



Faculty of Science

Max-linear Bayesian networks

Based on joint work with Carlos Améndola, Nadine Gissibl, Claudia Klüppelberg, and Ngoc Tran — in progress.

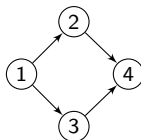
Steffen Lauritzen

Department of Mathematical Sciences



Structural equation models

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

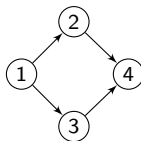


Each node $v \in V$ represents a *random variable* X_v .



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*

$$X_1 = \phi_1(Z_1)$$

$$X_2 = \phi_2(X_1, Z_2)$$

$$X_3 = \phi_3(X_1, Z_3)$$

$$X_4 = \phi_4(X_2, X_3, Z_4)$$

where Z_1, Z_2, Z_3, Z_4 are independent



Linear structural equations

A (recursive) linear structural equation system

$$X_v = \sum_{u \in \text{pa}(v)} c_{vu} X_u + c_{vv} Z_v, \quad v \in V, \quad (1)$$

where $Z_v, v \in V$ are independent noise variables and $c_{vu}, u \in \text{pa}(v), c_{vv}$ are *structural coefficients*.



Linear structural equations

A (recursive) linear structural equation system

$$X_v = \sum_{u \in \text{pa}(v)} c_{vu} X_u + c_{vv} Z_v, \quad v \in V, \quad (1)$$

where $Z_v, v \in V$ are independent noise variables and $c_{vu}, u \in \text{pa}(v), c_{vv}$ are *structural coefficients*.

For studying dependence among extreme events in a network, it could make sense to replace sum with maximum

$$X_v = \bigvee_{u \in \text{pa}(v)} c_{vu} X_u \vee c_{vv} Z_v$$



Linear structural equations

A (recursive) linear structural equation system

$$X_v = \sum_{u \in \text{pa}(v)} c_{vu} X_u + c_{vv} Z_v, \quad v \in V, \quad (1)$$

where $Z_v, v \in V$ are independent noise variables and $c_{vu}, u \in \text{pa}(v), c_{vv}$ are *structural coefficients*.

For studying dependence among extreme events in a network, it could make sense to replace sum with maximum

$$X_v = \bigvee_{u \in \text{pa}(v)} c_{vu} X_u \vee c_{vv} Z_v$$

Z 's are *innovations*. Noise ignored for this lecture.



Max-linear structural equations

We consider *recursive max-linear structural equation systems*, replacing $+$ in (1) with $\vee$, where $x \vee y = \max(x, y)$:

$$X_v = \bigvee_{u \in \text{pa}(v)} c_{vu} X_u \vee c_{vv} Z_v, \quad v \in V, \quad (2)$$

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



Max-linear structural equations

We consider *recursive max-linear structural equation systems*, replacing $+$ in (1) with $\vee$, where $x \vee y = \max(x, y)$:

$$X_v = \bigvee_{u \in \text{pa}(v)} c_{vu} X_u \vee c_{vv} Z_v, \quad v \in V, \quad (2)$$

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

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



Max-linear structural equations

We consider *recursive max-linear structural equation systems*, replacing $+$ in (1) with $\vee$, where $x \vee y = \max(x, y)$:

$$X_v = \bigvee_{u \in \text{pa}(v)} c_{vu} X_u \vee c_{vv} Z_v, \quad v \in V, \quad (2)$$

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

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

Models defined and studied in Gissibl and Klüppelberg (2018); Gissibl (2018); see also Klüppelberg and Lauritzen (2019).



