Causal Discovery Methods

Causal Discovery methods based on observational data

November 13, 2024

Introduction to Causality

Definition

- Causality is the influence one event (cause) has on another (effect).
- It implies that changes in the cause lead to changes in the effect, forming a non-random link.

Key Characteristics of Causality

- **Directionality**: A causes B, but B does not necessarily cause A.
- **Mechanism**: Changes in the cause generate changes in the effect.
- **Counterfactuals**: Considers what would happen to the effect if the cause did not occur.

Causality Discovery

Methods

- Experiment-based approach
  - Control experiment: Intervention causes changes in outcomes,
  - In many cases too expensive, too time-consuming, or even impossible.
- Data-based approach

Overview of Data-based Causal Discovery Methods

Methods

- Constraint-based methods
  - PC
  - FCI
- Score-based methods
  - GES (Greedy Equivalence Search)
- Functional Causal Models
  - Linear, Non-Gaussian Model
  - Non-linear Methods
- Hybrid Methods
  - SELF (Structural Equational Likelihood Framework)
  - FRITL (Functional Representation with Independent Triad and Likelihood)

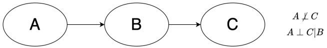
1. Introduction to Causality

2. Overview of Data-based Causal Discovery Methods
   - Constraint-Based Methods
   - Functional Causal Models (FCMs)
   - Hybrid Methods

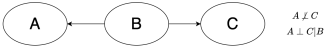
3. Comparison of Methods

Constraint-Based Method: Assumptions

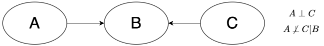
Causal Markov Assumption: A variable X is independent of every other variable (except X's effects) conditional on all of its direct causes.
Causal Faithfulness Assumption: for all observed variables, X_i is independent of X_j conditional on variables Z if and only if the Markov Assumption for $\mathcal{G}$ entails such conditional independencies.



Chain Structure



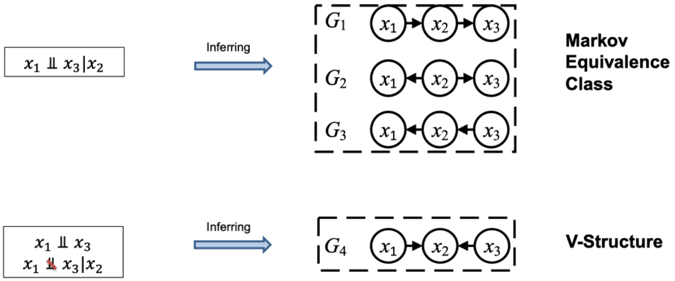
Fork Structure



Colider Structure

Constraint-Based Method

Limitations: DAGs within the same Markov Equivalence Class cannot be distinguished solely based on the conditional independence relationships observed in the data.

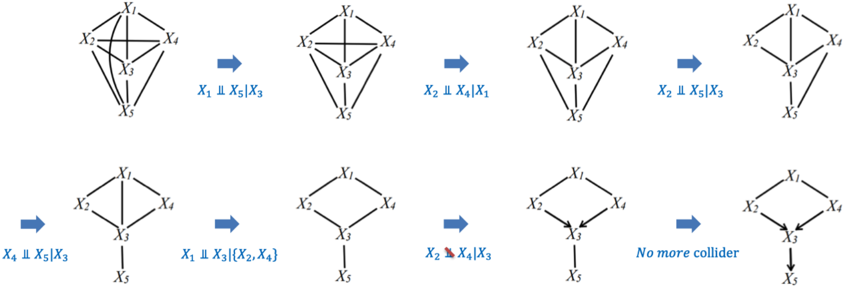


Using Conditional Independence Tests to Infer Causal Structure

Constraint-Based Method: PC Algorithm

1. **Initialize Graph**: Start with a fully connected undirected graph.
2. **Edge Removal**: Test conditional independence for each pair of variables given subsets of other variables. Remove edges where conditional independence is found.
3. **Identify Colliders**: Orient edges for v-structures ($X \rightarrow Z \leftarrow Y$) where X and Y are independent unless conditioned on Z.
4. **Orient Remaining Edges**: Use orientation rules to direct undetermined edges, leaving ambiguous edges undirected.
5. **Output CPDAG**: The result is a CPDAG representing the Markov Equivalence Class of the causal structure.

