Causal discovery: noise-based methods

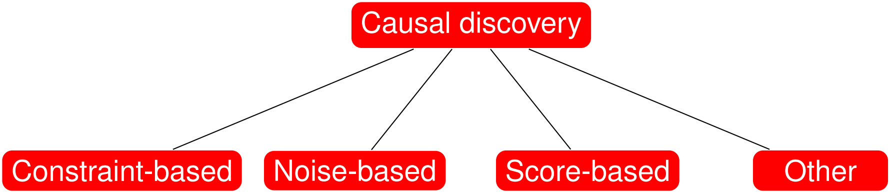
Charles Assaad, Emilie Devijver

charles.assaad@ens-lyon.fr

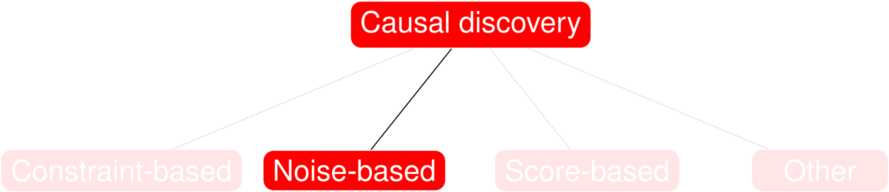
Table of content

Table of content

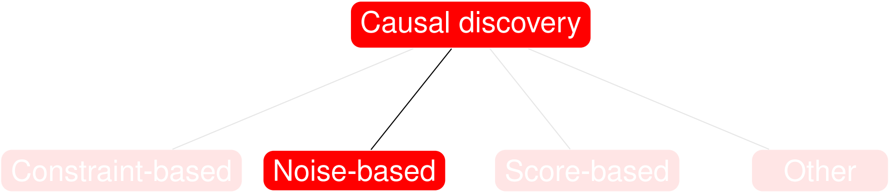
Causal discovery



Causal discovery



Causal discovery



Noise-based: find footprints in the noise that imply causal asymmetry.

Recap about causal graphical models

Causal sufficiency

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

Topological ordering: Consider a causal DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a topological ordering $\mathcal{T} = \{X_1, \cdots, X_p\}$. If $X_i \rightarrow X_j$ in $\mathcal{G}$ then $i < j$.

The intuition behind the noise (1/2)

$$\textit{Suppose} \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

The intuition behind the noise (1/2)

$$\textit{Suppose} \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

Given $P(X, Y)$, one can detect $X - Y$ but what about orientation?

The intuition behind the noise (1/2)

$$Suppose \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

Given $P(X, Y)$, one can detect $X - Y$ but what about orientation?

$Y := 2X + \xi_y$?
or
$X := \frac{Y}{2} + \hat{\xi}_x$?

The intuition behind the noise (1/2)

$$Suppose \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

Given $P(X, Y)$, one can detect $X - Y$ but what about orientation?

$Y := 2X + \xi_y$?
or
$X := \frac{Y}{2} + \hat{\xi}_x$?

Without further assumption we cannot know.

The intuition behind the noise (1/2)

$$\textit{Suppose} \begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

Given $P(X, Y)$, one can detect $X - Y$ but what about orientation?

$Y := 2X + \xi_y$?
or
$X := \frac{Y}{2} + \hat{\xi}_x$?

Without further assumption we cannot know.

Assume that the noise follow a uniform distribution on $\{-1, 0, 1\}$

The intuition behind the noise (1/2)

$$\textit{Suppose}\begin{cases} X := \xi_x \\ Y := 2X + \xi_y \end{cases}$$

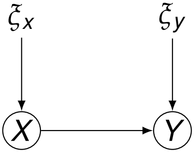
Given $P(X, Y)$, one can detect $X - Y$ but what about orientation?

$Y := 2X + \xi_y$?
or
$X := \frac{Y}{2} + \hat{\xi}_x$?

Without further assumption we cannot know.