A very simple example

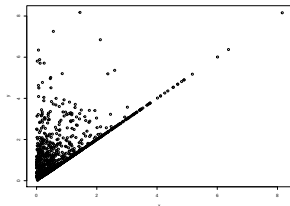
*Max-linear systems generate distributions without densities
w.r.t. a product measure:*



A very simple example

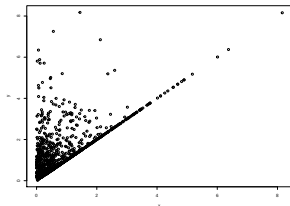
Max-linear systems generate distributions without densities w.r.t. a product measure:

$$\textcircled{X} \longrightarrow \textcircled{Y} \quad Y = cX \vee Z$$



A very simple example

Max-linear systems generate distributions without densities w.r.t. a product measure:

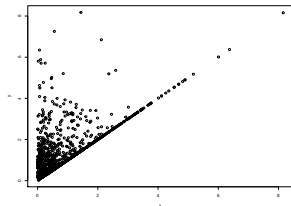


Easy in principle to find (geometrically) consistent estimates of graph, direction of edge, and coefficient c (Gissibl, 2018; Gissibl et al., 2019). Also works for more general networks.



A very simple example

Max-linear systems generate distributions without densities w.r.t. a product measure:



Easy in principle to find (geometrically) consistent estimates of graph, direction of edge, and coefficient c (Gissibl, 2018; Gissibl et al., 2019). Also works for more general networks.

Improvement using methods of Markus Reiss?



Markov properties of Bayesian networks

A joint distribution P satisfies the *well-ordered Markov property* (O) w.r.t. $\mathcal{D}$ if all variables are conditionally independent of their predecessors given their parents

$$i \perp\!\!\!\perp_P pr(i) \mid pa(i)$$

for all $i \in V = \{1, \dots, d\}$ and any well-ordering of V .



Markov properties of Bayesian networks

A joint distribution P satisfies the *well-ordered Markov property* (O) w.r.t. $\mathcal{D}$ if all variables are conditionally independent of their predecessors given their parents

$$i \perp\!\!\!\perp_P pr(i) \mid pa(i)$$

for all $i \in V = \{1, \dots, d\}$ and any well-ordering of V .

P obeys the *local Markov property* (L) w.r.t. $\mathcal{D}$ if every variable is conditionally independent of its non-descendants, given its parents:

$$v \perp\!\!\!\perp_P nd(v) \mid pa(v).$$



Markov properties of Bayesian networks

A joint distribution P satisfies the *well-ordered Markov property* (O) w.r.t. $\mathcal{D}$ if all variables are conditionally independent of their predecessors given their parents

$$i \perp\!\!\!\perp_P pr(i) \mid pa(i)$$

for all $i \in V = \{1, \dots, d\}$ and any well-ordering of V .

P obeys the *local Markov property* (L) w.r.t. $\mathcal{D}$ if every variable is conditionally independent of its non-descendants, given its parents:

$$v \perp\!\!\!\perp_P nd(v) \mid pa(v).$$

P satisfies the *global Markov property* (G) w.r.t. $\mathcal{D}$ if

$$A \perp_{\mathcal{D}} B \mid C \implies A \perp\!\!\!\perp_P B \mid C.$$

where $\perp_{\mathcal{D}}$ denotes *d-separation*.



Equivalence of Markov properties

We have

Theorem (Lauritzen et al. (1990))

Let $\mathcal{D}$ be a directed acyclic graph with $V = \{1, \dots, d\}$ well-ordered and P a probability measure on $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$. Then we have

$$(O) \iff (L) \iff (G).$$



Equivalence of Markov properties

We have

Theorem (Lauritzen et al. (1990))

Let $\mathcal{D}$ be a directed acyclic graph with $V = \{1, \dots, d\}$ well-ordered and P a probability measure on $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$. Then we have

$$(O) \iff (L) \iff (G).$$

Proof does not involve densities, so holds even though densities do not exist.



Equivalence of Markov properties

We have

Theorem (Lauritzen et al. (1990))

Let $\mathcal{D}$ be a directed acyclic graph with $V = \{1, \dots, d\}$ well-ordered and P a probability measure on $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$. Then we have

$$(O) \iff (L) \iff (G).$$

Proof does not involve densities, so holds even though densities do not exist.