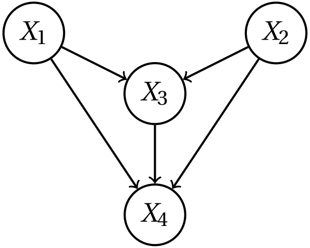
Constraint-Based Method: PC Algorithm Example

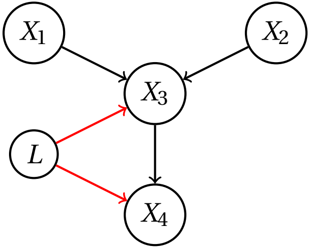


Example of PC(Peter-Clark)

PC Algorithm Limitation

- **Limitations**: Unable to deal with latent confounders





$$X_1 \perp\!\!\!\perp X_2$$

$$X_1 \not\!\perp\!\!\!\perp X_4 \mid X_3$$

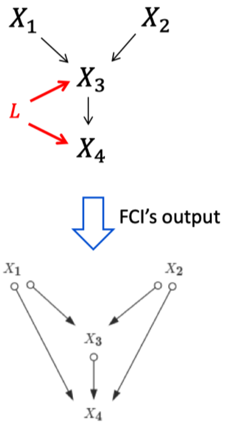
$$X_2 \not\!\perp\!\!\!\perp X_4 \mid X_3$$

Constraint-Based Method: FCI Algorithm Process

1. **Initialize Graph**: Start with a fully connected undirected graph over all observed variables.
2. **Edge Removal**: Test conditional independence between each pair of variables given subsets of other variables. Remove edges where conditional independence is found, accounting for possible latent confounders and selection bias.
3. **Identify Colliders**: Identify v-structures ($X \rightarrow Z \leftarrow Y$) where X and Y are not adjacent, and no conditioning set separates X and Y.
4. **Propagate Edge Orientations**: Apply orientation rules to propagate edge directions using partially oriented information, ensuring consistency and avoiding cycles.
5. **Handle Ambiguous Relationships**: Determine possible orientations considering latent variables and adjust edge marks to represent ambiguous causal relationships using open endpoints (e.g., $X \circ\!\rightarrow Y$).
6. **Output PAG**: The result is a Partial Ancestral Graph (PAG) that represents the Markov equivalence class, accounting for potential latent confounders and selection effects.

Constraint-Based Method: FCI Algorithm Process

$X_1 \circ\!\!\longrightarrow X_2$	X_2 is not an ancestor of X_1
$X_1 \circ\!\!-\!\!\circ X_2$	No set d-separates X_2 and X_1
$X_1 \longrightarrow X_2$	X_1 is a cause of X_2
$X_1 \longleftrightarrow X_2$	There is a latent common cause of X_1 and X_2



1. Introduction to Causality
2. Overview of Data-based Causal Discovery Methods
   - Constraint-Based Methods
   - Functional Causal Models (FCMs)
   - Hybrid Methods
3. Comparison of Methods

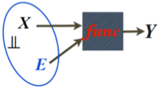
Functional-Based Method: Assumptions

Considering the data generating process, effect generated from causes and noises represented with functional causal model:

$$Y = f(X, E)$$

Introducing additional assumptions:

- Independent noise assumption: Independence between the causes X and noises E

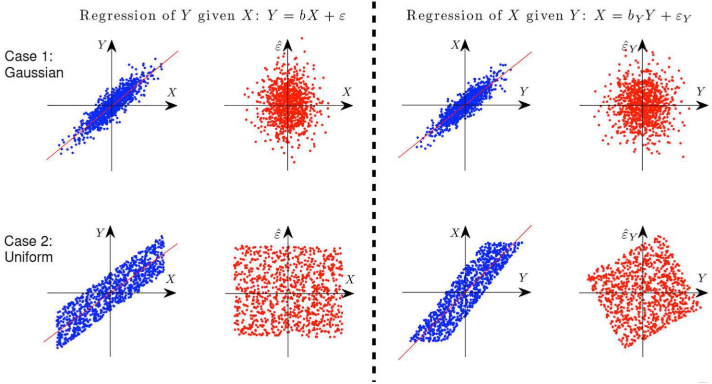


- Independent mechanism assumption: Independence between the causes X and process f

$$Y = f(x), X \perp f(x)$$

Functional-Based Methods: independent Noise (IN) Condition

- Causal Asymmetry in the Linear non-Gaussian Case
  Data Generated by $Y = \alpha X + E(i.e., X \rightarrow Y)$
  (X, Y) follows the IN condition iff regression residual $Y - \hat{w}^T X$ is independent from X



