mathcal{V}$:

(i) $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z} \Rightarrow \mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$ in every distribution P compatible with $\mathcal{G}$ (Also known as the global Markov property);

(ii) If $\mathcal{X} \not\perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$, then there exists a distribution P compatible with $\mathcal{G}$ such that $\mathcal{X} \not\perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$;

(ii) If $\mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} | \mathcal{Z}$ holds in all distributions compatible with $\mathcal{G}$, then $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} | \mathcal{Z}$.

Theorem (Markov equivalence for DAGs) Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent (have the same d-separations) *iff* they have the same skeleton and the same v-structures.

A characterization of Markov equivalence classes for DAGs (1/2)

Completed partially directed acyclic graph (CPDAG) Let $[\mathcal{G}]$ be the Markov equivalence class of a DAG $\mathcal{G}$. The CPDAG $\mathcal{G}^*$ of $\mathcal{G}$ is the graph:

- With the same skeleton as $\mathcal{G}$;
- Where an edge is directed in $\mathcal{G}^*$ iff it occurs as a directed edge with the same orientation in every graph in $[\mathcal{G}]$;
- All other edges are undirected.

A characterization of Markov equivalence classes for DAGs (1/2)

Completed partially directed acyclic graph (CPDAG) Let $[\mathcal{G}]$ be the Markov equivalence class of a DAG $\mathcal{G}$. The CPDAG $\mathcal{G}^*$ of $\mathcal{G}$ is the graph:

- With the same skeleton as $\mathcal{G}$;
- Where an edge is directed in $\mathcal{G}^*$ iff it occurs as a directed edge with the same orientation in every graph in $[\mathcal{G}]$;
- All other edges are undirected.

Proposition Let $\mathcal{G}$ be a DAG and $\mathcal{G}^*$ its CPDAG. Then $\mathcal{G}$ and $\mathcal{G}^*$ are Markov equivalent.

A characterization of Markov equivalence classes for DAGs (1/2)

Completed partially directed acyclic graph (CPDAG) Let $[\mathcal{G}]$ be the Markov equivalence class of a DAG $\mathcal{G}$. The CPDAG $\mathcal{G}^*$ of $\mathcal{G}$ is the graph:

- With the same skeleton as $\mathcal{G}$;
- Where an edge is directed in $\mathcal{G}^*$ iff it occurs as a directed edge with the same orientation in every graph in $[\mathcal{G}]$;
- All other edges are undirected.

Proposition Let $\mathcal{G}$ be a DAG and $\mathcal{G}^*$ its CPDAG. Then $\mathcal{G}$ and $\mathcal{G}^*$ are Markov equivalent.
Proof: Follows immediately by Theorem (Markov equivalence for DAGs) and by Definition of CPDAG.

A characterization of Markov equivalence classes for DAGs (2/2)

Lemma Let $\mathcal{G}_1^*$ and $\mathcal{G}_2^*$ denote two CPDAGs then $\mathcal{G}_1^* = \mathcal{G}_2^*$ iff $\mathcal{G}_1^*$ and $\mathcal{G}_2^*$ belong to the same Markov equivalent class.

A characterization of Markov equivalence classes for DAGs (2/2)

Lemma Let $\mathcal{G}_1^*$ and $\mathcal{G}_2^*$ denote two CPDAGs then $\mathcal{G}_1^* = \mathcal{G}_2^*$ iff $\mathcal{G}_1^*$ and $\mathcal{G}_2^*$ belong to the same Markov equivalent class.
Proof: Follows immediately by Theorem (Markov equivalence for DAGs) and by Definition of CPDAG.

A characterization of Markov equivalence classes for DAGs (2/2)

Lemma Let $\mathcal{G}_1^*$ and $\mathcal{G}_2^*$ denote two CPDAGs then $\mathcal{G}_1^* = \mathcal{G}_2^*$ iff $\mathcal{G}_1^*$ and $\mathcal{G}_2^*$ belong to the same Markov equivalent class.