It follows directly from the structural equation system that the joint distribution P satisfies the *well-ordered Markov property* (O) w.r.t. $\mathcal{D}$, hence all of them.



Faithfulness

The global Markov property gives a *sufficient* condition for conditional independence in terms of d -separation. Another central concept is that of faithfulness:

Definition

A probability distribution P on $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$ is said to be *faithful* to a DAG $\mathcal{D}$ if

$$A \perp_{\mathcal{D}} B \mid C \iff A \perp\!\!\!\perp_P B \mid C.$$

In other words, if d -separation is also necessary for conditional independence.



Faithfulness

The global Markov property gives a *sufficient* condition for conditional independence in terms of d -separation. Another central concept is that of faithfulness:

Definition

A probability distribution P on $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$ is said to be *faithful* to a DAG $\mathcal{D}$ if

$$A \perp_{\mathcal{D}} B \mid C \iff A \perp\!\!\!\perp_P B \mid C.$$

In other words, if d -separation is also necessary for conditional independence.

Typically, most probability distributions are faithful if they are Markov (Meek, 1995), but this is *not the case for max-linear Bayesian networks!*



Tropical linear algebra

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



Tropical linear algebra

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

Using the map $x \rightarrow \log x$ this is algebraically isomorphic to the *tropical semiring* $(\mathbb{R} \cup -\infty, \oplus, \otimes)$ where

$$a \oplus b = \max(a, b), \quad a \otimes b = a + b.$$



Tropical linear algebra

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

Using the map $x \rightarrow \log x$ this is algebraically isomorphic to the *tropical semiring* $(\mathbb{R} \cup -\infty, \oplus, \otimes)$ where

$$a \oplus b = \max(a, b), \quad a \otimes b = a + b.$$

We shall *stay in the max-times semiring, but use terminology from tropical linear algebra.*



Tropical linear algebra

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

Using the map $x \rightarrow \log x$ this is algebraically isomorphic to the *tropical semiring* $(\mathbb{R} \cup -\infty, \oplus, \otimes)$ where

$$a \oplus b = \max(a, b), \quad a \otimes b = a + b.$$

We shall *stay in the max-times semiring, but use terminology from tropical linear algebra.*

Extend to *semimodule* $(\bar{\mathbb{R}}_+^k, \vee, \odot)$ in the usual way so a max-linear map from $(\bar{\mathbb{R}}_+^n, \vee, \odot)$ to $(\bar{\mathbb{R}}_+^m, \vee, \odot)$ has the matrix representation $A \in \bar{\mathbb{R}}_+^{m \times n}$ and

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



Tropical linear algebra

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

Using the map $x \rightarrow \log x$ this is algebraically isomorphic to the *tropical semiring* $(\mathbb{R} \cup -\infty, \oplus, \otimes)$ where

$$a \oplus b = \max(a, b), \quad a \otimes b = a + b.$$

We shall *stay in the max-times semiring, but use terminology from tropical linear algebra.*

Extend to *semimodule* $(\bar{\mathbb{R}}_+^k, \vee, \odot)$ in the usual way so a max-linear map from $(\bar{\mathbb{R}}_+^n, \vee, \odot)$ to $(\bar{\mathbb{R}}_+^m, \vee, \odot)$ has the matrix representation $A \in \bar{\mathbb{R}}_+^{m \times n}$ and

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

Eigenvalues, eigenspaces etc., but results slightly different.



Matrix multiplication

As usual, composition of max-linear maps correspond to matrix multiplication:

$$(F \odot G)_{vu} = \left(\bigvee_{k=1}^n f_{vk} g_{ku} \right). \quad (3)$$

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

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



Matrix multiplication

As usual, composition of max-linear maps correspond to matrix multiplication:

$$(F \odot G)_{vu} = \left(\bigvee_{k=1}^n f_{vk} g_{ku} \right). \quad (3)$$

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

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

The system can also be represented as

$$X = B \odot Z$$

where the *idempotent max-linear coefficient matrix* B can be found as the *Kleene star* of C

$$B = C^* = I \vee C \vee C^{\odot 2} \vee \dots \vee C^{\odot (d-1)} = \lim_{k \rightarrow \infty} C^{\odot k}, \quad (4)$$



