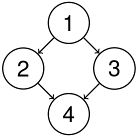
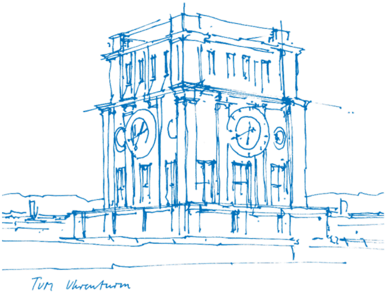
Introduction to max-linear models and tropical linear algebra

Carlos Enrique Améndola Cerón
with Claudia Klüppelberg, Steffen Lauritzen & Ngoc Tran

Technische Universität München

Graphical Models: CIAS Workshop
Garching bei München; October 23, 2019





Outline

Graphical Models

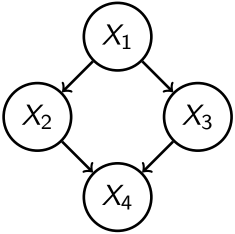
- Important family of statistical models that represent interaction structures between random variables.
- Consider $\mathcal{D} = (V, E)$ a directed acyclic graph (DAG).
- Each node $v \in V$ represents a random variable X_v.
- Edges $u \to v$ encode (conditional) dependence structure.
- The parents of $v \in V$ are $\text{pa}(v) = \{u \in V | u \to v\}$.
- The distribution $p(x)$ factors according to the graph $\mathcal{D}$:

$$p(x) = \prod_{v \in V} p(x_v | x_{pa(v)}).$$

- Usually nodes represent discrete or Gaussian distributions.

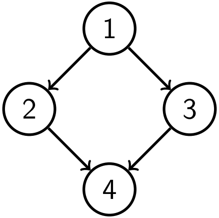
Directed Acyclic Graphs

- A *walk* from u to v of *length* n is a sequence of vertices $[u = u_0, u_1, \dots, u_n = v]$ so that $u_{i-1} \sim u_i$ for all $i = 1, \dots, n$.
- A walk is a *cycle* if $u = v$. A *path* is a walk with no repeated vertices.
- The walk/path is *directed* from u to v if $u_{i-1} \to u_i$ for all i.
- If all edges in a graph $\mathcal{D} = (V, E)$ are directed, $\mathcal{D}$ is a *directed graph*.
- A directed graph is *acyclic* if it has no directed cycles.
- Example. $V = \{1, 2, 3, 4\}$, $E = \{(1,2), (1,3), (2,4), (3,4)\}$



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

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

(*structural equations*), where Z_1, Z_2, Z_3, Z_4 are independent.

Linear structural equations

A (recursive) linear structural equation system

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

where $Z_v, v \in V$ are independent noise variables and all the c_{vu} with $u \in \mathrm{pa}(v)$ as well as the c_{vv} are *structural coefficients*.

For studying dependence among extreme events in a network, replace sum with *maximum*

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

where $x \vee y = \max(x, y)$.

Max-linear structural equations

- Setting is now *recursive max-linear structural equation systems*,

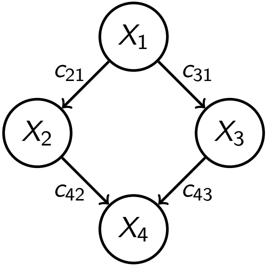
$$X_v = \bigvee_{u \in \mathrm{pa}(v)} c_{vu} X_u \vee c_{vv} Z_v, \quad v \in V,$$

- $Z_v, v \in V$ are independent *innovations* with *atom free* distributions having support $\mathbb{R}_+$
- $c_{vu}, u \in \mathrm{pa}(v), c_{vv}$ are *positive structural coefficients*, for simplicity we assume $c_{vv} = 1$ for all $v \in V$.
- Models defined and studied in [Gissibl and Klüppelberg (2018)] and [Klüppelberg and Lauritzen (2019)].

