

Interventions, Mendelian randomisation and causal discovery

M2 GENIOMHE – Session 3

Christophe Ambroise

M2 GENIOMHE – Université d'Évry

Sessions 1 and 2

- d-separation; Markov equivalence $\rightarrow$ CPDAG
- Gaussian graphical model: $K_{ij} = 0 \iff X_i \perp X_j \mid X_{\text{rest}}$
- Graphical lasso: sparse precision matrix
- PC algorithm: skeleton, v-structures, Meek rules $\rightarrow$ CPDAG
- Gaussian linear SEM: direction of $X_1 - X_2$ not identifiable

Prior knowledge

Structural causal models

Structural causal model

A structural causal model (SCM) generates each endogenous variable by a deterministic mechanism of its direct causes and of an exogenous noise.

$$\mathcal{M} = (U, V, F, P_U) : \quad V_j := f_j(V_{pa(j)}, U_j), \quad j = 1, \dots, p$$

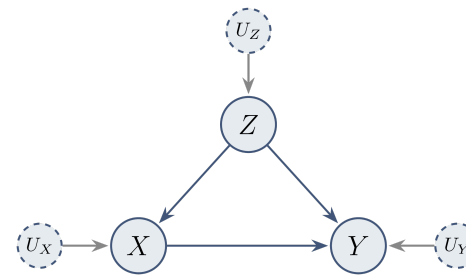
- V : endogenous variables, modelled
- U : exogenous variables, causes left outside the model
- U_j jointly independent $\rightarrow$ **Markovian** SCM

SCM example

$$Z := U_Z$$

$$X := 2Z + U_X$$

$$Y := 3X + Z + U_Y, \quad U_Z, U_X, U_Y \stackrel{iid}{\sim} \mathcal{N}(0, 1)$$



Causal DAG: from SCM to graph

- Edge $V_i \rightarrow V_j$ iff V_i argument of f_j
- Markovian SCM $\rightarrow$ observational law factorises along the DAG
- Same DAG as a Bayesian network, **stronger semantics**

Causal model vs graphical model

Graphical model: describes one distribution. Causal model: also predicts the distributions after interventions.

Interventions

Conditioning vs intervening

- **Conditioning** $P(Y \mid X = x)$: sub-population with $X = x$, graph intact
- **Intervening** $P(Y \mid do(X = x))$: X forced to x for everyone
- Confounder Z : association through $X \leftarrow Z \rightarrow Y$ only in the first case

$$P(Y \mid X = x) \neq P(Y \mid do(X = x)) \quad \text{in general}$$

Truncated factorisation

Under modularity, the interventional law is the observational factorisation with the factor of the intervened variable removed.

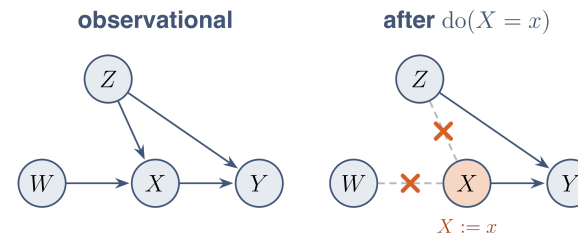
$$P(v \mid do(X = x)) = \mathbb{1}(x_v = x) \prod_{j: V_j \neq X} P(v_j \mid v_{pa(j)})$$

Causal effect on Y : marginalisation of the truncated product over $V \setminus \{X, Y\}$.

Modularity

Each mechanism f_j , or conditional $P(v_j \mid v_{pa(j)})$, is an autonomous module: an intervention on one variable leaves all other modules unchanged.

- $do(X = x)$: equation of X replaced by $X := x$
- Graph surgery: arrows **into** X deleted $\rightarrow G_{\bar{X}}$



