Causal Reasoning in Signal Processing

Pavel Loskot

pavelloskot@intl.zju.edu.cn

ZJU-UIUC INSTITUTE
Zhejiang University-University of Illinois at Urbana-Champaign Institute
浙江大学伊利诺伊大学厄巴纳香槟校区联合学院

IARIA

SIGNAL 2024: ***The Ninth International Conference on Advances in Signal, Image and Video Processing***

March 10–14, 2024, Athens, Greece

About Me

Pavel Loskot joined the ZJU-UIUC Institute as Associate Professor in January 2021. He received his PhD degree in Wireless Communications from the University of Alberta in Canada, and the MSc and BSc degrees in Radioelectronics and Biomedical Electronics, respectively, from the Czech Technical University of Prague. He is the Senior Member of the IEEE, Fellow of the HEA in the UK, and the Recognized Research Supervisor of the UKCGE.

In the past 25 years, he was involved in numerous industrial and academic collaborative projects in the Czech Republic, Finland, Canada, the UK, Turkey, and China. These projects concerned mainly wireless and optical telecommunication networks, but also genetic regulatory circuits, air transport services, and renewable energy systems. This experience allowed him to truly understand the interdisciplinary workings, and crossing the disciplines boundaries.

His current research focuses on statistical signal processing and importing methods from Telecommunication Engineering and Computer Science to model and analyze systems more efficiently and with greater information power.

Objective

Explore basic ideas in causal analysis

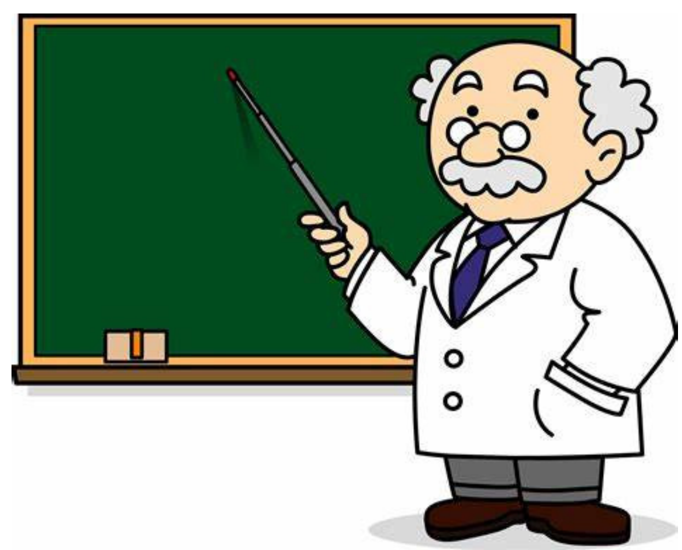
→ it involves data, models, experiments, and methods

→ it answers questions such as "*why*" and "*what if*"

→ it is becoming popular and included in university curricula

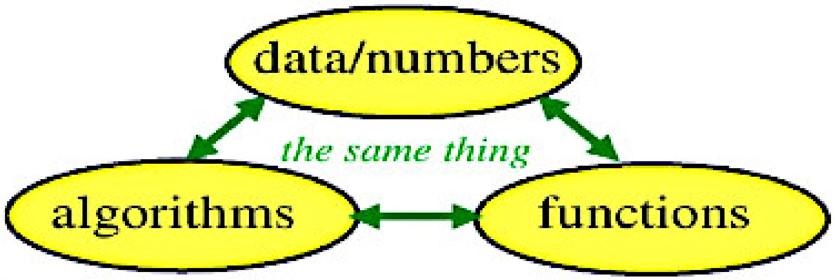
Topics

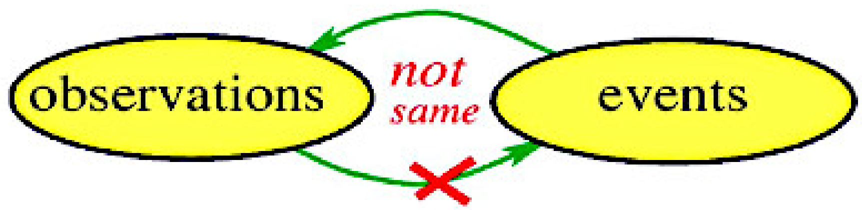
1. Causal associations
2. Experiment design
3. Structural causal models and do-calculus
4. Causality in time-series

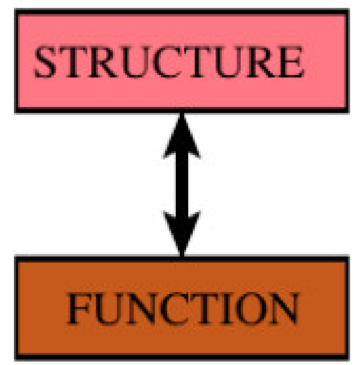


Fundamental Observations

Ignored equivalences



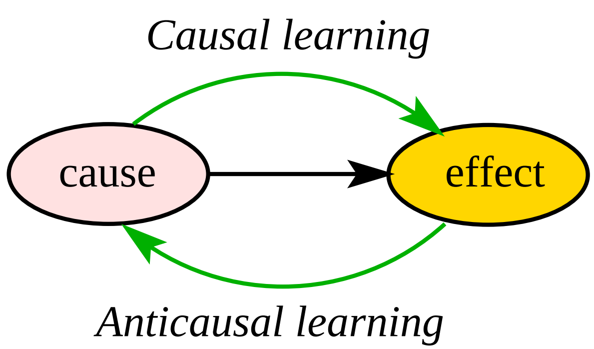


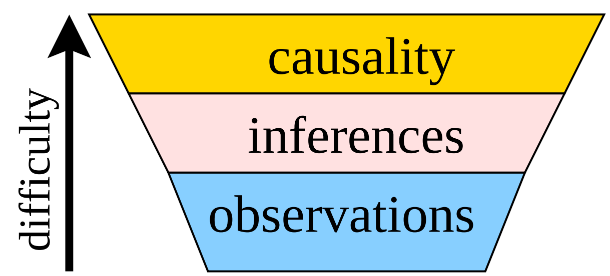


- machine learning models represented by datasets
  → input-&-output samples (=labeled data)
  → input-or-output samples (=unlabeled data)
- experiments are natural data/signal/info processing systems

Causal inference

- evidence-based explainability is becoming a ubiquitous task





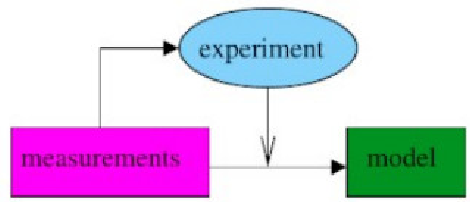
Data and Signal Processing

Aims

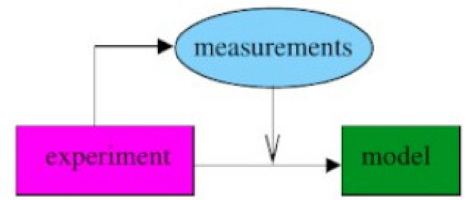
- unbiased and accurate
- sample and information efficient
- resources effective (effort, time)
- systematic, replicable, generalizable

Consideration

- data already available or not?
  → forward modeling
  → reverse modeling



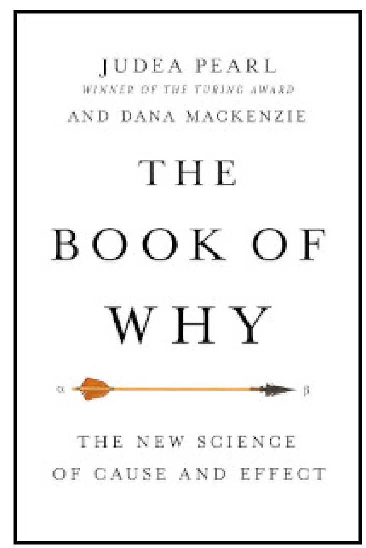
available measurements constrain possible applications



application determines required measurements

Explainability requires to

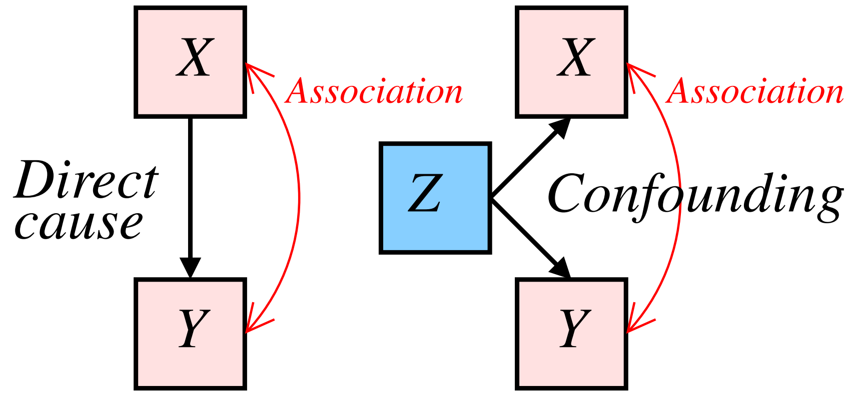
- decide about *the effect of [*independent variable*] on [*dependent variable*]*
- decide what causes change or variations in observed response
- predict unobserved response if ... (counterfactuals)
- compare responses under different settings



