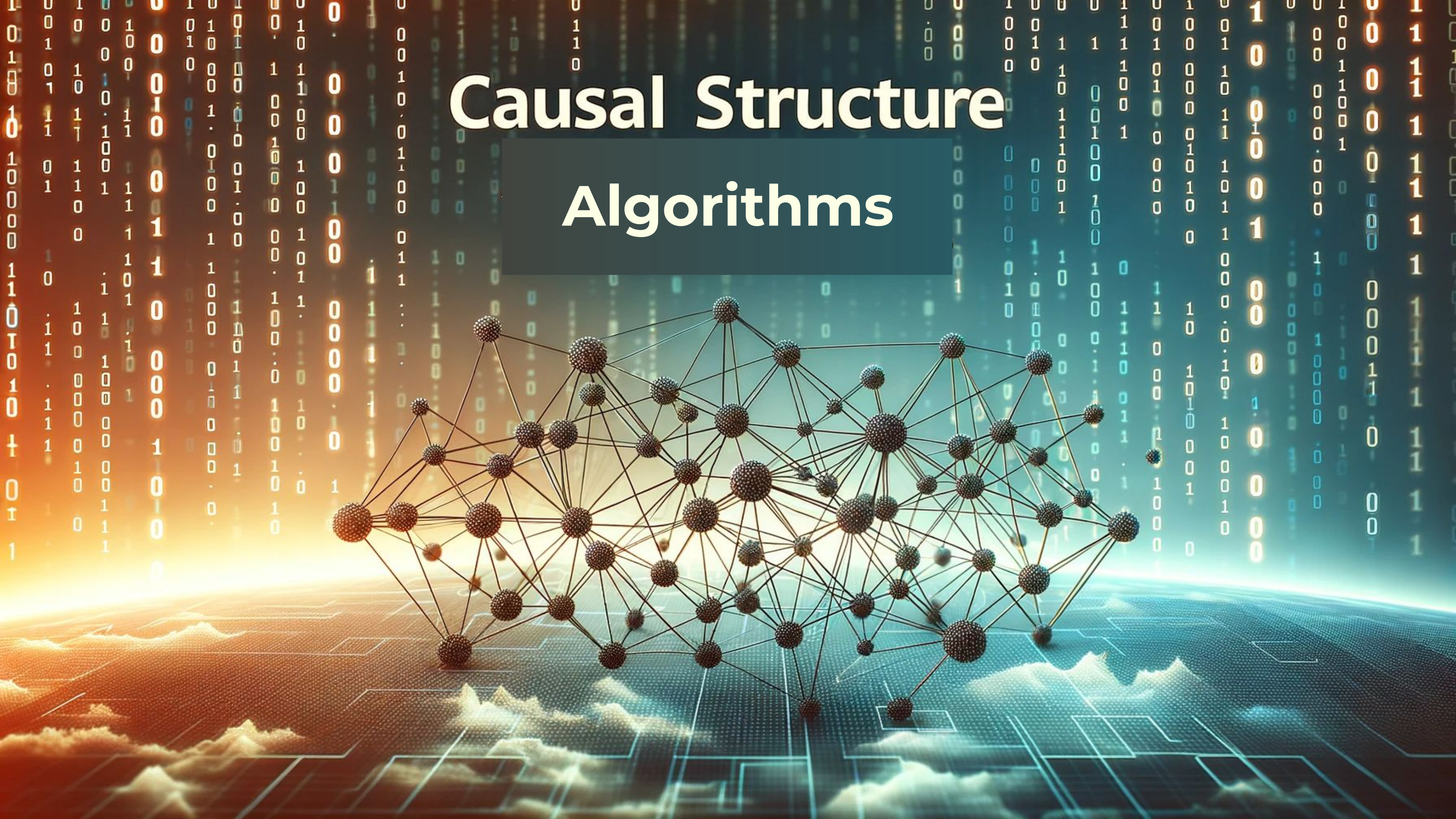
Causal Structure

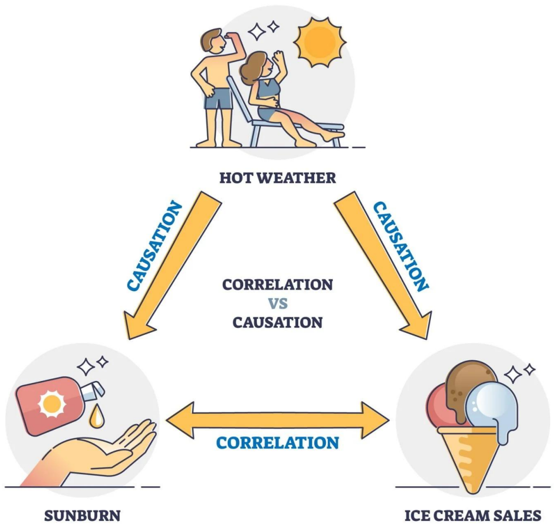
Algorithms

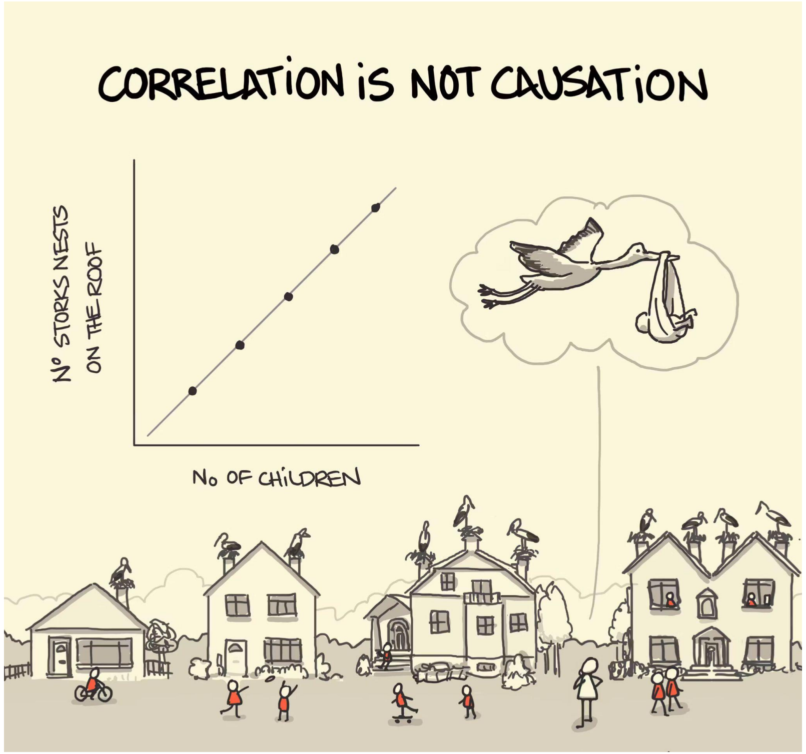


Outline

- What is **causality**?
- Problems to solve (effect and structure)
- Libraries
- Types of causal learning (structure) algorithms
- Examples (PC, GES, Lingam)
- Causal learning for time series

What is causality?





Causality refers to the relationship between two events or variables, where a change in one (the cause) produces a change in the other (the effect). The essence of causality is the generation of an effect by a cause.

Philosophical views

Regularity Theory

David Hume

Causality is understood as the regular succession of events.

Counterfactual Theories

David Lewis

Causation is understood in terms of counterfactuals.

Manipulationist or **Interventionist** Theories

Judea Pearl, James Woodward

A cause is something that can be manipulated to bring about a change in some effect.

Probabilistic Theories

Patrick Suppes, Ellery Eells

Causes raise the probabilities of their effects.

Aristotelian or Four Causes

Aristotle

Four types of causes:

Material Cause
Formal Cause
Efficient Cause
Final Cause

Process Theories

Wesley Salmon, Phil Dowe

Emphasizes physical processes and interactions.

Causal Powers and Dispositions

Rom Harré, Brian Ellis

Some entities have inherent powers or dispositions to produce certain effects.

Causal Pluralism

Nancy Cartwright

There isn't one single, unified concept of causation that applies in all contexts.