Critical paths

The element B_{ij} in the Kleene star is the maximal weight (product of coefficients) of a dipath from j to i . A dipath that attains this weight is a *critical dipath*



Critical paths

The element B_{ij} in the Kleene star 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 $X = B \odot Z$ implies that *the joint distribution of Z is completely determined by the critical paths.*



Critical paths

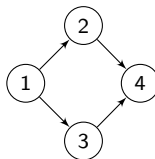
The element B_{ij} in the Kleene star is the maximal weight (product of coefficients) of a dipath from j to i . A dipath that attains this weight is a *critical dipath*

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

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



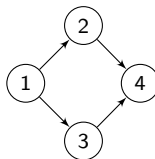
A simple example



The paths $[1 \rightarrow 2]$, $[1 \rightarrow 3]$, $[2 \rightarrow 4]$, and $[3 \rightarrow 4]$ are all critical as they are the only directed paths between their endpoints.



A simple example



The paths $[1 \rightarrow 2]$, $[1 \rightarrow 3]$, $[2 \rightarrow 4]$, and $[3 \rightarrow 4]$ are all critical as they are the only directed paths between their endpoints.

If we assume $c_{42}c_{21} > c_{43}c_{31}$, the path $[1 \rightarrow 2 \rightarrow 4]$ is critical whereas $[1 \rightarrow 3 \rightarrow 4]$ is not.



Example continued

The max-linear coefficient matrix becoms

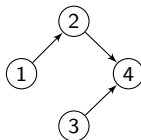
$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ c_{21} & 1 & 0 & 0 \\ c_{31} & 0 & 1 & 0 \\ c_{42}c_{21} & c_{42} & c_{43} & 1 \end{pmatrix}$$

since $\max(c_{42}c_{21}, c_{43}c_{31}) = c_{42}c_{21}$.



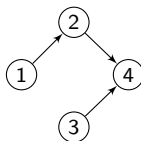
Example continued

If the path $[1 \rightarrow 2 \rightarrow 4]$ is critical, we can consider the subdag $\tilde{\mathcal{D}}$ obtained from $\mathcal{D}$ by removing the edge $1 \rightarrow 3$:



Example continued

If the path $[1 \rightarrow 2 \rightarrow 4]$ is critical, we can consider the subdag $\tilde{\mathcal{D}}$ obtained from $\mathcal{D}$ by removing the edge $1 \rightarrow 3$:



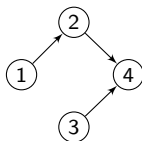
The new max-linear coefficient matrix becomes

$$\tilde{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ c_{21} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ c_{42}c_{21} & c_{42} & c_{43} & 1 \end{pmatrix}$$



Example continued

If the path $[1 \rightarrow 2 \rightarrow 4]$ is critical, we can consider the subdag $\tilde{\mathcal{D}}$ obtained from $\mathcal{D}$ by removing the edge $1 \rightarrow 3$:



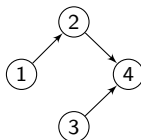
The new max-linear coefficient matrix becomes

$$\tilde{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ c_{21} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ c_{42}c_{21} & c_{42} & c_{43} & 1 \end{pmatrix}$$

From the relation $X = B \odot Z$ we realize that (X_1, X_2, X_4) *has the same joint distribution in the model determined by $\mathcal{D}$ as it has in the model determined by $\tilde{\mathcal{D}}$.*



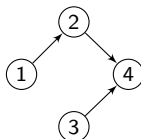
Example continued



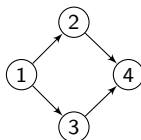
But we have $1 \perp_{\tilde{\mathcal{D}}} 4 \mid 2$, so the global Markov property implies $X_1 \perp\!\!\!\perp X_4 \mid X_2$ in the model determined by $\tilde{\mathcal{D}}$ and hence also by $\mathcal{D}$.