Associations and Causality

Observed correlations

- X and Y are correlated
- Z is confounder
- X and Y are independent conditioned on Z $\rightarrow$ associations can be accidental, spurious or conditional



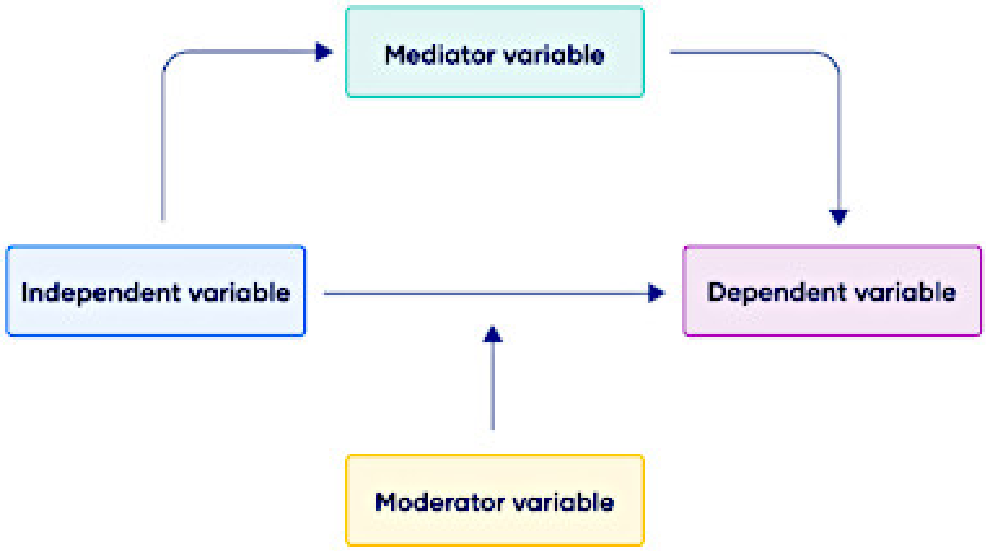
Strength of association

- strong association is neither necessary nor sufficient for causality
- weak association is neither necessary nor sufficient for absence of causality

Conditional independence

$$
\begin{array}{rccc}
 & X \perp\!\!\!\perp Y & \Rightarrow & \rho_{XY} = 0 \text{ (uncorrelated)} \\
1. & p(X,Y|Z) = p(X|Z)p(Y|Z) & \Leftrightarrow & X \perp\!\!\!\perp Y \,|\, Z \\
2. & X \perp\!\!\!\perp Y, W \,|\, Z & \Rightarrow & X \perp\!\!\!\perp Y \,|\, Z \ \vee\ X \perp\!\!\!\perp Y \,|\, W, Z \\
3. & X \perp\!\!\!\perp Y \,|\, Z \ \wedge\ X \perp\!\!\!\perp W \,|\, Y, Z & \Rightarrow & X \perp\!\!\!\perp Y, W \,|\, Z \\
4. & X \perp\!\!\!\perp Y \,|\, W, Z \ \wedge\ X \perp\!\!\!\perp W|Y, Z & \Rightarrow & X \perp\!\!\!\perp Y, W \,|\, Z
\end{array}
$$

Mechanisms Influencing Causal Relationships



Mediator

- caused by independent variable
- influences dependent variable
- can be full or partial
- increases correlations when taken into account

Moderator

- constrains the relationship between variables
- defines conditions for the relationship to exist
- influences level, direction, or presence of the relationship

Measuring Associations

Pearson correlation

$$\rho_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma_X} \right) \left(\frac{y_i - \bar{y}}{\sigma_Y} \right)$$

Spearman correlation

$$\rho_{XY} = 1 - \frac{6}{n(n^2-1)} \sum_i \text{rank\_difference}_i^2$$

$\rho_{XY} = 0 \quad : \quad X, Y$ uncorrelated
and also Gaussian $\Rightarrow$ independent

Partial correlation

$$\rho_{XY|Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{(1-\rho_{XZ}^2)(1-\rho_{YZ}^2)}}$$

$\rightarrow$ correlation between residuals of linear regression of X on Z and Y on Z

$\rho_{XY|Z} = 0 \quad : \quad X, Y$ partially uncorrelated given Z

$\rho_{XY|Z} = 0 \quad \not\Rightarrow \quad X \perp\!\!\!\perp Y | Z$

$\rho_{XY|Z} = 0 \quad \not\Leftarrow \quad X \perp\!\!\!\perp Y | Z$

Correlations in Multiple Dimensions

Problem

- measuring correlations for more than two random variables

Define

$|\boldsymbol{x}|_1 = |X_1 + X_2 + \cdots + X_N|$ (this is not l_1-norm $||\boldsymbol{x}||_1 = |X_1| + \cdots + |X_N|$)

m-th central sum-moment of random vector $\boldsymbol{X} \in \mathcal{R}^N$

$$\mu_m(|\boldsymbol{X}|_1) = \mathrm{E}\left[\left|\sum_{i=1}^{N}(X_i - \bar{X}_i)\right|^m\right], \quad m = 1, 2, \ldots$$

m-th central sum-moment for L random processes with N_l observations

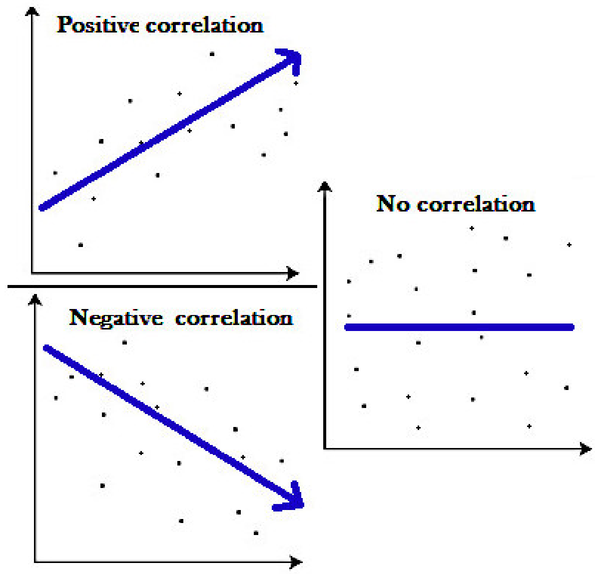
$$\mu_m(|\boldsymbol{X}_1|_1 + \ldots + |\boldsymbol{X}_L|_1) = \mathrm{E}\left[\left|\sum_{l=1}^{L}\sum_{i=1}^{N_l}(X_{li} - \bar{X}_{li})\right|^m\right]$$

P. Loskot, "Polynomial Representations of High-Dimensional Observations of Random Processes," Mathematics, 9(123), Jan. 2021.

Statistical Dependencies

Linear regression

$$
\begin{aligned}
Y = \quad & \beta_0 && \text{(constant)}\\
& +\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 && \text{(main effects)}\\
& +\beta_1 X_1^2 + \beta_2 X_2^2 + \beta_3 X_3^2 && \text{(curvature)}\\
& +\beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 && \text{(interactions)}\\
& +U && \text{(unobserved)}
\end{aligned}
$$



Instrumental variables

- induce change in explanatory variable, but no other effect on observations
- also useful when there are omitted variables affecting observations
- example:

$$Y \approx \beta X + \beta_0 + U \quad \xrightarrow{\text{LS}} \quad \hat{\beta} = \beta^* + \frac{\mathrm{cov}[X,U]}{\mathrm{var}[X]}$$

$\rightarrow$ if $\mathrm{cov}[X,U] \neq 0$, then $\hat{\beta}$ does not reflect true causal effect β^*

- assume instead

$$Y \approx \beta Z + \beta_0 + U$$

$\rightarrow$ such that $\mathrm{cov}[X,Z] \neq 0$ and $\mathrm{cov}[U,Z] = 0$

Testing for Independence

Task

- given time series $\{X_i\}$, and $\{Y_i\}$, decide if they are (conditionally) independent

Hypothesis test

- define hypotheses

$$\mathcal{H}_0 :\ X \perp\!\!\!\perp Y \quad \text{and} \quad \mathcal{H}_a :\ X \not\perp\!\!\!\perp Y$$

- choose test statistics S, and compare it to threshold, $S \lessgtr T_{\text{thr}}$

Statistics

- empirical correlation with t-test or Fisher's z-transform
- χ^2-test and odds ratios
  $\rightarrow$ use relative frequencies for conditional and marginal distributions
- non-parametric test: kernel projections/maps μ

$$S = \|\mu(p_{XY}) - \mu(p_X p_Y)\|, \quad \text{or} \quad S = \mathrm{E}[\mu_1(X)\mu_2(Y)] - \mathrm{E}[\mu_1(X)]\,\mathrm{E}[\mu_2(Y)]$$

