Dynamic Cause and Bayesian Explainability

Jim Q. Smith

University of Warwick & Alan Turing Institute

August 2022 - Greek Stochastics Corfu

Motivation for exploring Causality

1. Massive causal literature: role of a **methodological & applied statistician**?
2. Language & principles of cause - give **practical, extendable explanations** of how & why we see what we see.

3. "Be**causes**" semantics vehicle for domain experts to **communicate domain knowledge, + critique & own a statistical analysis**.

Generalisable statistical models to guide future acts

Will reflect on my **experiences of causal reasoning** in person centred AI/Stats over last 20yrs.

Format of these talks

1. **Review & reflection of recent historical development** of certain themes in causal reasoning - beginning with Bradford Hill, & Pearl.
2. Argue how **causal ideas & principles extend to a vast variety of domains** not just Bayesian Networks (BNs)/Structural Equation Models (SEMs)/standard experimental design.
3. **Illustrate** how causal discovery & causal algebras usefully **customised** to variety of other models.
4. Lens of **Subjective Bayesian** analysis of massive structured time series

Dynamic explainable causation to predict impact of future acts

5. Other central causal themes of *Design* of experiments,*counterfactuals* & *SEMs* left to others!

Aim: inspire extensions of causal models to new domains

Massive potential for major methodological & applied research.

Structure of talks

Ist. Session: **Basic Causal Discovery $\rightarrow$ dynamic framework**

1. Review some necessary conditions proposed for causal relationships.
2. Review BN: most developed graphical methodology supporting causal discovery.
3. Introduce Multiregresssion Dynamic Models (MDMs) & their illustrate application in causal discovery.
4. Where causal discovery needs to develop: putative/genuine causes.

Structure of talks

2nd session: Generating causal algebras to context

1. Review Bayesian Network (BN) Causal algebras
2. Emulation under control & relationship between causality & Integrating Decision Support System (IDSS)
3. Tree based formulations of causation & the chain event graph.
4. New causal algebras & dynamic models - regulatory graphs.
5. Current ongoing projects in causation.

How causal modelling adds to standard Bayesian inference

Causal methodology - stated generically from decision perspective (see e.g. Dawid (15) Didelez, Gineletti):

1. Build class of models $\mathcal{M}$ - **idle models** - & probability models $\mathcal{P}$ over them to capture **meaningfully causally related dependence relationships** over population Ω.
2. Use available data to **select** model $M^* \in \mathcal{M}$ describing idle domain well (**causal discovery**).
3. Construct decision space $d \in \mathcal{D}$ map $\mathcal{M} \rightarrow \mathcal{M}_d$: $M^* \mapsto M_d, \mathcal{P} \rightarrow \mathcal{P}_d : P \mapsto P_d$ - called a **causal algebra**. Let $\mathcal{M}_{\mathcal{D}} \triangleq \{\mathcal{M}_d : d \in \mathcal{D}\}$
4. Use inputs & causal reasoning to **explain & justify efficacy** of different potential controls $d \in \mathcal{D}$.

My focus: Causal discovery when can only **observe idle** $\mathcal{M}$: (for when also observe $\mathcal{M}_d$ for $d \in \mathcal{D}_0 \subset \mathcal{D}$ see e.g. Cooper & Yoo (99), Peters et al (16)).

Bayesian statistician view on current causal analysis.

1. Choose **families of sample distribution** & priors $\longrightarrow$ for each $M \in \mathcal{M}$ find posteriors $P \in \mathcal{P}$.
2. Use Bayesian model selection to choose **explainable** $M^* \in \mathcal{M}$ with posteriors $P^* \in \mathcal{P}$. Extract causal conjectures using **causal discovery algorithms**.
3. **Causal algebras** by link $d \in \mathcal{D}$ to "causes" to construct $(M_{\mathcal{D}}, P_{\mathcal{D}}) \triangleq \{M_d^*, P_d^* : d \in \mathcal{D}\}$.
4. **Assess** $d \in \mathcal{D}$ in terms of

$$\overline{U}(d) \triangleq \int U(d, \mathbf{e}) dP_d^* = \int U(d, \mathbf{e}) dP_{d \to \mathbf{e}}^*$$

where $\mathbf{e}$ vector of effects (responses) of interest $P_{d \to \mathbf{e}}^* \Rightarrow$ margin of $\mathbf{e}$ over P_d^* given causally related intervention $d \in \mathcal{D}$. Use **causal graph to justify best** d^* through comparing $\overline{U}(d^*)$ v $\overline{U}(d)$ (explained using $P_{d^* \to \mathbf{e}}^*$ v $P_{d \to \mathbf{e}}^*$).

Note Extension 3) = 1940's control theory - e.g. Kalman Filter - controls d replace system error in idle stochastic model: see below.

Genesis of Causal Inference and Statistics

In 1965 **Bradford Hill** - pioneer of randomised clinical trials & establishing causal link between smoking & cancer - **9 necessary properties** for cause $\longrightarrow$ effect. (List not complete - good starting point).
Four properties $\sim$ **statistical model of idle system**.

1. **Association** - an apparent **strong association** between putative cause & its effect ("no causation without association").
2. **Consistency** - association **holds for different subpopulations** within Ω & across different "parallel" Ω'. $\Longrightarrow$ form of **invariance** across effect of cause (see e.g. Peters et al(16)).
3. **Dose Response** - there is an **increasing** relationship between intensity of **cause** & intensity of **effect**, (e.g. more smoking $\Longrightarrow$ higher risk of cancer)
4. **Temporal**: Causes must happen before an effect.

Notes: Bullet 2 $\Rightarrow$ out of sample verification not sufficient! 3&4 not entailed by vanilla version of Pearl(00): see below.

Notion of Interventions Plausible: Causal Algebra

Causal effects require comparison of coherent worlds - cause present v not:

1. **Plausibility**: $\exists$ **coherent** event space + prob. **extension embedding causal map** - consistent with known evidence.
2. **Experiment** - *in principle* can **perform experiment** on $\omega \in \Omega$ where each ω could be exposed to a cause or not.
3. **Specificity** - *in principle* can disentangle effect of putative cause holding other " background" features fixed.

Note Underlying **plausibility**, **experiment** & **specificity** idea probability models $(\mathcal{M}, \mathcal{P})$ for Ω can in principle be **embeddable into a control space** $\mathcal{M}_{\mathcal{D}}$(or a counterfactual universe) where within $\mathcal{M}_{\mathcal{D}}$ extracted associations for $\mathcal{P}_{\mathcal{D}}$ still makes sense $\Rightarrow$ link causation to future $d \in \mathcal{D}$ (or such decisions you might have made) & new settings: see **Consistency**.

Two properties link causation to expert judgement not data science

1. **Coherence**: Experts in field acknowledge causal model $\{(M_d^*, P_d^*) : d \in \mathcal{D}\}$ plausible/ not absurd/ not totally countercultural.
2. **Analogy** - hypotheses compared with competing associations: chosen causal explanation proves "best" $\sim$ **model selection, testing**, **statistical diagnostics** but also its **compellingness**!

Note: 1), 2b) $\Longrightarrow$ expert judgements **must** enter at some stage of process. Subjective Bayesian $\Longrightarrow$ input prior structural information into model structure $\mathcal{M}$ & quantitative prior information into $\mathcal{P}$.

Some other general properties of Causation to demand

Data scientists supplement this list with other desirable properties. First two given here are associated with a **binary** relationship between components of measurement vectors, the last a **negative** property:

1. **Parsimony** - when c a cause of e then **the prediction of** e **from** c **can be specified more simply/consistently** (for example through no. bits needed to code it) than prediction of c from e (see e.g. Mitrovic et al,18)
2. **Explicability** - when c is a cause of e then **uncertainty of** e **given** c (as measured, for example, by some entropic or utility measure) is **strictly less than uncertainty of** c **given** e. e.g. if $e = f(c)$ where f a non invertible function satisfies this property.
3. **Invariance** Existence & strength of a causal relationship should be **invariant to monotone increasing transformations** of observed measurement variables.

There have been many other suggestions!

Some initial comments on these properties

1. BH properties as stated **can be ambiguous** especially in dynamic/multivariate settings!
2. Properties **all defined in natural language**. Need vehicle to translate properties & informed expert judgements into classes of probability models consistent with ideas and knowledge. Only then will classes be **explainable & interpretable**.
3. Several BH properties **not satisfied** in any sense by **current classes of models described as "causal"**!
4. Pearl, Spirtes et al,..**Cause** = **random vector**, Dawid, Didelez,...= **decision/plan** & S = **event sequence/ realised dynamic process**.
5. **Dynamics**: both a cause as an exposure & its effect often **felt over a period of time** (see smoking):⇒ **models of multivariate random processes.**

My focus: **embedding dynamics** explicitly in **causal** descriptions.

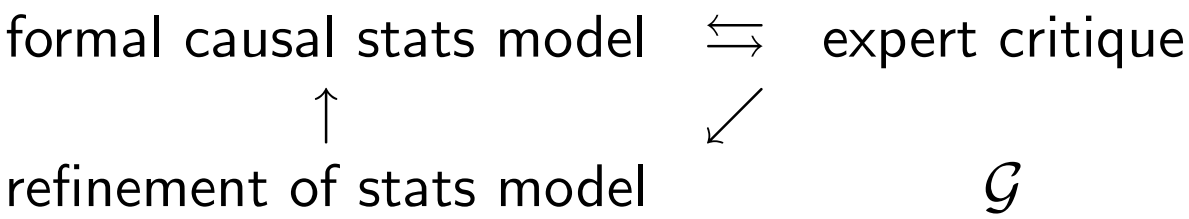
Viewing dynamic causal reasoning through a Bayesian Lens

Types of settings I analyse here using causal reasoning:

1. Plan to **act** on basis of **causal reasoning on effects that have a consequence**.
2. Purpose of **causal analysis** $\longrightarrow$ **improve predictions of likely consequences of each** $d \in D$ - linking directly or indirectly to causes c. Consequences measured through subjective expected utility (SEU). **Not scientific discovery but decision support**.
3. Causal modelling process either an Exploratory Data Analysis (**EDA**) for structuring $\mathcal{M}$ **or** part of process of **evaluating consequences** $\overline{U}(d)$ **explicitly**.
4. Focus on (bespoke) **graph based** methods for exploring, evaluating & communicating existence & strength of any causal arguments.

Graphs for Bayesian Causal Models

- **Elicit expert judgements** about multivariate descriptors for each $\omega \in \Omega$ using **natural language explanations**.
- Use graph + formal semantic to **translate** explanation into **hypotheses about idle statistical models** & **causal extensions.**
- Perform **model selection, estimation & prediction** using relevant data on $\omega \in \Omega$.
- **Feedback results to domain experts** using graph $\mathcal{G}$



Graphs $\longrightarrow$ domain experts **own** & **interact** with causal analysis!!!!

BNs to represent irrelevance statements

Enter Machine Learners/Data Scientists: Array of data $\{Y_1, Y_2, \ldots Y_n\}$ collected on $\omega \in \Omega$. Let $\mathbf{Y}_I \triangleq \{Y_i : i \in I\}$

- **Bayesian Network (BN):** (Pearl(86) Spirtes et al (90)) depicts c.i. relationships between $\{Y_1, Y_2, \ldots Y_n\}$

$$\mathbf{Y}_A \amalg \mathbf{Y}_B | \mathbf{Y}_C$$

- **Translates natural language statement** "If I had available $\mathbf{Y}_B$ & $\mathbf{Y}_C$ to help me forecast the value of measurement $\mathbf{Y}_A$ then I would need to use only $\mathbf{Y}_C$. Once I know the value $\mathbf{y}_C$ of $\mathbf{Y}_C$, the value $\mathbf{y}_B$ of $\mathbf{Y}_B$ would be irrelevant to the forecasts I would make concerning anything about the (as yet unknown) value of the measurement $\mathbf{Y}_A$".
- **BN** $\mathcal{G}$ simply **encodes** $n-1$ such **irrelevance statements**:

$$Y_k \amalg \mathbf{Y}_{R(k)} | \mathbf{Y}_{A(k)}$$

$k = 2, 3, \ldots, n$ where *parent* set
$A(k) \subseteq \{1, 2, \ldots, k-1\}$, $R(k) \triangleq \{1, 2, \ldots, k-1\} \setminus A(k) \Rightarrow \mathcal{G}$ has vertices $\{Y_1, Y_2, \ldots Y_n\}$ & directed edge from $Y_j \longrightarrow Y_k$ iff $j \in A(k)$.

Graphs, factorisations and modularity

BN maps NL explanations over $\omega \in \Omega$ about **connectedness/relevance** $\rightarrow (\mathcal{M}, \mathcal{P})$

$$\begin{array}{cccccc} \mathcal{G} & \mathbf{Y}_1 & \rightarrow & \mathbf{Y}_3 & \rightarrow & \mathbf{Y}_5 \\ & & & \downarrow & \searrow & \\ & \mathbf{Y}_2 & \rightarrow & \mathbf{Y}_4 & \rightarrow & \mathbf{Y}_6 \end{array}$$

$$\mathcal{G} \Rightarrow p(\mathbf{y}|\mathcal{G}) = \prod_{i-1}^{m} p(\mathbf{y}_i|\mathbf{y}_{A_i}, \mathcal{G})$$

So for $\mathcal{G}$ above $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_6|\mathcal{G})$

$$= p(\mathbf{y}_1|\mathcal{G})p(\mathbf{y}_2|\mathcal{G})p(\mathbf{y}_3|\mathbf{y}_1, \mathcal{G})p(\mathbf{y}_4|\mathbf{y}_2, \mathbf{y}_3, \mathcal{G})p(\mathbf{y}_5|\mathbf{y}_3, \mathcal{G})p(\mathbf{y}_6|\mathbf{y}_3, \mathbf{y}_4, \mathcal{G})$$

Key point: Parametrise $p(\mathbf{y}_i|\mathbf{y}_{A_i}, \mathcal{G})$ by $\boldsymbol{\theta}_i$, so e.g. $p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_6|\boldsymbol{\theta}, \mathcal{G})$

$$= p(\mathbf{y}_1|\boldsymbol{\theta}_1)p(\mathbf{y}_2|\boldsymbol{\theta}_2)p(\mathbf{y}_3|\mathbf{y}_1, \boldsymbol{\theta}_3)p(\mathbf{y}_4|\mathbf{y}_2, \mathbf{y}_3, \boldsymbol{\theta}_4)p(\mathbf{y}_5|\mathbf{y}_3, \boldsymbol{\theta}_5)p(\mathbf{y}_6|\mathbf{y}_3, \mathbf{y}_4, \boldsymbol{\theta}_6)$$