Proof: Follows immediately by Theorem (Markov equivalence for DAGs) and by Definition of CPDAG.

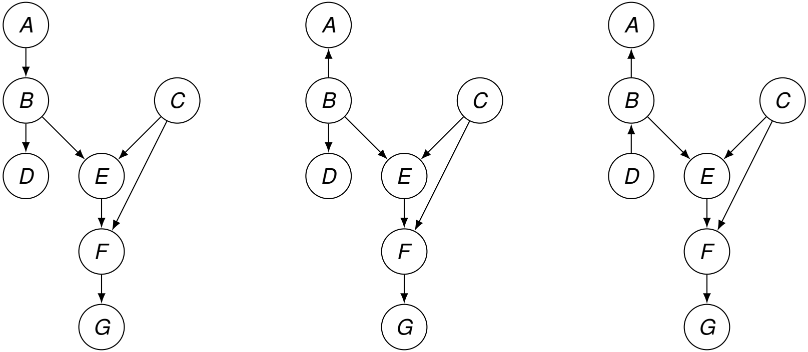
$\Longrightarrow$ All graphs in the same Markov equivalent class have the same CPDAG.

A characterization of Markov equivalence classes for DAGs (2/2)

Lemma Let $\mathcal{G}_1^*$ and $\mathcal{G}_2^*$ denote two CPDAGs then $\mathcal{G}_1^* = \mathcal{G}_2^*$ iff $\mathcal{G}_1^*$ and $\mathcal{G}_2^*$ belong to the same Markov equivalent class.

Proof: Follows immediately by Theorem (Markov equivalence for DAGs) and by Definition of CPDAG.

$\Longrightarrow$ All graphs in the same Markov equivalent class have the same CPDAG.

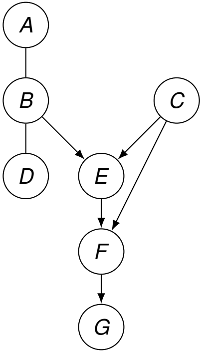


A characterization of Markov equivalence classes for DAGs (2/2)

Lemma Let $\mathcal{G}_1^*$ and $\mathcal{G}_2^*$ denote two CPDAGs then $\mathcal{G}_1^* = \mathcal{G}_2^*$ iff $\mathcal{G}_1^*$ and $\mathcal{G}_2^*$ belong to the same Markov equivalent class.

Proof: Follows immediately by Theorem (Markov equivalence for DAGs) and by Definition of CPDAG.

$\Longrightarrow$ All graphs in the same Markov equivalent class have the same CPDAG.



Constraint based question

Main question: Given $P(\mathcal{V})$ a compatible probability distribution of $\mathcal{G}$, can we discover $\mathcal{G}^*$ the CPDAG of $\mathcal{G}$?

Constraint based question

Main question: Given $P(\mathcal{V})$ a compatible probability distribution of $\mathcal{G}$, can we discover $\mathcal{G}^*$ the CPDAG of $\mathcal{G}$? No!

Constraint based question

Main question: Given $P(\mathcal{V})$ a compatible probability distribution of $\mathcal{G}$, can we discover $\mathcal{G}^*$ the CPDAG of $\mathcal{G}$? No!

Because $X \perp\!\!\!\perp_P Y \mid Z \not\Longrightarrow X \perp\!\!\!\perp_G Y \mid Z$.