Functional-Based Methods: LiNGAM

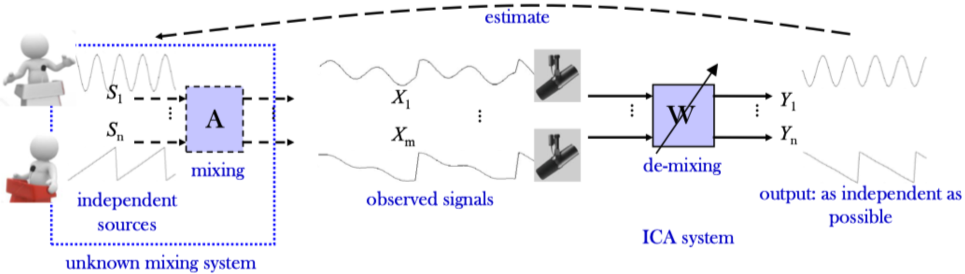
- Under the above assumptions, the LiNGAM can be expressed as:

$$\mathbf{X} = \mathbf{B}\mathbf{X} + \mathbf{E}$$

  - $\mathbf{X}$ is a p-dimensional random vector, representing the observed variables.
  - $\mathbf{B}$ is a $p \times p$-dimensional matrix, representing the connection weights between the observed variables.
  - $\mathbf{E}$ is a p-dimensional non-Gaussian random noise vector.
- Because of the DAG assumption, there exists a permutation matrix $P \in \mathbb{R}^{p \times p}$ such that $\mathbf{B}' = P\mathbf{B}P^T$ is a strict lower triangular matrix with diagonal elements all equal to 0.

LiNGAM: analysis by ICA

☑ ICA, **Independent Component Analysis**, can be used to solve the LiNGAM



$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S}$$

$$\mathbf{Y} = \mathbf{W} \cdot \mathbf{X}$$

☑ Assumptions in ICA

- At most one of S_i is Gaussian
- Size(X) >= Size(S), and **A** is of full column rank

Then **A** can be estimated up to column scale and permutation indeterminacies

LiNGAM: analysis by ICA

- **LiNGAM:**

$$X = BX + E$$

$$E = X - BX$$

$$E = (I - B)X$$

- **ICA:**

$$Y = WX$$

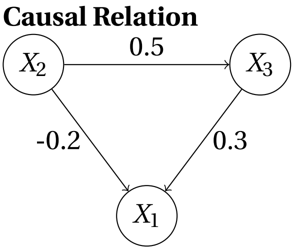
$$B = I - W$$

LiNGAM: Example

Example:

$$\begin{bmatrix} E_1 \\ E_3 \\ E_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -0.5 & 1 & 0 \\ 0.2 & -0.3 & 1 \end{bmatrix} \begin{bmatrix} X_2 \\ X_3 \\ X_1 \end{bmatrix}$$

$$\Rightarrow \begin{cases} X_2 = E_1 \\ X_3 = 0.5X_2 + E_3 \\ X_1 = -0.2X_2 + 0.3X_3 + E_2 \end{cases}$$

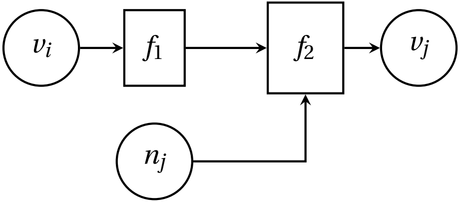


Functional-Based Methods: PNL (post-NonLinear method)

- LiNGAM algorithms can only solve linear problems.
- In the PNL model, assuming that there is a causal relationship $v_i \to v_j$, it can be expressed as

$$v_j = f_2(f_1(v_i) + n_j)$$

  - v_i and n_j are independent of each other
  - f_1 is a non-constant smooth function
  - f_2 is a reversible smooth function and $f_2' \neq 0$



1. Introduction to Causality

2. Overview of Data-based Causal Discovery Methods
   - Constraint-Based Methods
   - Functional Causal Models (FCMs)
   - Hybrid Methods

3. Comparison of Methods

Hybrid Methods

- Combine constraint-based and functional approaches
- **Examples**:
  - SELF (Structural Equational Likelihood Framework)
  - FRITL (Functional Representation with Independent Triad and Likelihood)
- **Strengths**: Addresses limitations in handling latent confounders

Comparison of Causal Discovery Methods

	PC	FCI	GES	LiNGAM/PNL/ANM	SELF	FRITL
Faithfulness assumption required?	Yes	Yes	Some weaker condition required (not totally clear yet)	No	No	No
Specific assumptions on data distributions required?	No	No	Yes (usually assumes linear-Gaussian models or multinomial distributions)	Yes	Yes	Yes
Properly handle confounders?	No	Yes	No	No	No	Yes
Output	Markov equivalence class	Partial ancestral graph	Markov equivalence class	DAG as well as causal model (under the respective identifiability conditions)	DAG with likelihood-based causal structure (assumes observed variables)	DAG or PAG, refined with ICA and Triad condition for latent confounders

Comparison of Causal Discovery Methods

Thank you for listening!