Bayesian consequences: searching over many models

Assume *global independence* $\amalg_{i=1}^{m} \boldsymbol{\theta}_i \Rightarrow$

$$p(\mathbf{y}, \boldsymbol{\theta}, \mathcal{G}) = \prod_{i-1}^{m} p(\mathbf{y}_i | \mathbf{y}_{A_i}, \boldsymbol{\theta}_i(\mathcal{G})) \pi(\boldsymbol{\theta}_i(\mathcal{G})) = \prod_{i-1}^{m} p(\mathbf{y}_i, \boldsymbol{\theta}_i(\mathcal{G}) | \mathbf{y}_{A_i})$$

$\Rightarrow$ posterior density separates $\Longrightarrow$ scores like MAP do too $\Longrightarrow$ logscores add by component. Assume *ancestral random samples* of $\mathbf{y}^+$ - i.e. observe unit's $\mathbf{y}_i \Rightarrow$ also observe $(\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_{i-1}) \Rightarrow \mathbf{y}_{A_i}$. Then likelihood $l(\boldsymbol{\theta} | \mathbf{y}^+)$ separates

$$l(\boldsymbol{\theta}(\mathcal{G}) | \mathbf{y}^+) = \prod_{i-1}^{m} l_i(\boldsymbol{\theta}_i(\mathcal{G}) | \mathbf{y}^+_{A_i}, \mathbf{y}^+_i)$$

$\Rightarrow$ posterior density given by

$$\pi(\boldsymbol{\theta}(G) | \mathbf{y}^+) = \prod_{i-1}^{m} \pi_i(\boldsymbol{\theta}_i(\mathcal{G}) | \mathbf{y}^+_{A_i}, \mathbf{y}^+_i)$$

Log predictive density $\log q(\mathbf{z} | \mathbf{y}^+, \mathcal{G})$ of unobserved $\omega \in \Omega$ **additive** (on local relationships over competing $\mathcal{G}$)

Bayesian consequences: searching over many models

$$
\begin{aligned}
\log q(\mathbf{z}|\mathbf{y}^+, \mathcal{G}) &= \log \left\{ \int_{\boldsymbol{\theta}} \prod_{i-1}^{m} q(\mathbf{z}_i|\mathbf{z}_{A_i}, \boldsymbol{\theta}_i, \mathcal{G}) \pi(\boldsymbol{\theta}_i|\mathbf{y}^+, \mathcal{G}) d\boldsymbol{\theta} \right\} \\
&= \log \left(\prod_{i-1}^{m} \left\{ \int_{\boldsymbol{\theta}_i} q(\mathbf{z}_i|\mathbf{z}_{A_i}, \boldsymbol{\theta}_i, \mathcal{G}) \pi(\boldsymbol{\theta}_i|\mathbf{y}^+, \mathcal{G}) d\boldsymbol{\theta}_i \right\} \right) \\
&\triangleq \sum_{i=1}^{m} \log q_i(\mathbf{z}_i|\mathbf{z}_{A_i}, \mathbf{y}^+, \mathcal{G})
\end{aligned}
$$

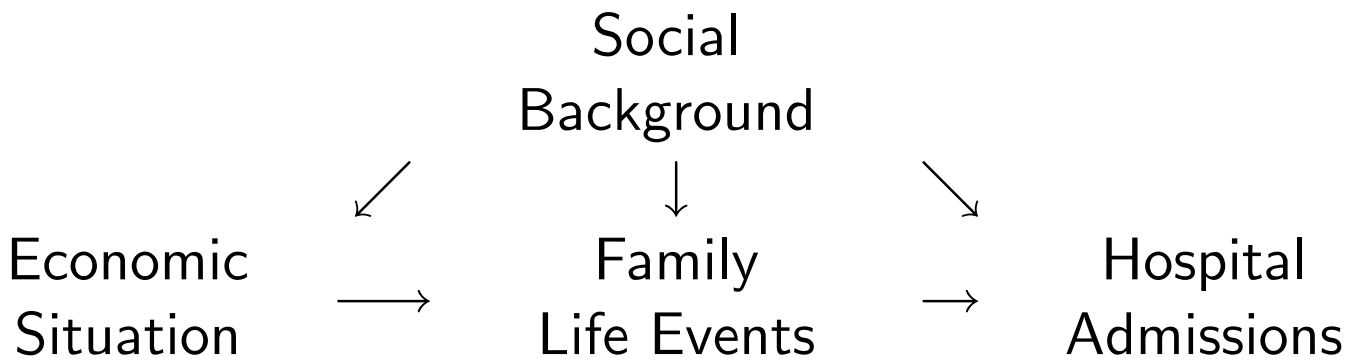
Additivity $\Rightarrow$ each term $\log q_i(\mathbf{z}_i|\mathbf{z}_{A_i}, \mathbf{y}^+, \mathcal{G})$ in **score shared** across $\mathcal{G}$ with same parent configurations on Z_i $\Rightarrow$ fast search when score separates on terms (e.g. MAP)

Easy to find best scoring BN even when 100's of variables.

Question Can causal conjectures be informed by an EDA on BNs?

A BN of child health (CHIDS) (Barclay et al, 2012)

Ferguson: longitudinal study in New Zealand: Survey categorised over 400 children + households using discrete classifiers. Here 4 (social background of household, stability of income, had certain stress inducing events. (e.g. divorce?, moving house?..) occurred & had child visited hospital?).
Best fitting BN $\mathcal{G}$ after exhaustive search of data set:



Irrelevance: "To predict whether or not child taken to hospital, sufficient to know whether or not life events had occurred & household social background. Given these the economic situation of family irrelevant" (missing edge $ES \longrightarrow HA$ in $\mathcal{G}$).

Comments in the interpretation of the graph

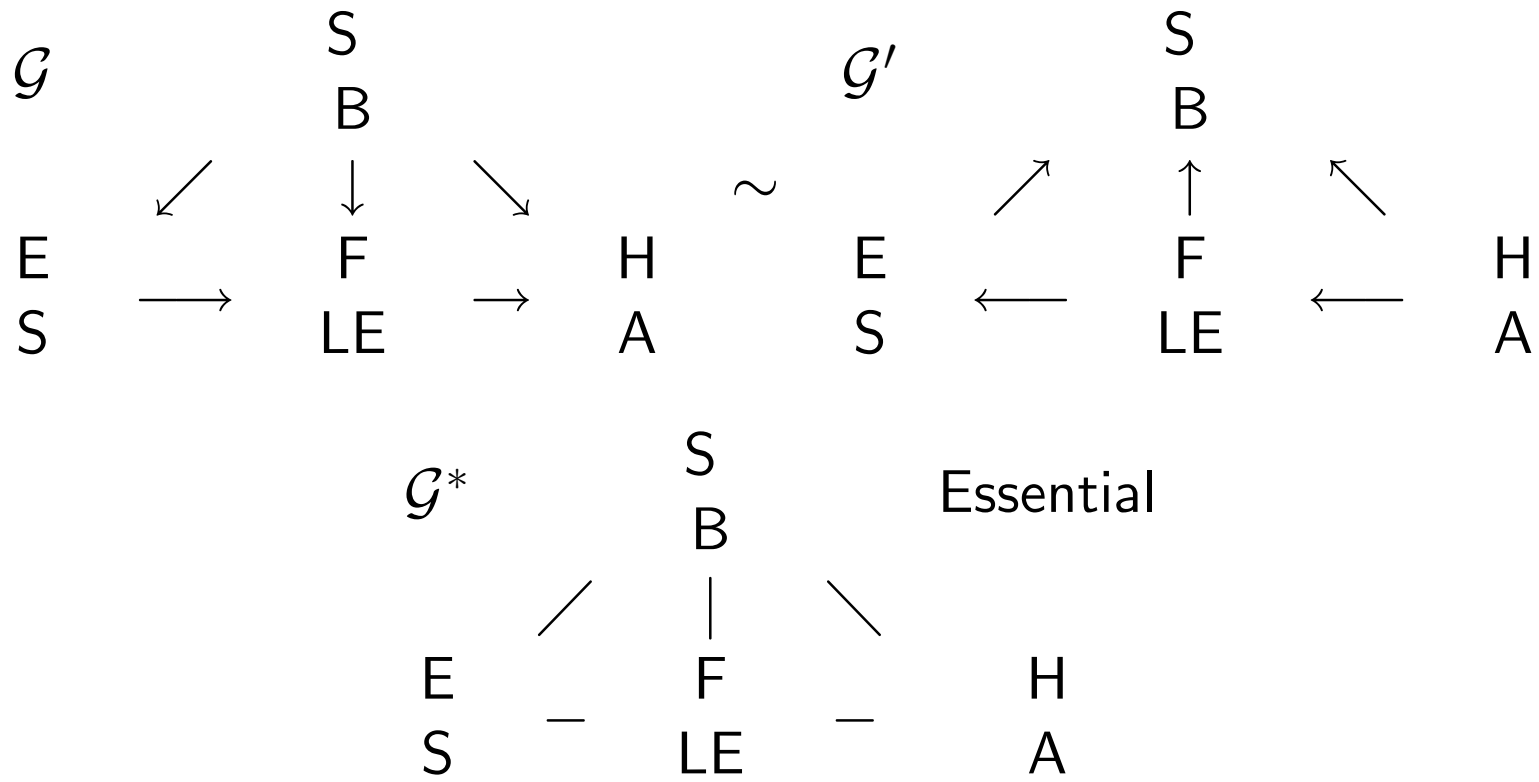
Recap: Performed favourite model selection algorithm over set of all BNs $\Rightarrow$ best fitting model M^* has graph $\mathcal{G}$.
Question Do arrows track causal relationships? *Parent* of Y_j, $j \in Q_i$ "direct cause" of Y_i.
Answer No!!!!!

Example

If $(Y_1, Y_2, \ , Y_n)$ were multivariate Gaussian then we can represent its joint distribution as a sequence of $n-1$ linear regressions. In this regression representation the parents $\{Y_j : j \in Q_i\}$ of Y_i are those independent variables that "explain" the variation in the dependent variable Y_i - but not causes, regression coefficients not the strengths of these causal relationships.

Because - even if BN perfectly estimated - **may not be unique** $\Longrightarrow$ identifying a direct cause with a parent only justified if possible cause was a parent in *all* graphs $\mathcal{G}$ representing associated density.

CHIDs example



- So no pairwise causal relationships hold!

Arrows in BN do not necessarily represent causation

Markov Equivalence Classes

See Smith(10) - construct *essential graph* $\mathcal{G}^*$ of $\mathcal{G}$: a mixed graph also on vertices $\{Y_1, Y_2, \ldots, Y_n\}$. Next deduce *putative causes:*

1. Model selection algorithm or expert judgement finds best $\mathcal{G}$ explaining relationships between $\{Y_1, Y_2, \ldots, Y_n\}$.
2. Essential graph $\mathcal{G}^*$ of $\mathcal{G}$ constructed.
3. Call Y_j a *putative cause* of Y_i when $\exists$ path of directed edges in $\mathcal{G}^*$ from Y_j to Y_i.

Exploratory data analysis (EDA) $\longrightarrow$ unambiguous causal direction when $\mathcal{G}$ "faithful" to density $p(\mathbf{y})$.

Why enduring arrows not enough - marginalisation.

Pearl argues many rvs expected missing $\Rightarrow$ we observe a margin. Suppose only observe (Y_1, Y_2, Y_3, Y_4) whose selected graph $\mathcal{G}_1 = \mathcal{G}_1^*$ with unknown additional variables.

$$\begin{array}{ccccccc} Y_1 & & & & & & \mathcal{G}_1 \\ & \searrow & & & & & \\ Y_2 & \rightarrow & Y_3 & \rightarrow & Y_4 & & \end{array}$$

Can check $\mathcal{G}_1$ same ci on (Y_1, Y_2, Y_3, YX_4) as $\mathcal{G}_+$ with hidden pair (H_1, H_2)

$$\begin{array}{ccccccc} Y_1 & \longleftarrow & H_1 & & & & \mathcal{G}_+ \\ & & & \searrow & & & \\ Y_2 & \leftarrow & H_2 & \rightarrow & Y_3 & \rightarrow & Y_4 \end{array}$$

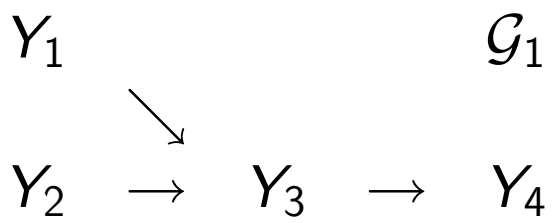
$\Rightarrow$ putative causal relationships $Y_1 \longrightarrow Y_3$ & $Y_2 \longrightarrow Y_3$ can't be read from $\mathcal{G}_+$.

But Pearl proves all possible $\mathcal{G}_+$ we must have $Y_3 \longrightarrow Y_4$ Pearl calls Y_3 *genuine direct cause* of Y_4.

Note Relies on data generating process being a BN!

Instruments - another alibi for deduction of genuine cause

Invariance of arrow in BN $\not\Rightarrow$ a *causal* relationship between tail & head!



Another argument: In $\mathcal{G}_1$ "cause" $Y_3 \longrightarrow Y_4$ "genuine" because Y_1 acts as an *instrument* - a randomising devise on Y_3 not affecting Y_4.
$\Longrightarrow$ observed $p(y_4|y_3)$ what would see if Y_1 used to allocate random values $\widehat{y}_3$ to Y_3 in designed experiment $\Rightarrow$ **implicit BH Experiment**

Terrorist Example

Example

Suppose Y_1 indicates people seen fleeing building, H_1 bomb placed in basement, Y_2 someone seen sitting in car parked outside building, H_2 man with detonator nearby, Y_3 bomb explodes in building & Y_4 people on roof die.

- Could not reasonably assert that people running away from building "caused" bomb to explode or someone parked outside building "caused" it to explode.
- Y_1 people running = **effect** of hidden cause - H_1 because they have seen a bomb, Y_2 someone sitting in car nearby **effect** of H_2 (hidden intent) to detonate bomb.
- OTOH Y_3 bomb explosion "causes" deaths Y_4.

H_i a *hidden common cause* of Y_i & $Y_3, i = 1, 2.$

Comments in the interpretation of BN discover algorithms

1. Example $\Rightarrow$ as part of EDA, **Causal discovery algorithms with BNs helpful**.
2. Best fitting model from model search often most **parsimonious** description of data - see supplementary properties $\Rightarrow$ simplest explanation "= because". Modelling relevance statements: e.g. $Y \leftarrow X \rightarrow Z$ "Y & Z **associated** to each other only be**cause** both associated to X" $\Rightarrow$ non association property applies across subpopulations of Ω whenever $X = x$ **consistency.**
3. However from BH **association, consistency is only 2 of 9 BH causal properties**.