Philosophical views

Regularity Theory

David Hume

Causality is understood as the regular succession of events. If event A regularly precedes event B, then A is the cause of B. This perspective denies the existence of any necessary connection between cause and effect.

Counterfactual Theories

David Lewis

Causation is understood in terms of counterfactuals. If event A had not occurred, event B would not have occurred either. In other words, A causes B if and only if, had A not happened, B would not have happened.

Manipulationist or Interventionist Theories
Judea Pearl, James Woodward
A cause is something that can be manipulated to bring about a change in some effect. This perspective is rooted in the experimental method where interventions or manipulations are made to ascertain causality.

Probabilistic Theories

Patrick Suppes, Ellery Eells

Causes raise the probabilities of their effects. Causation is characterized in terms of probability-raising relations.

Aristotelian or Four Causes

Aristotle
Aristotle delineated four types of causes:
Material Cause: The material out of which something is made.
Formal Cause: The defining characteristics of the thing.
Efficient Cause: The agent or process that brings something into existence.
Final Cause: The purpose or end for which something is done.

Process Theories

Wesley Salmon, Phil Dowe

Emphasizes physical processes and interactions. A causes B if there is a continuous process that runs from A to B, and there's a conservation of energy and momentum in the interaction.

Causal Powers and Dispositions

Rom Harré, Brian Ellis

Definition: Some entities have inherent powers or dispositions to produce certain effects. Causation involves the actualization of these dispositions under appropriate conditions.

Causal Pluralism

Nancy Cartwright

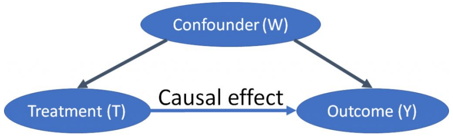
Definition: There isn't one single, unified concept of causation that applies in all contexts. Instead, there are many different types of causal relations that are relevant in different contexts.

Causality Definitions

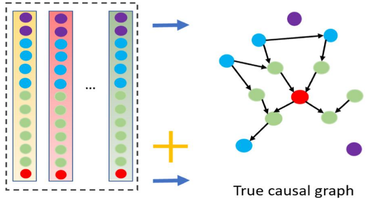
1. **Interventional Definition**: If we intervene to change X, and as a direct consequence Y changes, then X is a cause of Y. This is represented in the do-calculus as:
$$P(Y|\text{do}(X = x)) \neq P(Y) \tag{1}$$
If changing X changes the distribution of Y, then X has a causal effect on Y.

2. **Counterfactual Definition**: Consider two hypothetical worlds – one in which X takes a value x and another in which X takes a different value x'. If in these two worlds, Y takes different values, then X is a cause of Y. This is represented as:
$$Y_x \neq Y_{x'} \tag{2}$$
Where Y_x is the value Y would take if X were set to x, and $Y_{x'}$ is the value Y would take if X were set to x'.

Problems to solve



Identification: Causal effect → Observed effect conditioned on W, $\mathrm{E}[Y|T,W]$
Estimation: $\mathrm{E}[Y|T,W]$ → Propensity Score Stratification



Task	**Causal inference (effect)**	**Causal discovery (structure)**
Objective	determine the causal effect of one variable (treatment) on another variable (outcome) **"What happens to Y if I do X?"**	The goal is to uncover the entire causal structure or network among a set of variables **"What are the causal relationships among X,Y,Z...?"**
Prior Knowledge	the causal structure (or the relevant part of it) is known or hypothesized	Does not assume a known causal structure
Data Requirements	Can be based on observational data, but randomized controlled trials (RCTs) or experiments are the gold standard for drawing causal inferences as they help address confounding.	Typically begins with purely observational data, but there are advanced methods that can incorporate interventional data to refine and confirm the discovered structure
Output	causal effect estimate, which quantifies the change in the outcome variable for a unit change in the treatment variable, often with uncertainty bounds	causal graph or network, representing the relationships among variables. Edges in the graph represent potential causal relationships, though directionality might not always be definitive

Libraries and repositories | causal inference

DoWhy

Purpose: Framework to allow users to specify their model in terms of treatment, outcome, and graph, and then test the model with various methods.

