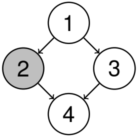
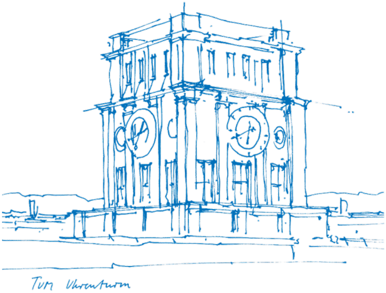
Max-linear Bayesian networks

Claudia Klüppelberg
Carlos Améndola, Steffen Lauritzen, and Ngoc M. Tran

Technical University of Munich

Banff, June 2022





Graphical models [Lauritzen (1996)]

- Represent multivariate distributions to facilitate statistical analysis.
- Describe high-dimensional distribution by a careful combination of lower dimensional factors.
- Use graphs as natural data structure models for algorithmic treatment.
- Use graphical models for causal interpretation through a recursive system on a directed acyclic graph (DAG) [Pearl (2009)].
- Conditional independence and Markov properties are essential features.

We present conditional independence properties of max-linear Bayesian networks, which emphasize the difference to Bayesian linear networks.

Max-linear Bayesian networks (MLBN) [Gissibl & K. (2018)]

Let $\mathcal{D} = (V, E)$ be a DAG and each node i represent a r.v. X_i.
Define the MLBN over $\mathcal{D}$ by the recursive ML structural equation system [Pearl (2009)]

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

for independent innovations $Z_1, \ldots, Z_d > 0$ with continuous distributions, coefficients $c_{ki} > 0$, and $\mathrm{pa}(i)$ (parents of i) denotes the set of nodes j with a directed edge from j to i ($j \to i$).
The system has solution

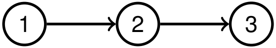
$$X_i = \bigvee_{j \in \mathrm{an}(i) \cup \{i\}} c^*_{ij} Z_j \quad i = 1, \ldots, d.$$

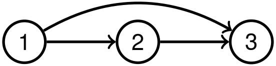
where $\mathrm{an}(i)$ (ancestors of i) denotes the set of nodes j with a directed path from j to i ($j \rightsquigarrow i$), and
c^*_{ij} is a maximum taken over all the products of the coefficient along $j \rightsquigarrow i$.
Any such path that realizes this maximum is called max-weighted under C.

Path notation

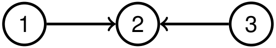
- Define $C^* = (c^*_{ij})$ such that c^*_{ij} is a maximum weight of all paths (weight = product of the coefficients) along $j \rightsquigarrow i$. Hence, C^* is a weighted reachability matrix, i.e., supported by the reachability DAG $\mathcal{D}^*$.

Example. $\mathcal{D} = (V, E) = (\{1, 2, 3\}, \{(1, 2), (2, 3)\})$

$\mathcal{D}$ 

$\mathcal{D}^*$ 

- A path in a DAG $\mathcal{D}$ is a sequence of nodes $i_0, i_1, \ldots, i_k$ such that $i_\ell \to i_{\ell+1}$ or $i_{\ell+1} \to i_\ell$ is an edge in $\mathcal{D}$ for each $\ell = 0, \ldots, k$.
- A directed path has edges $i_\ell \to i_{\ell+1}$ for all ℓ.
- A collider on a path is a node i_ℓ in a path such that $i_{\ell-1} \to i_\ell \leftarrow i_{\ell+1}$.



Tropical linear algebra [e.g. Butkovic (2010)]

Linear Bayesian networks are based on classical linear algebra; in contrast, max-linear Bayesian networks are based on tropical linear algebra in the max-times semiring $(\mathbb{R}_{\geq}, \odot, \cdot)$, defined by

$$a \odot b = a \vee b = \max(a, b), \quad a \cdot b = ab \quad \text{for } a, b \in \mathbb{R}_{\geq}.$$

These operations extend to $\mathbb{R}_{\geq}^{d}$ coordinate-wise and to corresponding matrix multiplication for $R \in \mathbb{R}_{\geq}^{m \times n}$ and $S \in \mathbb{R}_{\geq}^{n \times p}$ as

$$(R \odot S)_{ij} = \bigvee_{k=1}^{n} r_{ik} s_{kj}.$$

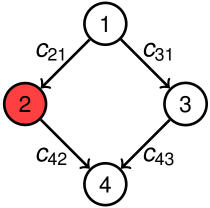
For max-linear Bayesian networks, X is Markov with respect to its DAG. However, the tropical linear algebra has various consequences concerning conditional independence properties and statistical analysis of the model.