BN Causal Discovery only extracts idle **association/consistency**

Comments in the interpretation of BN discover algorithms

- **Inferential Issues**: Causal hypotheses more compelling when **variables after conditioning are more uncertain than variables before**: see supplementary property **explication** "A cause reliably explains an effect". Currently not encoded into vanilla search of BNs. e.g. for discrete BNs usual MAP setting of Dirichlet priors over $\mathcal{M}$ imply reverse! New search algorithms of causal search needed!
- **Practical issues**: different model selection $\Longrightarrow$ very different graphs $\mathcal{G}$ (so $\mathcal{G}^*$)/even in known truth **need massive data sets** to identify generating process/ vanilla **code only for discrete or Gaussian** (or mixed).

Causal BN Discovery $\longrightarrow$ candidate causal hypotheses to consider (EDA)

Any Causal Discovery $\longrightarrow$ candidate causal hypotheses to consider (EDA)

What BN discovery misses

.

1. Temporal condition not embedded (BN only gives p.o. & formally does not totally ordered). **Causes should always happen before an effect** Sadly when data collected temporal information lost $\Rightarrow$ cross sectional study. But **temporal information** often **critical** to good causal inferences! See next example.
2. Like in example above - **causal hypotheses are about processes unfolding in time**.

One class of dynamic causal model discussed below

What BN discovery misses

.

1. **Cause** an **event that happened** (for example a flood) or an **action that was taken** (a bomb placed) $\Rightarrow$ cause not rv in non-experimental settings (e.g. drug trial where can set different concentrations of a drug) $\Longrightarrow$ e.g. tree based formulation: see next talk.
2. **Recorded random variables often not natural candidates** for putative causes.
3. **Causes often** not directly observed but **latent.** (scales example)

See next session for causal models addressing above

Discovery in Dynamic Model: Costa et al (15)

- **The Relational Model** MDM (Queen & Smith,1993) graphical model state space model

$$\begin{aligned} Y_t(i)|PaY_t(i) &= \mathbf{F}_t(PaY_t(i))\boldsymbol{\theta}_t(i) + v_t(i) \\ \boldsymbol{\theta}_t(i) &= G\boldsymbol{\theta}_t(i) + \mathbf{w}_t(i) \end{aligned}$$

- Regression but where **regression coefficients define a stochastic process** for each child on parents. $\amalg_{i=1}^{n}\mathbf{w}_t(i)$
- As for BNs, parents candidate direct causes - but now domain demands **strengths of relationships** stochastic.

Note: Embeds Gaussian BNs degenerate MDM $\Rightarrow$ causal analogues hold.

An Example

- Draw time slice of MDM, true conditionally on all parameters

$$\begin{array}{ccccc} Y(1) & \rightarrow & Y(3) & \rightarrow & Y(4) \\ & \nearrow & & & \\ Y(2) & & & & \end{array}$$

- Time series defined in terms of the stochastic change in regression coefficients. e.g. if

$$\boldsymbol{\theta}_t(3) = (\theta_t(3,0), \theta_t(3,1), \theta_t(3,2))$$

simple LMDM

$$\begin{aligned} Y_t(3)|Y_1(1), Y_1(2) &= \theta_t(0) + \theta_t(1)Y_1(1) + \theta_t(2)Y_t(2) + v_t(3) \\ \boldsymbol{\theta}_t(3) &= \boldsymbol{\theta}_t(3) + \mathbf{w}_t(3) \text{ where } \mathbf{w}_t(3) \text{ ind. error} \end{aligned}$$

Interpretation of an MDM.

- **Fast** (e.g. daily) **dynamic on time slice** - between components.
- **Slow** (yearly) **dynamic between time frames** - captured by systematic drift in regression parameters/ relationships.
- One step ahead distributions ($\Rightarrow$ so k steps ahead) remain **modular**.
- Conjugacy assumptions $\Rightarrow$ conditional on hyperparameters statistical model is **closed form** $\Rightarrow$ BF scores products of multivariate student t's $\Rightarrow$ greedy search over MDMs fast!

$$\begin{array}{ccccc} \mathbf{Y}_{1t}|\mathbf{y}^{t-1} & \rightarrow & \mathbf{Y}_{3t}|\mathbf{y}^{t-1} & \rightarrow & \mathbf{Y}_{5t}|\mathbf{y}^{t-1} \\ & & \downarrow & \searrow & \\ \mathbf{Y}_{2}|\mathbf{y}^{t-1} & \rightarrow & \mathbf{Y}_{4t}|\mathbf{y}^{t-1} & \rightarrow & \mathbf{Y}_{6t}|\mathbf{y}^{t-1} \end{array}$$

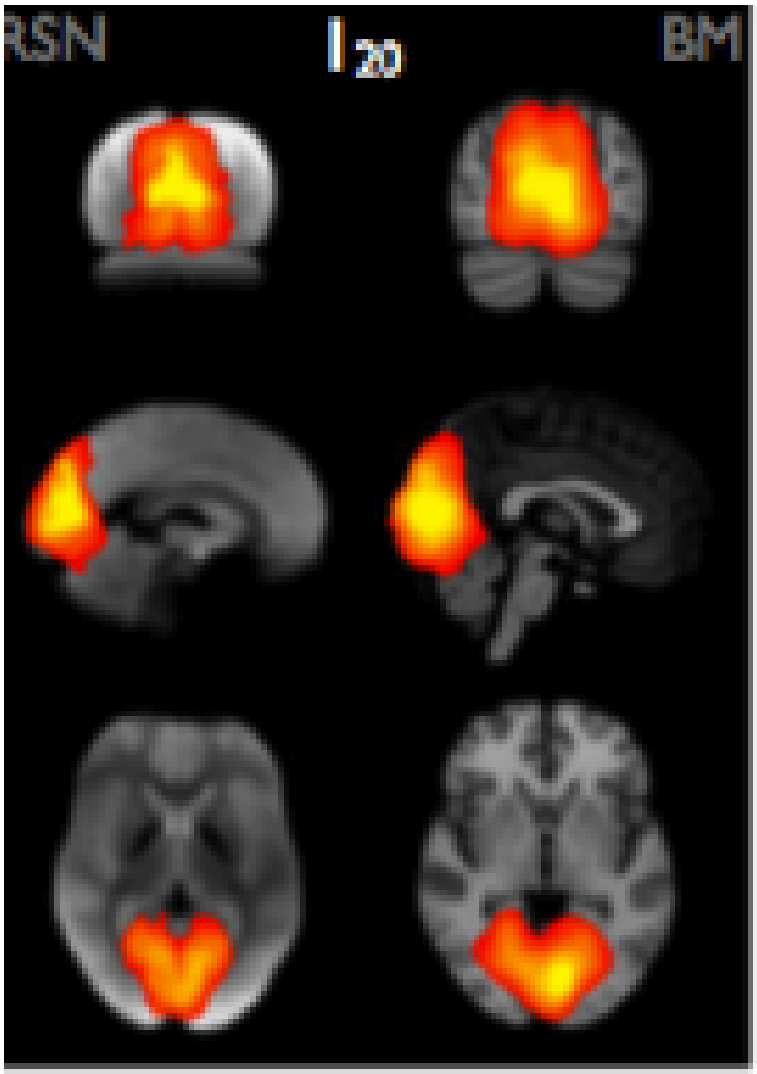
Causality and the MDM

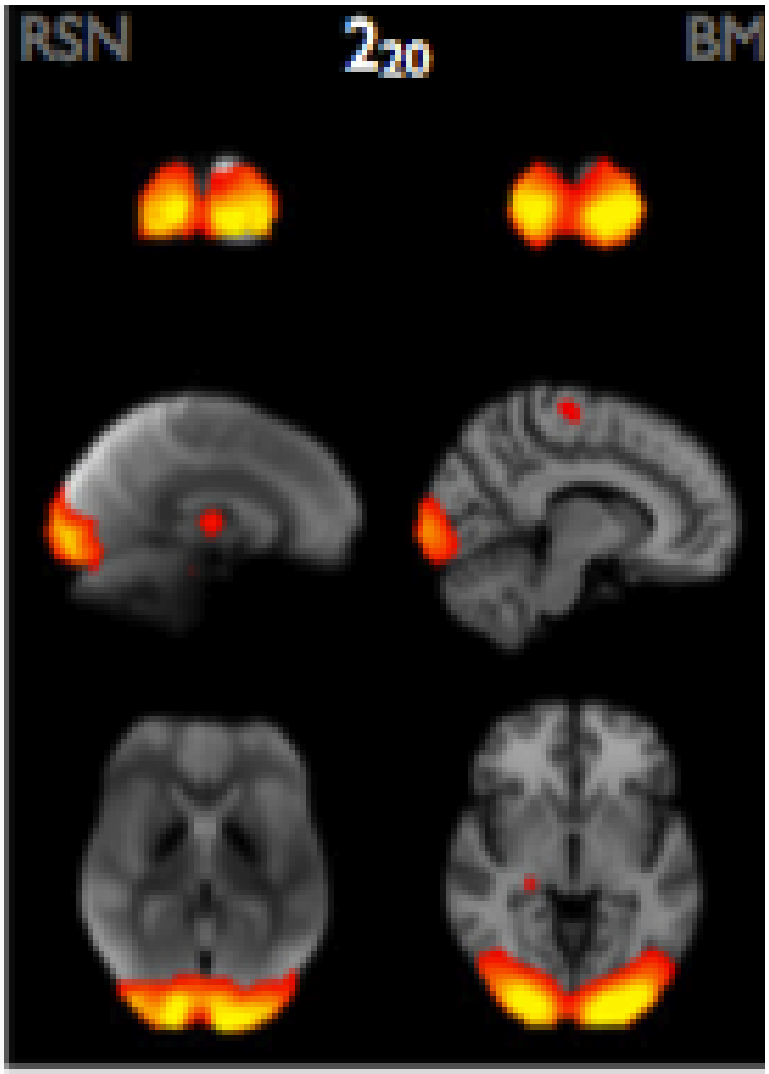
- Predictive distributions non-Gaussian if state drifts $\mathbf{w}_t(i)$ non zero $\Rightarrow$ each graph has different likelihood $\Rightarrow$ **always** $G^* = G$: see Wilkerson(20) $\Rightarrow$ all arrows give putative genuine causes.
- Directionality "causal" if Markov drifts $\{\mathbf{w}_t(i)\}$ in regression parameters independent of each other.

$$\begin{aligned} Y_t(i)|PaY_t(i) &= \mathbf{F}_t(PaY_t(i))\boldsymbol{\theta}_t(i) + v_t(i) \\ \boldsymbol{\theta}_t(i) &= G\boldsymbol{\theta}_t(i) + \mathbf{w}_t(i) \end{aligned}$$

- Causal manipulation of parents of effect enduring relationship despite changing regressions - **BH consistency**

Two regions of the Brain





fMRI data analysis the context

Science

- 10 -20 regions of brain identified to give **dynamic time series** associated with each.
- Activation in **one part of the** almost instantaneously **excites** (causes?) **activity in another** (like a blood flow).
- Connections enduring but **strength of excitation stochastic**.
- Interested in **directional network graph** e.g. to check differences between healthy & diseased individuals.

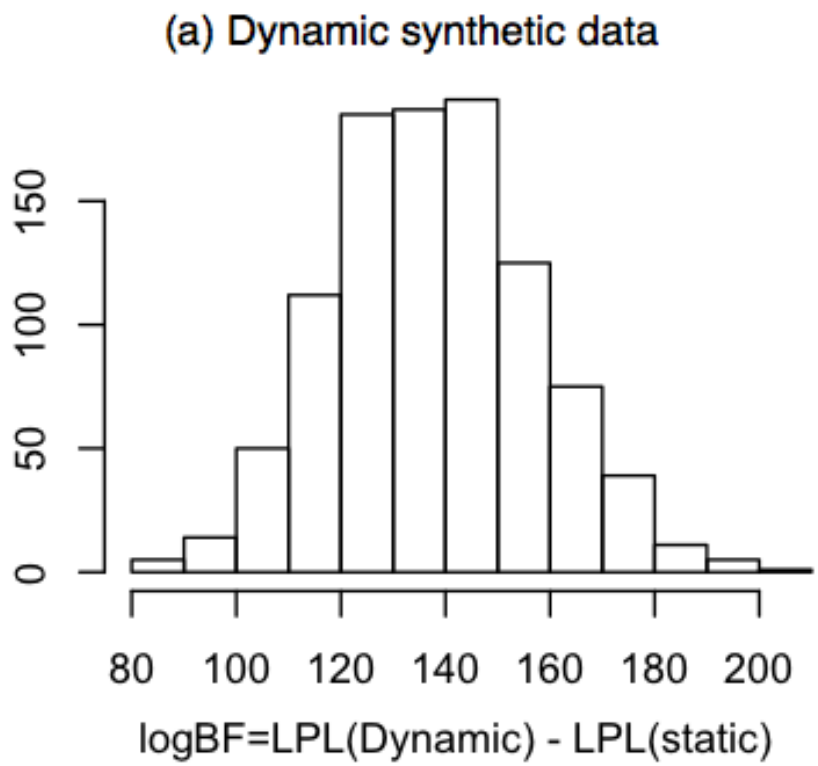
Common current practice

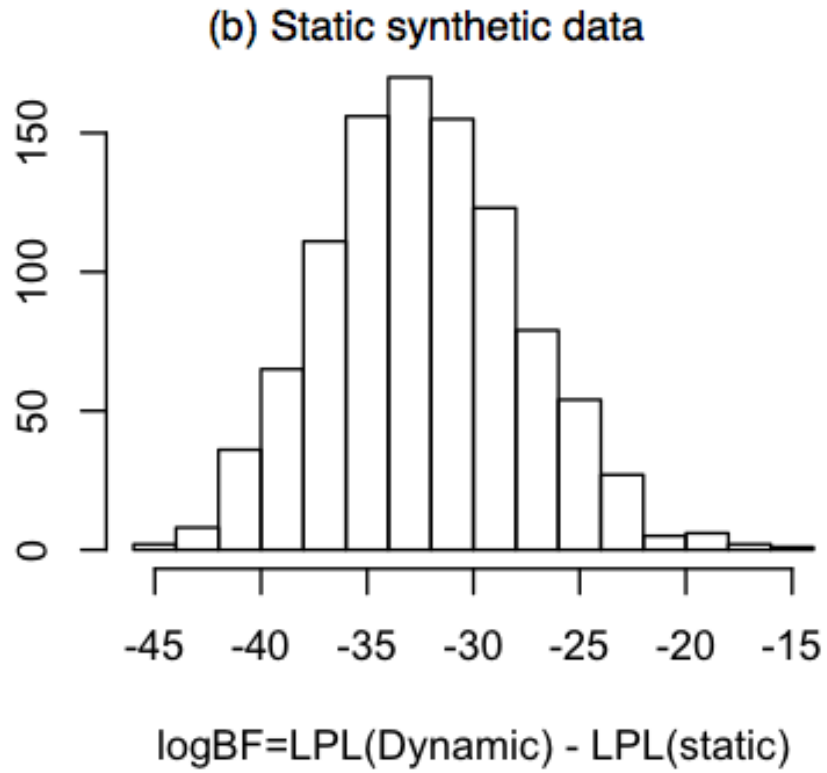
- Ignore known stochasticity in strengths & search system using standard BN algorithms.
- Select BN & interpret arrows as communications between regions of brain.
- Post hoc fit dynamic model (DCM) numerically so model drift of strengths in connectivity.