Repository: DoWhy on GitHub

CausalML

Purpose: Uplift modeling and causal inference using machine learning algorithms.

Repository: CausalML on GitHub

EconML

Purpose: Uses recent advancements in ML to estimate causal effects. Developed by Microsoft Research.

Repository: EconML on GitHub

CausalPy

Purpose: Developed by PyMC, library for causal inference and discovery. Tools for estimating causal effects using matching, weighting, and regression. Also offers sensitivity analysis and result visualization. beta stage

Repository: Github blog

statsmodels

Purpose: Provides classes and functions for the estimation of many different statistical models, including some tools for causal analysis.

Repository: statsmodels on GitHub

Libraries and repositories | causal discovery

causal-learn

Purpose: It includes various algorithms for learning the causal graph structure, such as the PC algorithm, the FCI algorithm, the GES algorithm, and more.

Repository: causal-learn on GitHub

gCastle (by Huawei)

Purpose: With a foundational emphasis on a graph-based approach, it facilitates the identification of causal relationships among variables within a dataset.
Repository: GitHub

Causal Discovery Toolbox

Purpose: It encompasses a range of state-of-the-art algorithms to derive causal graphs from data, integrating both constraint-based and score-based methods.

Repository: CausalDiscoveryToolbox on GitHub

CausalNex

CausalNex

Purpose: A toolkit that allows users to create, fit, and validate Bayesian Networks for both data exploration and predicting causal effects.

Repository: CausalNex on GitHub

pcalg (originally R)

Purpose: Provides functions to learn and work with causal graphs, specifically implementing the PC algorithm.

Repository: The original is in R, but Python is pcalg on GitHub

bnlearn (originally R)

Purpose: Provides tools to learn the structure of Bayesian networks, some of which can be interpreted causally.

Repository (originally R)
Repository on python: GitHub documentation

pyAgrum

Purpose: A library for probabilistic graphical models. It is used for Bayesian networks, dynamic Bayesian networks, and learning algorithms for these models.

Repository: pyAgrum on GitHub

py-causal (Python)

Purpose: Python wrapper for TETRAD for causal discovery.