- conditional independence on Z: test that, $X \perp\!\!\!\perp Y | Z = z$, $\forall z$
  $\rightarrow$ need to enumerate all values of Z
  $\rightarrow$ extensions for continuous Z exist

Identifying Independent Variables

PC algorithm [Peter Spirtes and Clark Glymour, parents and children]

- determines if association is causal using conditional independence tests
- find variables, so conditioning on them, remaining variables are independent

Input: data $\boldsymbol{D}$, predictor variables $\{X_i\}$, and target variable Y
Output: parent and children variables of Y

1. set $\mathrm{PC} = \{X_i\}$
2. iteratively remove variables from the PC set that are neither parents nor children of $Y \rightarrow$ test independence of Y and removed variables conditioned on the remaining variables
3. remaining variables in the PC set are parents or children of Y

False discoveries by PC algorithm

- removing variables from the PC set is sub-optimum (false positives) $\rightarrow$ wrong decisions propagate to next level
- conditional tests are done at given significance α
- several modifications of the original algorithm exist

Causal Reasoning

Causal specificity

- a cause leads to a single effect
- an effect has exactly one cause
- hypotheses: $\mathcal{H}_0$ can be causal, and $\mathcal{H}_a$ is non-causal or not specific

Temporality

- if X precedes Y, then X *could* be cause of Y
- if X cannot precede Y, then X cannot be cause of Y

Trend

- linear or monotonic dependence (regression) can be due to confounding

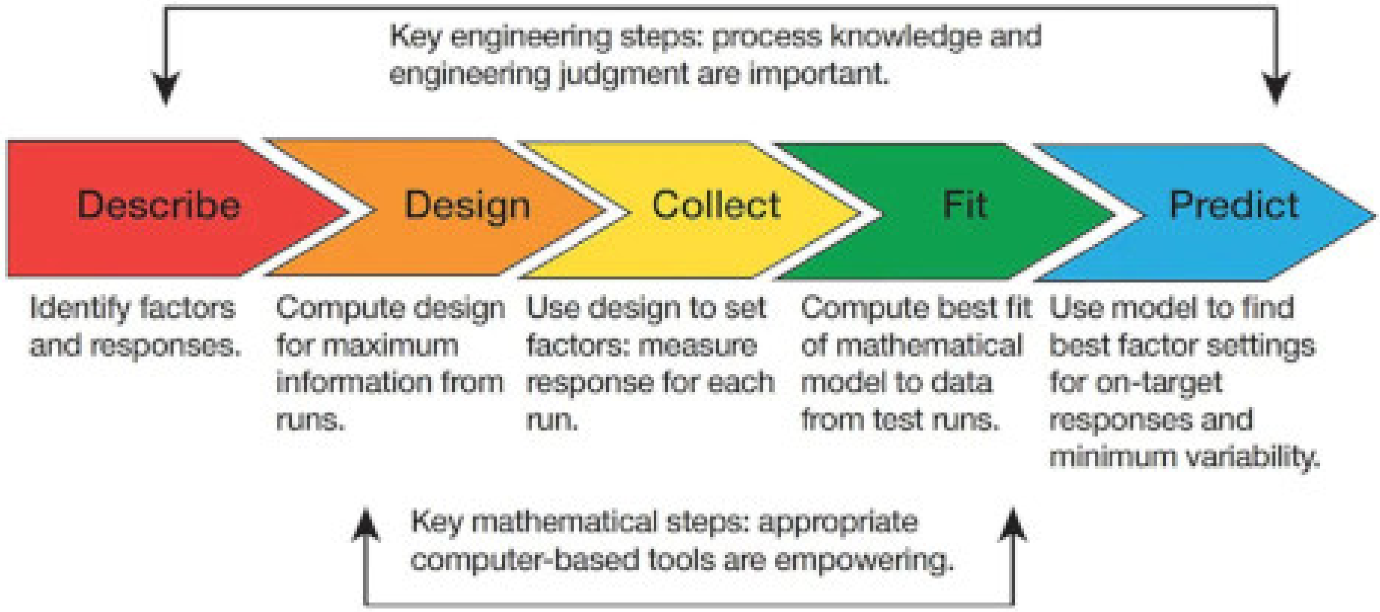
Sufficient cause

- sufficient conditions to cause or to prevent an effect

Necessary causal cause

- appears in every sufficient cause

Experiment Design



Key ideas

- manipulate (some) inputs to determine changes in observed response
- identify sources of variations
- comparative vs. observational experiments

Inputs and outputs

- controlled inputs: factors
- uncontrolled inputs: blocking variables, covariates, nuisance variables
- all inputs: predictors, independent variables
- all outputs: dependent, response variables

Subjects/Data Sampling

Objective

- contain both known and unknown confounding otherwise bias
- necessary for excessively large populations (data)

Basic methods

- simple random sampling
- stratified sampling
- cluster sampling
- systematic sampling

Sample size

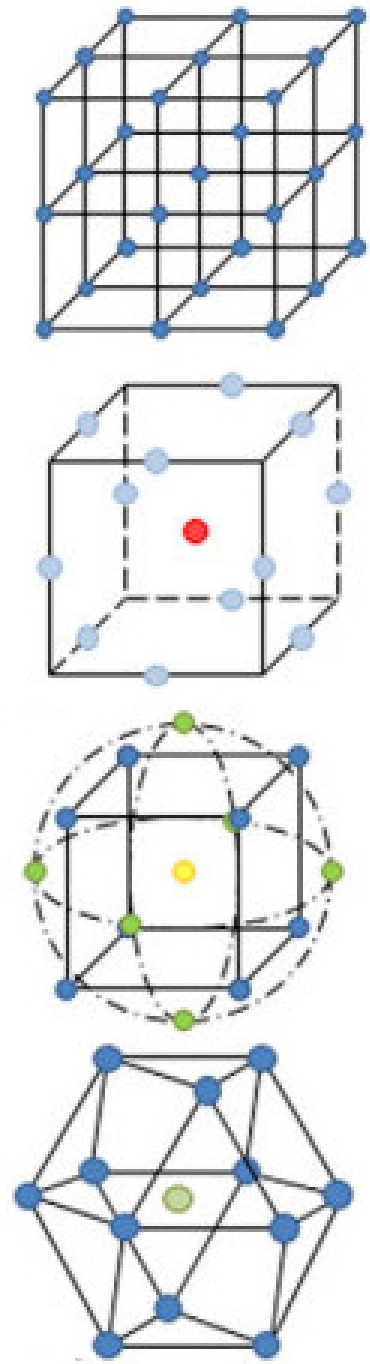
- too many samples
  → waste of resources, may not be statistically meaningful
- too few samples
  → not accurately represent population, not statistically significant

two experiments, or one experiment with twice as many samples?

Two basic strategies

- Pearson: as many samples as possible
  → more samples, more statistical power
- Fisher: fewer, but representative samples
  → detecting effect with less samples is statistically more powerful

Screening Models/Experiments



Objective

- identify key factors most affecting the outcome
- ideally, no confounding, and factors are independent
  → factor interactions may or not be statistically significant
  → separate the outcome effects from the factor interactions
- maximize/minimize outcome effects

Screening methods

- one-at-time: simple, but inefficient and unreliable
- factors quantization: low, average, high (still too complex)
- fractional factorial designs
  → experiments with selected factor settings
  → omitted experiments cause aliasing of effects
  → some outcomes can be predicted from other experiments

Remove/control confounding

- randomization, blocking, balancing
- can be also used for non-significant effects

Variance Expansions

Sobol's expansion

$$Y = f(X_1, X_2, \dots, X_N) = f_0 + \sum_i f_i(X_i) + \sum_{i<j} f_{ij}(X_i, X_j) + \cdots + \sum_{\text{except } i} f_{12\cdots N-1}(X_1, \dots, X_{N-1})$$

Variance expansion

$$V(Y) = \sum_i V_i + \sum_{i<j} V_{ij} + \sum_{i<j<l} V_{ijl} + \cdots + V_{12\ldots N}$$

where

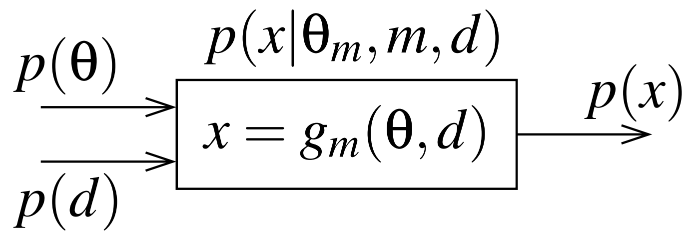
$$\begin{aligned} V_i &= V_{X_i}(\mathrm{E}_{X_{-i}}[Y|X_i]) \\ V_{ij} &= V_{X_iX_j}(\mathrm{E}_{X_{-i-j}}\left[Y|X_i, X_j\right]) - V_i - V_j \\ V_{ijl} &= V_{X_iX_jX_l}(\mathrm{E}_{X_{-i-j-l}}\left[Y|X_i, X_j, X_l\right]) - V_{ij} - V_{jl} - V_{il} - V_i - V_j - V_l \end{aligned}$$

