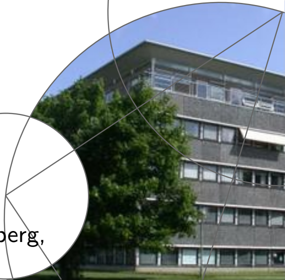
Conditional independence in max-linear Bayesian networks [1]

Steffen Lauritzen
Department of Mathematical Sciences

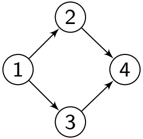
Munich October 2019



[1] Based on joint work with C. Améndola, C. Klüppelberg, and N. Tran — in progress.

Structural equation models

Consider a *directed acyclic graph* (DAG) $\mathcal{D} = (V, E)$:



Each node $v \in V$ represents a *random variable* X_v.

Joint distribution of $X = (X_1, X_2, X_3, X_4)$ is determined by a system of *structural equations*

$$\begin{aligned} X_1 &= \phi_1(Z_1) \\ X_2 &= \phi_2(X_1, Z_2) \\ X_3 &= \phi_3(X_1, Z_3) \\ X_4 &= \phi_4(X_2, X_3, Z_4) \end{aligned}$$

where Z_1, Z_2, Z_3, Z_4 are independent

Max-linear structural equations

We consider *recursive max-linear structural equation systems*, where $x \vee y = \max(x, y)$:

$$X_v = \bigvee_{u \in \mathrm{pa}(v)} c_{vu} X_u \vee c_{vv} Z_v, \quad v \in V, \tag{1}$$

where now $Z_v, v \in V$ are independent *innovations* with *atom free* distributions having support $\mathbb{R}_+$ and $c_{vu}, u \in \mathrm{pa}(v), c_{vv}$ are *positive structural coefficients*.

For simplicity we assume $c_{vv} = 1$ for all $v \in V$.

Tropical linear algebra

We work in the *max-times semiring* $(\bar{R}_+ = \mathbb{R}_+ \cup 0, \vee, \odot)$ where $\odot$ denotes ordinary multiplication.

A max-linear map from $(\bar{\mathbb{R}}_+^n, \vee, \odot)$ to $(\bar{\mathbb{R}}_+^m, \vee, \odot)$ has the matrix representation $A \in \bar{R}_+^{m \times n}$ and

$$(A \odot x)_i = \bigvee_{j=1}^{n} a_{ij} x_j.$$

If we collect the innovations into the column vector $Z = (Z_1, \ldots, Z_d)^\top$ the equation system becomes

$$X = (C \odot X) \vee Z.$$

Solving the structural equation

The system can also be represented as

$$X = C^* \odot Z$$

where the *idempotent* *Kleene star* matrix of C

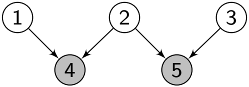
$$C^* = I \vee C \vee C^{\odot 2} \vee \cdots \vee C^{\odot (d-1)}, \qquad (2)$$

Elements C^*_{ij} in the Kleene star matrix is the *maximal* weight (product of coefficients) of a dipath from j to i. A dipath that attains this weight is a *critical dipath*

The equation $X = C^* \odot Z$ implies that *the joint distribution of Z is completely determined by the critical dipaths.*

Thus edges which do not form part of a critical path are redundant and can be removed.

Cassiopeia (an example)

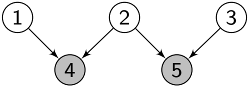


Surprisingly, it holds for any MLBN that $1 \perp\!\!\!\perp 3 \mid \{4,5\}$ whereas $\neg(1 \perp_{\mathcal{D}} 3 \mid \{4,5\})$!

To see this, let all coefficients equal to 1. Then the conditional distribution of (X_1, X_2, X_3) given $(X_4, X_5) = (x_4, x_5)$ is determined by imposing the following inequalities on (X_1, X_2, X_3):

$$\max(X_1, X_2) \leq x_4, \quad \max(X_2, X_3) \leq x_5.$$

Example continued



$$\max(X_1, X_2) \leq x_4, \quad \max(X_2, X_3) \leq x_5.$$

Now distinguish three cases:

1. If $x_4 < x_5$, the condition is equivalent to

$$\max(X_1, X_2) \leq x_4, \quad X_3 \leq x_5.$$

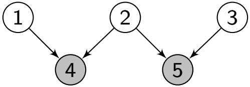
2. If $x_4 > x_5$, the condition is equivalent to

$$\max(X_1) \leq x_4, \quad \max(X_2, X_3) \leq x_5.$$

3. If $x_4 = x_5$ we must have $X_2 = x_4 = x_5$ and hence equiv

$$X_1 \leq x_4, \quad X_3 \leq x_5, \quad X_2 = x_4 = x_5.$$

Example continued



In all three (contexts) we have independence of X_1 and X_3 because they do not occur together in the same inequality

1. If $x_4 < x_5$, *x_5 cannot be caused by X_2 but only by X_3 or Z_5*.
2. If $x_4 > x_5$, *x_4 cannot be caused by X_2 but only by X_1 or Z_4*.
3. If $x_4 = x_5$ *both must be caused by X_2.*

So the key is to identify causes.

