THE UNIVERSITY of EDINBURGH

Methods for Causal Inference
Lecture 18: Additive Noise Models

Ava Khamseh

School of Informatics

2024-2025

Causal Discovery Methods (based on graphical models)

Class of Algorithm	Name	Assumptions	Short comings	Input
Constraint-based	PC (oldest)	Any distribution, No unobsv. confounders, Markov cond, faithfulness	Causal info only up to equivalence classes, Non bivariate	Complete undirected graph
	FCI	Any distribution, Asymptotically correct with confounders, Markov cond, faithfulness		
Score-based	GES	No unobsv. confounders	Non-bivariate	Empty graph, adds edges, removes some
Functional Causal Models (FCMs)	LinGAM/ ANM	Asymmetry in data	Requires additional assumptions (not general), harder for discrete data	Structural Equation Model

Glymour et al. (2019)

Constraint-based assumptions

- **Markov condition:**
    - Absent edge implies conditional independence (**CI**)
    - Observing conditional dependence implies an edge

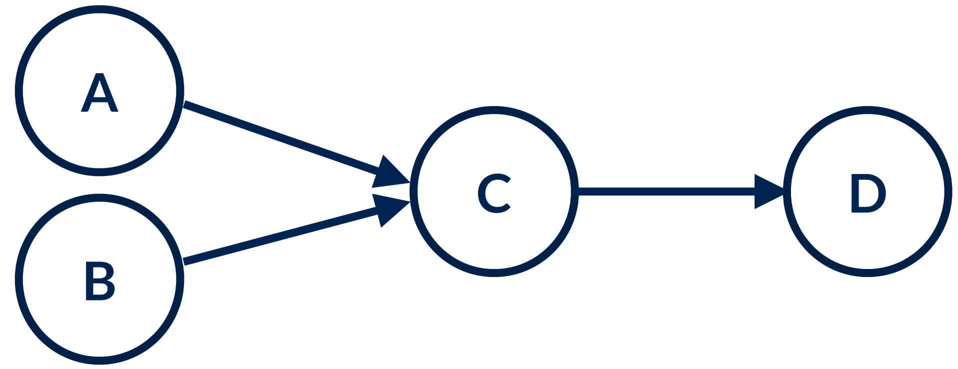
- **Causal sufficiency:** For any pair of variables X, Y, if there exists a variable Z which is a direct of cause of both X and Y, then Z is included in the causal graph (Z may be unobserved)

- **Faithfulness:**
    - Conjugate to the Markov condition
    - Edge implies conditional dependence
    - Observing CI implies absence of an edge

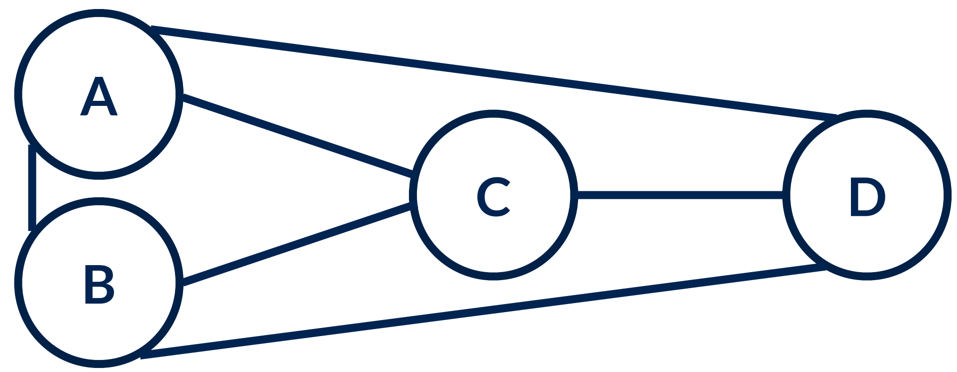
Could fail in regulatory systems, e.g., homeostasis.