Assume that the noise follow a uniform distribution on $\{-1, 0, 1\}$

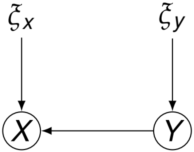
X	Y	$\xi_y = Y - 2X$	$\hat{\xi}_x = X - Y/2$
1	2	$0 \in \{-1, 0, 1\}$	$0 \in \{-1, 0, 1\}$
3	6	$0 \in \{-1, 0, 1\}$	$0 \in \{-1, 0, 1\}$
4	9	$1 \in \{-1, 0, 1\}$	$-0.5 \notin \{-1, 0, 1\}$



$$M_1 : \begin{cases} X := f_x(\xi_x) \\ Y := f_y(X, \xi_y) \end{cases}$$

- $X \perp\!\!\!\perp_G \xi_y$
- $Y \not\perp\!\!\!\perp_G \xi_x$

Backwards model:



$$M_2 : \begin{cases} Y := g_y(\xi_y) \\ X := g_x(Y, \xi_x) \end{cases}$$

- $X \not\perp\!\!\!\perp_G \xi_y$
- $Y \perp\!\!\!\perp_G \xi_x$

Noise based question

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

Noise based question

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

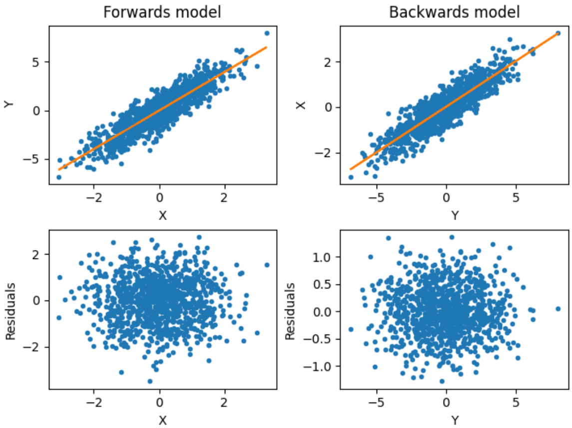
Noise based question

Main question: Given $P(\mathcal{V})$ a compatible probability distribution of $\mathcal{G}$, can we discover $\mathcal{G}$? No!
It is possible that $Y \perp\!\!\!\perp_P \hat{\xi}_x$.

Noise based question

Main question: Given $P(\mathcal{V})$ a compatible probability distribution of $\mathcal{G}$, can we discover $\mathcal{G}$? No!
It is possible that $Y \perp\!\!\!\perp_P \hat{\xi}_x$.
Example:

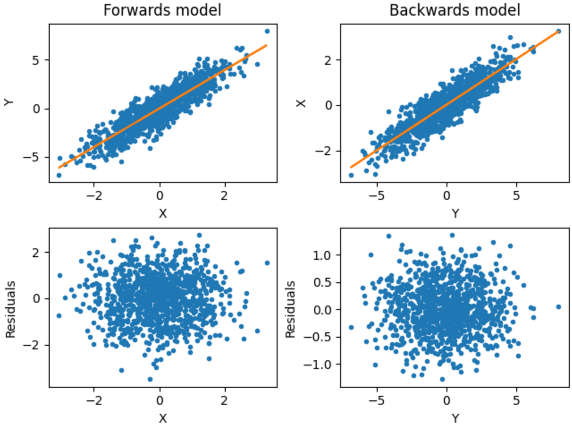
$$X \sim N(0,1)$$
$$\xi_y \sim N(0,1)$$
$$Y := 2X + \xi_y$$



Noise based question

Main question: Given $P(\mathcal{V})$ a compatible probability distribution of $\mathcal{G}$, can we discover $\mathcal{G}$? No!
It is possible that $Y \perp\!\!\!\perp_P \hat{\xi}_x$.
Example:

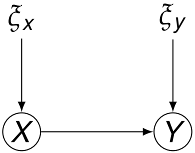
$X \sim N(0,1)$
$\xi_y \sim N(0,1)$
$Y := 2X + \xi_y$



$\Longrightarrow$ The Markov equivalence class is the best we can do!

Table of content

The linear case (1/2)

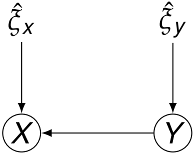


$$M_1 : \begin{cases} X := \xi_x \\ Y := aX + \xi_y \end{cases}$$

- $X \perp\!\!\!\perp_G \xi_y$
- $Y \not\perp\!\!\!\perp_G \xi_x$

When $Y \perp\!\!\!\perp_P \hat{\xi}_x$?