Faithfulness

Faithfulness We say that a graph $\mathcal{G}$ and a compatible probability distribution P are faithful to one another if all and only the conditional independence relations true in P are entailed by the Markov condition applied to $\mathcal{G}$.

faithfulness and d-sep

Theorem (implication of faithfulness on d-sep) $P(\mathcal{V})$ is faithful to directed acyclic graph $\mathcal{G}$ with vertex set $\mathcal{V}$ iff for all disjoint sets of vertices $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subset \mathcal{V}$, $\mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} \mid \mathcal{Z}$ iff $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} \mid \mathcal{Z}$.

faithfulness and d-sep

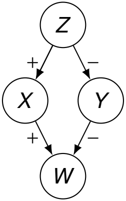
Theorem (implication of faithfulness on d-sep) $P(\mathcal{V})$ is faithful to directed acyclic graph $\mathcal{G}$ with vertex set $\mathcal{V}$ iff for all disjoint sets of vertices $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subset \mathcal{V}$, $\mathcal{X} \perp\!\!\!\perp_P \mathcal{Y} \mid \mathcal{Z}$ iff $\mathcal{X} \perp\!\!\!\perp_G \mathcal{Y} \mid \mathcal{Z}$.

Proof: Follows immediately by Theorem (probabilistic implication on d-separation) and by Definition of faithfulness.

Violation of faithfulness (1/2)

Example 1: Canceling out

Consider



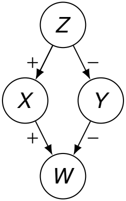
where

- $Z = \epsilon_z$
- $X = a_{zx} \times Z + \epsilon_x$
- $Y = a_{zy} \times Z + \epsilon_y$
- $W = a_{xw} \times X - \frac{a_{zx} a_{xw}}{a_{zy}} \times Y + \epsilon_w$

Violation of faithfulness (1/2)

Example 1: Canceling out

Consider



where

- $Z = \epsilon_z$
- $X = a_{zx} \times Z + \epsilon_x$
- $Y = a_{zy} \times Z + \epsilon_y$
- $W = a_{xw} \times X - \frac{a_{zx} a_{xw}}{a_{zy}} \times Y + \epsilon_w$

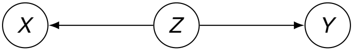
By canceling out

- $Z \perp\!\!\!\perp_P W$

Violation of faithfulness (2/2)

Example 2: Determinism

Consider



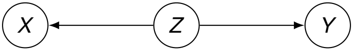
where

- $Z = \epsilon_z$
- $X = a_{zx} \times Z + \epsilon_x$
- $Y = a_{xy} \times Z$

Violation of faithfulness (2/2)

Example 2: Determinism

Consider



where

- $Z = \epsilon_Z$
- $X = a_{zx} \times Z + \epsilon_X$
- $Y = a_{xy} \times Z$

By determinism

- $X \perp\!\!\!\perp_P Z \mid Y$

Table of content

Preliminaries

Causal discovery with causal sufficiency

Tests

Conclusion

Finding skeleton and v-structures

Theorem (faithfulness, adjacencies and v-structures) If $P(\mathcal{V})$ is faithful to some directed acyclic graph, then $P(\mathcal{V})$ is faithful to directed acyclic graph $\mathcal{G}$ with vertex $\mathcal{V}$ iff:

- For $X, Y \in \mathcal{V}$, X and Y are adjacent iff $\forall \mathcal{S} \subseteq \mathcal{V} \backslash \{X, Y\}$, $X \not\perp\!\!\!\perp_P Y \mid \mathcal{S}$;
- For $X, Y, Z \in \mathcal{V}$ such that X is adjacent to Z and Z is adjacent to Y and X and Y are not adjacent, $X \rightarrow Z \leftarrow Y$ in $\mathcal{G}$ iff $\forall \mathcal{S} \in \mathcal{V} \backslash \{X, Y\}$ such that $Z \in \mathcal{S}$, $X \not\perp\!\!\!\perp_P Y \mid \mathcal{S}$.