Example of recursive max-linear model

For $Z_1, \ldots, Z_d > 0$ independent, continuous, unbounded support, and edge weights $c_{ik} > 0$, we define the **recursive max-linear model**

$$X_i := \bigvee_{k \in \mathrm{pa}(i)} c_{ik} X_k \vee Z_i \quad i = 1, \ldots, d$$



$$\begin{aligned}
X_1 &= Z_1 \\
X_2 &= c_{21} X_1 \vee Z_2 = c_{21} Z_1 \vee Z_2 \\
X_3 &= c_{31} X_1 \vee Z_3 = c_{31} Z_1 \vee Z_3 \\
X_4 &= c_{42} X_2 \vee c_{43} X_3 \vee Z_4 \\
&= c_{42}(c_{21} Z_1 \vee Z_2) \vee c_{43}(c_{31} Z_1 \vee Z_3) \vee Z_4 \\
&= (c_{21} c_{42} \vee c_{31} c_{43}) Z_1 \vee c_{42} Z_2 \vee c_{43} Z_3 \vee Z_4
\end{aligned}$$

Asadi, P., Davison, A.C. and Engelke, S. (2015) Extremes on river networks

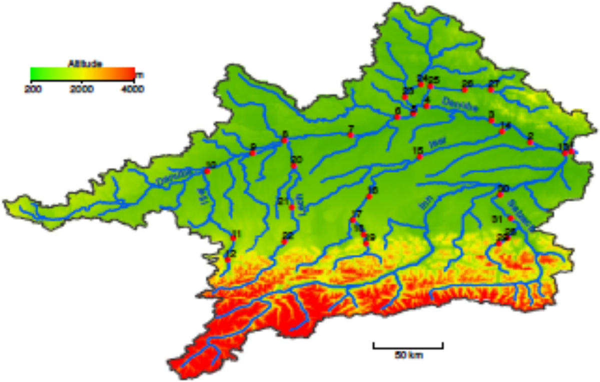
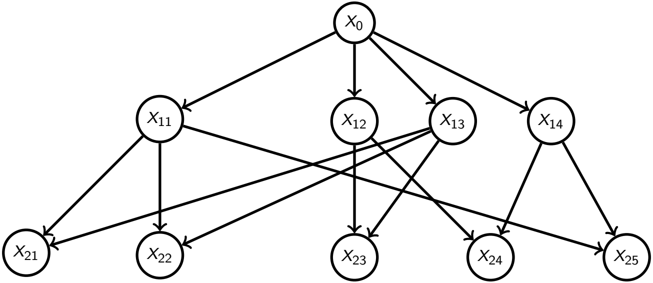


FIGURE 1. Topographic map of the upper Danube basin, showing sites of 31 gauging stations (red blobs) along the Danube and its tributaries. Water flows broadly from left to right.

Einmahl, Kiriliouk and Segers (2016) A continuous updating weighted least squares estimator of tail dependence in high dimensions.



X_0 (EURO STOXX 50),
$X_{11}, X_{12}, X_{13}, X_{14}$ (chemical industry, insurance, DAX, CAC40),
$X_{21}, X_{22}, X_{23}, X_{24}, X_{25}$ (Bayer, BASF, Allianz, Axa, Airliquide)

Markov properties

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

$$i \perp\!\!\!\perp pr(i) \,|\, \mathrm{pa}(i)$$

for all $i \in V = \{1, \ldots d\}$ 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:

$$v \perp\!\!\!\perp \mathrm{nd}(v) \,|\, \mathrm{pa}(v).$$

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

$$A \perp_{\mathcal{D}} B \,|\, C \implies A \perp\!\!\!\perp B \,|\, C.$$

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

Classical d-separation

A path π from i to j is *d-connecting* given $S \subseteq V$ if

1. All colliders v on π satisfy $v \in \mathrm{An}(S)$;
2. All non-colliders v on π satisfy $v \notin S$.

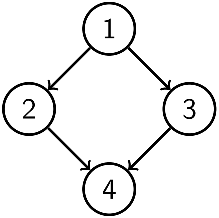
A *collider* has arrows meeting head-to-head $\rightarrow v \leftarrow$

Definition

A is *d-separated from B by S* if and only if there are no d-connecting paths from A to B.

We write $A \perp_{\mathcal{D}} B \mid S$.

A simple example (Diamond)



Here $2 \perp_{\mathcal{D}} 3 \,|\, 1$ but $\neg(2 \perp_{\mathcal{D}} 3 \,|\, 1, 4)$.

The path $[2 \rightarrow 4 \leftarrow 3]$ is *d*-connecting relative to $\{1, 4\}$ but not *d*-connecting relative to $\{1\}$.