Backwards model:



$$M_2 : \begin{cases} Y := \hat{\xi}_y \\ X := bY + \hat{\xi}_x \end{cases}$$

$$\begin{aligned} \hat{\xi}_x &= X - bY \\ &= X - b(aX + \xi_y) \\ &= (1 - ba)X - b\xi_y \end{aligned}$$

The linear case (2/2)

$$Y = aX + \xi_y$$
$$\hat{\xi}_x = (1 - ba)X - b\xi_y$$

When $Y \perp\!\!\!\perp_P \hat{\xi}_x$?

The linear case (2/2)

$$Y = aX + \xi_y$$
$$\hat{\xi}_x = (1 - ba)X - b\xi_y$$

When $Y \perp\!\!\!\perp_P \hat{\xi}_x$?

Theorem (Darmois-Skitovich): Let $X_1, \cdots, X_n$ be independent, non degenerate random variables. If for two linear combinations:

$$l_1 = a_1 X_1 + \cdots + a_n X_n$$
$$l_2 = b_1 X_1 + \cdots + b_n X_n$$

are independent, then each X_i is normally distributed.

The linear non gaussian case (1/2)

Theorem (identiability of linear non-Gaussian models): Assume that $P(X, Y)$ admits the linear model

$$Y := aX + \xi_y, \qquad X \perp\!\!\!\perp_P \xi_y,$$

with continuous random variables X, ξ_y, and Y. Then there exists $b \in \mathbb{R}$ and a random variable $\hat{\xi}_x$ such that

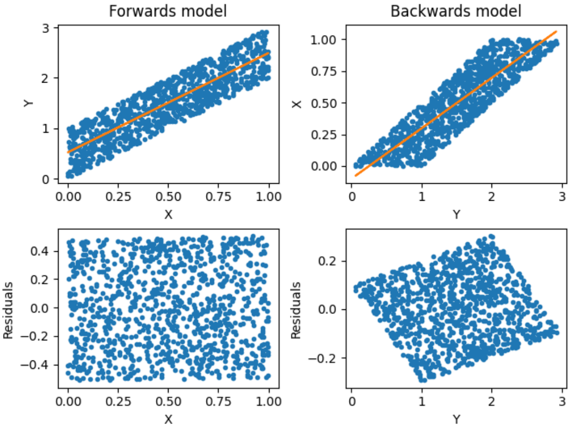
$$X := bY + \hat{\xi}_x, \qquad Y \perp\!\!\!\perp_P \hat{\xi}_x,$$

if and only if ξ_y and X are Gaussian.
(proof on board)

The linear non gaussian case (2/2)

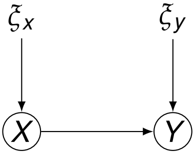
Example:

$$X \sim U(0,1)$$
$$\xi_y \sim U(0,1)$$
$$Y := 2X + \xi_y$$



The non linear case (1/3)

Continuous additive noise models



$$M_1 : \begin{cases} X := \xi_x \\ Y := f_y(X) + \xi_y \end{cases}$$

- $X \perp\!\!\!\perp_G \xi_y$
- $Y \not\perp\!\!\!\perp_G \xi_x$

When $Y \perp\!\!\!\perp_P \hat{\xi}_x$?

The non linear case (2/3)

Theorem (identiability of additive noise models): Assume that $P(X, Y)$ admits the non-linear additive noise model

$$Y := f_y(X) + \xi_y, \qquad X \perp\!\!\!\perp_P \xi_y,$$

with continuous random variables X, ξ_y, and Y. Then there exists $g()$ and random variable $\hat{\xi}_x$ such that

$$X := f_x(Y) + \hat{\xi}_x, \qquad Y \perp\!\!\!\perp_P \hat{\xi}_x,$$

if and only if *Complicated Condition* is satisfied.
(Hoyer et al, 2008)

The non linear case (2/3)

Theorem (identiability of additive noise models): Assume that $P(X, Y)$ admits the non-linear additive noise model

$$Y := f_y(X) + \xi_y, \qquad X \perp\!\!\!\perp_P \xi_y,$$

with continuous random variables X, ξ_y, and Y. Then there exists $g()$ and random variable $\hat{\xi}_x$ such that

$$X := f_x(Y) + \hat{\xi}_x, \qquad Y \perp\!\!\!\perp_P \hat{\xi}_x,$$

if and only if *Complicated Condition* is satisfied.
(Hoyer et al, 2008)

Complicated Condition: The triple $(f_y, P(X), P(\xi_y))$ solves the following differential equation for all x, y with $(\log P(\xi_y))''(y - f_y(x))f'(x) \neq 0$.