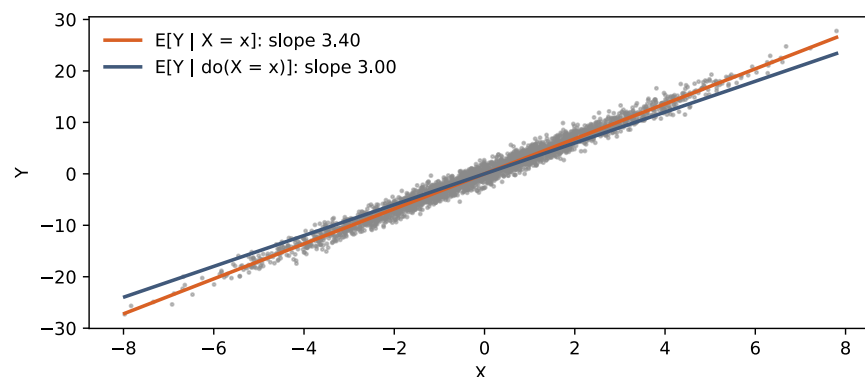
Confounder: adjustment by Z

DAG $Z \rightarrow X, Z \rightarrow Y, X \rightarrow Y$:

$$\begin{aligned} P(y \mid do(x)) &= \sum_z P(z) P(y \mid x, z) \\ P(y \mid x) &= \sum_z P(z \mid x) P(y \mid x, z) \end{aligned}$$

Difference: weights $P(z)$ vs $P(z \mid x)$, exactly the mechanism of Simpson's paradox.

Linear example: regression slope vs causal effect



Theory: $\mathbb{E}[Y | X = x] = 3x + \mathbb{E}[Z | X = x] = 3.4x$,
 $\mathbb{E}[Y | do(X = x)] = 3x$.

Backdoor path

- Path between X and Y starting with an arrow **into** X : $X \leftarrow \dots Y$
- Carrier of non-causal association
- Causal paths: directed paths $X \rightarrow \dots \rightarrow Y$

Identification by adjustment

Backdoor criterion

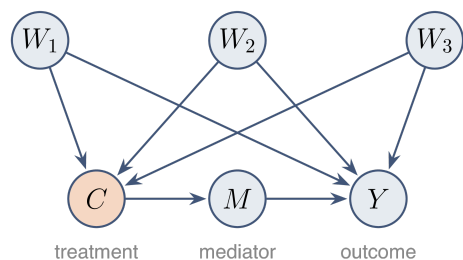
A set Z satisfies the backdoor criterion relative to (X, Y) when it contains no descendant of X and blocks every backdoor path between X and Y .

- (i) $Z \cap de(X) = \emptyset$,
- (ii) Z blocks every path $X \leftarrow \dots Y$

Backdoor adjustment formula

$$P(y | do(x)) = \sum_z P(y | x, z) P(z)$$

Several confounders



- Treatment C , outcome Y , mediator M
- Backdoor paths $C \leftarrow W_k \rightarrow Y$
- Valid set: $\{W_1, W_2, W_3\}$
- M : descendant of C , **never** adjusted

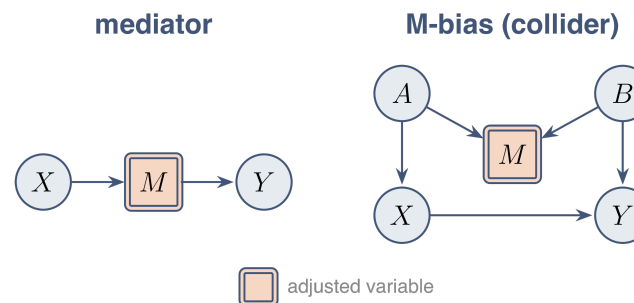
$$P(y \mid do(c)) = \sum_w P(y \mid c, w) P(w)$$

17

Instrumental variables and Mendelian randomisation

19

Bad controls



- Mediator M : part of the effect removed
- Collider M (M-bias): adjustment **opens** $X \leftarrow A \rightarrow M \leftarrow B \rightarrow Y$
- Rule: adjustment set chosen from the graph, not from correlations

18

Unobserved confounding

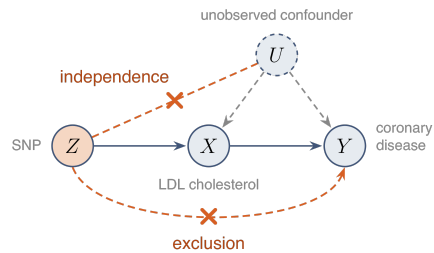
- Exposure X (LDL cholesterol), outcome Y (coronary disease)
- Confounder U (lifestyle, socio-economic status): **unobserved**
- No backdoor set available $\rightarrow$ adjustment impossible

20

Instrumental variable

A variable Z is an instrument for the effect of X on Y when it is associated with X , independent of the confounders, and acts on Y only through X .