- generally not unique, but unique if the terms are orthogonal:
  $\rightarrow$ several other strategies exist
- can be used for sensitivity analysis, factor screening and similar

Bayesian Experiment Design

design model

$$p(\theta|x,d,m) = \frac{p(x|\theta,d,m)\,p(\theta,m)}{p(x|d,m)}$$



New problem

- can choose from multiple models $m \in \{1,2,\dots,M\}$
- model (input) parameters

$$\begin{aligned} \theta \in \Omega&: \quad \text{uncontrolled unknown inputs} \\ d \in \mathcal{D}&: \quad \text{controlled known inputs} \end{aligned}$$

- the objective is to specify the optimum design d to aid estimation of θ and selection of model m from observations x

Strategy

- define average utility $U(d)$ for the experiment setting d
  $\rightarrow$ average over data x, model m, parameters θ
- the optimum experiment design

$$d^* = \mathrm{argmax}_{d\in\mathcal{D}}\ \bar{U}(d) = \mathrm{argmax}_{d\in\mathcal{D}}\ \mathrm{E}_{x,m,\theta}[U(d,x,m,\theta)]$$

Optimum Experiment Design

Objectives

- maximize utility, minimize variance, maximize information or entropy
- across whole design space while limiting computational complexity

Expected utility for single experiment

$$\bar{U}(d) = \int_{\mathcal{X}} \int_{\Theta} \underbrace{U(d,x,\theta)}_{D_{\mathrm{KL}}(p(\theta|x,d)\,||\,p(\theta))} \times \underbrace{p(\theta,x|d)}_{p(\theta|x,d)p(x|d)} \,\mathrm{d}\theta\,\mathrm{d}x \quad \Rightarrow \quad d^* = \mathrm{argmax}_{d\in\mathcal{D}}\,\bar{U}(d)$$

Batch experiment design

- expected utility of N experiments $\neq$ sum of their utilities
- perform N-times the single optimally designed experiment
  $\rightarrow$ combine outputs to reduce the variance

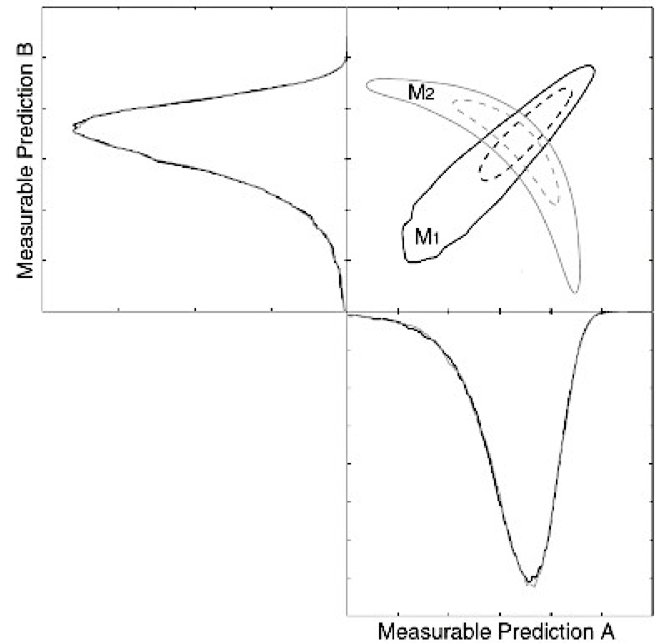
Sequential experiment design

- posterior $p(\theta, x_t|d_t)$ used as prior for the $(t+1)$-th experiment
  $\rightarrow$ can adapt sequence of models m_t
- greedy approach, the optimum design is a dynamic programming problem
  $\rightarrow$ may outperform batch design due to inherent adaptation

Model Selection

Multiple models

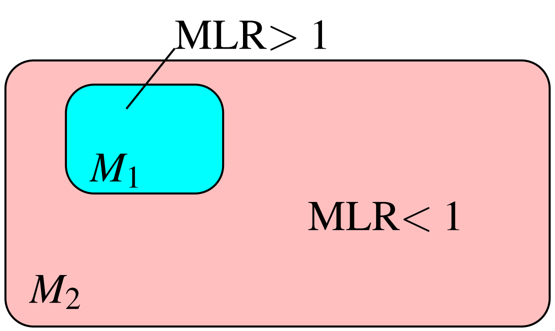
- two models M_1 and M_2 yield two predictions A and B, respectively
  $\rightarrow$ joint prediction has much larger discriminatory power than individual predictions



Vanlier et al., BMC Systems Biol., 8:20, 2014.

Comparing model confidence

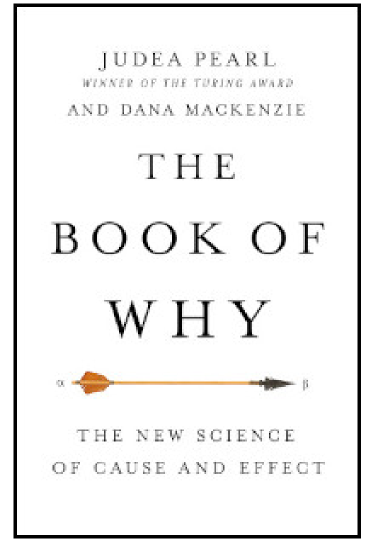
- model likelihood ratio given data X:
  $\text{MLR} =\propto \frac{p(X|M_1)}{p(X|M_2)} \lessgtr 1$
- however, M_2 is much more likely to better explain majority of random experiments



Structural Causal Models (SCM)

Key ideas

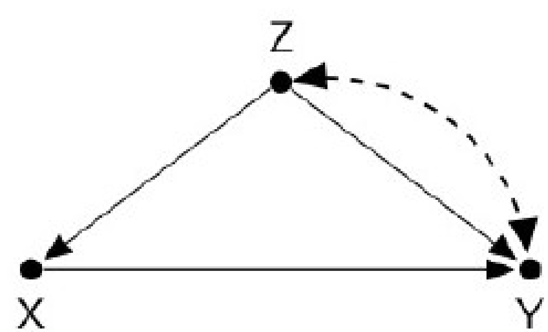
- Markov causal assumption and faithfulness condition
- extend/modify Bayesian networks
  $\rightarrow$ directed edges indicate causal effects rather than statistical dependencies
  $\rightarrow$ avoid cycles (avoid variable to be cause of itself)
- can accommodate interventions
  $\rightarrow$ $\mathrm{do}(\cdot)$ operator and do-calculus
- allow for non-linear dependencies



SCM rules

- endogenous noises are not shown explicitly
- u are exogenous unobserved variables/effects (effects outside the model)
- asymmetry: $x \leftarrow z|u \;\neq\; z \leftarrow x|u$
  symmetry: $x \perp\!\!\!\perp y|u \;=\; y \perp\!\!\!\perp x|u$
- $Z \leftarrow ---\rightarrow Y$ indicates there is unobserved common cause, i.e., $Z \leftarrow U \rightarrow Y$

$$z \leftarrow f_Z(u_z)$$
$$x \leftarrow f_X(z, u_x)$$
$$y \leftarrow f_Y(x, z, u_Y)$$



Learning SCM

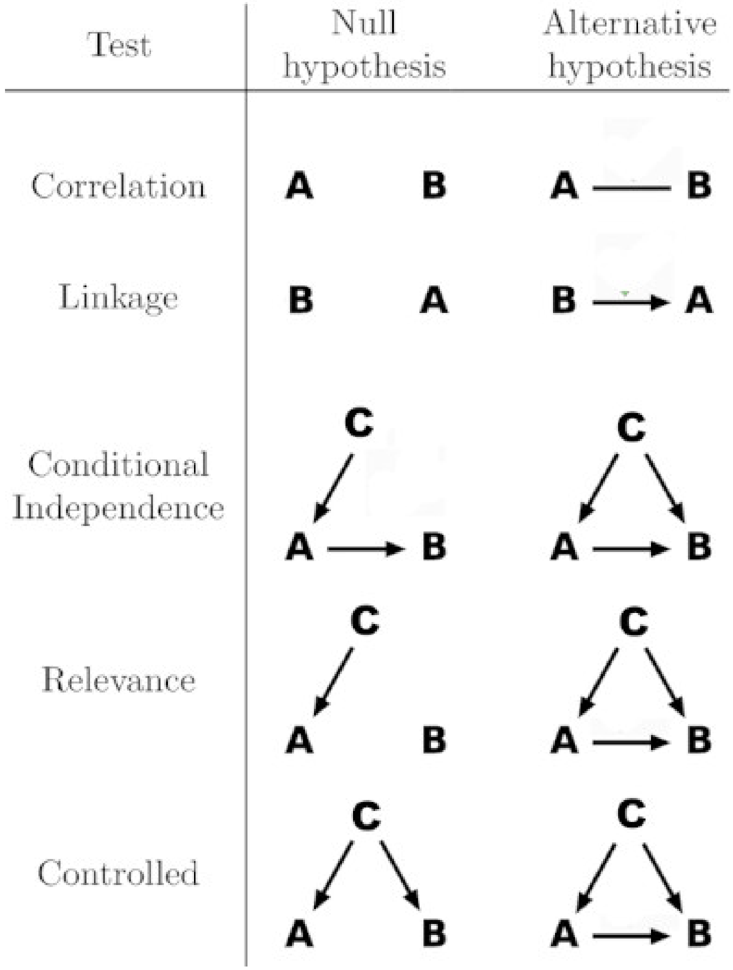
Fundamental question

can causality be discovered from observations?