The non linear case (3/3)

- The space that satisfy the condition is a 3-dimentional space;
  The space of continuous distributions is infinite dimensional;
  $\Longrightarrow$ we have identifiability for most distributions.
- If the noise is Gaussian, then the only functional form that satisfies Complicated Condition is linearity.
- If the function is linear and the noise is non-Gaussian, then one can't fit a linear backwards model **but** one can fit a non-linear backwards models.

Causal order discovery procedure in the bivariate case

Given $P(X, Y)$ and a dependence estimator $\hat{I}$

Procedure:

Given $P(X, Y)$ and a dependence estimator $\hat{I}$

Procedure:

1. Fit $\hat{f}_Y$ and $\hat{f}_X$:

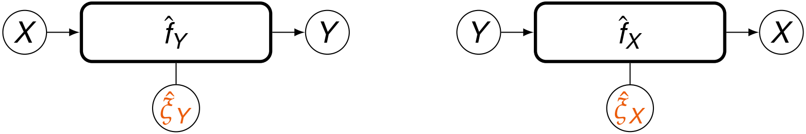
Causal order discovery procedure in the bivariate case

Given $P(X, Y)$ and a dependence estimator $\hat{I}$

Procedure:

1. Fit $\hat{f}_Y$ and $\hat{f}_X$:

2. Compute residuals $\hat{\xi}_Y$ and $\hat{\xi}_X$:



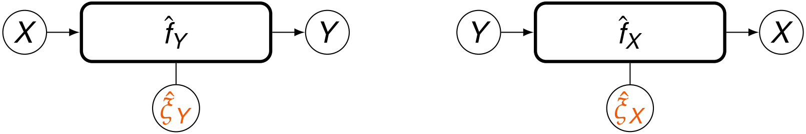
Causal order discovery procedure in the bivariate case

Given $P(X, Y)$ and a dependence estimator $\hat{I}$

Procedure:

1. Fit $\hat{f}_Y$ and $\hat{f}_X$:

2. Compute residuals $\hat{\xi}_Y$ and $\hat{\xi}_X$:



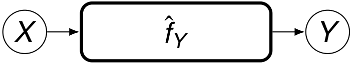
3. Order:
   - $\mathcal{T} = [X, Y]$ if $\hat{I}(x, \hat{\xi}_Y) < \hat{I}(y, \hat{\xi}_X)$
   - $\mathcal{T} = [Y, X]$ if $\hat{I}(y, \hat{\xi}_X) < \hat{I}(x, \hat{\xi}_Y)$

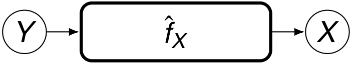
Causal order discovery procedure in the bivariate case

Given $P(X, Y)$ and a dependence estimator $\hat{I}$

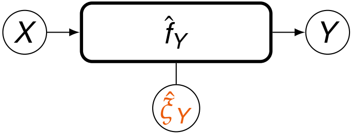
Procedure:

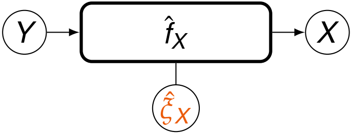
1. Fit $\hat{f}_Y$ and $\hat{f}_X$:





2. Compute residuals $\hat{\xi}_Y$ and $\hat{\xi}_X$:





3. Order:
   - $\mathcal{T} = [X, Y]$ if $\hat{I}(x, \hat{\xi}_Y) < \hat{I}(y, \hat{\xi}_X)$
   - $\mathcal{T} = [Y, X]$ if $\hat{I}(y, \hat{\xi}_X) < \hat{I}(x, \hat{\xi}_Y)$
4. Output (suppose $\mathcal{T} = [X, Y]$):
   - $X \rightarrow Y$ if $X \perp\!\!\!\perp_P \hat{\xi}_Y$ and $Y \not\perp\!\!\!\perp_P \hat{\xi}_X$

Table of content

Minimality

Minimality condition A DAG $\mathcal{G}$ compatible with a probability distribution P is said to satisfy the minimality condition if P is not compatible with any proper subgraph of $\mathcal{G}$.