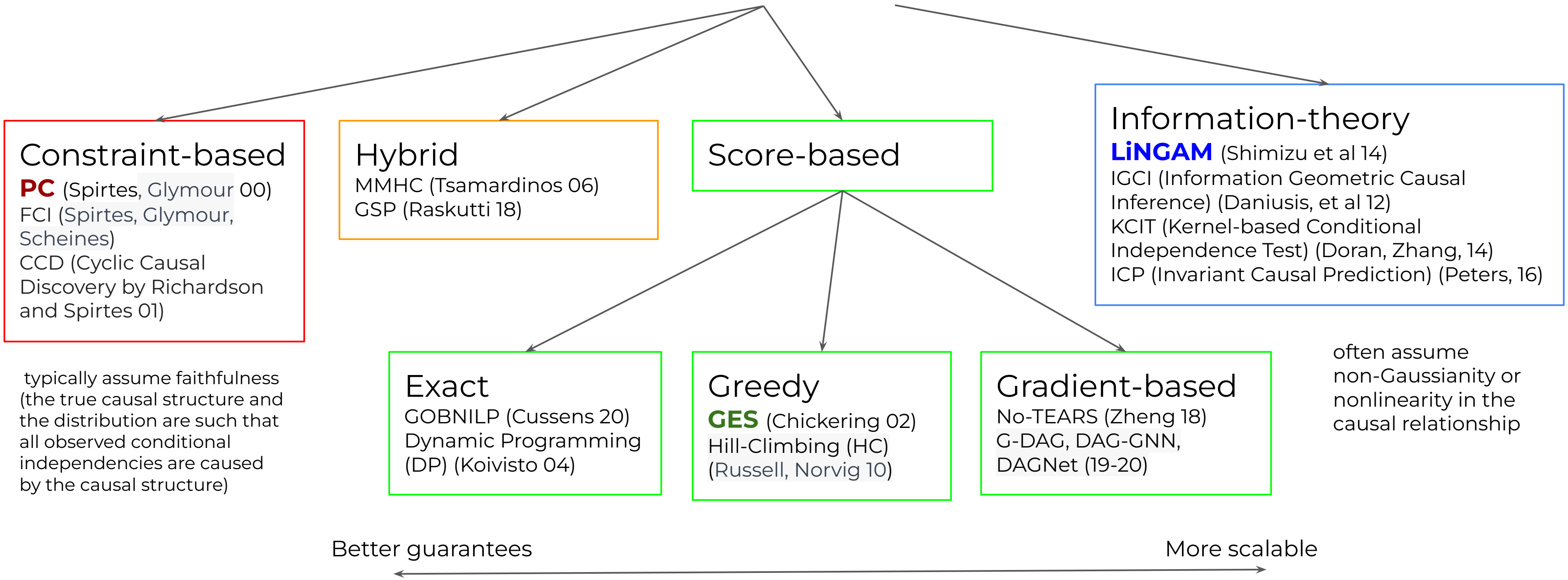
Repository: py-causal on GitHub

CLUSTERING IS SUBJECTIVE

CLUSTERING IS SUBJECTIVE

			
Simpson's family	School employees	Females	Males

Types of causal learning (structure) algorithms



Computational-statistical trade-off

https://link.springer.com/article/10.1007/s10208-022-09581-9

Examples: PC

1. Initialization:

Fully connected undirected graph G with nodes representing the variables in the dataset.

2. Skeleton Identification:

For each X and Y in G:
Test for conditional independence between X and Y, starting with an empty conditioning set (C), increase C.
If X and Y are found to be independent given any subset of the remaining variables, remove the edge between them.
Repeat

3. Edge Orientation:

Collider orientation (v-structure)
For any unshielded triplet X–Z–Y: If X and Y are not adjacent and are conditionally dependent given the set including Z and all subsets of remaining variables, then orient as: X→Z←Y.

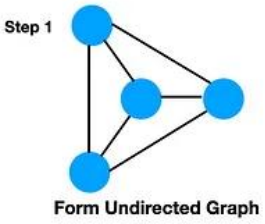
Avoiding New Colliders (propagation rule):
If A–B–C is an unshielded triplet and the edge between A and B has been oriented as A→B, and
A and C are not adjacent, then orient the edge between B and C as B→C to avoid creating a new collider.

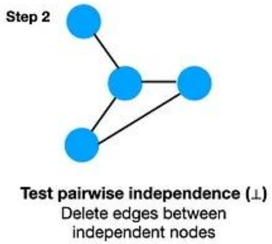
Avoiding Cycles (acyclic rule):
Orient A–B as A→B only if doing so does not create a directed cycle.

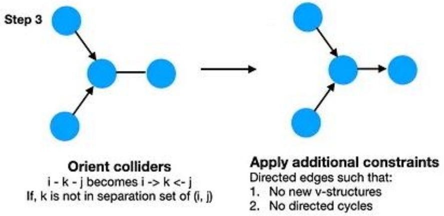
Avoiding New v-structures (consistency rule):
Do not orient an edge in a manner that would create a new v-structure (collider) that was not previously identified in the v-structure identification step.

Transitive Orientation:
If A→B and B–C are edges such that A and C are not adjacent, then orient B–C as B→C.

Propagation by Confluence:
If B has two incoming arrows, A→B and C→B, and there's an undirected edge between A and C, then orient A–C as A→C.







A.) Form the complete undirected graph C on the vertex set $\mathbf{V}$.

B.)

$n = 0$.

repeat

repeat

select an ordered pair of variables X and Y that are adjacent in C such that $\mathbf{Adjacencies}(C,X)\backslash\{Y\}$ has cardinality greater than or equal to n, and a subset $\mathbf{S}$ of $\mathbf{Adjacencies}(C,X)\backslash\{Y\}$ of cardinality n, and if X and Y are d-separated given $\mathbf{S}$ delete edge X - Y from C and record $\mathbf{S}$ in $\mathbf{Sepset}(X,Y)$ and $\mathbf{Sepset}(Y,X)$;

until all ordered pairs of adjacent variables X and Y such that $\mathbf{Adjacencies}(C,X)\backslash\{Y\}$ has cardinality greater than or equal to n and all subsets $\mathbf{S}$ of $\mathbf{Adjacencies}(C,X)\backslash\{Y\}$ of cardinality n have been tested for d-separation;

$n = n + 1$;

until for each ordered pair of adjacent vertices X, Y, $\mathbf{Adjacencies}(C,X)\backslash\{Y\}$ is of cardinality less than n.