Peter-Clark (PC) Algorithm

True causal graph:

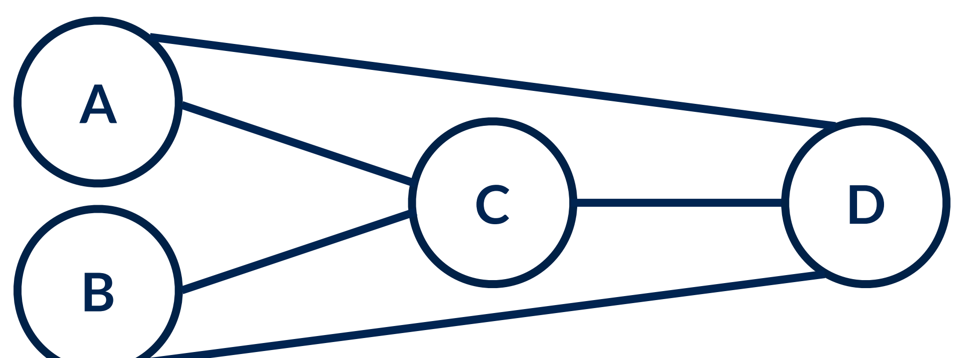


1. Start with the complete graph



2. Zeroth order CI, $A \perp\!\!\!\perp B$, by faithfulness:

Need statistical independence testing.

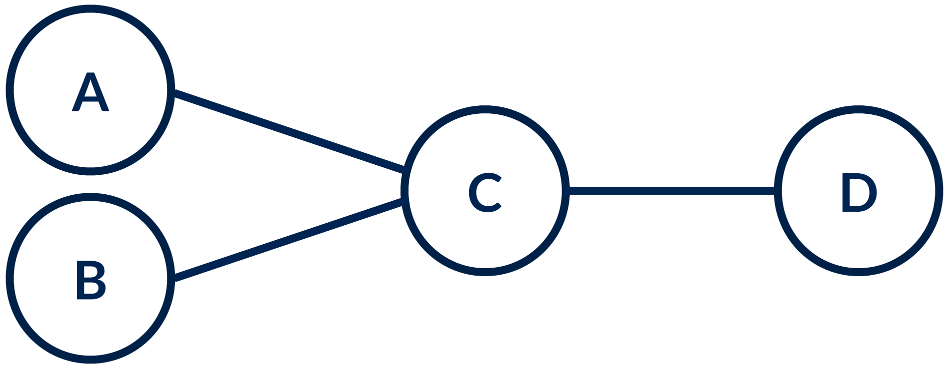


Glymour et al. (2019)

Peter-Clark (PC) Algorithm

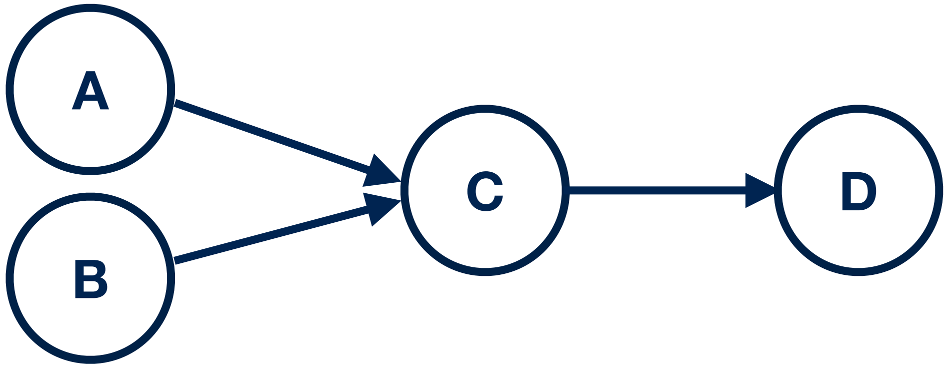
3. 1st order CI, $A \perp\!\!\!\perp D | C$, by faithfulness:

$B \perp\!\!\!\perp D | C$



4. No higher order CI observed. Notice that conditioning sets only need to contain **neighbours** for the two nodes due to the Markov condition. We do not know the parents but parents are a subsets of neighbours. As the graph becomes sparser, the number of tests to be performed decreases. This makes PC very efficient.