Minimality

Minimality condition A DAG $\mathcal{G}$ compatible with a probability distribution P is said to satisfy the minimality condition if P is not compatible with any proper subgraph of $\mathcal{G}$.

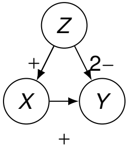
Remark: faithfulness $\implies$ minimality.

Minimality and d-sep

Theorem (implication of minimality on d-sep): Consider the random vector $\mathcal{V}$ and assume that the joint distribution has a density with respect to a product measure. Suppose that $P(\mathcal{V})$ is Markov with respect to $\mathcal{G}$. Then $P(\mathcal{V})$ satisfies the minimality condition iff $\forall X \in \mathcal{V}$ and $\forall Y \in Parents(X, \mathcal{G})$, $X \not\perp\!\!\!\perp_P Y \mid Parents(X, \mathcal{G}) \backslash \{Y\}$.
(proof on board)

Violation of minimality

Example 1: canceling out



Example 2: constant functions

Linear non gaussian

Theorem (LiNGAM) Assume a linear SCM with graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a compatible distribution $P(\mathcal{V})$ such that $\forall Y \in \mathcal{V}$

$$Y := \sum_{X \in Parents(Y, \mathcal{G})} a_{xy} X + \xi_y$$

where all ξ_y are jointly independent and non-Gaussian distributed. Additionally, we require that $\forall Y \in \mathcal{V}, X \in Parents(Y, \mathcal{G}), a_{xy} \neq 0$. Then, the graph $\mathcal{G}$ is identifiable from $P(\mathcal{V})$.
(proof in (Shimizu et al, 2011))

The LiNGAM algorithm

Algorithm 1 LiNGAM

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

Form an empty graph $\mathcal{G}$ on vertex set $\mathcal{V} = \{X_1, \cdots, X_p\}$
Let $S = \{1, \cdots, p\}$ and $\mathcal{T} = []$
repeat
  $H = []$
  for $i \in S$ **do**
    for $j \in S \backslash \{i\}$ **do**
      $\hat{\xi}_{ij} = X_j - \frac{cov(X_i, X_j)}{var(X_i)} X_i$
    end for
    $h = \sum_{j \in S \backslash \{i\}} \hat{I}(X_i, \hat{\xi}_{ij})$
    $H = [H, h]$
  end for
  $i^* = arg \min_{i \in S} H$
  $S = S \backslash \{i^*\}$
  $\mathcal{T} = [\mathcal{T}, i^*]$
  $\forall j \in S, X_j = \hat{\xi}_{i^* j}$
until $|S| = 0$
Append($\mathcal{T}$, S_0)
Construct a strictly lower triangular matrix by following the order in $\mathcal{T}$, and estimate the connection strengths $a_{i,j}$ by using some conventional covariance-based regression.
if $a_{i,j} > 0$ **then**
  Add $X_i \rightarrow X_j$ to $\mathcal{G}$
end if
Return $\mathcal{G}$

Additive noise models

Theorem (ANM) Assume a non-linear SCM with graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a compatible distribution $P(\mathcal{V})$ that satisfy the minimality condition with respect to $\mathcal{G}$. $\forall Y \in \mathcal{V}$

$$Y := f(Parents(Y, \mathcal{G})) + \xi_y$$

where all ξ_y are jointly independent. Then, the graph $\mathcal{G}$ is identifiable from $P(\mathcal{V})$.
(proof in (Peters et al, 2014))