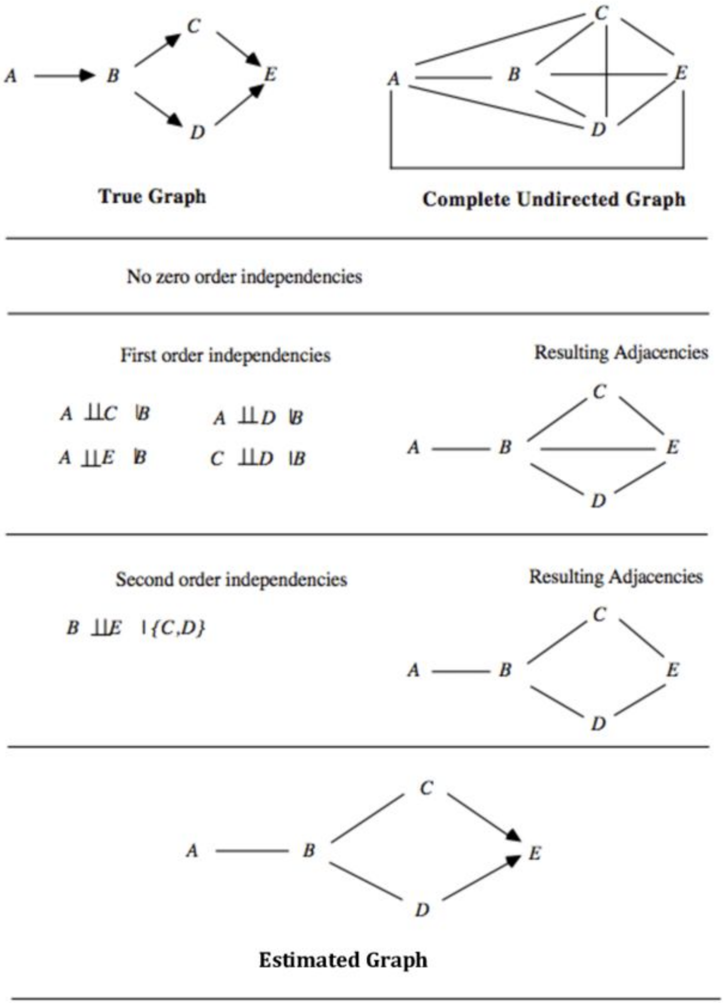
C.) For each triple of vertices X, Y, Z such that the pair X, Y and the pair Y, Z are each adjacent in C but the pair X, Z are not adjacent in C, orient X - Y - Z as X -> Y <- Z if and only if Y is not in $\mathbf{Sepset}(X,Z)$.

D. repeat

If A -> B, B and C are adjacent, A and C are not adjacent, and there is no arrowhead at B, then orient B - C as B -> C.

If there is a directed path from A to B, and an edge between A and B, then orient A - B as A -> B.

until no more edges can be oriented.



Examples: GES

1. Initialization, scoring setup:

Begin with an empty graph G that has no edges among the set of variables.

Choose of scoring setup: Bayesian Information Criterion (BIC), Bayesian Dirichlet equivalent (BDeu), Akaike Information Criterion (AIC), etc.

2. Forward Phase (Adding Edges):

Step 1: For each pair of non-adjacent nodes X and Y, evaluate the score of the graph obtained by adding the edge X→Y.

Step 2: Select the edge addition that results in the largest increase (or smallest decrease, if all additions reduce the score) in the scoring criterion.

Step 3: Add this edge to the graph.

Step 4: Repeat Steps 1-3 until adding any edge does not improve the score.

3. Backward Phase (Removing Edges):

Step 1: For each edge in the graph, evaluate the score of the graph obtained by removing the edge.