(proof in (Verma and Pearl, 1992))

Finding skeleton and v-structures

Theorem (faithfulness, adjacencies and v-structures) If $P(\mathcal{V})$ is faithful to some directed acyclic graph, then $P(\mathcal{V})$ is faithful to directed acyclic graph $\mathcal{G}$ with vertex $\mathcal{V}$ iff:

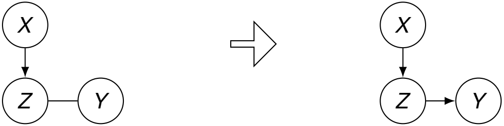
- For $X, Y \in \mathcal{V}$, X and Y are adjacent iff $\forall \mathcal{S} \subseteq \mathcal{V} \backslash \{X, Y\}$, $X \not\perp\!\!\!\perp_P Y \mid \mathcal{S}$;
- For $X, Y, Z \in \mathcal{V}$ such that X is adjacent to Z and Z is adjacent to Y and X and Y are not adjacent, $X \rightarrow Z \leftarrow Y$ in $\mathcal{G}$ iff $\forall \mathcal{S} \in \mathcal{V} \backslash \{X, Y\}$ such that $Z \in \mathcal{S}$, $X \not\perp\!\!\!\perp_P Y \mid \mathcal{S}$.

(proof in (Verma and Pearl, 1992))

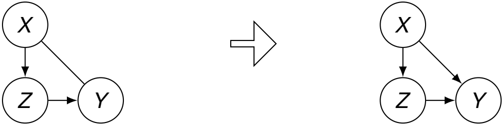
- Point 1 can be used to discover the skeleton of $\mathcal{G}$ from $P(\mathcal{V})$;
- Given the skeleton of $\mathcal{G}$, point 2 can be used to find all v-structures.

Orientation rules

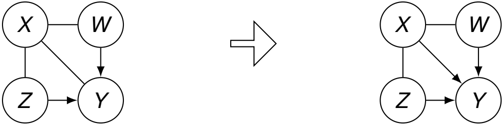
R1:



R2:



R3:



Orientation rules correctness

Pattern A pattern of a DAG $\mathcal{G}$ is a graph with the same skeleton as $\mathcal{G}$ but where only v-structures are oriented.

Orientation rules correctness

Pattern A pattern of a DAG $\mathcal{G}$ is a graph with the same skeleton as $\mathcal{G}$ but where only v-structures are oriented.

Theorem (orientation soundness) Given a pattern of some DAG, the three orientation rules $R1$, $R2$, $R3$ are sound.
(proof on board)

Orientation rules correctness

Pattern A pattern of a DAG $\mathcal{G}$ is a graph with the same skeleton as $\mathcal{G}$ but where only v-structures are oriented.

Theorem (orientation soundness) Given a pattern of some DAG, the three orientation rules $R1$, $R2$, $R3$ are sound.
(proof on board)

Theorem (orientation completeness) The result of recursively applying rules $R1$, $R2$, $R3$ to a pattern of some DAG is a CPDAG.
(proof in (Meek, 1995))

The SGS algorithm

Algorithm 1 SGS

Input: $P(\mathcal{V})$
Output: CPDAG $\mathcal{G}^*$

Form the complete undirected graph $\mathcal{G}^*$ on vertex set $\mathcal{V}$
for all $X - Y$ in $\mathcal{G}^*$
and subsets $\mathcal{S} \subseteq \mathcal{V} \backslash \{X, Y\}$ **do**
  if $\exists \mathcal{S} \subseteq \mathcal{V} \backslash \{X, Y\}$ such that $X \perp\!\!\!\perp_P Y \mid \mathcal{S}$ **then**
    Delete edge $X - Y$ from $\mathcal{G}^*$
  end if
end for
for all $X - Z - Y$ in $\mathcal{G}^*$ such that $X \notin Adj(Y, \mathcal{G})$ **do**
  if $\nexists\, \mathcal{S} \subseteq \mathcal{V} \backslash \{X, Y\}$ such that $Z \in \mathcal{S}$ and $X \perp\!\!\!\perp_P Y \mid \mathcal{S}$ **then**
    Orient $X \rightarrow Z \leftarrow Y$ in $\mathcal{G}^*$
  end if
end for
Recursively apply rules R1-R3 until no more edges can be oriented
Return $\mathcal{G}^*$

$Adj(Y, \mathcal{G})$: Adjacencies of Y in $\mathcal{G}$

Correctness of SGS

Theorem (correctness) Assume the distribution $P(\mathcal{V})$ is Markov and faithful to some DAG $\mathcal{G}$ and assume that we are given perfect conditional independence information about all pairs of variables. Let $\mathcal{G}^*$ be the CPDAG of $\mathcal{G}$. The SGS algorithm returns $\mathcal{G}^*$.