Conditional independence

Linear graphical models identify conditional independence relations through separation criteria applied to a graph.
The standard separation criteria is given by the following definition.
Definition Two nodes $i, j \in V$ are *d*-connected given a set $K \in V \setminus \{i, j\}$, if there is a path $\pi : j \rightsquigarrow i$ such that all colliders on π are in $K \cup \text{an}(K)$ and no non-collider on π is in K. For three disjoint subsets I, J, K of the node set V, the node set K *d*-separates I and J, if no pair of nodes $i \in I$ and $j \in J$ is *d*-connected relative to K.
Note:

- Conditional independence properties for max-linear Bayesian networks are very different from those in linear Bayesian networks. In particular, they are often not faithful to their underlying DAG $\mathcal{D}$.
- Hence, the above *d*-separation criterion on the DAG typically will not identify all valid conditional independence relations.
- This is in contrast to the situation for most Bayesian networks based on discrete random variables or linear structural equations.

Diamond DAG: max-weighted and other paths



$1 \to 2 \to 4$ is max-weighted $\Leftrightarrow c_{42}c_{21} \geq c_{43}c_{31}$.
Then

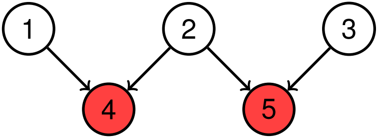
$$
\begin{aligned}
X_1 &= Z_1, \quad X_2 = c_{21}X_1 \vee Z_2, \\
X_4 &= c_{42}X_2 \vee Z_4 \vee c_{43}X_3 \\
&= c_{42}(Z_2 \vee c_{21}Z_1) \vee Z_4 \vee c_{43}(Z_3 \vee c_{31}Z_1) \\
&= c_{42}Z_2 \vee c_{42}c_{21}Z_1 \vee Z_4 \vee c_{43}Z_3 \vee c_{43}c_{31}Z_1 \\
&= c_{42}Z_2 \vee c_{42}c_{21}Z_1 \vee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

$$\Rightarrow \quad X_1 \perp\!\!\!\perp X_4 \,|\, X_2.$$

This does not follow from the *d*-separation criterion.
Here, the fact that $1 \to 2 \to 4$ is max-weighted renders the path $1 \to 3 \to 4$ unimportant for the conditional independence $X_1 \perp\!\!\!\perp X_4 \,|\, X_2$, even if $1 \to 3 \to 4$ were also max-weighted (that is, if $c_{42}c_{21} = c_{43}c_{31}$).

Cassiopeia DAG: double colliders along a path

For the sake of the argument, I set all $c_{ij} = c^*_{ij} = 1$.



$$X_1 = Z_1 \quad X_2 = Z_2 \quad X_3 = Z_3$$
$$X_4 = Z_1 \vee Z_2 \vee Z_4$$
$$X_5 = \qquad Z_2 \vee Z_3 \vee Z_5$$

Indeed: $X_1 \perp\!\!\!\perp X_3 \mid \{X_4 = x_4, X_5 = x_5\}$ for all coefficient matrices C:

Let $x_K = (x_4, x_5)$ and recall that all Z_i are a.s. different.
Then

$$\begin{bmatrix} x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} Z_1 \vee Z_2 \vee Z_4 \\ Z_2 \vee Z_3 \vee Z_5 \end{bmatrix} \geq \begin{bmatrix} Z_4 \\ Z_5 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x_4 \\ x_5 \end{bmatrix} \geq \begin{bmatrix} Z_1 \vee Z_2 \\ Z_2 \vee Z_3 \end{bmatrix}$$

Cassiopeia continued

We have three situations for (x_4, x_5) corresponding to

$$x_4 > x_5, \qquad x_4 < x_5, \qquad x_4 = x_5$$

$$\begin{bmatrix} x_4 \\ x_5 \end{bmatrix} \geq \begin{bmatrix} Z_1 \\ Z_2 \vee Z_3 \end{bmatrix}, \qquad \begin{bmatrix} x_4 \\ x_5 \end{bmatrix} \geq \begin{bmatrix} Z_1 \vee Z_2 \\ Z_3 \end{bmatrix}, \qquad \begin{bmatrix} x_4 \\ x_5 \end{bmatrix} \geq \begin{bmatrix} Z_1 \\ Z_3 \end{bmatrix} \text{ and } Z_2 = x_4 = x_5.$$

Hence, all Z_i are bounded in all three cases. Moreover, Z_1 and Z_3 never occur together in any inequality, rendering $X_1 \perp\!\!\!\perp X_3 \mid X_{\{4,5\}}$.
$x_4 > x_5$: then the causal source of 4 is 1, and of 5 they are 2,3
$x_4 < x_5$, then the situation is reversed
$x_4 = x_5$, then $X_2 = x_4 = x_5$ is a fixed node

Note: Conditional independence does not follow from the d-separation criterion, since the path from 1 to 3 is d-connecting relative to $\{4, 5\}$.