The ANM algorithm

Algorithm 2 ANM

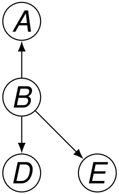
Input: $P(\mathcal{V})$

Output: $\mathcal{G}$

Form an empty graph $\mathcal{G}$ on vertex set $\mathcal{V} = \{X_1, \cdots, X_p\}$
Let $S = \{1, \cdots, p\}$ and $\mathcal{T} = []$
repeat
  $H = []$
  for $j \in S$ **do**
    $\hat{f}_{.j}$: Regress X^j on $\{X_i\}_{i \in S \setminus \{j\}}$
    $\hat{\xi}_{.j} = X_j - \hat{f}_{.j}(X_i)$
    $h = \hat{I}(\{X_i\}_{i \in S \setminus \{j\}}, \xi_{.j})$
    $H = [H, h]$
  end for
  $i^* = arg\min_{i \in S} H$
  $S = S \setminus \{i^*\}$
  $\mathcal{T} = [i^*, \mathcal{T}]$
until $|S| = 0$
for $j \in \{2, \cdots, p\}$ **do**
  for $i \in \{\mathcal{T}_1, \cdots, \mathcal{T}_{j-1}\}$ **do**
    $\hat{f}_{.j}$: Regress X^j on $\{X_k\}_{k \in \{\mathcal{T}_1, \cdots, \mathcal{T}_{j-1}\} \setminus \{i\}}$
    $\hat{\xi}_{.j} = X_j - \hat{f}_{.j}(X_i)$
    if $\{X_k\}_{k \in \{\mathcal{T}_1, \cdots, \mathcal{T}_{j-1}\} \setminus \{i\}} \not\perp\!\!\!\perp_P \xi_{.j}$ **then**
      Add $X_i \rightarrow X_j$ to $\mathcal{G}$
    end if
  end for
end for
Return $\mathcal{G}$

ANM in action (1/4)

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



ANM in action (2/4)

- Estimate $A, B, D \mapsto E$ and $\hat{\xi}_e$
  - $H_1 = \hat{I}(\{A, B, D\}, \hat{\xi}_e)$
- Estimate $A, D, E \mapsto B$ and $\hat{\xi}_b$
  - $H_3 = \hat{I}(\{A, D, E\}, \hat{\xi}_b)$
- Estimate $A, B, E \mapsto D$ and $\hat{\xi}_d$
  - $H_2 = \hat{I}(\{A, B, E\}, \hat{\xi}_d)$
- Estimate $B, D, E \mapsto A$ and $\hat{\xi}_a$
  - $H_4 = \hat{I}(\{B, D, E\}, \hat{\xi}_a)$

ANM in action (2/4)

- Estimate $A, B, D \mapsto E$ and $\hat{\xi}_e$
  - $H_1 = \hat{I}(\{A, B, D\}, \hat{\xi}_e)$
- Estimate $A, D, E \mapsto B$ and $\hat{\xi}_b$
  - $H_3 = \hat{I}(\{A, D, E\}, \hat{\xi}_b)$
- Estimate $A, B, E \mapsto D$ and $\hat{\xi}_d$
  - $H_2 = \hat{I}(\{A, B, E\}, \hat{\xi}_d)$
- Estimate $B, D, E \mapsto A$ and $\hat{\xi}_a$
  - $H_4 = \hat{I}(\{B, D, E\}, \hat{\xi}_a)$

$$4 = Argmin(H)$$
$$\mathcal{T} = [A]$$

ANM in action (3/4)

- Estimate $B, D \mapsto E$ and $\hat{\xi}_e$
  - $H_1 = \hat{I}(\{B, D\}, \hat{\xi}_e)$
- Estimate $D, E \mapsto B$ and $\hat{\xi}_b$
  - $H_3 = \hat{I}(\{D, E\}, \hat{\xi}_b)$
- Estimate $B, E \mapsto D$ and $\hat{\xi}_d$
  - $H_2 = \hat{I}(\{B, E\}, \hat{\xi}_d)$

ANM in action (3/4)

- Estimate $B, D \mapsto E$ and $\hat{\xi}_e$
  - $H_1 = \hat{I}(\{B, D\}, \hat{\xi}_e)$
- Estimate $B, E \mapsto D$ and $\hat{\xi}_d$
  - $H_2 = \hat{I}(\{B, E\}, \hat{\xi}_d)$
- Estimate $D, E \mapsto B$ and $\hat{\xi}_b$
  - $H_3 = \hat{I}(\{D, E\}, \hat{\xi}_b)$

$$1 = Argmin(H)$$
$$\mathcal{T} = [E, A]$$

ANM in action (3/4)

- Estimate $B, D \mapsto E$ and $\hat{\xi}_e$
  - $H_1 = \hat{I}(\{B, D\}, \hat{\xi}_e)$
- Estimate $B, E \mapsto D$ and $\hat{\xi}_d$
  - $H_2 = \hat{I}(\{B, E\}, \hat{\xi}_d)$
- Estimate $D, E \mapsto B$ and $\hat{\xi}_b$
  - $H_3 = \hat{I}(\{D, E\}, \hat{\xi}_b)$

$$1 = Argmin(H)$$
$$\mathcal{T} = [E, A]$$

- Estimate $D \mapsto B$ and $\hat{\xi}_b$
  - $H_1 = \hat{I}(D, \hat{\xi}_b)$
- Estimate $B \mapsto D$ and $\hat{\xi}_d$
  - $H_2 = \hat{I}(B, \hat{\xi}_d)$