Step 2: Select the edge removal that results in the largest increase (or smallest decrease) in the scoring criterion.

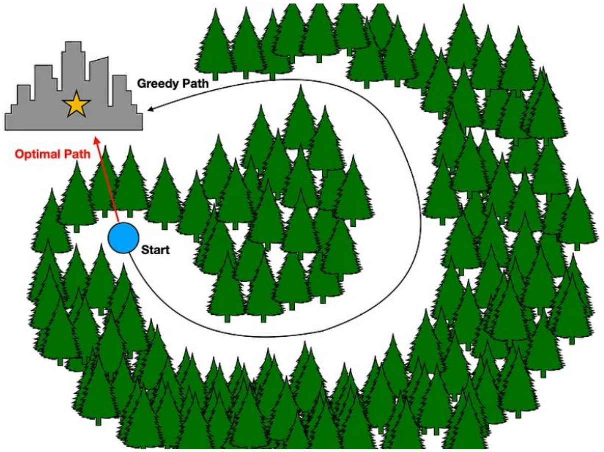
Step 3: Remove this edge from the graph.

Step 4: Repeat Steps 1-3 until removing any edge does not improve the score.

The algorithm stops when no single-edge addition or removal increases the score in the respective phases.

$$\mathrm{BIC} = \ell(\hat{\theta}|D) - \frac{p}{2} \cdot \log(n)$$

- $\ell(\hat{\theta}|D)$ is the log-likelihood of the observed data D given the estimated parameters $\hat{\theta}$,
- p is the number of parameters in the model,
- n is the sample size,



Examples: LiNGAM (Linear Non-Gaussian Acyclic Model)

1. Initialization and Assumptions:

Assumption that the observed data is generated by a mixture of **non-Gaussian** independent latent **variables** (causes),
and that these **relationships** are **linear**.

The data generating process forms a Directed Acyclic Graph (DAG) with no feedback loops.

2. Independent Component Analysis (ICA):

Apply ICA to the data to recover the independent components and estimate the mixing matrix.

ICA is chosen specifically because it can separate non-Gaussian source signals from their mixtures, which is critical for the subsequent steps in LiNGAM.

3. Ordering and Final DAG Construction:

Determine causal order using the mixing matrix's properties.

Permute independent components for a lower triangular mixing matrix.

Non-zero entries in the matrix indicate direct causal effects.

Construct the final DAG:
Represent variables as nodes.
Draw directed edges from cause to effect as indicated by the matrix.

Causal learning for time series

Granger Causality

Granger Causality: Tests if past values of one variable help predict future values of another.

Geweke's Granger Causality: A modification of traditional Granger causality that includes instantaneous feedback, allowing for contemporaneous effects.

Directed Granger Causality: Focuses on the directionality of the prediction, examining the specific influence one time series has on another, potentially within a multivariate framework.

Frequency-Domain Methods:

Partial Directed Coherence (PDC): Assesses direct influences by excluding indirect effects in frequency domain.

Directed Transfer Function (DTF): Quantifies the directional flow of information between time series at different frequencies.

Model-Based Methods:

Vector Autoregression (VAR): Captures linear interdependencies among multiple time series.

Dynamic Causal Modeling (DCM): Incorporates models of system dynamics, particularly in neuroscience.

State-Space Models: Uses latent variables to model dynamics, allowing for both linear and nonlinear causal inference.

Information-Theoretic Methods:

Transfer Entropy: Measures the reduction of uncertainty in one time series due to knowledge of another.

Directed Information Criterion (DIC): Quantifies the directed information flow, especially in non-linear and non-Gaussian time series.

Nonlinear Dynamics and Synchronization Methods:

Generalized Synchronization: Looks for synchronization in nonlinear dynamics as a sign of causality.

Phase Synchronization Methods: Examines the phase relations in oscillatory systems to infer causality.

Convergent Cross Mapping (CCM): Utilizes state-space reconstruction to detect causality in complex systems.

Nonlinear Granger Causality: Extends Granger causality framework to accommodate nonlinear relationships.

LiNGAM for Time Series:

Applies independent component analysis (ICA) for causal discovery in time series data.