5. Orient V-structures (colliders): take triplets where 2 nodes are connected to the 3rd: $A \not\perp\!\!\!\perp B | C$ only.

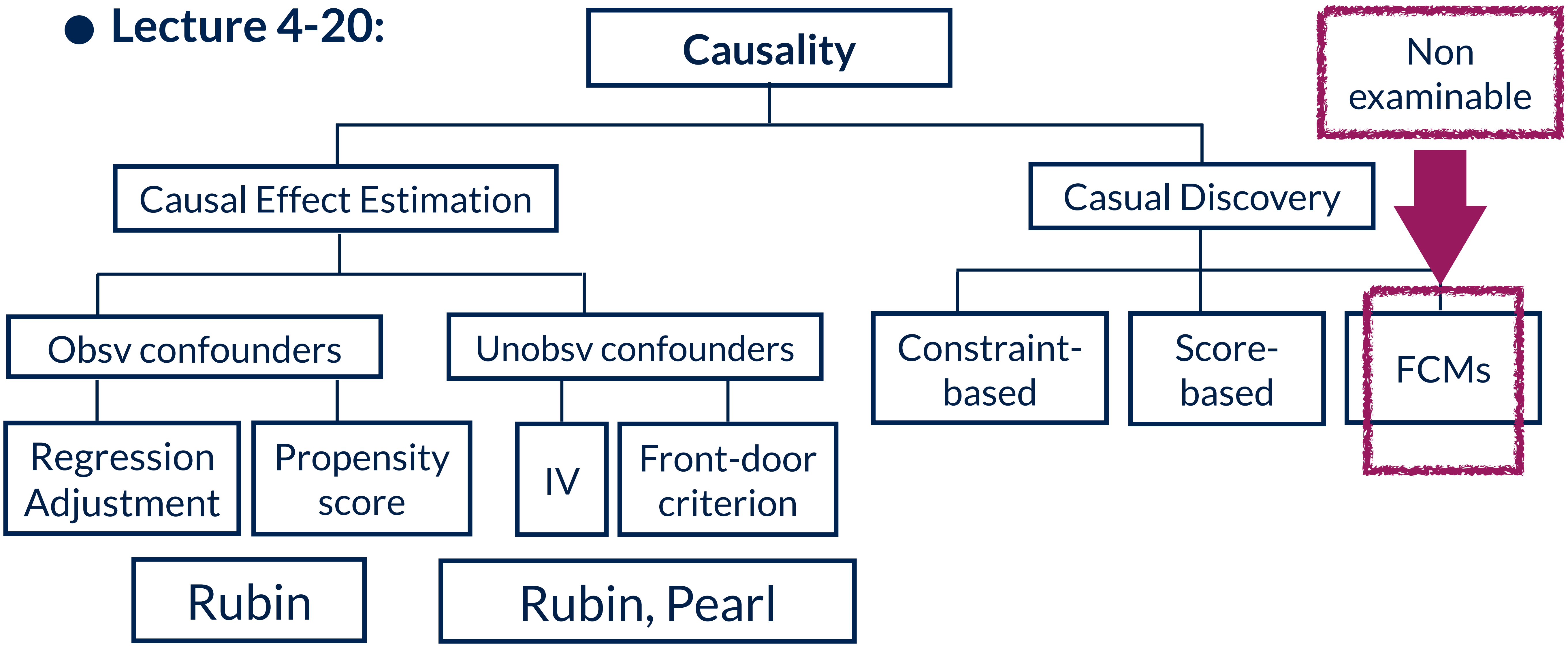


Note $C \leftarrow D$ cannot be as it would have been a collider (not detected in 5)