What we did: Embed the dynamics in model search

- MAP search **customised to MDMs** $\Rightarrow$ discover directionality of data generating process with given time slice DAG.
- For fixed observation/ system equations BN gives **unique equivalence class** for directionality (unlike for BN).
- **A new property for causation**! X conjectured cause of Y if when an **outlier** in X **usually excites outlier** in Y but outlier in Y does not usually excite outlier in X - uses prior error independence in stochastic process formulation!

Note Flexible: statistical diagnostics/ embellish model- e.g. change points in stochastics. Simulations BF recover processes over typical lengths of series/ real data: models scientifically plausible.
Note Scientist wanted +ve $\theta_t(i)$ - BH **dose response.** In practice mainly these found!


(a) Dynamic synthetic data
0
50
100
150
80
100
120
140
160
180
200
logBF=LPL(Dynamic) - LPL(static)


(b) Static synthetic data
0
50
100
150
-45
-40
-35
-30
-25
-20
-15
logBF=LPL(Dynamic) - LPL(static)

Recap of the first part

- Causal relationships sometimes suggested by Pearl's vanilla model search over BNs.
- Conjectures improved if select across better descriptive classes of models bespoke. MDMs - see also Flow Graphs Figuero &S(05), DCEGs Shenvi & S(20), regulatory graphs Liverani(16)).
- Methods speculative! Respect **Association**, **Consistency, Plausibility, Coherence** & **Analogue.** Vanilla methods not respecting **Dose Response** & (even in dynamic form) only partially **Temporal** property.
- Causal relationships often about **latent states** not measured directly, **causal events/decisions** not causal variables.

Next Session: Causal discovery via causal algebras $\longrightarrow$ causal predictions under control - **Experiment & Specificity** - map

$$(\mathcal{M}, \mathcal{P}) \rightarrow (\mathcal{M}_D, \mathcal{P}_D)$$

For second session

In non experimental settings using Bayesian methods to:

- **construct causal algebras** for BNs to guide map $(\mathcal{M}, \mathcal{P}) \rightarrow (\mathcal{M}_D, \mathcal{P}_D)$.
- illustrate algebras for some dynamic **causal models of massive coupled systems.**
- describe **CEG** where causal hypotheses expressed by trees.
- illustrate extraction **latent causes** using Bayesian model selection over explainable models.

Then discuss current research embedding **dose response** relationships & hierarchical structures into causal search.

Structure of talks

2nd session: Generaling causal algebras to context

1. Review BN Causal algebras
2. Emulation under control & relationship between causality & Integrating Decision Support System
3. Tree based formulations of causation & the chain event graph.
4. New causal algebras & dynamic models - regulatory graphs.
5. Current ongoing projects in causation.

Review of causal models in Bayesian inference

So far illustrated how:

1. Build class of models $\mathcal{M}$ - **idle models** - & probability models $\mathcal{P}$ over them to capture **meaningfully causally related dependence relationships** over population Ω.
2. Use available data to **select** model $M^* \in \mathcal{M}$ describing idle domain well (**causal discovery**). Seen this for BNs and MDMs.

Next move on to:

1. Construct decision space $d \in \mathcal{D}$ map $\mathcal{M} \rightarrow \mathcal{M}_d$: $M^* \mapsto M_d, \mathcal{P} \rightarrow \mathcal{P}_d : P \mapsto P_d$ - called a **causal algebra**. Let $\mathcal{M}_{\mathcal{D}} \triangleq \{\mathcal{M}_d : d \in \mathcal{D}\}$
2. Use inputs & causal reasoning to **explain & justify efficacy** of different potential controls $d \in \mathcal{D}$.

Begin by brief review of construction of maps for BNs

Constructing a Causal Algebra for BNs

In any given context: elicited or discovered **model** (M^*, P^*)(or several candidates) using expert judgement + any available observational data: **Must now find plausible** $(M^*, P^*) \longmapsto (M_d, P_d)$ to calculate scores $\overline{U}(d)$ for $d \in \mathcal{D}$. Let (M^*, P^*) be a BN with graph $\mathcal{G}$ on $\{Y_1, Y_2, .., Y_n\}$

- Suppose Y measures directly intensity of putative cause.
- Consider impact of d forcing Y to take value $\widehat{Y}$. Control called (*atomic/singular*) doing $Y = \widehat{y}$. Density of effect on other measurements

$$p_{d=\widehat{x}} \triangleq p\left(.||Y = \widehat{y}\right)$$

Note Experimental designs force = do covariates $Y = \widehat{y}$: here **guess** result of experiment from observational data (heroic!!).
Note Real inteventions $d \in \mathcal{D}$ often **compositions of atomic intervention** - applied only to certain subpopulations: see below.

Explicit causal maps of Bayesian Network

Assume $p(\mathbf{y}) > 0$. Let density $p(\mathbf{y}||y_j) \triangleq p(y_1, y_2, .., y_{j-1}, y_{j+1}, ., y_n||y_j)$ be of remaining rvs when variable Y_j set to value y_j. Pearl's formula $\Rightarrow$

$$p(\mathbf{y}||y_1) = \prod_{i=2}^{n} p_i(y_i|\mathbf{y}_{A_i}) = p(\mathbf{y}) \{p_1(y_1)\}^{-1}$$

& for $j = 2, 3, \ldots, n$

$$p(\mathbf{y}||y_j) = p_1(y_1) \prod_{i=2, i\neq j}^{n} p_i(y_i|\mathbf{y}_{A_i}) = p(\mathbf{y}) \left\{p_j(y_j|\mathbf{y}_{A_j})\right\}^{-1}$$

1. Except for founder variables, $p(\mathbf{y}||y_j) \neq p(\mathbf{y}|y_j)$.
2. **Importing prior contextual knowledge** critical to judging whether map appropriate: Causal Discover Algorithms just EDA

Some comments on Causal BNs: Invariance

$$p(\mathbf{y}||y_j) \triangleq p(\mathbf{y}) \left\{ p_j(y_j|\mathbf{y}_{A_j}) \right\}^{-1}$$

Marginal density of any Y_i unaffected by manipulation of Y_j whenever Y_i not downstream of Y_j in $\mathcal{G}$. Suppose measurements $\{Y_1, Y_2, .., Y_n\}$ perfectly measure underlying "causes" & are indexed consistently with order putative causes happen: so consistent with (but weaker than) BN **Temporal** property - "a cause Y_j after effect Y_i does not impact on Y_i" $\Rightarrow p(y_i||y_j) = p(y_i)$.

Definition

Call a BN with DAG $\mathcal{G}$ on $\{Y_1, \ldots, Y_n\}$ a *Causal Bayesian Network* (CBN) if $\mathcal{G}$ matches BN of data generating process of idle process & also, if whenever we force Y_j to take any of its possible values y_j $j = 1, 2, \ldots, n$, resulting prob. density $p(\mathbf{y}||y_j)$ satisfies formula above.

Latent variables & causes: an aside

1. Causal discovery algorithms need to entertain possibility of explanatory hidden variables.
2. Pearl assumes - although data generating process is governed by a BN - mortals **only able to observe a subvector of the measurement vector**.

Question: Can we still discover an **explicit formula** for $(M_d, P_d(e))$ as a fn. of (M^*, P^*) when d fn. of observed causal variables & e an effect of interest when **only observe a subvector of variables**
Answer: Sometimes! See e.g. Pearl(00) - called causal *identifiability results*.
Many theorems now for BNs - usually concerning solutions to algebraic equations - see e.g. Backdoor & Frontdoor Theorems
Analogous Thms. also now for other graphical models also now extant: see e.g. Thwaites et al(10) for CEGs, Yu &2(20)
Note Results assume densities given & not uncertain. But Bayes predictive/likelihood factorisations help!

Interpretation of the graph

1. In most applications not *all* variables need manipulation: sometimes absurd to entertain this $\Rightarrow$ **only need demand formula holds for certain variables** in system linked to $d \in \mathcal{D}$
2. Only need consider **subset of values doing** $\widehat{x}_k$ (don't force someone to smoke?) linked to considered $d \in \mathcal{D}$.

Restrict use of causal algebra to apply only where needed

Nevertheless **formula very substantive** & needs checking! Implication of setting $X_k = \widehat{x}_k$ e.g. same as conditioning for its children & has no direct effect on other descendants: invariance assertion

$\Rightarrow$ manipulation ($\Rightarrow d \in \mathcal{D}$) does not change mechanism of system.

Illustration

Example

In attack if made impossible to park car near building - do = "no car parked near building" - then surely determined adversary would simply park car elsewhere & return on foot to be near enough to detonate bomb.

$$\begin{array}{ccccccc} X_1 & & & & & & \mathcal{G}_1 \\ & \searrow & & & & & \\ X_2 & \rightarrow & X_3 & \rightarrow & X_4 & & \end{array}$$

Graph $\mathcal{G}_1$ not a CBN. e.g. $p(x_3, x_4 || x_1, \widehat{x}_2 = 0) \neq p(x_3, x_4 | x_1, x_2 = 0)$.

- **Standard causal algebras** typically $\Longrightarrow$ no **rational alternatives** sought by $\omega \in \Omega$.
- OTOH "Bomb in place" has same consequence whether this has happened naturally $X_3 \rightarrow X_4$ = e.g. adversary prevented from placing it/ does not place it.

CBNs and emulated networks (S,10)

Question: Why is the CBN a natural extension to data scientists?
Answer: Extension automatic if we simulate/ emulate data through a network (see also SEMs)!

- CBNs formula, assume underlying measurement vector network of n independent emulators, i^{th} emulator generates an outcome y_i for each Y_i density $p_i(y_i|\mathbf{y}_{Q_i})$, $i = 2, 3, \ldots, n$.
- Easily seen **emulator network a BN** with directed graph $\mathcal{G}$ on $(Y_1, \ldots, Y_n)$ where edges from Y_j to Y_i iff y_j component of $\mathbf{y}_{Q_i}$. Draws from this network = draws from $p(\mathbf{y})$.
- But also **implicitly makes predictions under control**: about when output Y_j of j^{th} emulator *forced* to take value $\widehat{y}_j$, $j = 1, 2, \ldots, n$. Simply replace j^{th} emulator with one that returns value $Y_j = \widehat{y}_j$ with prob. $1 \Longrightarrow$ Pearl's causal formulae.

CBNs, dynamic emulators & the IDSS (Leonelli& S,15))

- (CBN) $\Longleftrightarrow$ composite model coupling massive stochastic emulators together: very current.
- m possible massive totally ordered models $\{M_1, M_2, \ldots, M_m\}$ delivering *output vectors* $\{\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m\}$
- each M_i needs to receive as inputs/independent variables $\mathbf{Y}_{A_i}$ which is a subvector of $(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{i-1})$ where *parent indices* $\mathbf{A}_i$ subset of indices $\{1, 2, \ldots, i-1\}$
- e.g. time slice

$$\begin{array}{ccccc} \mathbf{Y}_1 & \rightarrow & \mathbf{Y}_3 & \rightarrow & \mathbf{Y}_5 \\ & & \downarrow & \searrow & \\ \mathbf{Y}_2 & \rightarrow & \mathbf{Y}_4 & \rightarrow & \mathbf{Y}_6 \end{array}$$

has $m = 6$ and e.g. $\mathbf{Y}_{Ai} = (\mathbf{Y}_2, \mathbf{Y}_3)$. Network response $(\mathbf{Y}_5, \mathbf{Y}_6)$.

Such coupled systems implicitly CBNs - causal extensions apply

Idle models for nuclear counter measures, food security, energy for carbon zero and flood risk.

A proposal of general causal modelling

Illustrated above BN causal algebra but usually need to **customise** to specific $\mathcal{M}$**:**

1. A deduced **directionality** of bespoke relation with customised formal graph $\mathcal{G}$.
2. A **plausible argument** to embed this directionality under composite acts $d \in D$.

- *Customise* semantics of a graphical model to faithfully reflect **types of relationships** seen in given domain (bullet 1 above).
- *Search* over $\mathcal{M}$ to find dependence model(s) M^***consistent with knowledge & available data.**(as in BN above)
- Use $\mathcal{G}$ (Bullet 1) & causal algebra to guide customised *map* $M^* \mapsto M_D$ **controlled analogues**: (bullet 2 above)

End point: Putative *causal hypotheses* & experiments to test them

From causal algebras for BNs to other customised graphs.

- **Networks of emulators** behave **very like causal BNs** & causal algebras often quite similar.
- Emulators not reality!! - so **efficacy of causal models needs** to be judged against **real systems**.
- Different bespoke models $\mathcal{M}$ & algebras for causal explanations emphasise different BH (& other) causal properties depending on underlying mechanisms.
- Explanations behind $\mathcal{M}$ & such extensions using **bespoke graphical representations $\mathcal{G}$ can often import natural language explanations** into a causal analysis

Customisation of $\mathcal{M}$ through $\mathcal{G}$ leads to conversation with experts

The IDSS and causal algebras on dynamic processes

- Components necessarily emulated $\Rightarrow$ any coupled model **automatically causal Bayesian** composite: now dynamic.
- Each component **conditional stochastic process updated autonomously** by different expert panels.
- Theory (Leonelli + S(15)) + several applications e.g. Leonelli + S(13), Barons et al(22), Volodina el al(22), Shenvi et al (22).