Equivalence of Markov properties

We have

Theorem (Lauritzen et al. (1990))

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

$$\text{(O)} \iff \text{(L)} \iff \text{(G)}.$$

It follows 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.
A natural concept is thus the one 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 $\mathcal{D}$ iff

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

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

Typically, *most Markov distributions are faithful* (Meek, 1995);

but (as we'll see) this is *not the case for max-linear Bayesian networks!*

Tropical Arithmetic

- Tropical: $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty\}$.
- Max-plus arithmetic operations:

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

- $(\overline{\mathbb{R}}, \oplus, \otimes, -\infty, 0)$ is the tropical semiring.
- Example of distributivity:

$$2 \odot (3 \oplus 7) = 2 \odot 7 = 9 = 5 \oplus 9 = 2 \odot 3 \ \oplus \ 2 \odot 7$$

- $(\overline{\mathbb{R}}^d, \oplus, \otimes)$ with tropical operations is semimodule over $\overline{\mathbb{R}}$:

$$\lambda \otimes x = (\lambda + x_1, \ldots, \lambda + x_d)$$

for $\lambda \in \overline{\mathbb{R}}$ and $x \in \overline{\mathbb{R}}^d$. Similarly for matrix multiplication in $\overline{\mathbb{R}}^{d \times d}$.

$$\begin{pmatrix} 2 & 3 \\ -1 & 4 \end{pmatrix} \otimes \begin{pmatrix} 5 & 1 \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} 7 \oplus 3 & 3 \oplus 1 \\ 4 \oplus 4 & 0 \oplus 2 \end{pmatrix} = \begin{pmatrix} 7 & 3 \\ 4 & 2 \end{pmatrix}$$

From max-plus to max-times

- We interpret recursive max-linear graphical models tropically.
- A matrix $C = \{c_{ij}\} \in \mathbb{R}_{\geq}^{d\times d}$ with $c_{ii} = 0$ for diagonal elements defines edge weights on a directed graph $\mathcal{D} = \mathcal{D}(C)$.
- This graph has node set $V = [d]$, with edge $j \to i \in E$ iff $c_{ij} > 0$.
- We work in the *max-times semiring* $(\bar{R}_+ = \mathbb{R}_+ \cup 0, \vee, \odot)$ where $\odot$ denotes ordinary multiplication.
- Using the map $x \to \log x$ this is algebraically isomorphic to the *tropical semiring* $(\mathbb{R} \cup -\infty, \oplus, \otimes)$.
- Concepts transfer *verbatim* from one structure to the other.

Tropical Linear Algebra

- For $A = (a_{ij}) \in \mathbb{R}_{\geq}^{d \times d}$, $\lambda \geq 0$ and $x \in \mathbb{R}_{>}^{d}$, we say that (λ, x) is a tropical eigenvalue-vector pair of A if it satisfies the equation

$$A \odot x = \lambda \odot x.$$

- In usual arithmetic, this is

$$\max_{j=1,\ldots,d} a_{ij} x_j = \lambda x_i \text{ for all } i \in V$$

.

- Example

$$\begin{pmatrix} 1 & 8 \\ 2 & 3 \end{pmatrix} \odot \begin{pmatrix} 2 \\ 1 \end{pmatrix} =$$

eigenvalue is $\lambda =$

Tropical Linear Algebra

- For $A = (a_{ij}) \in \mathbb{R}_{\geq}^{d \times d}$, $\lambda \geq 0$ and $x \in \mathbb{R}_{>}^{d}$, we say that (λ, x) is a tropical eigenvalue-vector pair of A if it satisfies the equation

$$A \odot x = \lambda \odot x.$$

- In usual arithmetic, this is

$$\max_{j=1,\dots,d} a_{ij} x_j = \lambda x_i \text{ for all } i \in V$$

.

- Example

$$\begin{pmatrix} 1 & 8 \\ 2 & 3 \end{pmatrix} \odot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 4 \end{pmatrix} = 4 \odot \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

eigenvalue is $\lambda = 4$.

The principal eigenvalue

- Let $\mathcal{D}_A$ be the directed graph determined by the positive entries of A.
- For a directed cycle $[j = k_0, k_1, \ldots, k_n = j]$ in $\mathcal{D}_A$, its *(geometric) mean* is defined as

$$\sqrt[n]{a_{k_0 k_1} a_{k_1 k_2} \cdots a_{k_{n-1} k_n}}.$$

- The maximum mean over all directed cycles is the *maximum cycle mean* of A, denoted $\lambda(A)$.
- If $\mathcal{D}_A$ is acyclic then $\lambda(A) = 0$.
- The number $\lambda(A) \geq 0$ is *always* a tropical eigenvalue of A, and if $\mathcal{D}_A$ is strongly connected, it is the *unique* tropical eigenvalue!