Glymour et al. (2019)

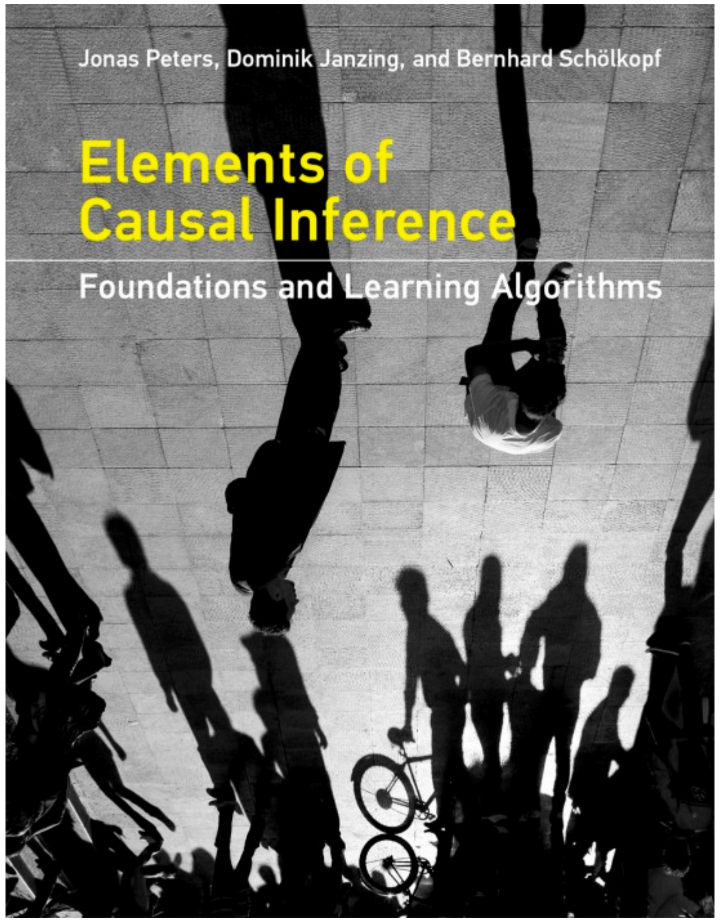
Overview of the course

- **Lecture 1**: Introduction & Motivation, why do we care about causality? Why deriving causality from observational data is non-trivial.
- **Lecture 2:** Recap of probability theory, variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3**: Recap of regression, multiple regression, graphs, SCM
- **Lecture 4-20:**



FCMs/LiNGAMs/ANMs/IGCI

- **Functional Causal Models (FCMs):** Utilising asymmetry in data for causal discovery
- **LiNGAMs:** Linear non-gaussian acyclic models, allow for new approaches for causal learning from observational data
- **ANM:** Additive noise models and **causal identifiablity**
- **IGCI:** Information Geometric Causal Inference



Jonas Peters et al, Elements of Causal Inference (2017)

Causal Structure Identifiability

- **LiNGAMs:** Linear non-gaussian acyclic models, allow for new approaches for causal learning from observational data.

- Focusing on 2 variables only, we wish to distinguish between:

$$x \to y \text{ or } y \to x$$

- from **observational data**.

- Assumption: The effect on E is a linear function of C up to additive noise:

$$E = \alpha C + N_E, \quad N_E \perp\!\!\!\perp C$$

These assumptions are not enough to identify cause/effect.

Theorem: Identifiability of LiNGAMs

i.e., non-identifiability of gaussian Cause and Effect. If:

$$Y = \alpha X + N_Y, \quad N_Y \perp\!\!\!\perp X$$

There exists a β and a random variable N_X s.t.:

$$X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y$$

if and only if $(X, N_Y) \sim \mathcal{N}$ are gaussian.

i.e., it is sufficient that for X (Y) or N_Y (N_X) to be **non-gaussian** to render the causal direction identifiable.