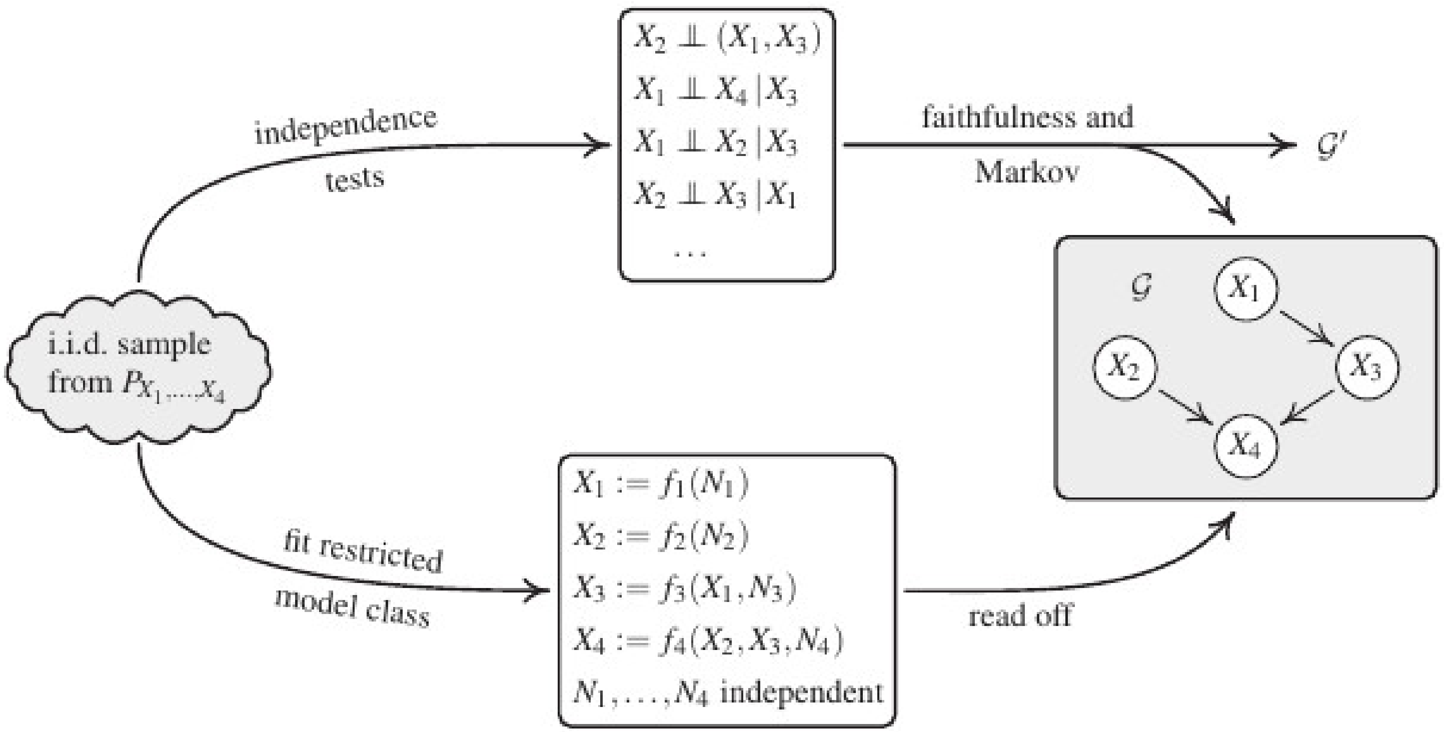
Identifiability

- can causal graph be identified from the joint distribution?
  → is this graph unique?
- basic rules
  → X and Y adjacent in graph *iff they cannot be d-separated*
  → *non-adjacent* X *and* Y *can be d-separated*
- *for Gaussian noises* N_j

(General) SCM:	$X_j := f_j(X_{\mathbf{PA}_j}, N_j)$	✗
ANM:	$X_j := f_j(X_{\mathbf{PA}_j}) + N_j$	✓
CAM:	$X_j := \sum_{k \in \mathbf{PA}_j} f_{jk}(X_k) + N_j$	✓
Linear Gaussian:	$X_j := \sum_{k \in \mathbf{PA}_j} \beta_{jk} X_k + N_j$	✗
Lin. G., eq. error var.:	$X_j := \sum_{k \in \mathbf{PA}_j} \beta_{jk} X_k + N_j$	✓



Learning SCM from Data



Main strategies

1. testing conditional independence in data
   → graph structure implied by Markov condition and faithfulness
   → (often) may not yield a unique graph
2. define the model structure and fit SCM directly
   → identify model with the best score/likelihood to fit the data

Working with SCM

D-separation

- conditional independence relationships

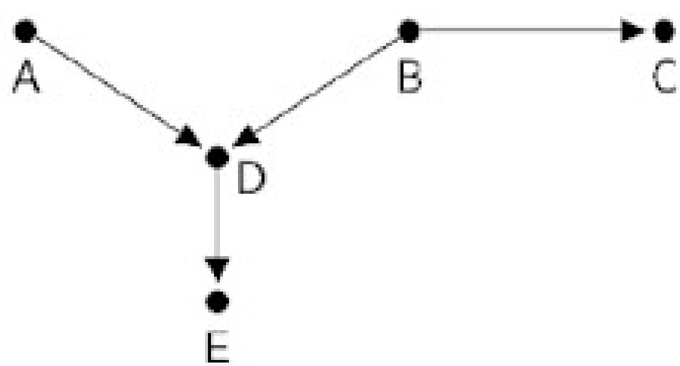
$$
\begin{array}{rlcl}
\text{chain:} & A \rightarrow B \rightarrow C & \Rightarrow & A \not\perp\!\!\!\perp C \text{ and } A \perp\!\!\!\perp C|B \\
\text{fork:} & A \leftarrow B \rightarrow C & \Rightarrow & A \not\perp\!\!\!\perp C \text{ and } A \perp\!\!\!\perp C|B \\
\text{collider:} & A \rightarrow B \leftarrow C & \Rightarrow & A \perp\!\!\!\perp C \text{ and } A \not\perp\!\!\!\perp C|B
\end{array}
$$

$\rightarrow$ conditioning on B, chain and fork block (d-separate) path $A \leftrightarrow C$

$\rightarrow$ conditioned on B, collider opens the path $A \leftrightarrow C$

SCM example

- D-separation implies the following testable causal independences
  $\rightarrow$ causal discovery: conditional independence learned from data



$$
\begin{array}{ll}
A \perp\!\!\!\perp B & A \perp\!\!\!\perp C \\
A \perp\!\!\!\perp E|D & B \perp\!\!\!\perp E|D \\
C \perp\!\!\!\perp D|B & C \perp\!\!\!\perp E|D \\
C \perp\!\!\!\perp E|B &
\end{array}
$$

Do-Calculus (1)

Pearl's rules of do-calculus

1. observations can be inserted/deleted in conditional probabilities
2. actions and observations can be exchanged in conditional probabilities
3. actions can be inserted/deleted in conditional probabilities

Inference by do-calculus

- if causal effect is identifiable, the causal effect statement can be transformed into probability expressions containing only observable variables
  $\rightarrow$ prone to automation
- unknown causal dependencies can be replaced with conditional distributions

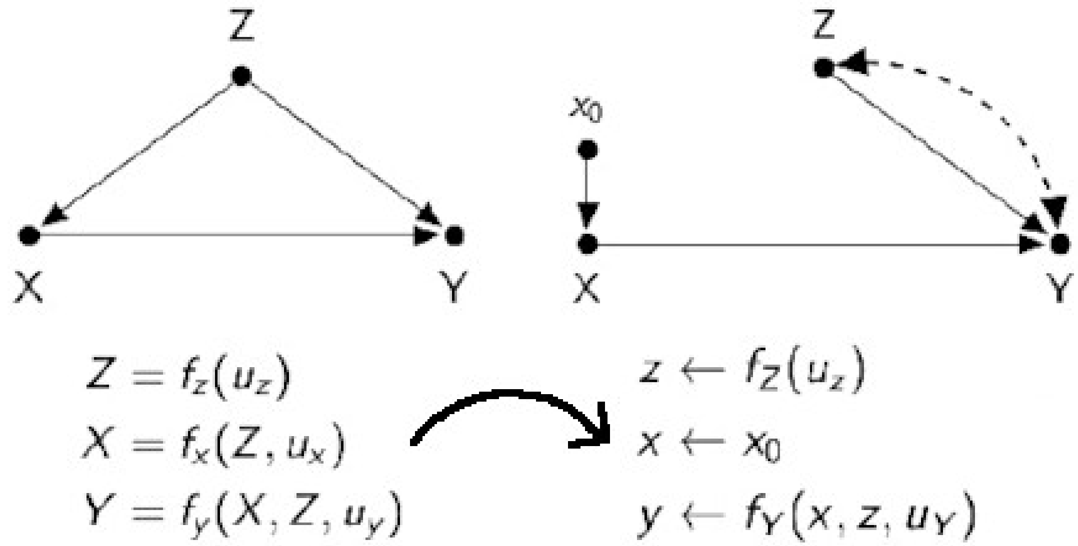
Typical applications of do-calculus

- removing confounding bias
- define surrogate experiments
- recovery from selection bias
- extrapolating causal knowledge to other scenarios