(relevance) $Z \not\perp X$, (independence) $Z \perp U$, (exclusion) $Z \perp_{G_X} Y$



Linear model: Wald ratio

$$X = \beta_{ZX}Z + \gamma_X U + \varepsilon_X, \quad Y = \alpha X + \gamma_Y U + \varepsilon_Y$$

- $\text{Cov}(Z, X) = \beta_{ZX} \text{Var } Z$
- $\text{Cov}(Z, Y) = \alpha \beta_{ZX} \text{Var } Z$, since $Z \perp U$ and exclusion

Wald ratio

$$\alpha = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, X)} = \frac{\beta_{ZY}}{\beta_{ZX}}$$

Mendelian randomisation

- Instrument Z : genetic variant (SNP, 0/1/2 risk alleles)
- Alleles randomly allocated at meiosis → **natural randomised trial**
- Independence: genotype fixed at conception, before U
- Threats: pleiotropy (exclusion), population stratification, linkage disequilibrium

21

22

Naive regression vs instrument

```
1 import numpy as np
2 rng = np.random.default_rng(7); N = 20_000
3 Z = rng.binomial(2, 0.3, N) # genotype
4 U = rng.standard_normal(N) # unobserved confounder
5 X = 0.5 * Z + U + rng.standard_normal(N) # exposure
6 Y = 0.2 * X + 1.5 * U + rng.standard_normal(N) # outcome, true effect 0.2
7 ols = np.cov(X, Y)[0, 1] / np.var(X, ddof=1)
8 wald = np.cov(Z, Y)[0, 1] / np.cov(Z, X)[0, 1]
9 print(f"OLS: {ols:.3f} Wald ratio: {wald:.3f}")
```

OLS: 0.912 Wald ratio: 0.185

- OLS: biased by U
- Wald ratio: consistent; variance large if weak instrument

23

24

Beyond Markov equivalence: LiNGAM

Linear non-Gaussian acyclic model

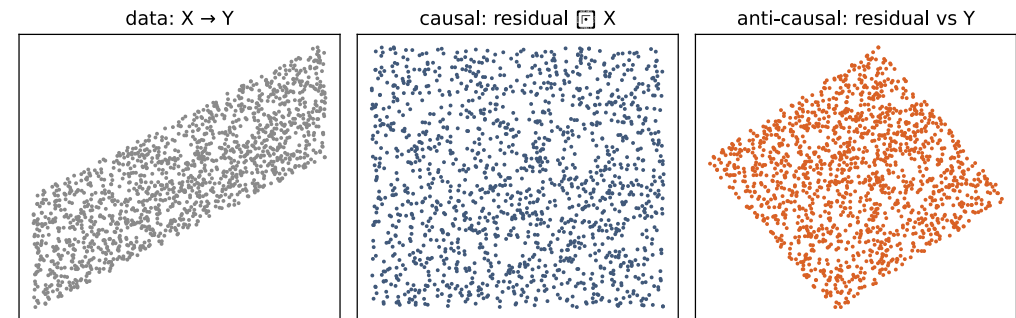
In LiNGAM, the data follow a linear SEM on a DAG, with mutually independent noises of which at most one is Gaussian.

$$X = BX + \varepsilon, \quad B \text{ acyclic (strictly lower triangular up to permutation),}$$

Identifiability (Shimizu *et al.*, 2006)

Under these assumptions, B and the causal order are identifiable from the distribution of X alone.

Two variables, one direction?



- Both residuals **uncorrelated** with the regressor
- Only the causal one **independent**: non-Gaussianity breaks the symmetry

DirectLiNGAM

X_j is exogenous (a source) if and only if every other variable, regressed on X_j , leaves a residual independent of X_j .

$$X_j \text{ source} \iff \forall i \neq j: \quad X_i - \frac{\text{Cov}(X_i, X_j)}{\text{Var } X_j} X_j \perp X_j$$

1. Most independent candidate $\rightarrow$ next variable of the order
2. Its effect regressed out of all remaining variables
3. Repeat; then B by adaptive lasso along the order

Shimizu *et al.* (2011), *JMLR*. Python package [lingam](#).