Correctness of SGS

Theorem (correctness) Assume the distribution $P(\mathcal{V})$ is Markov and faithful to some DAG $\mathcal{G}$ and assume that we are given perfect conditional independence information about all pairs of variables. Let $\mathcal{G}^*$ be the CPDAG of $\mathcal{G}$. The SGS algorithm returns $\mathcal{G}^*$.

Proof: By Theorem (faithfulness, adjacencies and v-structures), Theorem (orientation soundness) and Theorem (orientation completness).

Computational complexity of SGS

Running time of SGS depends *exponentially* on the *number of vertices* in the graph:

- For all pairs check all subsets;
- For all triples check all subsets.

A better approach?

Optimizing the procedure for skeleton construction

A better approach?

Optimizing the procedure for skeleton construction

By the Parental Markov condition:

$X \notin Adj(Y, \mathcal{G})$ iff $X \perp\!\!\!\perp_P Y \mid Parents(X, \mathcal{G})$ or $X \perp\!\!\!\perp_P Y \mid Parents(Y, \mathcal{G})$

A better approach?

Optimizing the procedure for skeleton construction

By the Parental Markov condition:

$X \notin Adj(Y, \mathcal{G})$ iff $X \perp\!\!\!\perp_P Y \mid Parents(X, \mathcal{G})$ or $X \perp\!\!\!\perp_P Y \mid Parents(Y, \mathcal{G})$

Since the graph $\mathcal{G}$ is unknown:

- The parent set is unknown ahead of time;
- We look at $\mathcal{S} \subseteq Adj(X, \mathcal{G}')$ and $\mathcal{S}' \subseteq Adj(Y, \mathcal{G}')$ for some G' which is a supergraph of the true unknown skeleton;
- We can pursue an iterative strategy such that we increase the size of $\mathcal{S}$ iteratively.

A better approach?

Optimizing the procedure for finding v-structures

A better approach?

Optimizing the procedure for finding v-structures

Lemma (either d-sep or d-connect) Given the distribution $P(V)$ that is Markov and faithful to some DAG $\mathcal{G}$, if $Z \in Adj(X, \mathcal{G})$, $Z \in Adj(Y, \mathcal{G})$ and $Y \notin Adj(X, \mathcal{G})$, then either Z is in every set of variables that d-separates X and Y or it is in no set of variables that d-separates X and Y.
(proof on board)

A better approach?

Optimizing the procedure for finding v-structures

Lemma (either d-sep or d-connect) Given the distribution $P(V)$ that is Markov and faithful to some DAG $\mathcal{G}$, if $Z \in Adj(X, \mathcal{G})$, $Z \in Adj(Y, \mathcal{G})$ and $Y \notin Adj(X, \mathcal{G})$, then either Z is in every set of variables that d-separates X and Y or it is in no set of variables that d-separates X and Y.
(proof on board)

$sepset(X, Y)$: subset that permitted the separation of X and Y during the skeleton construction.

A better approach?

Optimizing the procedure for finding v-structures

Lemma (either d-sep or d-connect) Given the distribution $P(V)$ that is Markov and faithful to some DAG $\mathcal{G}$, if $Z \in Adj(X, \mathcal{G})$, $Z \in Adj(Y, \mathcal{G})$ and $Y \notin Adj(X, \mathcal{G})$, then either Z is in every set of variables that d-separates X and Y or it is in no set of variables that d-separates X and Y.
(proof on board)

$sepset(X, Y)$: subset that permitted the separation of X and Y during the skeleton construction.