[The proof is non-examinable]

Theorem: Identifiability of LiNGAMs

Proof:

1. Theorem (Darmois-Skitvic): Let $x_1, \cdots, x_d$ be independent, non-degenerate random variable. If there exists non-vanishing coefficients $a_1, \cdots, a_d$ and $b_1, \cdots, b_d$ such that the two linear combinations:

$$l_1 = a_1 x_1 + \cdots + a_d x_d$$
$$l_2 = b_1 x_1 + \cdots + b_d x_d$$

$l_1 \perp\!\!\!\perp l_2$ are independent, then each x_i is normally distributed

Theorem: Identifiability of LiNGAMs

Proof:

1. Theorem (Darmois-Skitvic): Let $x_1, \cdots, x_d$ be independent, non-degenerate random variable. If there exists non-vanishing coefficients $a_1, \cdots, a_d$ and $b_1, \cdots, b_d$ such that the two linear combinations:

$$l_1 = a_1 x_1 + \cdots + a_d x_d$$
$$l_2 = b_1 x_1 + \cdots + b_d x_d$$

$l_1 \perp\!\!\!\perp l_2$ are independent, then each x_i is normally distributed

2. Lemma (Peters 2008): Let $X \perp\!\!\!\perp N$. Then $N \not\perp\!\!\!\perp (X + N)$

Theorem: Identifiability of LiNGAMs

Proof:

1. Theorem (Darmois-Skitvic): Let $x_1, \cdots, x_d$ be independent, non-degenerate random variable. If there exists non-vanishing coefficients $a_1, \cdots, a_d$ and $b_1, \cdots, b_d$ such that the two linear combinations:

$$l_1 = a_1 x_1 + \cdots + a_d x_d$$
$$l_2 = b_1 x_1 + \cdots + b_d x_d$$

$l_1 \perp\!\!\!\perp l_2$ are independent, then each x_i is normally distributed

2. Lemma (Peters 2008): Let $X \perp\!\!\!\perp N$. Then $N \not\!\perp\!\!\!\perp (X + N)$

3. We prove that $\underset{N_Y \perp\!\!\!\perp X}{Y = \alpha X + N_Y} \Rightarrow X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y$
iff $(X, N_Y) \sim \mathcal{N}$

Theorem: Identifiability of LiNGAMs

Proof:

3 We prove that if $(X, N_Y) \sim \mathcal{N}$ and $Y = \alpha X + N_Y, N_Y \perp\!\!\!\perp X$

$$\Longrightarrow X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y$$

Theorem: Identifiability of LiNGAMs

Proof:

3 We prove that if $(X, N_Y) \sim \mathcal{N}$ and $Y = \alpha X + N_Y, N_Y \perp\!\!\!\perp X$

$$\Longrightarrow X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y$$

Define:

$$\beta := \frac{Cov[X,Y]}{Cov[Y,Y]} = \frac{\alpha Var[X]}{\alpha^2 Var[X] + Var[N_Y]}$$

$$X = \beta Y + N_X \Rightarrow N_X = X - \beta Y$$

Theorem: Identifiability of LiNGAMs

Proof:

3 We prove that if $(X, N_Y) \sim \mathcal{N}$ and $Y = \alpha X + N_Y, N_Y \perp\!\!\!\perp X$

$$\Longrightarrow X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y$$

Define:

$$\beta := \frac{Cov[X,Y]}{Cov[Y,Y]} = \frac{\alpha Var[X]}{\alpha^2 Var[X] + Var[N_Y]}$$

$$X = \beta Y + N_X \Rightarrow N_X = X - \beta Y$$

$$\begin{aligned} Cov[N_X, Y] &= Cov[X - \beta Y, Y] = Cov[X,Y] - \beta Cov[Y,Y] \\ &= Cov[X,Y]\left(1 - \beta \frac{Cov[Y,Y]}{Cov[X,Y]}\right) \\ &= Cov[X,Y]\left(1 - \beta \times \beta^{-1}\right) = 0 \end{aligned}$$