DirectLiNGAM in practice

```
1 import numpy as np, lingam
2 rng = np.random.default_rng(1); n = 2000
3 e = rng.laplace(size=(n, 4))          # non-Gaussian noises
4 x0 = e[:, 0]
5 x2 = 0.8 * x0 + e[:, 2]
6 x1 = -0.7 * x2 + e[:, 1]
7 x3 = 0.5 * x0 + 0.6 * x1 + e[:, 3]
8 X = np.c_[x0, x1, x2, x3]
9 model = lingam.DirectLiNGAM().fit(X)
10 print("order:", [int(i) for i in model.causal_order_])
11 print(np.round(model.adjacency_matrix_, 2)) # B[j, k] : effect k -> j
```

```
order: [0, 2, 1, 3]
[[ 0.   0.   0.   0. ]
 [ 0.   0.  -0.69  0. ]
 [ 0.79  0.   0.   0. ]
 [ 0.51  0.6   0.   0. ]]
```

Measuring non-Gaussianity

- Classical contrasts: kurtosis, negentropy approximations, likelihood surrogates
- Choice of a nonlinearity or moment order → method tuned to source shape
- Wanted: **one** criterion, no tuning, only finite second moments

Laplante, Ambroise & Humbert, *Contrast-free ICA and causal inference via Wasserstein distances to the Gaussian*, preprint. Package [otlingam](#).

OT-LiNGAM: non-Gaussianity by optimal transport

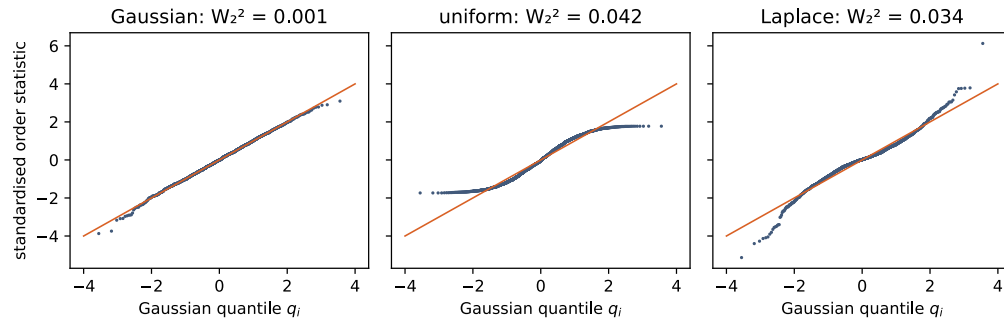
2-Wasserstein distance

The squared 2-Wasserstein distance is the smallest mean squared displacement among all couplings of the two distributions.

$$\mathcal{W}_2(\mu, \nu)^2 = \inf_{X \sim \mu, Y \sim \nu} \mathbb{E}[(X - Y)^2] \stackrel{\text{dim 1}}{=} \int_0^1 (F_\mu^{-1}(u) - F_\nu^{-1}(u))^2 du$$

- Dimension 1: optimal coupling = matching of quantiles
- Non-Gaussianity of X : $\mathcal{W}_2(\text{std}(X), \mathcal{N}(0, 1))^2$, zero iff X Gaussian

Non-Gaussianity of three distributions



Mixtures are closer to the Gaussian

Under independence and standardisation, any non-trivial normalised mixture of sources is strictly closer to the Gaussian than the corresponding weighted average of the sources' distances.

$$\|\alpha\|_2 = 1, \#\{j : \alpha_j \neq 0\} > 1 \implies \mathcal{W}_2\left(\sum_j \alpha_j \mathcal{S}_j, \mathcal{N}(0, 1)\right)^2 < \sum_j \alpha_j^2 \mathcal{W}_2(\mathcal{S}_j, \mathcal{N}(0, 1))^2$$

Assumptions: $\mathcal{S}_j$ independent, centred, unit variance, at most one Gaussian.

Wasserstein order objective

For a candidate order σ , regress each variable on its predecessors and score the non-Gaussianity of the standardised residuals.

$$G(\sigma) = \sum_{j=1}^d \mathcal{W}_2\left(\text{std}(R_j(\sigma)), \mathcal{N}(0, 1)\right)^2, \quad R_j(\sigma) = X_j - \text{proj}_{\text{span}\{X_k : k \text{ before } j \text{ in } \sigma\}}$$

- Causal order: residuals = structural noises ε_j
- Wrong order: some residual mixes several noises $\rightarrow$ closer to Gaussian

Identification of causal orders

Theorem (Laplante, Ambroise & Humbert)

Under independent, non-degenerate noises with at most one Gaussian, the maximisers of G are exactly the causal orders of the model.

$$\arg \max_{\sigma \in \mathfrak{S}_d} G(\sigma) = \mathcal{I}^* = \{\sigma : \Pi^\sigma B(\Pi^\sigma)^\top \text{ strictly lower triangular}\}$$

Key steps: $\text{std}(R(\sigma)) = Q^\sigma \text{std}(\varepsilon)$ with Q^σ orthogonal; strict subadditivity; Q^σ signed permutation $\Rightarrow \sigma$ causal.