R0: For all triples $X - Z - Y \in \mathcal{G}^*$ such that $Y \notin Adj(X, \mathcal{G}^*)$, if $Z \notin sepset(X, Y)$ then orient $X \rightarrow Z \leftarrow Y$ in $\mathcal{G}^*$.

The PC algorithm

Algorithm 2 PC

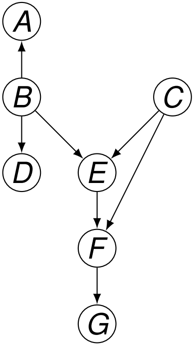
Input: $P(\mathcal{V})$

Output: CPDAG $\mathcal{G}^*$

Form the complete undirected graph $\mathcal{G}^*$ on vertex set $\mathcal{V}$
Let $n = 0$
repeat
    for all $X - Y$ in $\mathcal{G}^*$ such that $|Adj(X, \mathcal{G}^*)| \geq n$
    and subsets $\mathcal{S} \subseteq Adj(X, \mathcal{G}^*) \setminus \{Y\}$ such that $|\mathcal{S}| = n$ **do**
        if $X \perp\!\!\!\perp_P Y \mid \mathcal{S}$ **then**
            Delete edge $X - Y$ from $\mathcal{G}^*$
            Let $sepset(X, Y) = sepset(Y, X) = \mathcal{S}$
        end if
    end for
    Let $n = n + 1$
until for each pair of adjacent vertices (X, Y), $|Adj(X, \mathcal{G}^*) \setminus \{Y\}| \leq n$
Apply R0
Recursively apply rules R1-R3 until no more edges can be oriented
Return $\mathcal{G}^*$

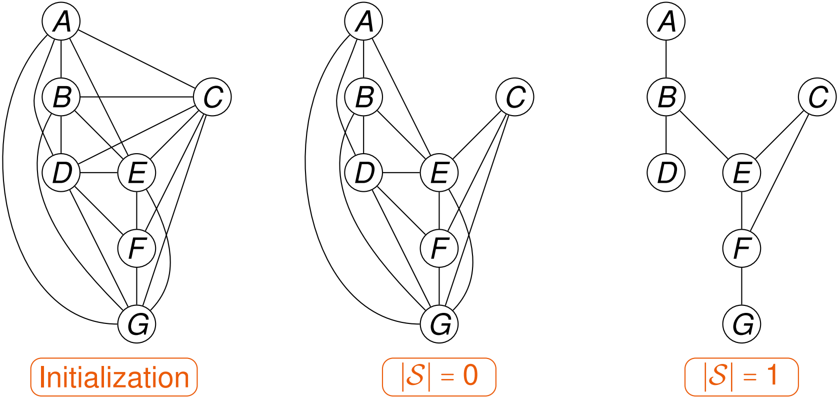
PC in action (1/3)

- Suppose the true graph on right;
- Assumptions: CMC, faithfulness, causal sufficiency.



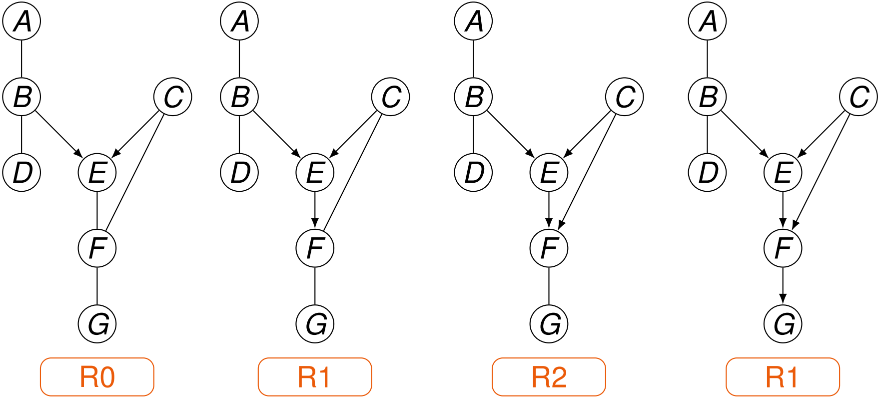
PC in action (2/3)