Tropical representation

- The *Kleene star of A*, denoted A^*.

$$A^* = \mathcal{I} \oplus \bigoplus_{i=1}^{d-1} A^{\odot i} = \bigoplus_{i=0}^{d-1} A^{\odot i}$$

- Consider a tropical equation of the form

$$X = A \odot X \oplus Z.$$

Lemma (e.g. BCOQ, Thm 3.17)

If $\lambda(A) < 1$, then the unique solution to $X = A \odot X \oplus Z$ is

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

Tropical Representation

- If we collect the innovations into the column vector $Z = (Z_1, \ldots, Z_d)^t$ the max-linear recursive equation becomes precisely

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

- Since $\mathcal{D}_C$ is a DAG, $\lambda(C) = 0$. By the Lemma, the system can also represented as

$$X = B \odot Z$$

  where $B = C^*$.
- The *idempotent* matrix B is the *max-linear coefficient matrix*, as defined by Gissibl & Klüppelberg.

Compare with the standard linear case

Here we have

$$X = CX + Z \implies (I - C)X = Z$$

Hence, since C is lower triangular

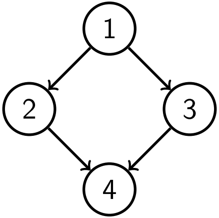
$$\begin{aligned} X &= (I - C)^{-1} Z \\ &= (I + C + C^2 + \cdots) Z \\ &= (I + C + C^2 + \cdots + C^{d-1}) Z = \tilde{C} Z \end{aligned}$$

Note that elements of $\tilde{C}_{ij}$ are *sums* of products of coefficients along dipaths from j to i.

Critical paths

- Now, B_{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 $X = B \odot Z$ 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.

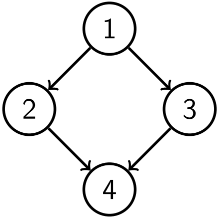
A simple example (Diamond)



The paths $[1 \to 2]$, $[1 \to 3]$, $[2 \to 4]$, and $[3 \to 4]$ 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 $[1 \to 2 \to 4]$ is critical whereas $[1 \to 3 \to 4]$ is not.

Example continued



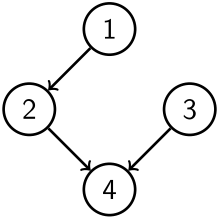
The max-linear coefficient matrix becomes

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

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

Example continued

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

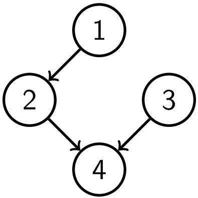


The new max-linear coefficient matrix becomes

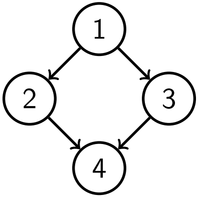
$$\tilde{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ c_{21} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ c_{42}c_{21} & c_{42} & c_{43} & 1 \end{pmatrix}$$

From the relation $X = B \odot Z$ we see 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 by* $\tilde{\mathcal{D}}$.

Example continued



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



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

Preview: Theorem

We define a notion of *$*$-connecting paths* as d-connecting paths with the modification that

1. we only take into account *critical paths*
2. we only allow *at most one collider* on paths

Then $*$-separation accordingly.

Theorem (Am., Klüppelberg, Lauritzen, Tran (2019))

Let $\mathcal{D}$ be a directed acyclic graph and C a generic coefficient matrix for a recursive max-linear model on $\mathcal{D}$. Then

$$X_I \perp\!\!\!\perp X_J | X_K \iff I \perp_{C^*} J | K$$

Hence, *any generic max-linear BN is faithful to $*$-separation.*

Conclusion

- *Max-linear graphical models* as a way of modeling events in extreme value theory.
- Extension from classical recursive linear models by *tropicalizing* the structural equation.
- Refined (nontrivial) version of *d*-separation, $*$*-separation*, characterizes CI statements.
- Key ingredient: *Tropical Geometry*. Marginal and conditional distributions can be expressed by *tropical* relations.
- Connections still emerging and lots to explore...
- More details on conditional independence and the Theorem in

  Steffen Lauritzen's talk, **next!**