Do-Calculus (2)

Model intervention with the do-operator

- change data generation process from $P(Y|X)$ to $P(Y|\mathrm{do}(X))$
  $\rightarrow$ replace causal mechanism $f_x(x)$ with setting x to a constant x_0



Hypothetical performance improvement

- actual performance

$$Y = \int_X f(X)\, p(X)\, \mathrm{d}X$$

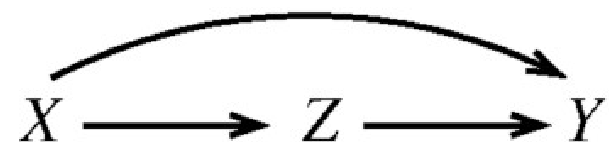
- inferred hypothetical (counterfactual) performance due to intervention

$$Y^* = \int_X f(X)\, p^*(X)\, \mathrm{d}X = \int_X f(X) \frac{p^*(X)}{p(X)}\, p(X)\, \mathrm{d}X$$

Causal Graph Examples

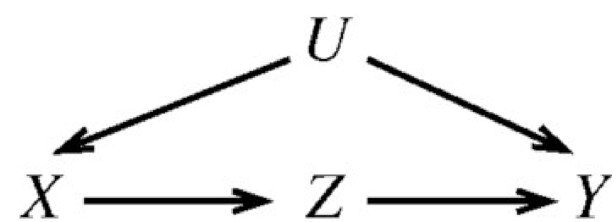
Example 1

- direct causal paths: $X \to Z$, $Z \to Y$, $X \to Y$
- backdoor path between Z and Y: $Z \leftarrow X \to Y$ (X is common cause, confounder)
- conditioned on X blocks the backdoor path and allows causal inference



Example 2

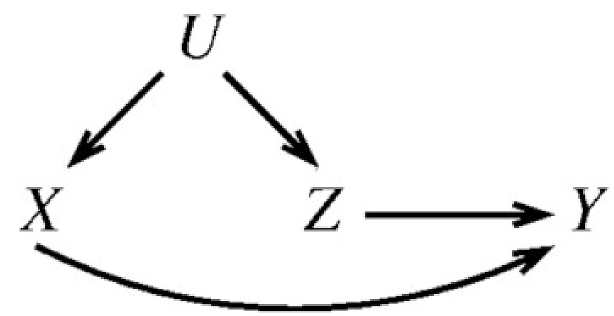
- U is unmeasured/unobserved statistics (confounding by indication)
- confounded associations: $X \to Z \to Y$, $U \to X \to Z \to Y$
- conditioning on U is not possible whilst conditioning on X removes any unmeasured confounding



Causal Graph Examples (cont.)

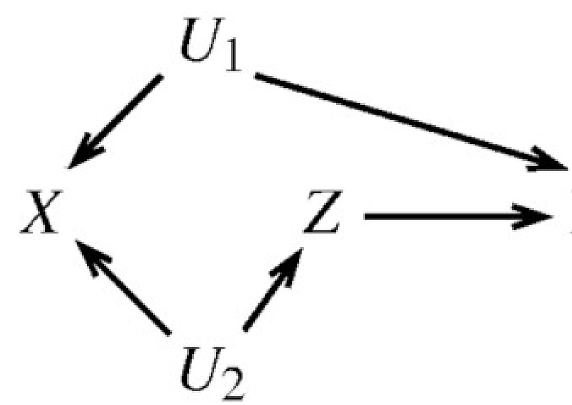
Example 3

- U is unmeasured/unobserved statistics
- conditioning on X is sufficient to block backdoor path



Example 4

- U_1 and U_2 are unmeasured, and without any conditioning, there is no bias
- fixing X will induce selection bias by opening backdoor path $Z \leftarrow U_2 \rightarrow X \leftarrow U_1 \rightarrow Y$ between Z and Y
- conditioning on X will create direct and backdoor associations between Z and Y



Working with Bayesian Networks

Key idea

- convert SCM to a Bayesian network using do-calculus
- define queries as inference tasks

$$\bar{f}_k(x_k) = \sum_{\substack{x_1,\dots,x_n \\ \text{except } x_k}} f(x_1,\dots,x_n) \quad \text{(marginalization)}$$

$$\hat{f}_k(x_k) = \max_{\substack{x_1,\dots,x_n \\ \text{except } x_k}} f(x_1,\dots,x_n) \quad \text{(maximization)}$$

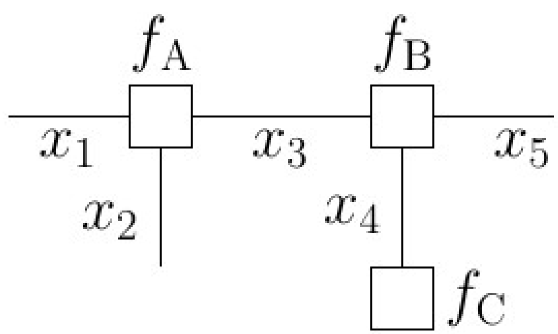
Factor graphs

$$f(X_1,X_2,\dots,X_n) = \prod_{j=1}^{m} f_j(S_j), \quad S_j \subseteq \{X_1,\dots,X_n\}$$

$$f(X_1,X_2,\dots,X_n) = \sum_{j=1}^{m} f_j(S_j), \quad S_j \subseteq \{X_1,\dots,X_n\}$$

Example

$$f(x_1,x_2,x_3,x_4,x_5) = f_A(x_1,x_2,x_3)\cdot f_B(x_3,x_4,x_5)\cdot f_C(x_4)$$



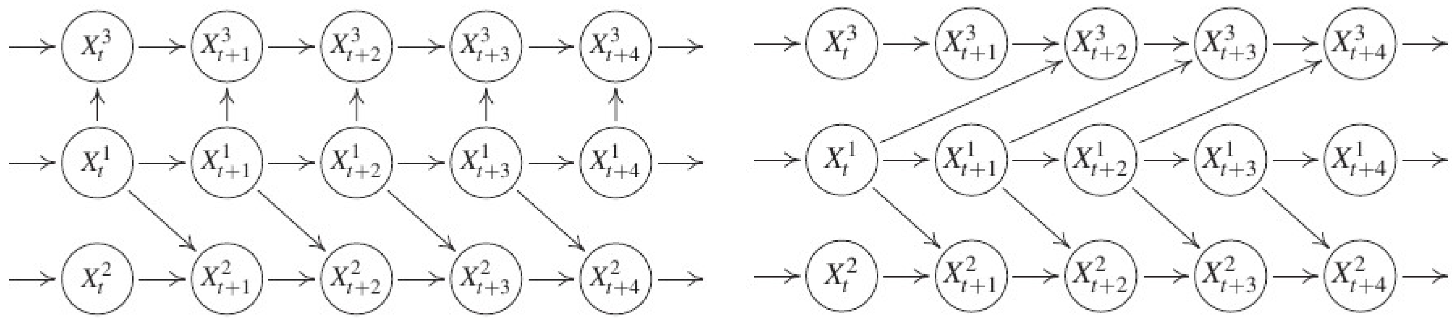
Potential Outcomes Framework

Key ideas

- evaluate potential outcome of an action or intervention
  → whatever action, counterfactual outcome is never known
  $$O(A=1) \quad \text{and} \quad O(A=0)$$
- quantify the factor change (type, duration, frequency) to cause the outcome
  → i.e., not to determine if the factor is a cause
  → counterfactual inference is agnostic to identifying actual causes
- compute the average causal effect
  $$\boxed{\text{ACE} = \mathrm{E}[O(A=1) - O(A=0)]}$$
  → add comparison group to represent a counterfactual scenario
  → many variations how to define/choose comparison group
- potential outcomes and actual interventions must be independent
  → control for confounding (randomization, ...)
  → but the more variables to control, the more difficult experiment design

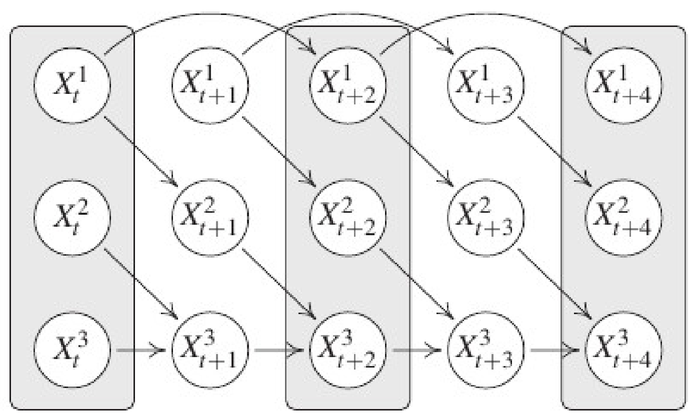
Causality in Time-Series (1)