Granger causality

1. Collect and model time series:

Gather time series data for X and Y, ensuring that the data is stationary or has been appropriately differenced to achieve stationarity.

Build two models to predict the current value of Y. One model uses only the past values of Y (the autoregressive model), and the other uses past values of both Y and X.

2. Calculate Predictive Accuracy:

Estimate both models and evaluate their predictive accuracy, typically by comparing the sum of squared residuals from each model.

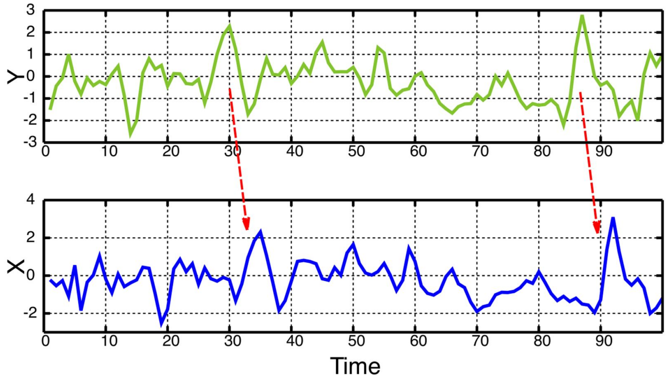
Perform a hypothesis test (often an F-test) to determine whether the inclusion of past values of X significantly improves the prediction of Y. The null hypothesis is that X does not Granger-cause Y, meaning that the coefficients on the lagged values of X are zero.

3. Statistical Testing:

Perform a hypothesis test (often an F-test) to determine whether the inclusion of past values of X significantly improves the prediction of Y.

The null hypothesis: X does not Granger-cause Y, meaning that the coefficients on the lagged values of X are zero.

If the test rejects the null hypothesis, then you conclude that X Granger-causes Y. This means that the historical values of X provide statistically significant information about future values of Y.

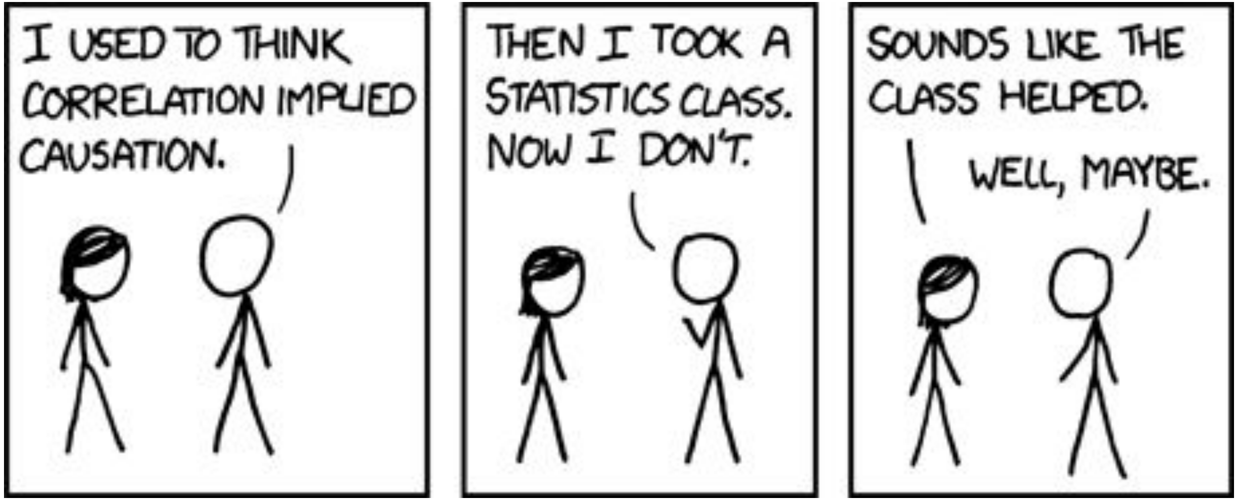


The simplest test is to estimate the regression which is based on

$$x_t = c_1 + \sum_{i=0}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{p} \beta_j y_{t-j} + u_t$$

using OLS and then conduct a F-test of the null hypothesis

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0.$$



I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.