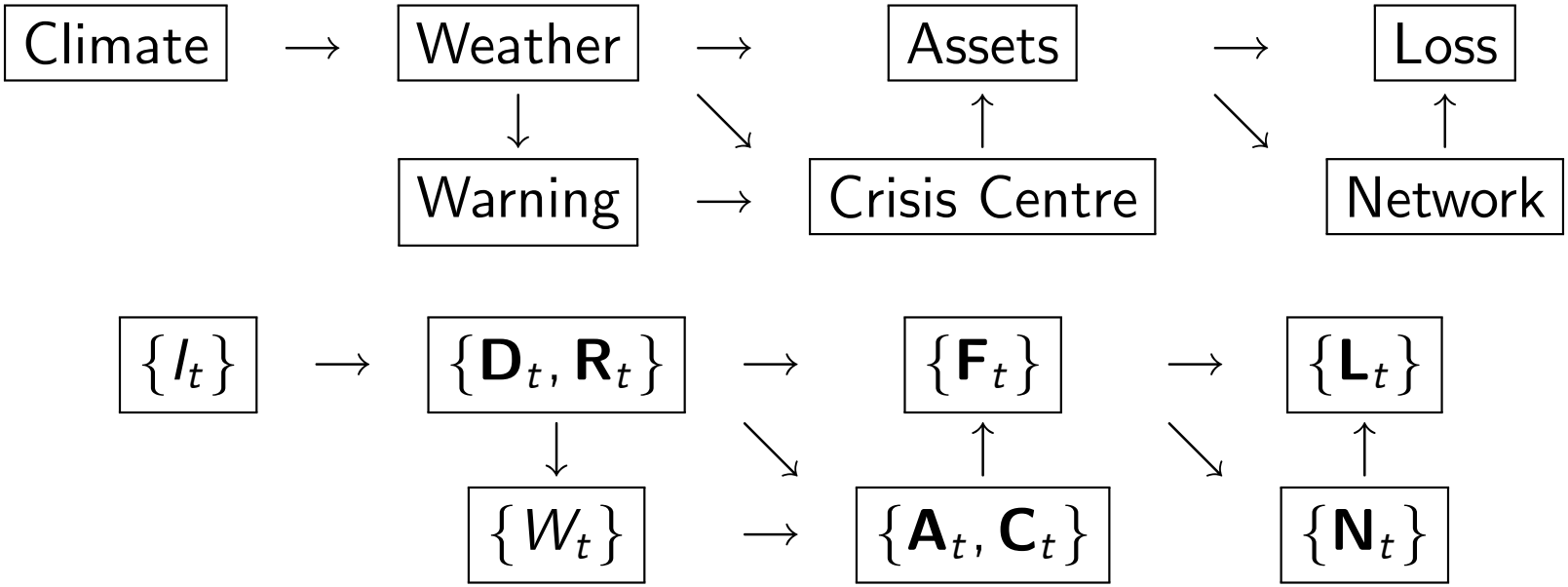
$$\begin{aligned} p_T(\mathbf{z}^{(T)}|\mathbf{y}) &= \prod_{t=1}^{T} p(\mathbf{z}_t|\mathbf{z}^{(t-1)},\mathbf{y}) = \prod_{t=1}^{T}\left\{\prod_{k=1}^{m} p_k(\mathbf{z}_{kt}|\mathbf{z}_{A_k}^{(t-1)},\mathbf{y})\right\} \\ &= \prod_{k=1}^{m}\left\{\prod_{t=1}^{T} p_k(\mathbf{z}_{kt}|\mathbf{z}_{A_k}^{(t-1)},\mathbf{y})\right\} = \prod_{k=1}^{m} p_k(\mathbf{z}_k^{(t)}|\mathbf{z}_{A_k}^{(t-1)},\mathbf{y}) \end{aligned}$$

Fact (Can divide and conquer)

m panels + data $\mathbf{y} \longrightarrow$ *time slice* $p_k(\mathbf{z}_{kt}|\mathbf{z}_{A_k}^{(t-1)},\mathbf{y}) \Rightarrow$ *full prob. spec.!!!*

CReDo and its IDSS

IDSS = DBN (e.g. Korb & Nicholson(10), S(10)): gives parents A_k on time slice of component $1 \leq k \leq m = 7$.



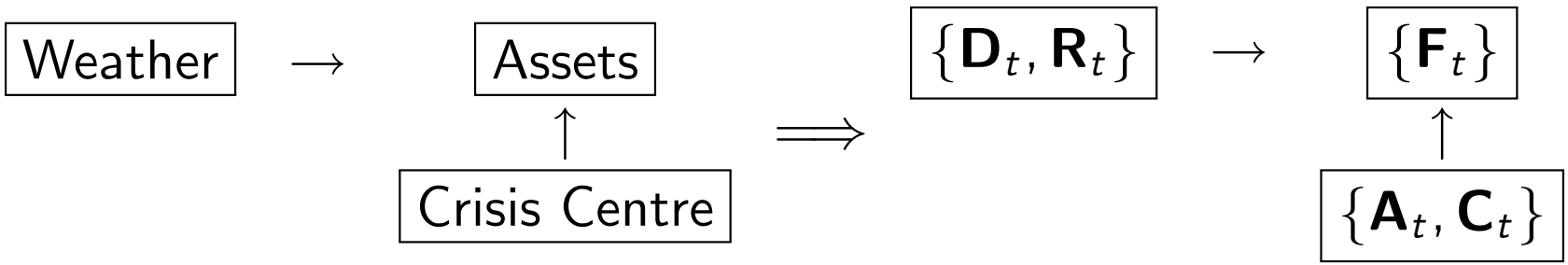
Fact (No need to couple all possible outputs!)

Here IDSS interface variables only needed to inform asset owner's losses!!

The CReDo project: stochastic transfer of incident

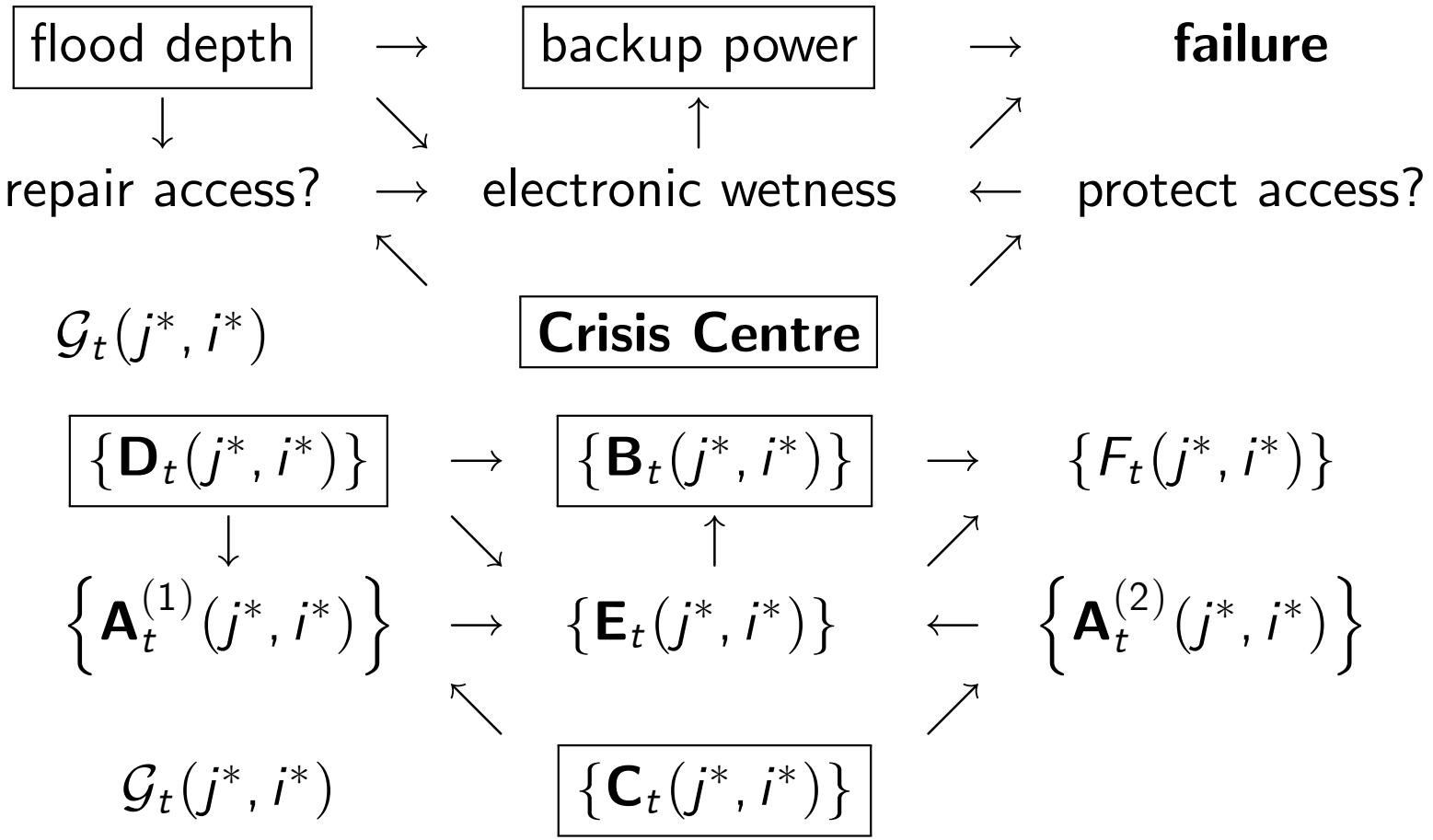
- Marginalise over weather to produce output $\longrightarrow$ OR module (e.g. optimiser - mitigating rerouting options).
- Vectors of failures across all assets with mass function $\left\{p_t^*(\mathbf{f}_t(i)|\mathbf{f}^{t-1}(i))\right\}_{t\in\mathbb{T}}$ sampled for each chosen incident $i \in I$. Jt. mass fn.

$$p_T^*(\mathbf{f}(i)) = \prod_{t=1}^{T} p_t^*(\mathbf{f}_t(i)|\mathbf{f}^{t-1}(i))$$



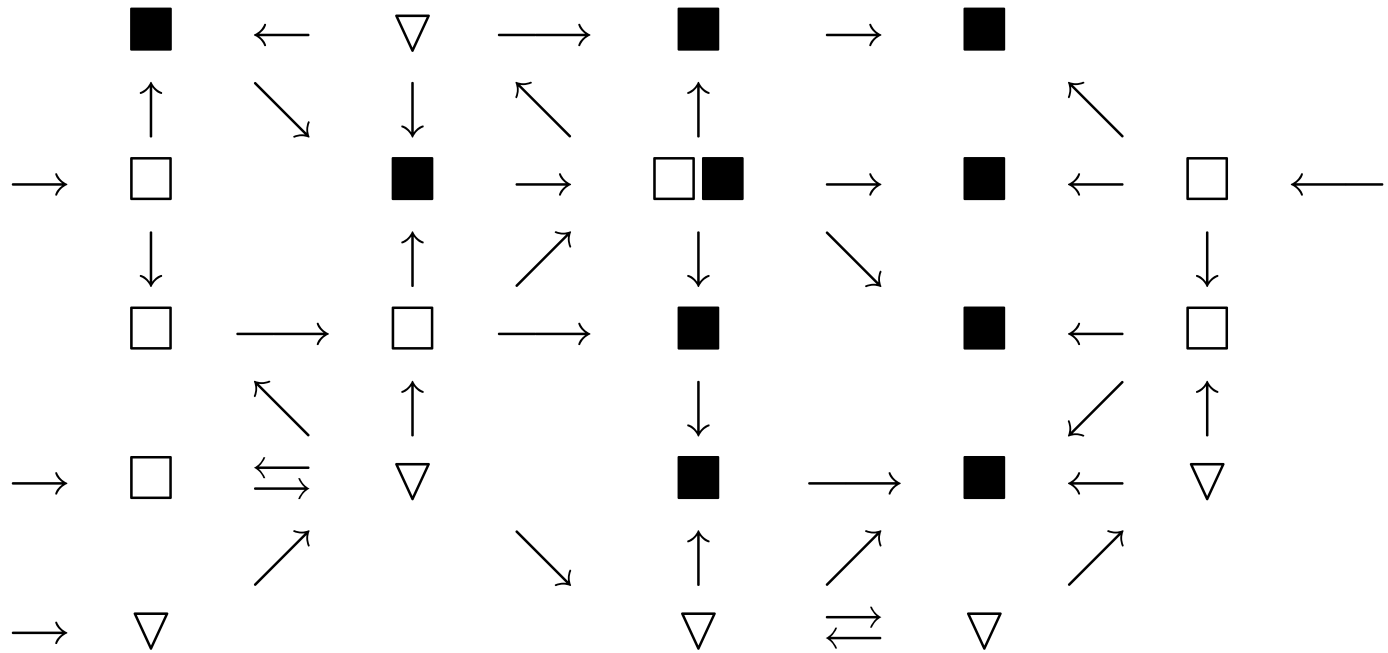
Specific DBN for Pumping Station failure

Next we customise to specific assets:



CReDo project: Digital Twin of Flood Threat

100's of diverse types of asset: overlay different connections
□ - asset of power company, ■ - asset of water company, ▽ - asset of communications company.



Note Connection types depend on 2 assets even within same company (substations etc.)

Structural Elicitation and causality in composite models

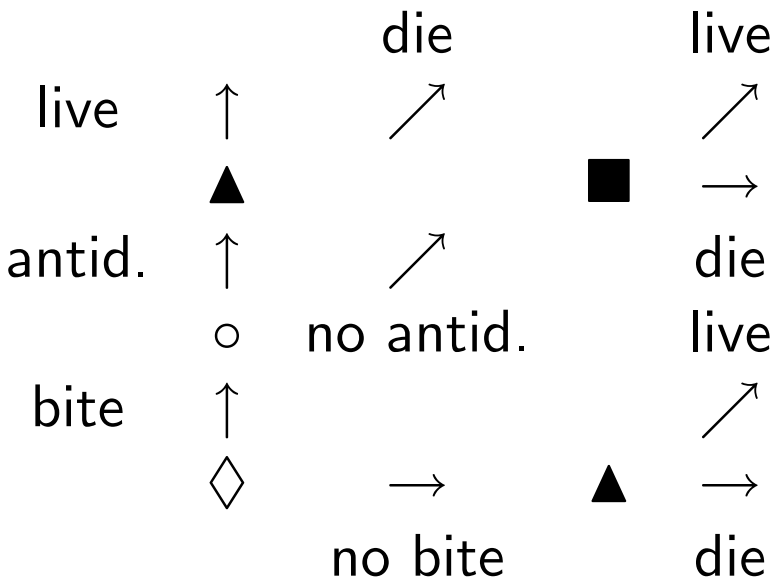
- Small no. candidate models to searched across. Model selection driven by structural expert judgements: intense sessions unpacked both how crises unfolded & remedial steps possible $\Rightarrow$ why **Bayes is needed** 1!
- **Causal** extensions essential to construct model $\Rightarrow$ digital twin is causal: admits a **causal algebra** appropriate to possible interventions $\longrightarrow$ increased reliability.
- Domain experts inform explicit quantitative causal modelling $\Rightarrow$ faithful causal models $\Rightarrow$ data into $M^* \longmapsto M_D$ prediction under control $\Rightarrow$ **why Bayes needed** 2!!
- Embedding utility means that effect measurements appropriate - **why Bayes needed** 3!!!
- Deep theory needed to appropriately model these systems at customised levels - e.g. **coupling theory** & stochastic search & optimisation. Leads to **great stochastic modelling challenges**!