Instantaneous effects



- summary graph: $X^3 \leftarrow X^1 \rightarrow X^2$
  $\rightarrow$ summary graphs can be cyclic

Subsampling



- summary graph: $X^1 \rightarrow X^2 \rightarrow X^3$
- interventions during observed instances: no causal effect from X^1 to X^2
  $\rightarrow$ there are ways to account for hidden causal effects

Causality in Time-Series (2)

Identifiability

- different for time series with/without instantaneous effects
- identify if cause-effect exists, and preferably also its direction

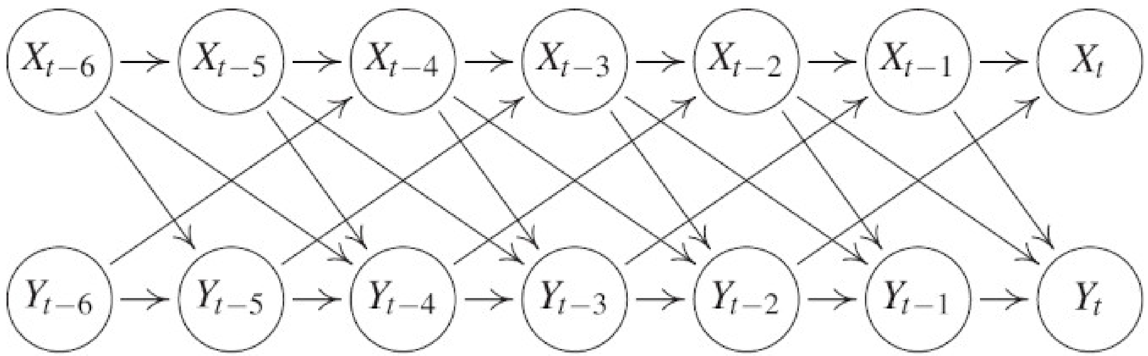
Identifiability theorems

1. Two SCM without instantaneous effects are equal, if the corresponding full time graphs are Markov equivalent.
2. Two SCM are equal, if the corresponding full time graphs are Markov equivalent and their summary graphs are acyclic.
3. Justification of Granger causality: If SCM does not have instantaneous effects, and the joint distribution has faithful property, then summary graph contains $X^i \to X^j$ *iff* $X_t^j \not\perp\!\!\!\perp X^i_{\text{past}(t)} \,|\, X^{-i}_{\text{past}(t)}$.

Examples

- if $Y_t \not\perp\!\!\!\perp X_{\text{past}(t)} \,|\, Y_{\text{past}(t)}$, then $X \to Y$
- if there are no instantaneous effects, and $Y_t \perp\!\!\!\perp X_{\text{past}(t)} \,|\, Y_{\text{past}(t)}$, then X does not cause Y

Causality in Time-Series (3)



Granger causality

- test independence $Y_t \perp\!\!\!\perp X_{\text{past}(t)} \,|\, Y_{\text{past}(t)}$ to infer summary graphs: X and Y, $X \to Y$, $X \leftarrow Y$, $X \leftrightarrows Y$
- formally, X Granger-causes $Y \iff Y_t \not\perp\!\!\!\perp X_{\text{past}(t)} \,|\, Y_{\text{past}(t)}$ $\to$ past history of X_t helps to predict Y_t
- alternatively, time series X_t Granger-causes time series Y_t, if
$$\text{var}[Y_t|Y_{t-\tau}, X_{t-\tau}] < \text{var}[Y_t|Y_{t-\tau}]$$
- the lag τ can be determined using information criteria (Akaike, Schwartz)

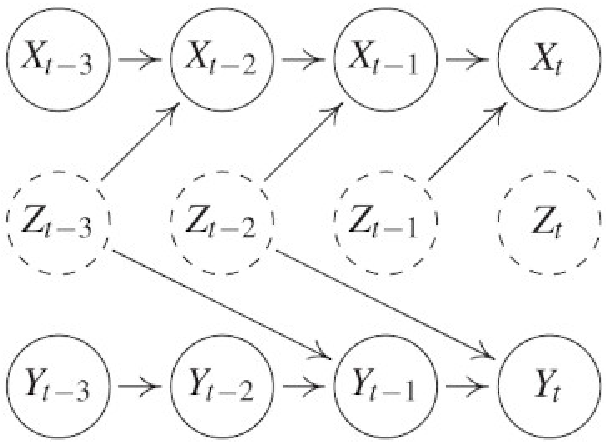
In practice

$$Y_t \approx \sum_i a_i X_{t-i} + \sum_j b_j Y_{t-j} + u_t$$

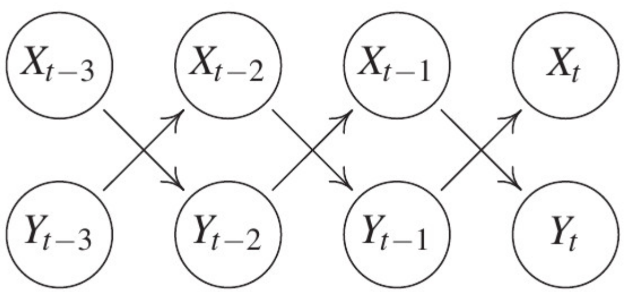
- hypothesis $\mathcal{H}_0$: $a_i = 0$ is a better model (then, X_t does not cause Y_t)
- use F-statistics in the modified Wald test

Causality in Time-Series (4)

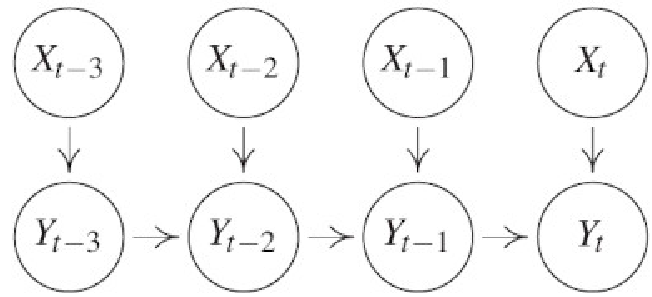
Limitations of Granger causality



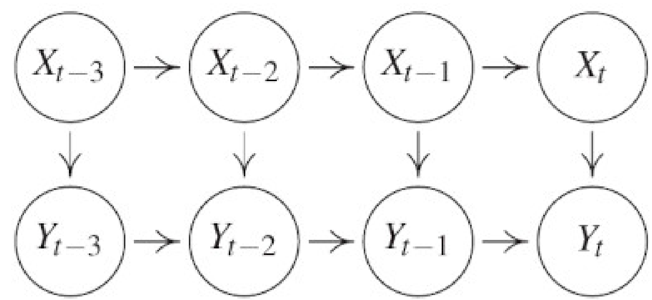
- due to common cause Z, X and Y are (erroneously) detected as Granger-causal



- for deterministic influences, they cannot be detected by Granger causality



- Granger causality cannot detect influence of X on Y



- Granger causality correctly detects influence of X on Y

Causality in Time-Series (5)

Intervention causality

- idle regime: no intervention to X_t
- atomic intervention: $X_t = X^*$
- conditional intervention: $X_t = g_t(X_{1:t-1})$
- random intervention: $X_t \sim p_t(X_t|X_{1:t-1})$

Average causal effect (ACE)

- assume intervention σ_t in X_t at time t, then

$$\mathrm{ACE}(t+\tau;\sigma_t) = \mathrm{E}_{\sigma_t}[X_{t+\tau}] - \mathrm{E}[X_{t+\tau}]\,,\ \tau > 0$$

- difference of difference

$$\mathrm{DoD}(t+\tau;\sigma_t,\sigma_t') = \mathrm{ACE}(t+\tau;\sigma_t) - \mathrm{ACE}(t+\tau;\sigma_t')$$

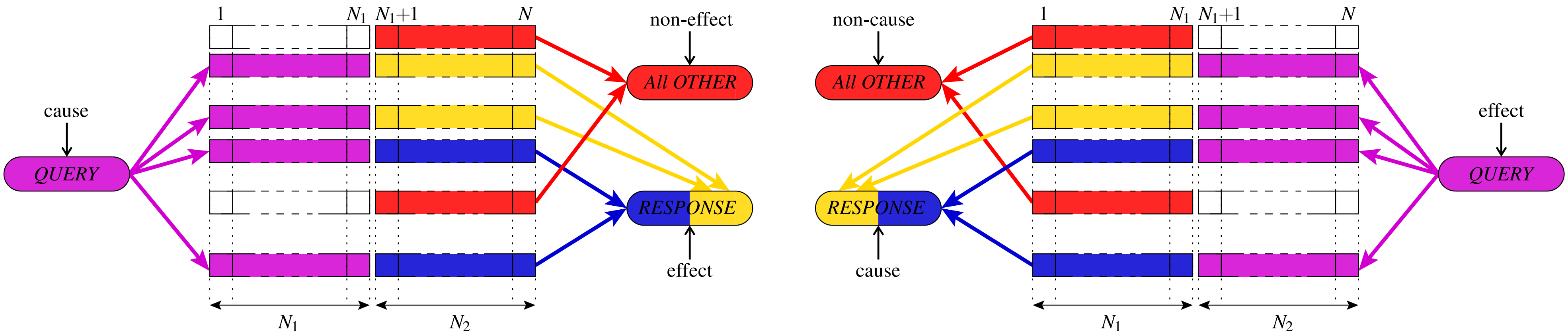
- can assume other statistics e.g. variance