Greedy solver and full pipeline

Greedy OT-LiNGAM: most non-Gaussian standardised residual $\rightarrow$ next source; its effect regressed out; repeat. Cost $O(d^2 n \log n)$.

Solver	Cost	Global optimum
Exhaustive OT-LiNGAM	$O(d^2 (nd^2 + n \log n))$	yes
Greedy OT-LiNGAM	$O(d^2 n \log n)$	no
OT-ICA-LiNGAM	$O(nd^2 + dn \log n)$ / iteration	no

Order known $\rightarrow B$ by adaptive lasso along the order.

OT-LiNGAM in practice

```

1 import numpy as np
2 from otlingam import ExhaustiveOTLiNGAM, GreedyOTLiNGAM
3 from otlingam.utils import disorder
4 rng = np.random.default_rng(1); n = 2000
5 e = rng.laplace(size=(n, 4))
6 x0 = e[:, 0]; x2 = 0.8 * x0 + e[:, 2]
7 x1 = -0.7 * x2 + e[:, 1]; x3 = 0.5 * x0 + 0.6 * x1 + e[:, 3]
8 X = np.c_[x0, x1, x2, x3]
9 B_true = np.zeros((4, 4)); B_true[2, 0], B_true[1, 2], B_true[3, 0], B_true[3, 1] = .8, -.7, .5, .6
10 for M in (ExhaustiveOTLiNGAM(), GreedyOTLiNGAM()):
11     M.fit(X)
12     print(type(M).__name__, [int(i) for i in M.causal_order_], "disorder:", disorder(B_true, M.causal_order_))

```

ExhaustiveOTLiNGAM [0, 2, 1, 3] disorder: 0
 GreedyOTLiNGAM [0, 2, 1, 3] disorder: 0

Evaluation: disorder of an order

Number of true edges reversed by the estimated order:

$$\text{dis}(\hat{\sigma}) = \#\{(k, j) : B_{jk} \neq 0, \hat{\sigma}^{-1}(k) > \hat{\sigma}^{-1}(j)\}$$

- $\text{dis}(\hat{\sigma}) = 0 \iff \hat{\sigma}$ causal order
- Simulations of the paper (ER and scale-free graphs, Laplace or Student noises): exhaustive OT-LiNGAM close to zero disorder, robust when some noises approach Gaussianity
- Continuous methods (DAGMA): degraded under heteroscedastic noises

37

38

Case study: chloroplast RNA maturation

39

40

Biological context

- Chloroplast transcripts of *Arabidopsis thaliana*: genes *ndhB*, *ndhD*
- Post-transcriptional maturation: C-to-U **editing** at specific sites, intron **splicing**
- Nanopore long reads: one read = one full-length RNA molecule
- Each read: snapshot of one molecule at some stage of its maturation

Data: Guilcher *et al.* (2021), *Int. J. Mol. Sci.* 22, 11297.

Data: one binary vector per read

ed1	ed2	ed3	ed4	ed5
False	False	False	True	.
True	False	.	.	.
False	False	False	False	.
False	True	False	True	.
True	False	False	True	True
False	True	True	True	.

Gene	Reads	Events	Complete reads
<i>ndhB</i>	1 899	12 editing sites + intron splicing	117
<i>ndhD</i>	7 574	5 editing sites	930

True: event observed; **False**: not observed; **Err**: sequencing error; empty: not covered.

Question: chronology of maturation events

- Order in which sites are edited and the intron spliced
- Chronology as a **causal order** of the event variables
- Snapshot data only: no time stamps

Open problem

Can a causal order learned from snapshots of molecules reveal the maturation chronology?

Methodological difficulties

- **Binary** variables: linear non-Gaussian model only an approximation
- **Missing blocks**: reads truncated, contiguous sites unobserved
- **Errors** (**Err**): measurement noise on some sites
- Rare events (sites edited in < 10 % of reads)
- Validation: no ground-truth chronology