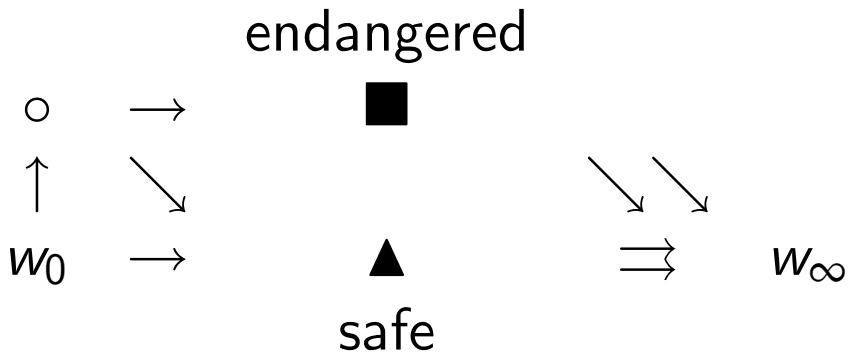
The CEG instead of BN for causal discovery/ causal reasoning

- A **CEG generalises a discrete BN**. BUT shares with BN nearly all of its desirable properties including accommodation of causal discovery algorithms and its own causal algebra.
- CEGs particularly useful for **idle systems describing unfolding events**: for single people or populations. Examples include: Biological **cell attacks** S & Anderson (08), **Educational progress** through university programmes (Freeman & S 10,12), **Migration** processes, & victims of **Modern Slavery** (Strong & S 22), **Cerebral palsy**, **epilepsy** (Barclay et al, 14), **falls in elderly** (Shenvi & S(2019)), **child hospitalisation** (Barclay 12, Cowell & S, 15 Collazo et al 18) **Forensic evidence** in court cases - Amanda Knox case, drugs on bank notes cases, (Wilson et al 22,22a)
- Admit **simple causal algebras**, more expressive than is possible through vanilla BNs.

Snake Bite Example: A toy example

$X_1 \sim$ Bitten by snake, $X_2 \sim$ Carry and apply perfect antidote,$X_3 \sim$Die tomorrow..





$X \sim$ not bitten/bitten but apply antidote, $Y \sim (= X_3)$ live/die, $Z \sim$ safe/endangered.

Note Variables exhibiting conditional independence read from CEG. Effect variable indicator "being safe".

Constructing a CEG

Event tree $\longrightarrow$ Staged tree $\longrightarrow$ CEG [by positions and stages]

- Start with an event tree as illustrated above.
- Colour vertices of tree to rep its stages ($\triangleq$ staged tree).
- Identify positions ($+\ w_\infty$) give vertices of CEG.
- Construct CEG by inheriting edges in obvious way from tree - then attach all leaves to w_∞.

Chain Event Graphs in General: Recap

- Derived from probability trees but often **topologically much simpler**.
- Like a tree embed collections of hypotheses about **how things might have happen:** so link directly to **BH temporality**.
- Model non homogeneous semi Markov processes graphically.
- Like a BN **able to express many hypothesised independences** within story: read from its **cuts** link to causal story in graph Collazo et al (18).
- Like BNs provide a **framework for conjugate inference** & model selection. see Collazo et al (18). So can be used to set things up for Causal Discovery from sampled observational data set.
- Equivalence classes now known Gorgen & S (18)- so supports a causal discovery algorithm.

Drawing experimental and sample evidence into CEG's

- Likelihood separates! $\Rightarrow$ class of regular CEG's admits simple conjugate learning.
- For example likelihood under complete random sampling given by

$$\begin{aligned} l(\boldsymbol{\pi}) &= \prod_{u \in U} l_u(\boldsymbol{\pi}_u) \\ l_u(\boldsymbol{\pi}_u) &= \prod_{i \in u} \pi_{i,u}^{x(i,u)} \end{aligned}$$

where $x(i,u)$ # units entering stage u & proceeding along edge labelled (i,u), $\sum_i \pi_{u,i} = 1$ in sample.

- From Bayesian perspective e.g. independent Dirichlet priors $D(\boldsymbol{\beta}(u))$ on the vectors $\boldsymbol{\pi}_u$ leads to independent Dirichlet $D(\boldsymbol{\beta}^*(u))$ posteriors where

$$\beta^*(i,u) = \beta(i,u) + x(i,u)$$

Learning the topology of a CEG

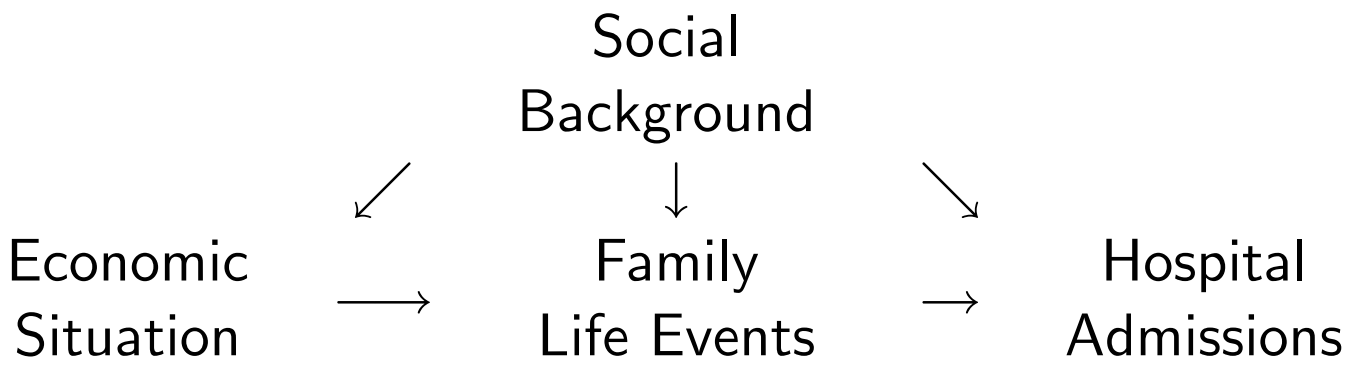
.

- Appropriate priors & modular parameter priors over CEGs $\Rightarrow$ log marginal likelihood score ancestral data *linear* in CEG stage components.
- Explicitly for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$, let $s(\boldsymbol{\alpha}) = \log \Gamma(\sum_{i=1}^{k} \alpha_i)$ and $t(\boldsymbol{\alpha}) = \sum_{i=1}^{k} \log \Gamma(\alpha_i)$

$$\begin{aligned} \Psi(C) &= \log p(C) = \sum_{u \in C} \Psi_{u(c)} \\ \Psi_{u(c)} &= \sum s(\alpha(i,u)) - s(\alpha^*(i,u)) + t^*(\alpha(i,u)) - t(\alpha(i,u)) \end{aligned}$$

- e.g. MAP model selection using AHC , Dynamic Prog.,Integer Prog, simple & fast over vast space of CEG's (see Cowell & S,15). Well documented search codes in R and Python.

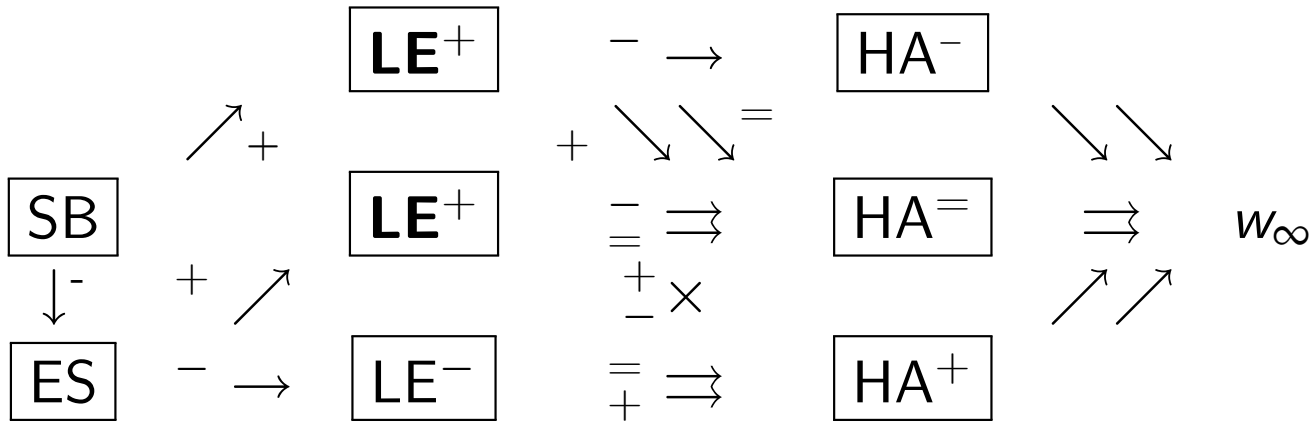
CEG's of idel systems better than BN's (Barclay et al, 2012)



- Search over all CEGs whose trees consistent with hyothesis of order happened.
- Search discovers CEG whose MAP score 80 times better than best BN.

The MAP CEG

Best explanatory model of our data with HA the response.

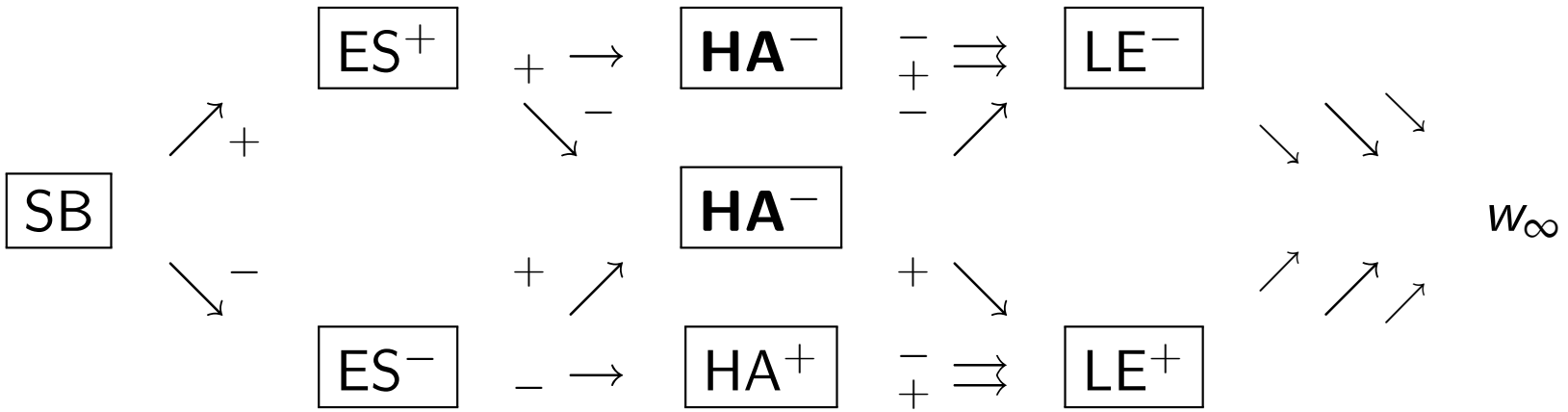


Paths define different routes to low,HA^{-} medium $HA^{=}$& high HA^{+} proportions of hospital admissions.

- ES. not "cause" of life events or hospital admissions for High SB household.
- High SB & low LE uniquely "causes" child good HA. (**access to credit = new putative cause**).

Example CHIDS a different CEG

Dynamic Programming search optimum changes response variable.



- This model explains *life events as a result of poor child health*. To discriminate look at timings embedded in surveys.
- High hospital admissions HA^+ due to access to poverty (2 categories).
- High life events unaffected by Hospital Admissions unless when exactly one of SB or ES is low then poor child health can shift into lower life event category.

Chain Event Graphs and Causal Hypotheses

- CEG's automatically **constructs explanatory variables** = cuts from best fitting stories $\longrightarrow$ can provide putative causes for use in a causal algebra. For BNs discovery only "do" variables = potential causes collections of **variables specified a priori**..
- We can **"do" events in a story not just variables** see Thwaites et al (10)
- Directly use extracted story of how thing happen **BH temporality automatic** $\Rightarrow$ a causal event can only affect events in subtree rooted at that putative cause.
- Atomic doing just sets an edge prob to 1. Decisions $d \in D$ typically compositions of these atomic acts.

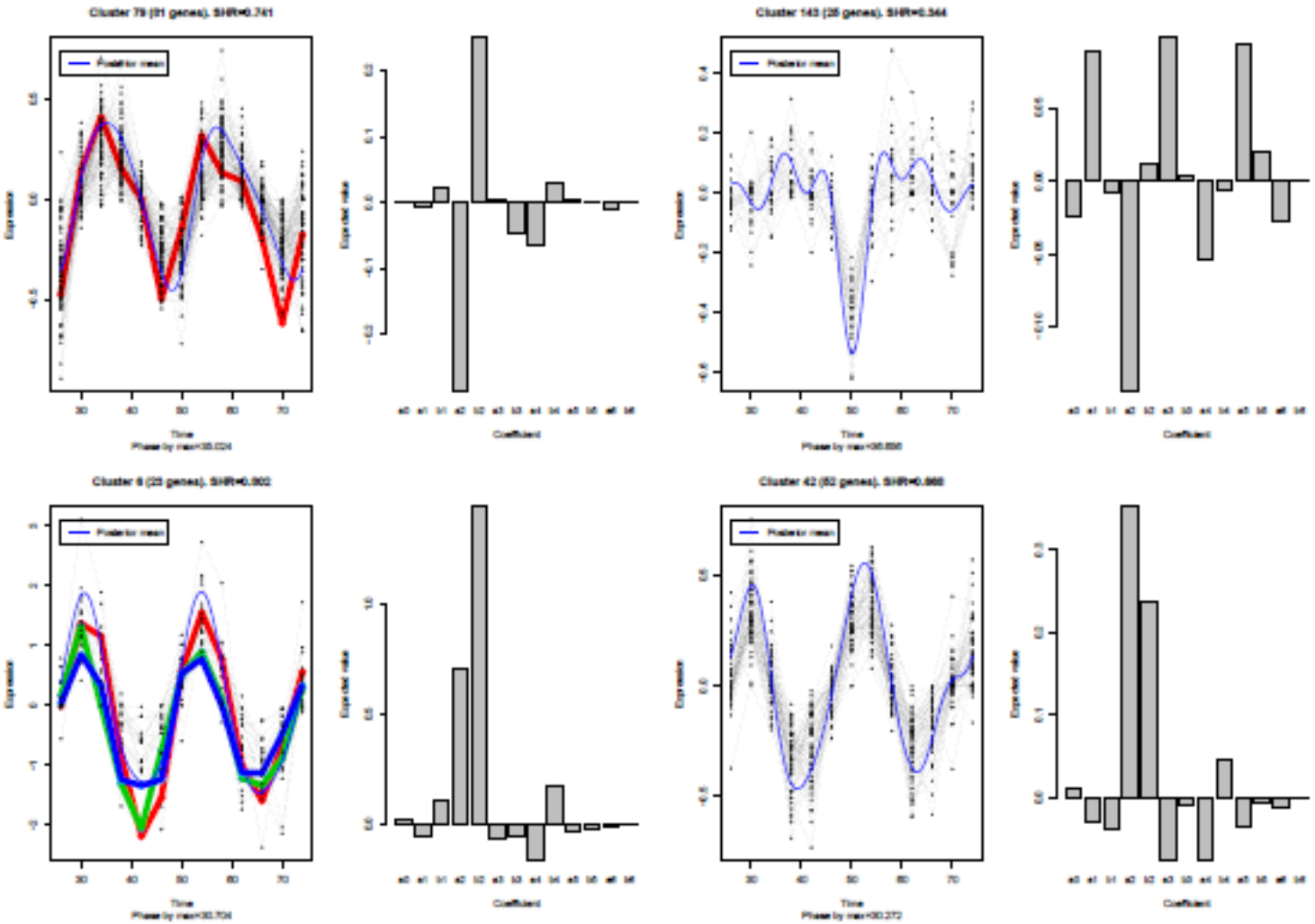
Mapping explicable through graph of idle model!!