Representing conditional distributions

Key is representation of conditional distribution of $X_L \mid X_K = x_K$ as

$$X_L = C^*_{LK} \odot x_K \vee C^*_{LL} \odot Z_L$$

with conditional distribution of Z_L given $X_K = x_K$ *determined by restriction*

$$x_K \geq C^*_{KL} \odot Z_L$$

and *removing redundant terms* (as in the "Cassiopeia" example) to obtain a *compact representation*

First we need to go closer into the structure of max-linear Bayesian networks.

The impact graph

Definition

Let $\mathcal{D} = (V, E)$ be a DAG and C be a coefficient matrix supported by $\mathcal{D}$. The *impact graph* is a random graph $G = G(Z)$ on V consisting of the following edges:

$$j \to i \iff X_i = c^*_{ij} Z_j$$

and we let $\mathcal{E}(g) = \{z : G(z) = g\}$.

Note that with probability one *it holds that $G(Z)$ is a forest*: each node has at most one parent because the distributions of Z are atom free.

We say that $j \to i$ in g means that X_i i *realized* by Z_j. Recall

$$X_i = \bigvee_j c^*_{ij} Z_j.$$

Impact exchange matrix

Before we give a full characterization of the possible impact graphs, we need

Definition

Consider a DAG $\mathcal{D}$ with coefficient matrix C and Kleene star C^* and let g be a forest with root set $R = R(g)$. The *impact exchange matrix* $M = M(g) = M(g, C^*)$ of g with respect to C^* is an $R \times R$ matrix with entries defined by $m_{rr} = 0$ for all $r \in R$, and for $r \neq r'$:

$$m_{rr'} := \max_{i \in \mathrm{ch}_g(r)} \frac{c^*_{ir'}}{c^*_{ir}}. \tag{3}$$

Here $\mathrm{ch}_g(r)$ denotes the children of r in g.

Structure of impact graphs

Theorem

Consider a max-linear BN with Kleene star C^*. It then holds that $P(\mathcal{E}(g)) > 0$ if and only if the following four conditions hold:

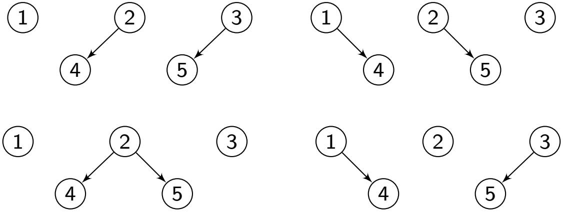
- a *g is a subgraph of $\mathcal{D}^* = \mathcal{D}(C^*)$*
- b *g is a galaxy, i.e. a forest of stars*
- c *If $j \to i$ in g and $c^*_{ij} = c^*_{ik} c^*_{kj}$ then $k \not\to i$ and $j \to k$ in g.*
- d *$\lambda(M(g)) < 1$.*

Here a *star* is a tree of height at most one, and and $\lambda(M(g))$ denotes the *max-times tropical eigenvalue* of $M(g)$.

This theorem gives a *complete control of how extreme events (corresponding to roots in g) spread deterministically* to other parts of the network.

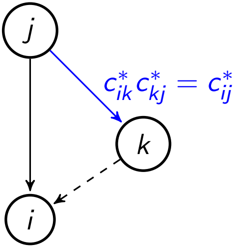
Impact graphs for Cassiopeia

For the Cassiopeia example the possible impact graphs are: the empty graph, all subgraphs with a single edge, and the following four subgraphs with two edges:



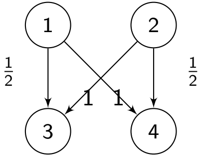
The triangle condition (c) and eigenvalue condition (d) are not relevant for this example.

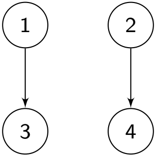
Illustration of the triangle condition



If the innovation at j is responsible for the value at i it must also be responsible for the value at k, indicated in blue.

Illustration of the eigenvalue condition





The graph to the right cannot be an impact graph for the max-linear BN to the left because

$$M(g) = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$$

and then $\lambda(M(g)) = 2 > 1$.

For if $1 \to 3$ we must have $\frac{1}{2}X_1 > X_2$ and since $2 \to 4$ we must also have $\frac{1}{2}X_2 > X_1$ implying $X_2 > 4X_2$ which is not possible.

Piecewise linearity

A max-linear map is *piecewise linear* and the the impact graph index the linear pieces.

Indeed the pieces are the maps $L_g : \mathbb{R}_+^{R(g)} \to R_+^V$ given as

$$L_g(z)_r = z_r, r \in R(g); \quad L_g(z)_i = c^*_{ir} z_r \text{ iff } r \to i \text{ in } g$$

and $L_g(z)_i = 0$ otherwise. So we actually have

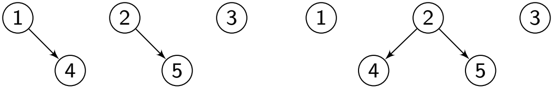
$$X = C^* \odot Z \stackrel{\text{a.s.}}{=} L_{G(Z)}(Z).$$

This insight is important for the next concept.

