[1] University of Amsterdam

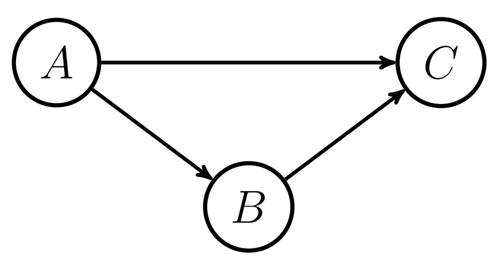
Nonparametric Bayesian networks are typically faithful in the total variation metric

[2] Booking.com

Philip Boeken[12] Patrick Forré[1] Joris M. Mooij[1]

Bayesian networks

A *Bayesian network* is a Directed Acyclic Graph (DAG) $G = (V, E)$, with for all $v \in V$ a Markov kernel $\mathbb{P}(X_v \,|\, X_{\mathrm{pa}(v)})$, e.g.



$\mathbb{P}(X_A)$

$\mathbb{P}(X_B \,|\, X_A)$

$\mathbb{P}(X_C \,|\, X_A, X_B)$,

which induces a joint distribution

$$\mathbb{P}(X_V) := \bigotimes_{v \in V} \mathbb{P}(X_v \,|\, X_{\mathrm{pa}(v)}).$$

Markov property

For any Bayesian network over DAG G with observational distribution $\mathbb{P}$ the *global Markov property* holds:

$$A \overset{d}{\underset{G}{\perp}} B \,|\, C \implies X_A \underset{\mathbb{P}}{\perp\!\!\!\perp} X_B \,|\, X_C$$

for all $A, B, C \subseteq V$ [5].

Faithfulness

A Bayesian network is called *faithful* if for all $A, B, C \subseteq V$ we have

$$A \overset{d}{\underset{G}{\not\perp}} B \,|\, C \implies X_A \underset{\mathbb{P}}{\not\perp\!\!\!\perp} X_B \,|\, X_C.$$

Faithfulness violations

Consider the Bayesian network with the graph at the top of the page, and the Markov kernels induced by

$$X_A = \varepsilon_A$$
$$X_B = \beta_{AB} X_A + \varepsilon_B$$
$$X_C = \beta_{AC} X_A + \beta_{BC} X_B + \varepsilon_C$$

with $\varepsilon_A, \varepsilon_B, \varepsilon_C$ independent. We have $A \not\perp^d_G C$. If $\beta_{AC} = -\beta_{AB}\beta_{BC}$, then $X_A \perp\!\!\!\perp X_C$ in which case the Bayesian network is *unfaithful*.

Other types of faithfulness violations exist, e.g. because of deterministic variables, or deterministic relations.

Why care about faithfulness?

Given a finite sample from $\mathbb{P}(X_V)$ from a Bayesian network with DAG G, *constraint-based causal discovery* methods test for all conditional (in)dependencies in the data. Assuming faithfulness, this *characterises* the set of all d-separations in the underlying DAG:

$$\{(A, B, C) : X_A \underset{\mathbb{P}}{\perp\!\!\!\perp} X_B \,|\, X_C\} = \{(A, B, C) : A \overset{d}{\underset{G}{\perp}} B \,|\, C\}.$$

From this set of d-separations, the graph G is reconstructed (up to a certain equivalence). If faithfulness does not hold, one may draw wrong causal conclusions from the constructed graph.

Genericity results

Faithfulness is an untestable assumption. In practice it is often *motivated* by the following results:

Theorem (Spirtes et al., 1993) Unfaithful parameters of linear Gaussian Bayesian networks, as a subset of $\mathbb{R}^m$, have Lebesgue measure zero. [4]

Theorem (Meek, 1995) Unfaithful parameters of discrete Bayesian networks, as a subset of $\mathbb{R}^m$, have Lebesgue measure zero. [3]

To our knowledge, no nonparametric extensions or analogues exist in the literature.

Main result

Let a DAG G be given, let $\mathcal{X}_v$ be a standard Borel space for every $v \in V$, and let $\mathcal{P}(\mathcal{X}_V)$ be the space of all probability measures on $\mathcal{X}_V$. Define

$$M_G := \{\mathbb{P} \in \mathcal{P}(\mathcal{X}_V) : \mathbb{P} \text{ satisfies the global Markov property w.r.t. } G\}$$
$$F_G := \{\mathbb{P} \in M_G : \mathbb{P} \text{ is faithful w.r.t. } G\}$$
$$U_G := \{\mathbb{P} \in M_G : \mathbb{P} \text{ is unfaithful w.r.t. } G\}$$

Equip M_G with the total variation metric

$$d_{TV}(\mathbb{P}, \mathbb{Q}) := \sup_{A \in \mathcal{B}(\mathcal{X}_V)} |\mathbb{P}(A) - \mathbb{Q}(A)|.$$

Theorem:
The set of faithful distributions F_G is a non-empty, dense and open set, and the unfaithful distributions U_G are nowhere dense.