ANM in action (3/4)

- Estimate $B, D \mapsto E$ and $\hat{\xi}_e$
  - $H_1 = \hat{I}(\{B, D\}, \hat{\xi}_e)$
- Estimate $B, E \mapsto D$ and $\hat{\xi}_d$
  - $H_2 = \hat{I}(\{B, E\}, \hat{\xi}_d)$
- Estimate $D, E \mapsto B$ and $\hat{\xi}_b$
  - $H_3 = \hat{I}(\{D, E\}, \hat{\xi}_b)$

$$1 = Argmin(H)$$
$$\mathcal{T} = [E, A]$$

- Estimate $D \mapsto B$ and $\hat{\xi}_b$
  - $H_1 = \hat{I}(D, \hat{\xi}_b)$
- Estimate $B \mapsto D$ and $\hat{\xi}_d$
  - $H_2 = \hat{I}(B, \hat{\xi}_d)$

$$2 = Argmin(H)$$
$$\mathcal{T} = [D, E, A]$$

ANM in action (3/4)

- Estimate $B, D \mapsto E$ and $\hat{\xi}_e$
  - $H_1 = \hat{I}(\{B, D\}, \hat{\xi}_e)$
- Estimate $B, E \mapsto D$ and $\hat{\xi}_d$
  - $H_2 = \hat{I}(\{B, E\}, \hat{\xi}_d)$
- Estimate $D, E \mapsto B$ and $\hat{\xi}_b$
  - $H_3 = \hat{I}(\{D, E\}, \hat{\xi}_b)$

$$1 = Argmin(H)$$
$$\mathcal{T} = [E, A]$$

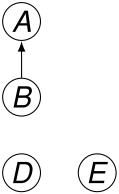
- Estimate $D \mapsto B$ and $\hat{\xi}_b$
  - $H_1 = \hat{I}(D, \hat{\xi}_b)$
- Estimate $B \mapsto D$ and $\hat{\xi}_d$
  - $H_2 = \hat{I}(B, \hat{\xi}_d)$

$$2 = Argmin(H)$$
$$\mathcal{T} = [D, E, A]$$

$$\mathcal{T} = [B, D, E, A]$$

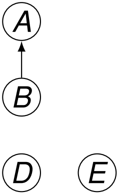
ANM in action (4/4)

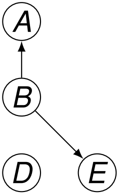
$$\mathcal{T} = [B, D, E, A]$$



ANM in action (4/4)

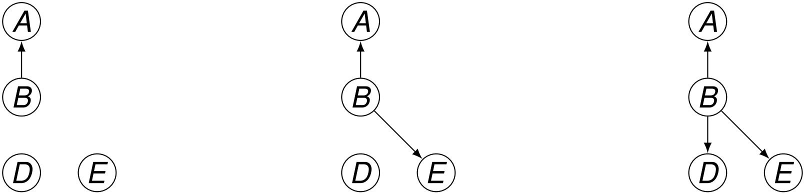
$$\mathcal{T} = [B, D, E, A]$$





ANM in action (4/4)

$$\mathcal{T} = [B, D, E, A]$$



Exercise 1

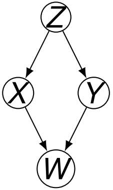
Why is faithfulness needed for constraint-based methods whereas noise-based methods only need minimality?

Exercise 2

After applying LiNGAM, how can you know if causal sufficiency is not respected?

Exercise 3

- Suppose the true graph on right;
- Assumptions: CMC, causal sufficiency, minimality;
- Generative process:

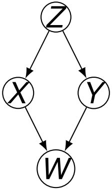


$$
\begin{aligned}
Z &= \xi_z & \xi_z &\sim U(0,1);\\
X &= a * Z + \xi_x & \xi_x &\sim U(0,1);\\
Y &= b * Z + \xi_y & \xi_y &\sim U(0,1);\\
W &= c * X - d * Y + \xi_w & \xi_w &\sim N(0,1).
\end{aligned}
$$

- Given a compatible distribution what would be the output of the LiNGAM algorithm? And what about the ANM algorithm?