Causal Algebras for dynamic models: stochastic analogues

- Seen MDM has **straightforward causal discovery algorithms** to prompt hypotheses about potential causal algebras e.g. Wilkerson (20).
- Once graph decided **simple causal algebra for instantaneous "doing" variables** on specific time slices. Almost precise copy of CBN but now on a time slice. But **can also "do" on states** often natural too (e.g. "do" $\theta_{1t} = 0$ says an edge does not exist at time t)
- Furthermore **realistic interventions often composed** over many time slice, contingent plans & have often have a *direct* effect on variables over time. So modified causal algebra often needed. **Must be elicited from experts**! Map to decision space non-trivial.
- **Optimisation algorithms** to investigate optimal acts. Aglietti et al(20)

New Dynamic Causal algebras not just for MDMs

Fourier Gene Expression over time Liverani & S (15)



Circadian Causal Models (Liverani & S, 08, 10, 15)

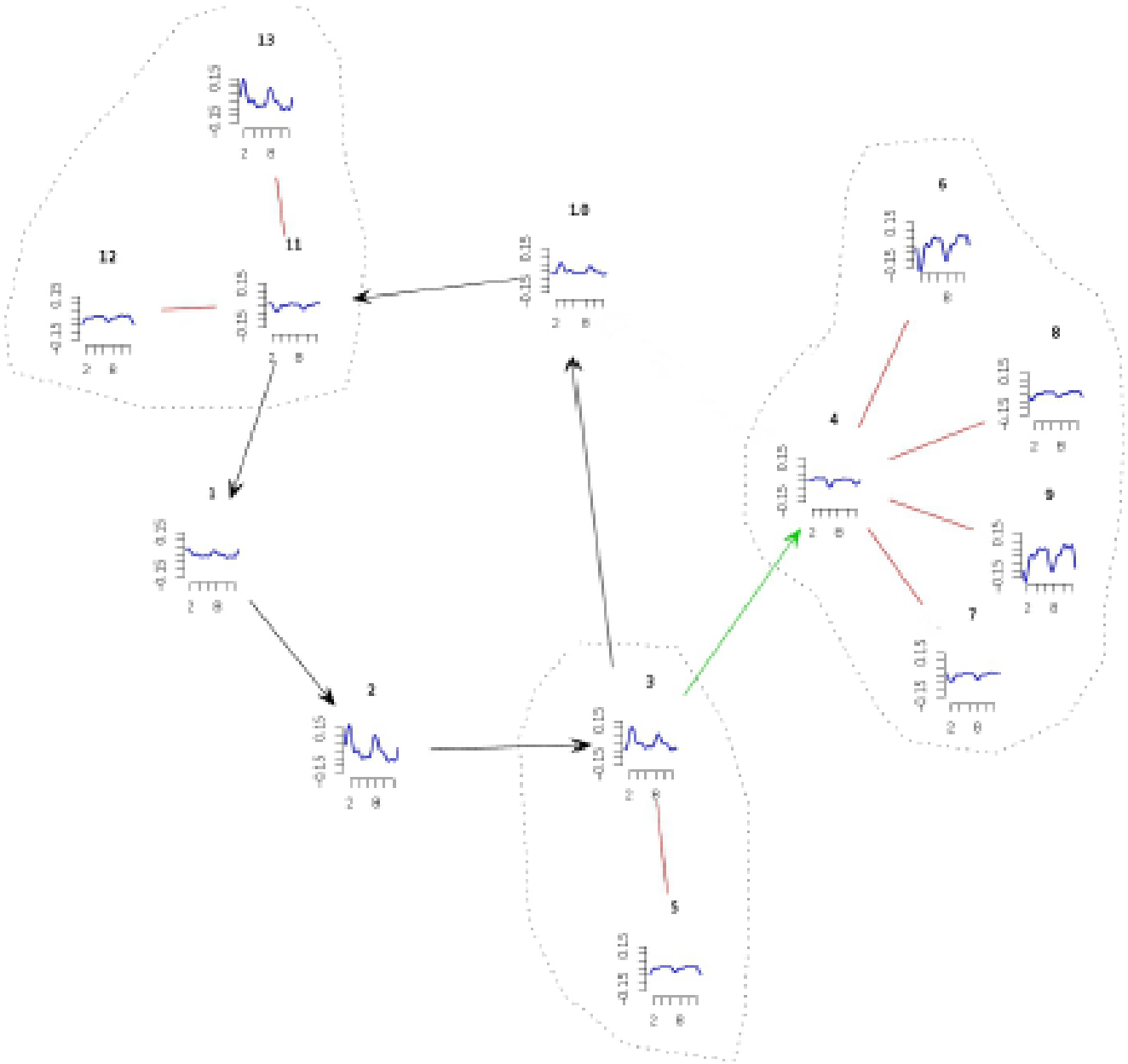
- Circadian regulation plausible $\Rightarrow$ **shape** of (subset of) regulated cluster relates its shape to shape of regulating cluster. e.g. a short positive phase change (& damping?). So objects of interest are **clusters** Potential causal relationships indicated by damped translations.
- **Search observed system** Find coexpression clusters of longitudinal **shapes**: over partition space of 20,000 different time courses using BF selection . Conditional conjugacy gives product student t scores $\Rightarrow$ greedy search feasible. Biologists extract 100-200 clusters whose profiles might link to regulation.
- Extend search for **causal relationships** between identified cluster whose shape dependence customised to damped translations! Quick expressed a group & **retains closed form**. **Graph defined** & can be annotated & **causal control defined**!

What we do

- Use BF search over a **group** of data transformations acting on cluster shape. Orbit of such groups coarsen partition of *supraclusters*.
- BF score tends to support coarser model if evidence against it not strong:then combines.
- Clusters in different supraclusters hypothesised to relate to **different regulatory cycles**:
- For clusters in same supracluster node joined by directed edge to another depending on value of parameters of "best explaining" group transformation.
- Then a **directed edge** corresponds to a potential "causal" relationship within a connected component.

Note Standard **DBN wrong form** of relationship - not appropriate for this class!

The MAP Regulation Graph for Supracluster of Aribidopsis



A new bespoke causal algebra for gene regulation

- **Tearing** a cluster $\Rightarrow$ ripping out these genes from system. Predicted effect is to take out relevant edges & vertices retain vertex & edge distributions from idle system .
- **Doing** $\Rightarrow$ forcing all genes in cluster to produce a signal.
- Testable because each graph has different predicted effects under these controls .

Some Asides

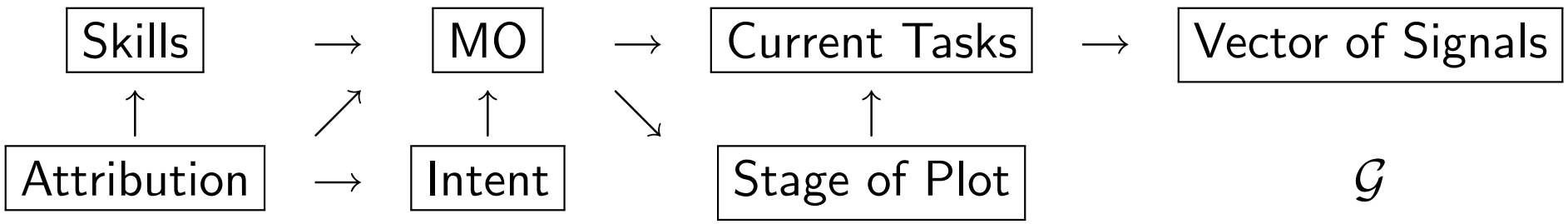
- Typically believed regulatory relationships sparse $\Rightarrow$ search for minimum edge length graphs satisfying known constraints -**Parsimony**.
- Distances expressible as fn of parameters of transformations/edge lengths

Current developments in Causality from our team

- Now applying search with positivity constraints on subsets of variables MTP2 - embedding BH **Dose Response**
- Causal stories often happen in **hierarchical dynamic processes** - see Bunnin & S (21) with DCEGs at base level: applied to illicit drug production, exfiltration of documents.
- Extraction algorithms for **causal reasoning from engineers' natural language reliaibility reports** from engineers Yu & S (20,22)
- Modern Slavery & Migration Models Strong & S (22) $\rightarrow$ **Dynamic Interacting Agents & Agent Based Models**.

Dynamic hybrid graphical structures

Time Slice of generic Causal Model of Criminal Attack



1. All but last variable in $\mathcal{G}$ **hidden** from surveillance - BUT intelligence informs other variables.
2. Dynamic causal algebras project intelligence into "stage of plot" & current tasks.
3. Vector of signals observational data informs what tasks performed.
4. Police interventions **do** Stage of Plot (catch gang red-handed) or Current Tasks (frustrate essential task).

Note Generic Causal Models formulation $\longrightarrow$ **protocol for explainable, data based analysis** $\longrightarrow$ wiser decision making.

Conclusions

- All causal hypotheses to be articulated through bespoke semantics:
  1. **A set of objects** to be identified as plausible causal and effect features.
  2. Well defined **dependence relationships** between these consistent with causal relationships within nature.
  3. A plausible **embedding** of this directionality **into a controlled domain** where causes freely enacted.
- **Model selection method** bespoke to these relational models to perform causal discovery - using **characterised statistical equivalent models** invariant to directionality of association to lead construction of putative causal hypotheses.
- **Different** forms of **causal representation** need to be **customised** to specific context & science.
- Great fun developing **new formal graphical systems to help clients formulate their particular brand of causal conjecture**.

Many thanks for your attention!!!!

Selected Publications

Shenvi, A, Bunnin FO & S(22) “A Bayesian decision support system for counteracting activities of terrorist groups” Math ArXiv
Strong P., McAlpine, A. & S (22) ’Towards A Bayesian Analysis of Migration Pathways using Chain Event Graphs of Agent Based Models’ Proc BaYSM (to appear)
Barons ,M Fonseca T, Davies, A & S(22) "An Integrating decision support system for addressing food security in the UK" RSSA,(to appear)
Volodina, V Sonenberg, N., S, Dent, CJ Challenor, P & Wynn, HP (22) "Propagating uncertainty in a network of energy models" IEEE Explore
Yu, X. & S(22) "Hierarchical Causal Analysis of Natural Languages on a Chain Event Graph" arXiv (invited revision)
Dent, CJ, Mawdsley, B, S, & Wilson KJ (22) " Assessing Asset Vulnerability" CReDo Tech. Paper 3, National Digital Twin Programme online digitaltwinhub.co.uk
Yu, X. & S(21) "Causal Algebras on Chain Event Graphs with Informed Missingness for System Failure", Entropy, 23, 10, 1308

Selected Publications

Wilkerson, RL & S(21) "Customised Structural Elicitation" In Expert Judgement in Risk and Decision Analysis Springer

Leonelli, M, Riccomagno,E & S(20) “Coherent combination of probabilistic outputs for group decision making: an algebraic approach” OR Spectrum

Wilkerson, R. (20) "Thesis" Uni of Warwick

Aglietti, V. Lu,X. Paleyes A & González J (20) "Causal Bayesian Optimisation" Proc 23rd Int Con AI & Stats, PMLR 108:3155-3164,

Shenvi, A. Smith, J.Q., Walton,R.,.& Eldridge, S. (19) “Modelling with Non-Stratified Chain Event Graphs (Proc. BAYSM 2018)

Barons MJ, Wright SK, & S(18) "Eliciting Probabilistic Judgments for Integrating Decision Support Systems", in: Dias LC, Morton A, Quigley J (eds), Elicitation: The Science and Art of Structuring Judgement. Springer, New York, Chapter 17 p 445-478

Gorgen, C. , Bigatti,A. ,Riccomagno, E. & S(18) "Discovery of statistical equivalence classes using computer algebra" International J. of Approximate Reasoning 95 . pp. 167-184. .doi:10.1016/j.ijar.2018.01.003

Selected Publications

Mitrovic,J.,Sejdinovic, D. & Teh Y.W.(18) "Causal Inference via Kernel Deviance Measures" NeurIPS 31

Collazo, R.A., Gorgen,C. & S(18) "Chain Event Graphs" Chapman and Hall

Gorgen, C. & S(18) "Equivalence classes of Staged Trees" Bernoulli, Vol. 24, No. 4A, 2676-2692.

Collazo, R.A. & S(17) "The Dynamic Chain Event Graph" Proc. of ISI Marrakech,

Thwaites P.A. & S(17) "A Graphical method for simplifying Bayesian Games" Reliability Enginereering &Safety Systems

Collazo, R.A. & S(16) "A new family of Non-local Priors for Chain Event Graph model selection" Bayesian Analysis 11, 4, 1165 - 1201

Peters, J., Buhlmann N & Meinshaussen, N(16) "Causal Inference by using invariant prediction: identification and confidence intervals" JRSSB 78,5, 947-1012

Barclay, L.M., R. Collazo, S, Thwaites , P. & Nicholson , A. (15) "Dynamic Chain Event Graphs" EJ of Stats 9, 2, 2130-2169.