Skeleton construction:



PC in action (3/3)

Orientation:



Correctness of PC

Theorem (correctness) Assume the distribution $P(\mathcal{V})$ is Markov and faithful to some DAG $\mathcal{G}$ and assume that we are given perfect conditional independence information about all pairs of variables. Let $\mathcal{G}^*$ be the CPDAG of $\mathcal{G}$. The PC algorithm returns $\mathcal{G}^*$.
(proof in (Spirtes and Glymour, 1991))

Computational complexity of PC

Running time of PC depends *exponentially* on the *maximal degree* of the graph **but** for a fixed maximal degree running time over the *number of vertices* is *polynomial*.

Exercise 1

Consider data that are generated from a chain $X \rightarrow Y \rightarrow Z$. Assuming that all assumptions are satisfied, which CPDAG would a constraint based causal discovery algorithm report?

If you could supply prior knowledge to the algorithm on only one arc that is required to be present, what arc (if any) would allow the entire structure to be learned? Explain briefly.

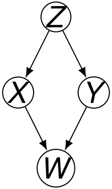
Exercise 2

Consider data that truly come from a fork $X \leftarrow Y \rightarrow Z$.
Assuming that all assumptions are satisfied, which CPDAG would a constraint based causal discovery algorithm report?

If you could supply prior knowledge to the algorithm on only one arc that is required to be present, what arc (if any) would allow the entire structure to be learned? Explain briefly.

Exercise 3

- Suppose the true graph on right;
- Assumptions: CMC, causal sufficiency, no deterministic relations;
- Generative process:

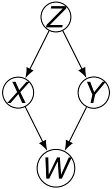


$$Z = \xi_z \qquad \xi_z \sim N(0,1);$$
$$X = a * Z + \xi_x \qquad \xi_x \sim N(0,1);$$
$$Y = b * Z + \xi_y \qquad \xi_y \sim N(0,1);$$
$$W = c * X - \frac{a * c}{b} * Y + \xi_w \qquad \xi_w \sim N(0,1).$$

- Given a compatible distribution what would be the output of the PC algorithm?

Exercise 4

- Suppose the true graph on right;
- Assumptions: CMC, causal sufficiency, deterministic relations, no canceling out paths;
- Given a compatible distribution what would be the output of the PC algorithm?



Exercise 5

- Suppose the true graph on right;
- Assumptions: CMC, faithfulness;
- Given a compatible distribution what would be the output of the PC algorithm if Z is unobserved?

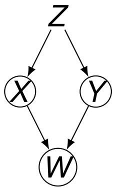


Table of content

Preliminaries

Causal discovery with causal sufficiency

Tests

Conclusion

Conditional independence tests

With finit data, SGS and PC needs a procedure for deciding whether $X \perp\!\!\!\perp_P Y \mid \mathcal{S}$.

In practice, test the null hypothesis:

$$H_0 : X \perp\!\!\!\perp_P Y \mid \mathcal{S}$$

and reject the null hypothesis if some test statistic $T(x) < \alpha$, where α is a user-specified significance threshold. That is, if we reject the null hypothesis, we keep the edge, and if we fail to reject, we remove the edge.

Examples of conditional independence tests

Tests	Assumptions
Fisher Z-transform	Linear, gaussian
χ^2 test	Multinomial discrete
Kernel-based CI test	-
Local permutation test	-

Consistency

Theorem (consistency) Assume the distribution $P(\mathcal{V})$ is Markov and faithful to some DAG $\mathcal{G}$. Let $\mathcal{G}^*$ be the CPDAG of $\mathcal{G}$ and let $\hat{\mathcal{G}}^*$ be the output of SGS, PC with some consistent conditional independence test and significative level α. Then there is a sequence of $\alpha_n \to 0 (n \to \infty)$ such that $\lim_{n\to\infty} \Pr(\hat{\mathcal{G}}^* = \mathcal{G}^*) = 1$.
(proof in (Spirtes et al, 2000))

Exercise 7

As the significance level is lowered to 0, what would you expect to happen to the graph skeleton learned by constraint based causal discovery algorithms? As the significance level is increased to 1? Explain.