Example continued



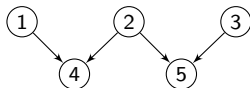
But we have $1 \perp_{\tilde{\mathcal{D}}} 4 \mid 2$, so the global Markov property implies $X_1 \perp\!\!\!\perp X_4 \mid X_2$ in the model determined by $\tilde{\mathcal{D}}$ and hence also by $\mathcal{D}$.



But since $\neg(1 \perp_{\mathcal{D}} 4 \mid 2)$, *the distribution is not faithful to $\mathcal{D}$* .



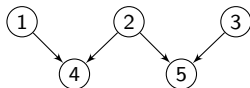
Another example



Here it holds that $1 \perp\!\!\!\perp 3 \mid \{4, 5\}$!



Another example



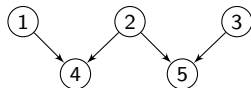
Here it holds that $1 \perp\!\!\!\perp 3 \mid \{4, 5\}$!

Assume 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



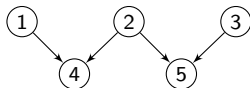
Now distinguish three cases:

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

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



Example continued



Now distinguish three cases:

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

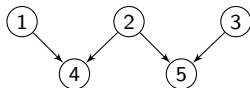
$$\max(X_1, X_2) \leq x_4, \quad X_3 \leq x_5.$$

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

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



Example continued



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

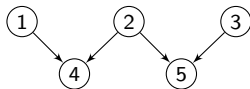
$$\max(X_1) \leq x_4, \quad \max(X_2, X_3) \leq x_5.$$

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

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



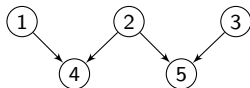
Example continued



In all three (contexts) we have independence of X_1 and X_3



Example continued

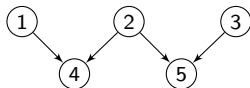


In all three (contexts) we have independence of X_1 and X_3

- 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 .



Example continued



In all three (contexts) we have independence of X_1 and X_3

- 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.



A conjecture

We define a notion of **-connecting paths* by only taking into account critical paths and allowing at most one collider on the path.



A conjecture

We define a notion of **-connecting paths* by only taking into account critical paths and allowing at most one collider on the path.

Then *-separation accordingly.



A conjecture

We define a notion of **-connecting paths* by only taking into account critical paths and allowing at most one collider on the path.

Then *-separation accordingly.

Now I believe that **-separation is sound and complete for independence in max-linear Bayesian networks*, and that *any generic max-linear BN is faithful to *-separation*.



A conjecture

We define a notion of **-connecting paths* by only taking into account critical paths and allowing at most one collider on the path.

Then *-separation accordingly.

Now I believe that **-separation is sound and complete for independence in max-linear Bayesian networks*, and that *any generic max-linear BN is faithful to *-separation*.

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

$$X_L = B_{LK} \odot x_K \vee B_{LL} \odot Z_L$$

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

$$x_K \geq B_{KL} \odot Z_L$$

and *removing redundant terms*



HAPPY BIRTHDAY, AAD!



References

- Gissibl, N. (2018). *Graphical Modeling of Extremes: Max-linear Models on Directed Acyclic Graphs*. PhD thesis, Technical University of Munich.
- Gissibl, N. and Klüppelberg, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli*, 24:2693–2720.
- Gissibl, N., Klüppelberg, C., and Lauritzen, S. L. (2019). Identifiability and estimation of recursive max-linear models. arXiv:1901.03556.
- Klüppelberg, C. and Lauritzen, S. (2019). Max-linear models for Bayesian networks. arXiv:1901.03948.
- Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks*, 20:491–505.
- Meek, C. (1995). Strong completeness and faithfulness in Bayesian networks. In Besnard, P. and Hanks, S., editors, *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pages 411–418. Morgan Kaufman Publishers, San Francisco, CA.