Selected Publications

Leonelli, M, & S (15) "Bayesian Decision Support for Complex Systems with Many Distributed Experts" Ann Oper Res 235 (1) 517 - 542

Dawid AP(15) Statistical Causality from a decision theoretic perspective" A.Rev.Statist. Appl. ,2,273 - 303

Costa L, S. Nicholls, T. & Cussens J. (15) "Searching Multiregression Dynamic Models of Resting-State fMRI Networks Using Integer Programming" Bayesian Analysis 10, 2, 441-478

Liverani, S. & S(16) "Bayesian selection of graphical regulatory models". International J. of Approximate Reasoning, 77, 87–104. URL: doi:10.1016/j.ijar.2016.05.007

Cowell, R.G. & S(14) "Causal discovery through MAP selection of stratified chain event graphs" EJ of Stats,8, 965 - 997

Barclay, L.M., Hutton, J.L. & S (14) "Chain Event Graphs for Informed Missingness" Bayesian Analysis, 9,1, 53-76

Barclay, L.M. , Hutton, J.L. & S.(13) "Refining a Bayesian Network using a Chain Event Graph" IJ of Approximate Reasoning 54, 1300-1309.

Selected Publications

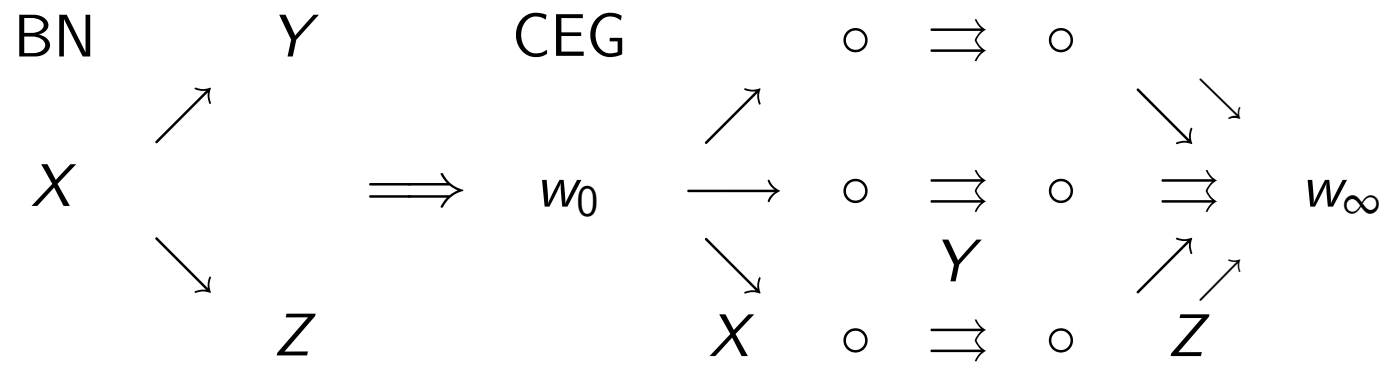
Leonelli, M & S (13) "Dynamic Uncertainty Handling for Coherent Decision Making in Nuclear Emergency Response" Proc. 2013 Amer Nuclear Society
Freeman, G. & S(11) "Dynamic Staged Trees for Discrete Multivariate Time Series: Forecasting, Model Selection & Causal Analysis", Bayesian Analysis,6,2, 279 - 306,Thwaites, P. S. & Riccomagno, E. (10) "Causal Analysis with Chain Event Graphs" Artificial Intelligence, 174, 889–909
Freeman, G. & S (11a) " Bayesian MAP Selection of Chain Event graphs" J. Multivariate Analysis, 102, 1152 -1165
S(10) "Bayesian Decision Analysis" CUP
Korb K. & Nicholson A. E. (09) Bayesian Artifical Intelligence CUP
S& Anderson P.E. (08) "Conditional independence and Chain Event Graphs" Artificial Intelligence, 172, 1, 42 - 68
S & Figueroa L.J.(07) "A Causal Algebra for Dynamic Flow Networks" in "Advances in Probabilistic Graphical Models" Eds P. Lucas, J.A.Gamez, and A. Salmeron, Springer, 39 -54

Selected Publications

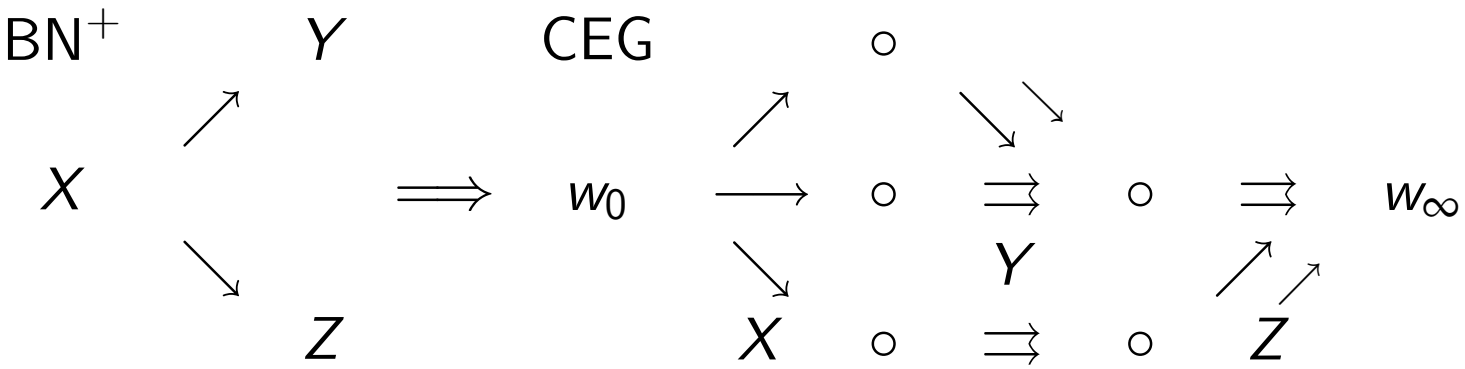
G. Shafer (03). "The Art of Causal Conjecture". Cambridge, MA, MIT Press
Pearl, J.(00). "Causality: Models, Reasoning, & inference". CUP
Cooper G & Yoo, C (99) "Causal discovery from a mixture of experimental and observational data" Proc 5 Conf UAI 116-125 Morgan Kauffman
Queen, CM & S (93) "Multiregression Dynamic Models" JRSS B 55(4) 849 - 870
P. Spirtes, C. Glymour & R. Scheines (93). "Causation, Prediction, and Search". Springer-Verlag, New York

Aside: How CEG extends a BN

.



but context specific BN^+ fits much better



(distribution of Z same whether or not X takes medium or large value)

Read CI from CEG?(Smith & Anderson,08, Thwaites & Smith,15)

Theorem

If the random variables $X_1, X_2, \ldots, X_n$ with known sample spaces are fully expressed as a BN, G, or as a context specific BN G, and you know its CEG, C, then the random variables $X_1, X_2, \ldots, X_n$ and all their conditional independence structure together with their sample spaces can be retrieved from C.

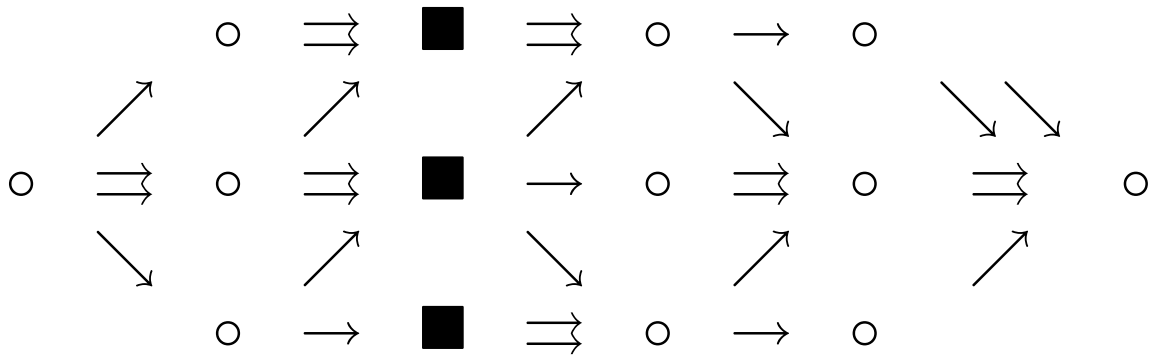
Theorem

$Downstream \amalg Upstream | \ w-Cut$

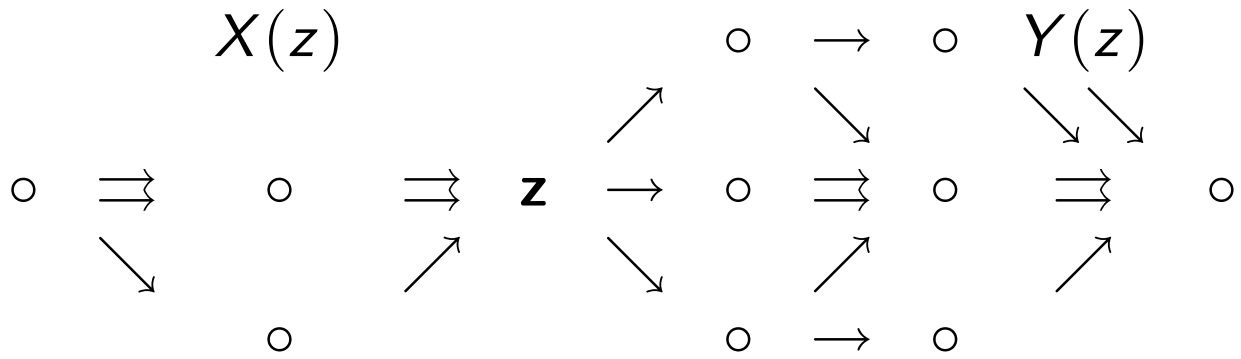
Theorem

$Children \amalg Upstream | u-Cut$

Example of a CEG with Cuts



Downstream $Y(z)$ independent of upstream $X(z)$ given cut $Z = z$.Cuts need not be orthogonal. So can construct dependence through functional relationships.



Observation & System equation

$$\begin{aligned} Y_t &= \mathbf{F}_t(\mathbf{x}_t)\boldsymbol{\theta}_t + v_t \\ \boldsymbol{\theta}_t &= G_t\boldsymbol{\theta}_{t-1} + \mathbf{w}_t \end{aligned}$$

where $\{v_t, \mathbf{w}_t : t = 1, 2, 3, \ldots\}$ II with 0 mean - observation variance V_t & system error covariance matrix $\mathbf{W}_t$. Let $\mathbf{y}^t$ denote $(y_1, y_2, \ldots, y_t)$.
$\mathbf{F}_t(x_t)$ is a vector fn. of known coeffs. of covariates $\mathbf{x}_t$.
When all errors Gaussian and $\boldsymbol{\theta}_0 \backsim N(m_0, C_0)$ then conditional on $\{\mathbf{F}_t(\mathbf{x}_t), G_t, V_t, W_t\}$, $\boldsymbol{\theta}_t|\mathbf{y}^t \sim N(\mathbf{m}_t, C_t)$ & $Y_t|\mathbf{y}^{t-1} \sim N(f_t, Q_t)$ where

$$\begin{aligned} \mathbf{m}_t &= G_t\mathbf{m}_{t-1} + \mathbf{A}_t(y_t - f_t) \\ C_t &= G_t C_{t-1} G_t^T + W_t - \mathbf{A}_t Q_t \mathbf{A}_t^T \end{aligned}$$

where $f_t = \mathbf{F}_t G_t \mathbf{m}_{t-1}$,&

$$\begin{aligned} \mathbf{A}_t &= [G_t C_{t-1} G_t^T + W_t]\mathbf{F}_t^T Q_t^{-1} \\ Q_t &= \mathbf{F}_t\left(C_t + \mathbf{A}_t W_t \mathbf{A}_t^T\right)\mathbf{F}_t^T + V_t \end{aligned}$$

Multiregression Dynamic Models (Queen & S,1993)

1. Structured so different components can be estimated & scored independently in closed form (conditional on hyperparameters): so **modular** like its non dynamic twin.
2. Univariate RDLM transferred seamlessly.

Simple example of typical LMDM on a 4 - vector time series $\mathbf{Y}_t = (Y_t(1), Y_t(2), Y_t(3), Y_t(4)), t = 1, 2, \ldots$Denote by

$$\begin{array}{ccc} Y_t(1) & \rightarrow & Y_t(3) \\ & \nearrow & \downarrow \\ Y_t(2) & \rightarrow & Y_t(4) \end{array}$$

- Graph says $\{Y_t(1)\}_{t\geq 1}$, $\{Y_t(2)\}_{t\geq 1}$ independent DLMs (no edge).
- $\{Y_t(3)\}_{t\geq 1}$ a RDLM $\{\mathbf{F}_t(\mathbf{y}^t(1), \mathbf{y}^t(2)), G_t, V_t, W_t\}$ but $\mathbf{F}_t(\mathbf{y}^t(1), \mathbf{y}^t(2))$ fn. of $Y_t(1), Y_t(2)$. (edges!)
- $\{Y_t(4)\}_{t\geq 1}$ a RDLM $\{\mathbf{F}_t(\mathbf{y}^t(2), \mathbf{y}^t(3)), G_t, V_t, W_t\}$ but $\mathbf{F}_t(\mathbf{y}^t(2), \mathbf{y}^t(3))$ fn. of $Y_t(2)$, $Y_t(3)$,(not $\{Y_t(1)\}_{t\geq 1}$).

Useful Properties of MDMs

Processes $\{\boldsymbol{\theta}_t(i)\}_{t\geq 0}$ $i = 1, 2, 3, 4$ a priori $\amalg \Rightarrow$ with ancestral data each

1. $\{\boldsymbol{\theta}_t(1)|\mathbf{y}^{t-1}, \boldsymbol{\theta}_t(2)|\mathbf{y}^{t-1}, \boldsymbol{\theta}_t(3)|\mathbf{y}^{t-1}, \boldsymbol{\theta}_t(4)|\mathbf{y}^{t-1}\}$ $\amalg$ $t = 1, 2, \ldots$.
2. $\{\boldsymbol{\theta}_t(1)|\mathbf{y}^{t}, \boldsymbol{\theta}_t(2)|\mathbf{y}^{t}, \boldsymbol{\theta}_t(3)|\mathbf{y}^{t}, \boldsymbol{\theta}_t(4)|\mathbf{y}^{t}\}$ $\amalg$ $t = 1, 2, \ldots$.
3. $\{\boldsymbol{\theta}_t(i)\}_{t\geq 1}$ $i = 1, 2, 3, 4$ - & their observation variances can be updated given $\{\mathbf{F}_t(\mathbf{y}^t(2), \mathbf{y}^t(3)), G_t, W_t/V_t\}$.

Note $\{\mathbf{F}_t(\mathbf{y}^t(2), \mathbf{y}^t(3)), G_t\}$ usually known Estimate W_t/V_t e.g maximise marginal likelihood numerically over a subspace.

- MDM is **not multivariate Gaussian** $\Rightarrow$ Marginal 1 step ahead distributions on terminal vertices (our attribute vectors) non-Gaussian because conditional 1 step ahead variances depend on values of parents.