Structural causality

$$\begin{aligned} X_t &= f(X_{1:t-1}, Y_{1:t-1}, Z_{1:t-1}, U_t) \\ Y_t &= g(X_{1:t-1}, Y_{1:t-1}, Z_{1:t-1}, V_t) \end{aligned}$$

- f and g are known
- Z are all observed variables, (U,V) are unobserved variables
- if X does not structurally cause Y, then $\mathrm{E}_{\sigma_t}[h(Y_{t+\tau})] = \mathrm{E}[h(Y_{t+\tau})]$

Causality in Chemical Reaction Networks



Task

- identify causal associations between subsequences $\boldsymbol{e}_i$ of reaction events
- exploit empirical conditional probabilities (a.k.a. attentions)
- ordering of reactions within $\boldsymbol{e}_i$ is irrelevant

Define causality as

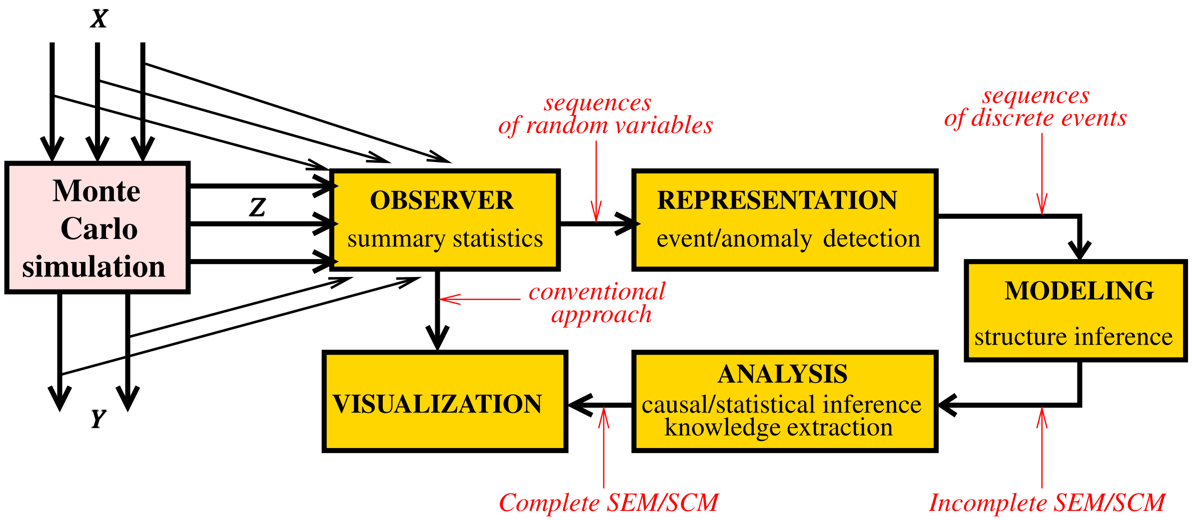
1. $\boldsymbol{e}_i$ causes $\boldsymbol{e}_j$, if $\Pr\left(\boldsymbol{e}_j|\boldsymbol{e}_i\right) \to 1$ (certain conditional event)
2. $\boldsymbol{e}_i$ does not cause $\boldsymbol{e}_j$, if $\Pr\left(\boldsymbol{e}_j|\boldsymbol{e}_i\right) \to 0$ (uncertain conditional event)

P. Loskot, Computational Biology and Chemistry, 2024.

Causal Learning

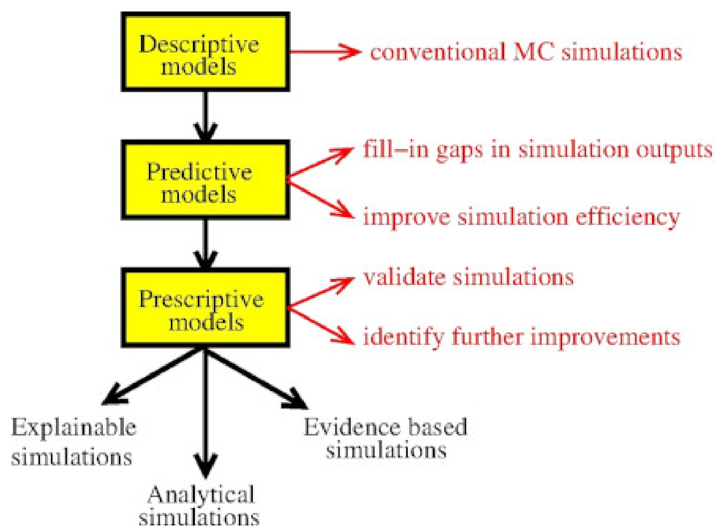
$$\begin{array}{lcl} \Pr(\text{cause}) & \perp\!\!\!\perp & \Pr(\text{effect}\,|\,\text{cause}) \\ \Pr(\text{effect}) & \not\perp\!\!\!\perp & \Pr(\text{cause}\,|\,\text{effect}) \end{array}$$

Augmented Monte-Carlo Simulations



SEM/SEC model

- inputs X, outputs Y
- augmented outputs Z
- inferred latent statistics U
- identified events E
- identified associations A



Take-Home Messages

Causality

- relies on statistical inferences and probabilistic models
- can expand capabilities of machine learning
- is intimately connected with explainability
- is also required for replicable outcomes, automate knowledge discovery

Key ideas

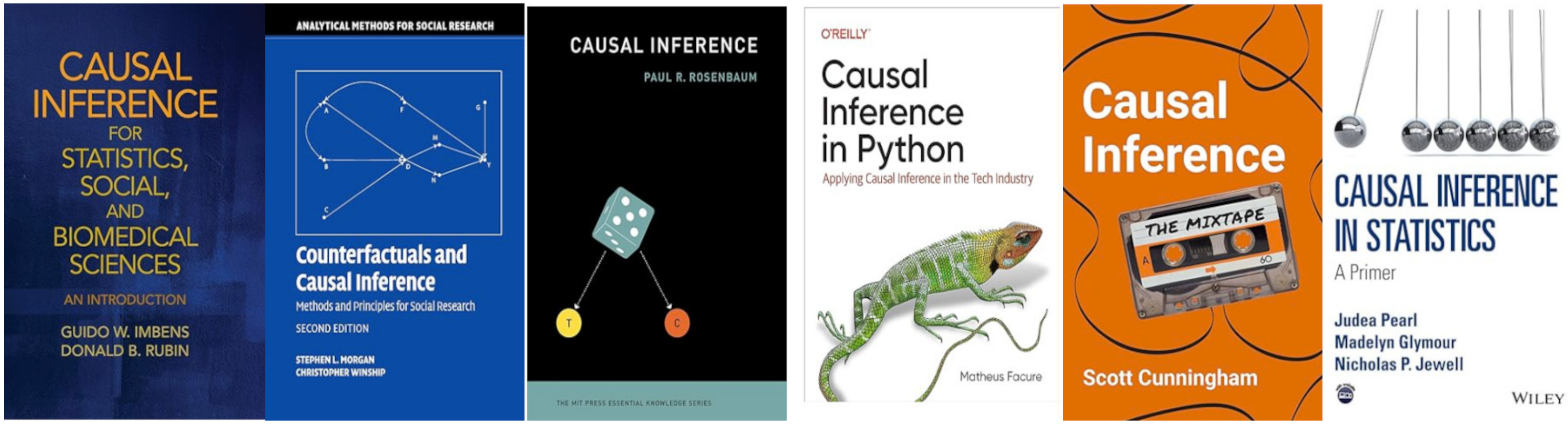
- identify cause-effect relationships (direction, strength)
- association does not imply causality
- causality can be learned via independence testing
- learning causal relationship from data is often difficult
- methods, data and experiments are equivalent representations

Inferring causality

- interventions and counterfactuals
- SCM, do-calculus, and do-operator

$$P(Y|X) \neq P(Y|\mathrm{do}(X))$$

Textbooks on Causal Inference



Python libraries for causal analysis

PC algorithm	`https://github.com/topics/pc-algorithm`
GES	`https://github.com/juangamella/ges`
LiNGAM	`https://sites.google.com/view/sshimizu06/lingam`
GOLEM	`https://github.com/ignavierng/golem`
gCastel	`https://github.com/huawei-noah/trustworthyAI/tree/master/gcastle`
CD Toolbox	`https://github.com/FenTechSolutions/CausalDiscoveryToolbox`

Thank you!

pavelloskot@intl.zju.edu.cn