Theorem: Identifiability of LiNGAMs

Proof:

3 We prove that if $(X, N_Y) \sim \mathcal{N}$ and $Y = \alpha X + N_Y, N_Y \perp\!\!\!\perp X$

$$\implies X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y$$

Define:

$$\beta := \frac{Cov[X,Y]}{Cov[Y,Y]} = \frac{\alpha Var[X]}{\alpha^2 Var[X] + Var[N_Y]}$$

$$X = \beta Y + N_X \Rightarrow N_X = X - \beta Y$$

Then N_X, Y are uncorrelated by construction,

Theorem: Identifiability of LiNGAMs

Proof:

3 We prove that if $(X, N_Y) \sim \mathcal{N}$ and $Y = \alpha X + N_Y, N_Y \perp\!\!\!\perp X$

$\Longrightarrow X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y$

Define:

$$\beta := \frac{Cov[X,Y]}{Cov[Y,Y]} = \frac{\alpha Var[X]}{\alpha^2 Var[X] + Var[N_Y]}$$

$$X = \beta Y + N_X \Rightarrow N_X = X - \beta Y$$

Then N_X, Y are uncorrelated by construction,

Moreover, Y is gaussian because $(X, N_Y) \sim \mathcal{N}$

Therefore, N_X is also gaussian.

Hence, N_X, Y are uncorrelated & gaussian, i.e., **independent**.

Theorem: Identifiability of LiNGAMs

Proof:

3 We prove the reverse: If

$$Y = \alpha X + N_Y, \quad N_Y \perp\!\!\!\perp X$$
$$X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y$$

$(X, N_Y) \sim \mathcal{N}$ $\Rightarrow$

Since $N_X \perp\!\!\!\perp Y$, we have: $N_X = X - \beta(\alpha X + N_Y) = (1 - \alpha\beta)X - \beta N_Y$

Theorem: Identifiability of LiNGAMs

Proof:

3 We prove the reverse: If $\boxed{Y = \alpha X + N_Y,}\quad N_Y \perp\!\!\!\perp X$
$X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y$ $\Rightarrow$ $(X, N_Y) \sim \mathcal{N}$

Since $N_X \perp\!\!\!\perp Y$, we have: $\boxed{N_X} = X - \beta(\alpha X + N_Y) = \boxed{(1 - \alpha\beta)X - \beta N_Y}$

There are 3 cases:

(i) $(1 - \alpha\beta) \neq 0 \ \& \ \beta \neq 0$

Then, given $N_X \perp\!\!\!\perp Y$, DS theorem implies $X, N_Y \sim \mathcal{N}$

Theorem: Identifiability of LiNGAMs

Proof:

3 We prove the reverse: If $\begin{array}{l} Y = \alpha X + N_Y, \quad N_Y \perp\!\!\!\perp X \\ X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y \end{array} \Rightarrow (X, N_Y) \sim \mathcal{N}$

Since $N_X \perp\!\!\!\perp Y$, we have: $N_X = X - \beta(\alpha X + N_Y) = (1 - \alpha\beta)X - \beta N_Y$

There are 3 cases:

(i) $(1 - \alpha\beta) \neq 0 \ \& \ \beta \neq 0$

Then, given $N_X \perp\!\!\!\perp Y$, DS theorem implies $X, N_Y \sim \mathcal{N}$

(ii) $(1 - \alpha\beta) \neq 0 \ \& \ \beta = 0$

Then, since $N_X \perp\!\!\!\perp Y$, and $N_X = X$, then $X \perp\!\!\!\perp \alpha X + N_Y$ in contradiction with Peters' lemma

Theorem: Identifiability of LiNGAMs

Proof:

3 We prove the reverse: If
$$\begin{aligned} Y &= \alpha X + N_Y, \quad N_Y \perp\!\!\!\perp X \\ X &= \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y \end{aligned} \Rightarrow$$
$(X, N_Y) \sim \mathcal{N}$

Since $N_X \perp\!\!\!\perp Y$, we have: $N_X = X - \beta(\alpha X + N_Y) = (1 - \alpha\beta)X - \beta N_Y$

There are 3 cases:

(iii) $1 - \alpha\beta = 0 \;\&\; \beta \neq 0$

Then, since $N_X \perp\!\!\!\perp \boxed{Y}$, and $N_X = -\beta N_Y$, $N_Y \perp\!\!\!\perp \boxed{\alpha X + N_Y}$ again in contradiction with Peters' lemma

Jonas Peters et al, Elements of Causal Inference (2017)

Theorem: Identifiability of LiNGAMs

Proof:

3 We prove the reverse: If $(X, N_Y) \sim \mathcal{N}$

$$\begin{aligned} Y &= \alpha X + N_Y, \quad N_Y \perp\!\!\!\perp X \\ X &= \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y \end{aligned} \quad \Rightarrow$$

Since $N_X \perp\!\!\!\perp Y$, we have: $N_X = X - \beta(\alpha X + N_Y) = (1 - \alpha\beta)X - \beta N_Y$

There are 3 cases:

(iii) $1 - \alpha\beta = 0 \;\&\; \beta \neq 0$

Then, since $N_X \perp\!\!\!\perp \boxed{Y}$, and $N_X = -\beta N_Y$, $N_Y \perp\!\!\!\perp \boxed{\alpha X + N_Y}$ again in contradiction with Peters' lemma

Therefore, as long as one of X, N_Y, Y, N_X is not gaussian, the causal direction is **identifiable** from **observational data**!

Jonas Peters et al, Elements of Causal Inference (2017)

Linear Additive Noise Models (ANMs)

ANM: The joint distribution $P_{X,Y}$ is said to admit an ANM for $X \rightarrow Y$ if there exists a measurable function f_Y and a noise variable N_Y s.t.

$$Y = f_Y(X) + N_Y, N_Y \perp\!\!\!\perp X$$

For this model, using convolution of probabilities we have:

$$p(x, y) = p_{N_Y}(y - f_Y(x))p_X(x)$$

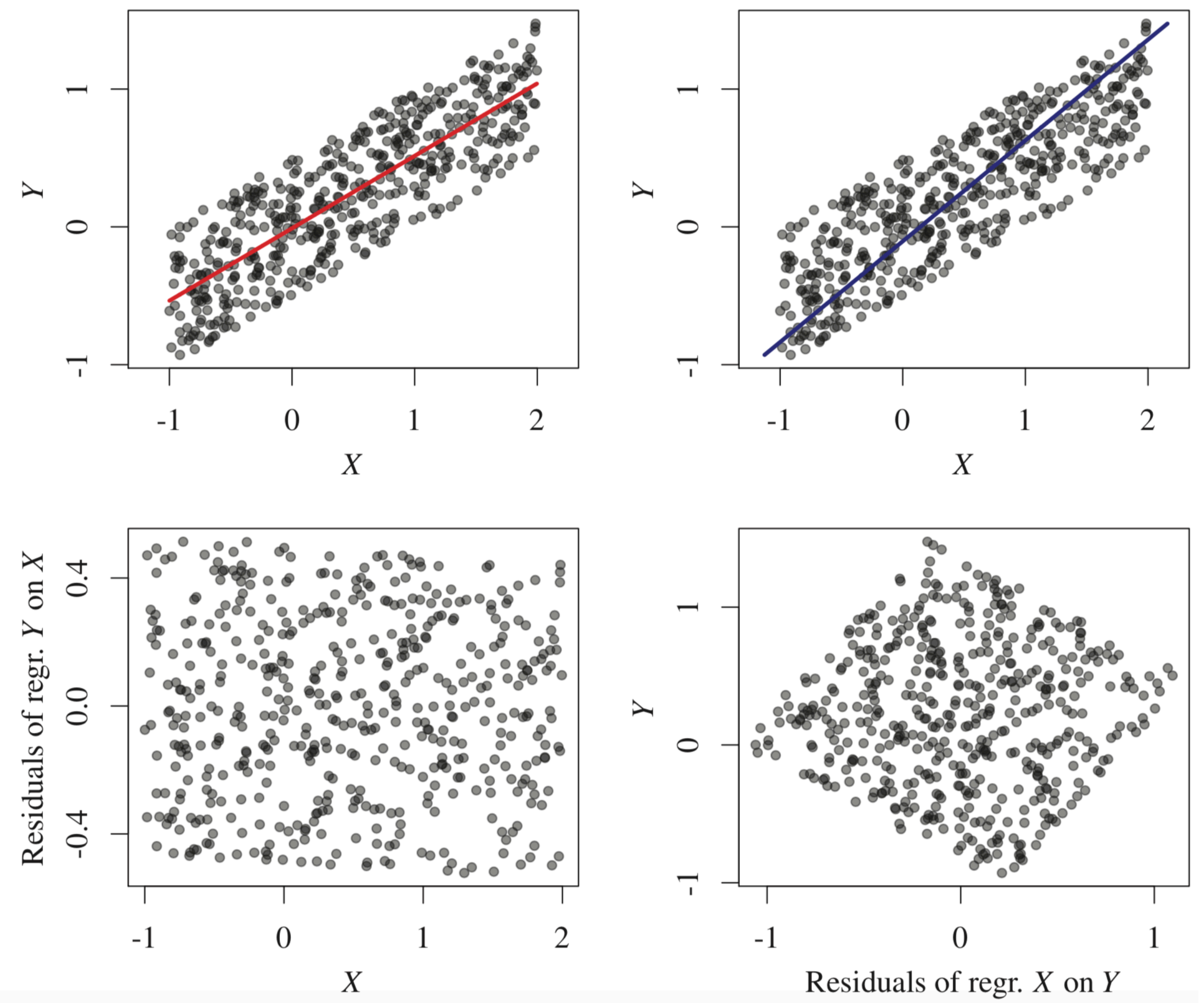
Similarly, **<u>if</u>** a backward model exists:

$$p(x, y) = p_{N_X}(x - f_X(y))p_Y(y)$$

It turns out: This imposes very strong conditions on $\log(p_X)$ for which $p_{X,Y}$ admits a smooth ANM from Y to X (backward model).

In practice

1. Regress Y on X
2. **Test** whether $Y - \hat{f}_Y$ is independent of X
3. Repeat, swapping X and Y
4. If the independence is accepted for one direction and rejected for the other, infer the former as the causal direction,



Statistical Test of Independence: Choose one that accounts for higher order statistic rather than testing correlations only, e.g. HSIC

Jonas Peters et al, Elements of Causal Inference (2017)

In practice

```
library(dHSIC)
library(mgcv)
#
# generate data set
set.seed(1)
X <- rnorm(200)
Y <- X^3 + rnorm(200)
#
# fit models
modelforw <- gam(Y ~ s(X))
modelbackw <- gam(X ~ s(Y))
#
# independence tests
dhsic.test(modelforw$residuals, X)$p.value
# [1] 0.7628932
dhsic.test(modelbackw$residuals, Y)$p.value
# [1] 0.004221031
#
# computing likelihoods
- log(var(X)) - log(var(modelforw$residuals))
# [1] 0.1420063
- log(var(modelbackw$residuals)) - log(var(Y))
# [1] -1.014013
```


Jonas Peters et al, Elements of Causal Inference (2017)

THE UNIVERSITY of EDINBURGH

Methods for Causal Inference
Lecture 18: Additive Noise Models

Ava Khamseh

School of Informatics

2024-2025