Impact graph compatible with $\{X_K = x_K\}$

We say an impact graph g is *compatible* with $\{X_K = x_K\}$ if

- x_K is in the image of the map $\Pi_K \circ L_g$;
- $\Pi_K \circ L_g$ has minimal rank among those who satisfy the image condition



In the Cassiopeia DAG the impact graph to the left is compatible with an event of the form $X_{\{4,5\}} = (x_4, x_5)$ if $x_4 > x_5$. Because then

$$x_4 = x_1 \vee x_2 = x_1 > x_2 = x_2 \vee x_3 = x_5$$

whereas it is only compatible with the graph to the right if $x_4 = x_5$ since the latter has rank 1 and the former rank 2.

The source graph

Whereas the impact graph describes the way extreme events spread in the network, the *source graph* $\mathcal{C}(X_K = x_K)$ tracks the possible sources for a given event $X_K = x_K$.

We abstain from giving the details of how to construct the source graph, but it involves forming the *total impact graph* $\mathcal{I}(X_K = x_K)$ compatible with $\{X_k = x_K\}$:

$$\mathcal{I}(X_K = x_K) = \bigcup_{g \in \mathfrak{G}(X_K = x_K)} g$$

where $\mathfrak{G}(X_K = x_K)$ denotes the set of impact graphs that are compatible with $\{X_K = x_K\}$.

Subsequently we identify *redundant nodes, redundant edges*, and remove them from the total impact graph.

Eventually, the source graph yields a *reduced representation* of the conditional distribution given $X_K = x_K$.

More auxiliary DAGs

Definition

The *conditional reachability DAG* $\mathcal{D}^*_K$ is the graph on V defined as: $j \to i \in \mathcal{D}^*_K$ if and only if there exists a directed path from j to i that circumvents K.

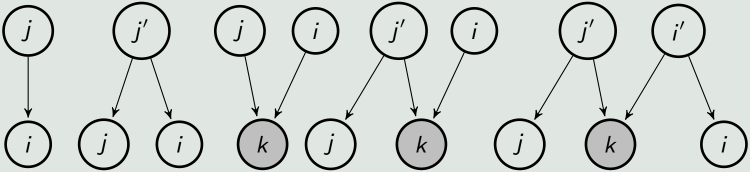
Definition

Let $\mathcal{D} = (V, E)$ be a DAG and C be a coefficient matrix supported by $\mathcal{D}$. The *critical DAG* $\mathcal{D}^*_K(C)$ is the graph on V defined as: $j \to i \in \mathcal{D}^*_K(C)$ iff $c^*_{ij} > 0$ and all critical directed paths π from j to i circumvent K.

$*$-separation

Definition

A path π between j and i in a DAG is *$*$-connecting* relative to K if and only if is one of the following, where shaded nodes are in K



A path that is not $*$-connecting relative to K is said to be *$*$-blocked* by K. We also say that I and J are *$*$-separated by K* if all paths between I and J are $*$-blocked and we then write $I \perp_* J \mid K$.

Context free conditional independence

Our main conditional are the following three theorems

Theorem (Independent of C)

Let $\mathcal{D} = (V, E)$ be a directed acyclic graph. If X follows a recursive max-linear model with support $\mathcal{D}$, then for all mutually disjoint $I, J, K \subseteq V$,

$$X_I \perp\!\!\!\perp X_J \,|\, X_K \text{ for all } C \text{ adapted to } \mathcal{D} \iff I \perp_* J \,|\, K \text{ in } \mathcal{D}^*_K.$$

Theorem (Context-free for given C)

Let $\mathcal{D} = (V, E)$ be a directed acyclic graph and C a fixed coefficient matrix with support $\mathcal{D}$. If X follows a recursive max-linear model with coefficient matrix C, then for all mutually disjoint $I, J, K \subseteq V$,

$$X_I \perp\!\!\!\perp X_J \,|\, X_K \iff I \perp_* J \,|\, K \text{ in } \mathcal{D}^*_K(C).$$

Context-dependent CI

Third major CI result:

Theorem (Context-dependent)

Let $\mathcal{D} = (V, E)$ be a directed acyclic graph and C a fixed coefficient matrix with support $\mathcal{D}$. Let X follow a recursive max-linear model with coefficient matrix C. Let $K \subseteq V$ and $\mathcal{C}(X_K = x_K)$ be the source graph of the event $\{X_K = x_K\}$. For all mutually disjoint subsets $I, J, K \subseteq V$

$$X_I \perp\!\!\!\perp X_J \,|\, X_K = x_K \iff I \perp_* J \,|\, K \text{ in } \mathcal{C}(X_K = x_K).$$

Summary and conclusion

- The *impact graph* describes exactly how extreme events spread deterministically;
- The *impact graph compatible with* $\{X_K = x_K\}$ does the same in the given context $\{X_K = x_K\}$
- The *source graph* identifies potential sources in the given context and provides a compact representation of the conditional distribution;
- As a consequence, basic CI theorems follow, based on resp. the *DAG alone,* the specific *coefficient matrix* C, or the latter in conjunction with the *context* $\{X_K = x_K\}$

Watch this space: there is more to come...