Tent DAG: context specific conditional independence

The source graph $\mathcal{C}(X_K = x_K)$:

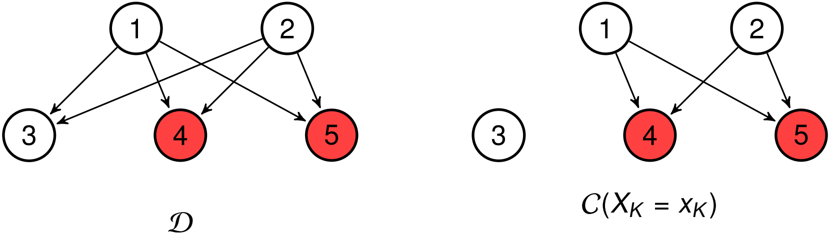
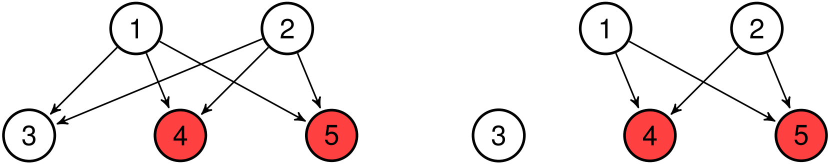


Figure: Left: Tent DAG $\mathcal{D}$. Right: For all coefficients equal to 1, the source graph $\mathcal{C}(X_K = x_K)$ with observed values $x_4 = x_5 = 2$ is obtained from $\mathcal{D}$ by removing the edges $1 \to 3$ and $2 \to 3$, which become redundant in the context $\{X_4 = X_5 = 2\}$.

Continuing the Tent DAG with $\{X_4 = X_5 = 2\}$



$$X_1 = Z_1, \quad X_2 = Z_2, \quad X_3 = Z_3 \vee X_1 \vee X_2$$
$$X_4 = Z_4 \vee X_1 \vee X_2 = 2, \quad X_5 = Z_5 \vee X_1 \vee X_2 = 2.$$

Since $Z_1, \ldots, Z_5$ are a.s. different, it holds outside a null-set that $X_1 \vee X_2 = Z_1 \vee Z_2 = 2$. This introduces bounds on the innovations; we must have $Z_1, Z_2, Z_4, Z_5 \leq 2$ and it also holds that $X_3 \geq 2$. Further, we then have

$$X_1 = Z_1, \quad X_2 = Z_2, \quad X_1 \vee X_2 = 2, \quad X_3 = Z_3 \vee 2,$$
$$X_4 = Z_4 \vee 2 = 2 \quad X_5 = Z_5 \vee 2 = 2,$$

Now, the dependence of X_3 on X_1, X_2 has disappeared, hence $X_3 \perp\!\!\!\perp (X_1, X_2) \,|\, X_4 = X_5 = 2$. This independence statement is reflected in the lack of edges $1 \to 3$ and $2 \to 3$ in the source DAG $\mathcal{C}(X_4 = X_5 = 2)$.

Summary and conclusion

- MLBN models have very different conditional independence properties than LBN models.
- I have given 3 examples, where the classical *d*-separation criterion fails.
- For an observed set of nodes K, a representation of $X_{\overline{K}} \mid X_K = x_k$ guides us to find a reduced representation of $X_{\overline{K}}$ taking deterministic features of a MLBN into account.
- Impact graphs describe how extreme events spread in the MLBN.
- The union of all impact graphs compatible with $\{X_K = x_K\}$ is the starting point for tracking possible sources of $\{X_K = x_K\}$.
- Cleaning up this union of graphs for fixed and redundant nodes and redundant edges yields the source graph $\mathcal{C}(X_K = x_K)$ giving a compact representation of the condional distribution given $X_K = x_K$.
- A new $*$-separation criterion is equivalent to CI statements in context-free and context-dependent settings, which we formulate as $*$-separation in different derived DAGs.

References

- Améndola, C., Klüppelberg, C., Lauritzen, S. and Tran, N.M. (2020) Conditional independence in max-linear Bayesian networks. *Annals of Applied Probability* **32**(1), 1-45.
- Buck, J., Klüppelberg, C. and Tran, N.M. (2020) Learning max-linear Bayesian trees under measurement errors. Submitted.
- Engelke, S. and Hitz, A.S. (2020) Graphical models for extremes. *JRSS B* **92**, 981-932.
- Gissibl, N. and Klüppelberg, C. (2018) Max-linear models on directed acyclic graphs. *Bernoulli* **24**(4A), 2693-2720.
- Gissibl, N., Klüppelberg, C. and Lauritzen, S. (2020) Estimation, identifiability, and structure learning of recursive max-linear models. *Scand. J. Statistics* **47**.
- Klüppelberg, C. and Krali, M. (2021) Estimating an extreme Bayesian network via scalings. *J. Mult. Anal.* **181**, 204672.
- Lauritzen, S.L. (1996). *Graphical Models.* Oxford University Press.
- Pearl, J. (2009) *Causality: Models, Reasoning, and Inference (2nd Ed.)* Cambridge University Press.