Table of content

Preliminaries

Causal discovery with causal sufficiency

Tests

Conclusion

Conclusion

- Under faithfulness and causal sufficiency constraint-based methods can discover a CPDAG (SGS, PC).
- Advantages:
  - Nonparametric (in principle);
  - PC is relatively scalable;
  - Lots of work on improvements.
- Drawbacks:
  - Cannot discover the entire true graph;
  - Faithfulness is not testable;
  - Cannot parallelize;
  - No confidence intervals;
  - Individual errors may propagate.

Conclusion

- Under faithfulness and causal sufficiency constraint-based methods can discover a CPDAG (SGS, PC).
- Advantages:
  - Nonparametric (in principle);
  - PC is relatively scalable;
  - Lots of work on improvements.
- Drawbacks:
  - Cannot discover the entire true graph;
  - Faithfulness is not testable;
  - Cannot parallelize;
  - No confidence intervals;
  - Individual errors may propagate.

Conclusion

- Under faithfulness and causal sufficiency constraint-based methods can discover a CPDAG (SGS, PC).
- Advantages:
  - Nonparametric (in principle);
  - PC is relatively scalable;
  - Lots of work on improvements.
- Drawbacks:
  - Cannot discover the entire true graph;
  - Faithfulness is not testable;
  - Cannot parallelize;
  - No confidence intervals;
  - Individual errors may propagate.

Some extensions

- Causal discovery without causal sufficiency;
- Incorporating background knowledge;
- Order independent;
- Time series.

Some extensions

- Causal discovery without causal sufficiency;
- Incorporating background knowledge;
- Order independent;
- Time series.

Some extensions

- Causal discovery without causal sufficiency;
- Incorporating background knowledge;
- Order independent;
- Time series.

Some extensions

- Causal discovery without causal sufficiency;
- Incorporating background knowledge;
- Order independent;
- Time series.

References (1/2)

Direct inspirations for part 1

1. *Causation, Prediction, and Search*, P. Spirtes, C. Glymour , R. Scheines. MIT Press, 2nd edition, 2000
2. *An Algorithm for Fast Recovery of Sparse Causal Graphs*, P. Spirtes , C. Glymour. Social Science Computer Review, 1991
3. *An algorithm for deciding if a set of observed independencies has a causal explanation*, T. Verma, J. Pearl. Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence, 1992
4. *Causal inference and causal explanation with background knowledge*, C. Meek. Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, 1995
5. *A characterization of Markov equivalence classes for acyclic digraphs*, S. A. Andersson and D. Madigan and M. D. Perlman. Annals of Statistics, 1997
6. *Learning equivalence classes of bayesian-network structures*, D. M. Chickering. JMLR, 2002

References (2/2)

Additional readings

1. *Weakening faithfulness: some heuristic causal discovery algorithms*, Zhalma, J. Zhang, W. Mayer. International Journal of Data Science and Analytics, 2017
2. *Order-Independent Constraint-Based Causal Structure Learning*, D. Colombo, M. Maathuis. JMLR, 2014
3. *Learning high-dimensional directed acyclic graphs with latent and selection variables*, D. Colombo, M. Maathuis, M. Kalisch, T. Richardson. The Annals of Statistics, 2011
4. *Uniform Consistency in Causal Inference*, J. Robins. Biometrika, 2003
5. *Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm*, M. Kalisch, P. Bühlmann. JMLR, 2007
6. *Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets*, J. Runge. Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence, 2020