Definition Given a topological space P a set $F \subseteq P$ is *typical* if it is the complement of a countable union of nowhere dense sets [1].

So faithful Bayesian networks are typical with respect to the total variation metric.

Proof sketch

We can write F_G as a countable intersection

$$F_G = \bigcap_{A \not\perp^d_G B \,|\, C} (M_G \setminus I_{AB|C})$$
$$U_G = M_G \setminus F_G$$

where $I_{AB|C} := \{\mathbb{P} \in \mathcal{P}(\mathcal{X}_V) : X_A \perp\!\!\!\perp_{\mathbb{P}} X_B \,|\, X_C\}$.

Theorem (Lauritzen, 2024) $I_{AB|C}$ is closed in the total variation metric [2].

In particular, F_G *is open*. Now we show that each $M_G \setminus I_{AB|C}$ is dense:

Theorem Let $A \not\perp^d_G B \,|\, C$, then for any $\mathbb{P}_0 \in I_{AB|C}$ there exists a net $(\mathbb{P}_\lambda)_{\lambda \in (0, \lambda^*)} \subseteq M_G \setminus I_{AB|C}$ such that $\mathbb{P}_\lambda \to \mathbb{P}_0$ as $\lambda \to 0$.

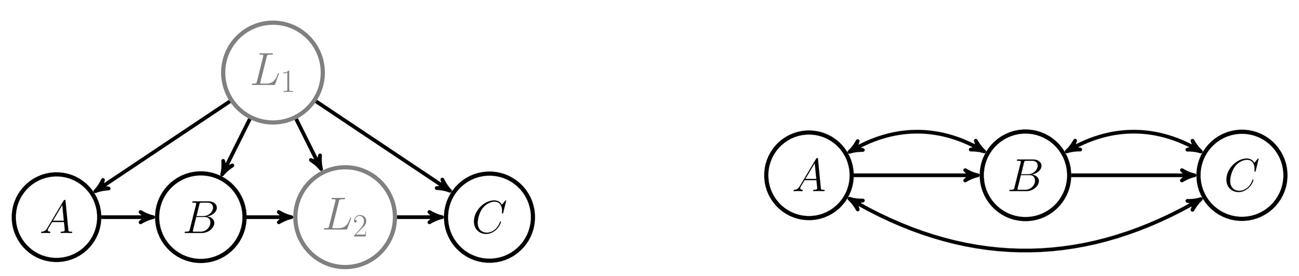
1. There exists a $\mathbb{P}_1 \in M_G \setminus I_{AB|C}$.
2. Define
$$\mathbb{P}_\lambda(X_V) := \bigotimes_{v \in V} \left((1-\lambda)\mathbb{P}_0(X_v \,|\, X_{\mathrm{pa}(v)}) + \lambda \mathbb{P}_1(X_v \,|\, X_{\mathrm{pa}(v)}) \right),$$
then $\mathbb{P}_\lambda \in M_G$.
3. There exists a $\lambda^* \in (0, 1)$ such that $X_A \not\perp\!\!\!\perp_{\mathbb{P}_\lambda} X_B \,|\, X_C$ for all $\lambda \in (0, \lambda^*)$, i.e. $(\mathbb{P}_\lambda)_{\lambda \in (0,\lambda^*)} \subseteq M_G \setminus I_{AB|C}$.
4. We have $\mathbb{P}_\lambda \to \mathbb{P}_0$ as $\lambda \to 0$.

Finite intersections of dense open sets are dense, so F_G *is dense*. Complements of dense, open sets are nowhere dense, so U_G *is nowhere dense*.

Further results

Theorem Unfaithful parameters of linear Gaussian or discrete Bayesian networks are nowhere dense in $\mathbb{R}^m$.

Often there are unobserved variables, and the (causal) relations between the observed variables are modelled with an acyclic directed mixed graph called the *latent projection*:



(a) DAG G (b) Latent projection G^p

Theorem Given an acyclic directed mixed graph G^p with vertices V, for any DAG G with vertices $V \cup W$ such that G^p is the latent projection of G onto V, the set of the Bayesian networks over G that are unfaithful with respect to G^p are nowhere dense.

Theorem The set of parameters of linear Gaussian or discrete Bayesian networks with latent variables that are unfaithful to latent projection G^p is nowhere dense and measure-zero.

[1] Alexander Kechris. *Classical descriptive set theory*. Springer, 1995.
[2] Steffen Lauritzen. Total variation convergence preserves conditional independence. *Statistics & Probability Letters*, 214:110200, November 2024.
[3] Christopher Meek. Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, page 411–418, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.
[4] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*, volume 81 of *Lecture Notes in Statistics*. Springer, New York, NY, 1993.
[5] Thomas Verma and Judea Pearl. Causal Networks: Semantics and Expressiveness. In Ross D. Shachter, Tod S. Levitt, Laveen N. Kanal, and John F. Lemmer, editors, *Machine Intelligence and Pattern Recognition*, volume 9 of *Uncertainty in Artificial Intelligence*, pages 69–76. North-Holland, January 1990.