Exercise 4

- Suppose the true graph on right;
- Assumptions: CMC, causal sufficiency, minimality;
- Generative process:



$$
\begin{aligned}
Z &= \xi_z & \xi_z &\sim U(0,1); \\
X &= Z^2 + \xi_x & \xi_x &\sim U(0,1); \\
Y &= Z^3 + \xi_y & \xi_y &\sim U(0,1); \\
W &= XY + \xi_w & \xi_w &\sim U(0,1).
\end{aligned}
$$

- Given a compatible distribution what would be the output of the LiNGAM algorithm? And what about the ANM algorithm?

Table of content

Conclusion

- Under linear non gaussian models noise-based methods can discover the causal graph.
- Under non-linear additive noise models noise-based methods can discover the causal graph.
- Advantages:
  - Can discovery the true graph;
  - Faithfulness is not needed.
- Drawbacks:
  - Semi parametric assumptions;
  - Need large sample size.

Conclusion

- Under linear non gaussian models noise-based methods can discover the causal graph.
- Under non-linear additive noise models noise-based methods can discover the causal graph.
- Advantages:
  - Can discovery the true graph;
  - Faithfulness is not needed.
- Drawbacks:
  - Semi parametric assumptions;
  - Need large sample size.

Conclusion

- Under linear non gaussian models noise-based methods can discover the causal graph.
- Under non-linear additive noise models noise-based methods can discover the causal graph.
- Advantages:
  - Can discovery the true graph;
  - Faithfulness is not needed.
- Drawbacks:
  - Semi parametric assumptions;
  - Need large sample size.

Conclusion

- Under linear non gaussian models noise-based methods can discover the causal graph.
- Under non-linear additive noise models noise-based methods can discover the causal graph.
- Advantages:
  - Can discovery the true graph;
  - Faithfulness is not needed.
- Drawbacks:
  - Semi parametric assumptions;
  - Need large sample size.

Some extensions

- Without causal sufficiency if linear relations;
- Extension to discrete additive noise models;
- Post non linear relations;
- Time series.

Some extensions

- Without causal sufficiency if linear relations;
- Extension to discrete additive noise models;
- Post non linear relations;
- Time series.

Some extensions

- Without causal sufficiency if linear relations;
- Extension to discrete additive noise models;
- Post non linear relations;
- Time series.

Some extensions

- Without causal sufficiency if linear relations;
- Extension to discrete additive noise models;
- Post non linear relations;
- Time series.

References (1/2)

Direct inspirations

1. *Elements of causal inference*, J. Peters, D. Janzing , B. Schölkopf. MIT Press, 2nd edition, 2017
2. *DirectLiNGAM: A Direct Method for Learning a Linear Non-Gaussian Structural Equation Model*, S. Shimazu, T. Inazumi, Y. Sogawa, A. Hyvarinen, Y. Kawahara, T. Washio, P. Hoyer, K. Bollen. JMLR, 2011
3. *Nonlinear causal discovery with additive noise models*, P. Hoyer, D. Janzing, J. Mooij, J. Peters, B. Schölkopf. Neurips, 2008
4. *Causal Discovery with Continuous Additive Noise Models*, J. Peters, J. Mooij, D. Janzing, B. Schölkopf. JMLR, 2014

References (2/2)

Additional readings

1. *Causal inference from noise*, N. Climenhaga, L. DesAutels, G. Ramsey. Noûs, 2019
2. *On the logic of causal models*, D. Geiger, J. Pearl. In Proceedings of the Fourth Annual Conference on Uncertainty in Artificial Intelligence, 1990
3. *A Linear Non-Gaussian Acyclic Model for Causal Discovery*, S. Shimazu, P. Hoyer, A. Hyvarinen, A. Kerminen. JMLR, 2006
4. *Analyse générale des liaisons stochastiques.*, G. Darmois. Review of the International Statistical Institute, 1953
5. *On a property of the normal distribution*, W. P. Skitovitch. Doklady Akademii Nauk SSSR, 89:217–219, 1953
6. *Causal Inference on Time Series using Restricted Structural Equation Models*, J. Peters, D. Janzing, B. Schölkopf. Neurips, 2013