1 VU VRIJE UNIVERSITEIT AMSTERDAM

Are Bayesian networks typically faithful?

2 UNIVERSITY OF AMSTERDAM

Philip Boeken[1] Patrick Forré[2] Joris M. Mooij[2]

Constraint-based causal discovery (PC, FCI, ...) assumes **faithfulness**. Untestable, so why assume it?

This work: faithfulness is open and dense, hence a topologically *typical* property.

Auxiliary result: CI is closed in weak topology, hence consistently testable.

Consequence: PC and FCI are consistent on a dense, open set of Bayesian networks.

Definitions

A **causal Bayesian network** is a **DAG** $G = (V, E)$ together with a **Markov kernel** $\mathbb{P}(X_v \mid X_{\mathrm{pa}(v)})$ for every $v \in V$. These kernels factorise into a joint distribution:

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

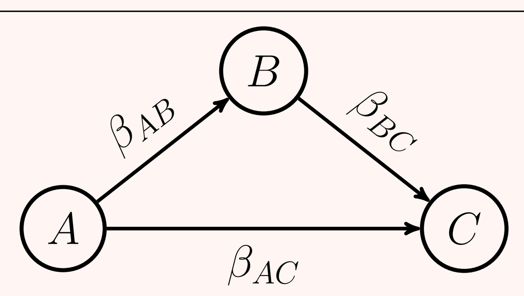
Two conditions on the pair $(G, \mathbb{P})$:

Markov $A \perp^d_G B \mid C \implies X_A \perp\!\!\!\perp_{\mathbb{P}} X_B \mid X_C$ (always holds)

Faithful $A \not\perp^d_G B \mid C \implies X_A \not\perp\!\!\!\perp_{\mathbb{P}} X_B \mid X_C$ (can fail)

Together they say: *d-separations in G are exactly the conditional independencies in $\mathbb{P}$.* Constraint-based causal discovery exploits this equivalence to reconstruct G from data.

Example of a faithfulness violation



Linear-Gaussian network. The direct effect β_{AC} and the indirect effect $\beta_{AB}\beta_{BC}$ cancel if

$$\beta_{AC} = -\beta_{AB}\beta_{BC}.$$

Data shows $X_A \perp\!\!\!\perp X_C$ even though $A \not\perp^d_G C$.

Other violations: deterministic variables, deterministic relations.

Faithfulness fails on a thin, exceptional set — but in what sense, and in *which* model classes?

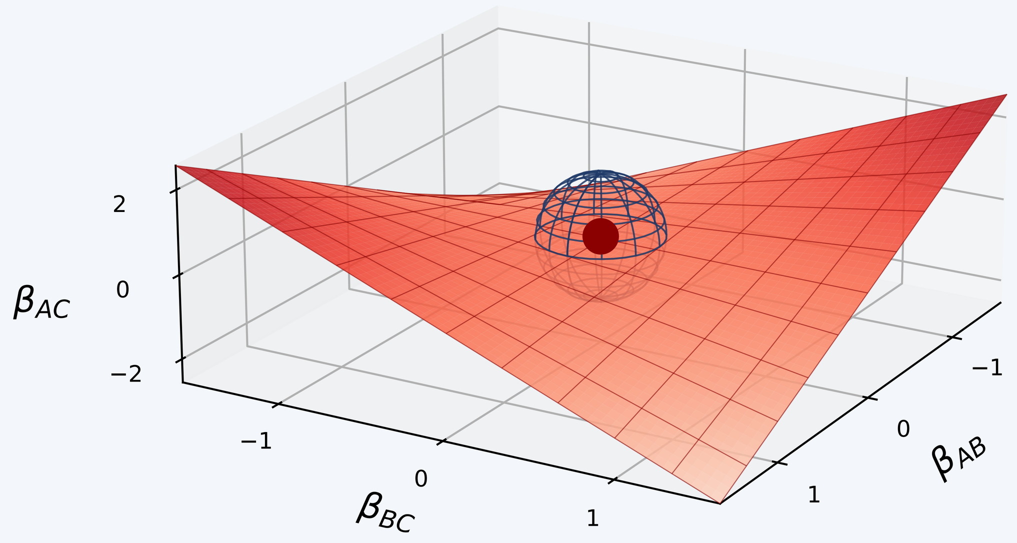
Existing results

Spirtes et al. 1993 [6]	Linear-Gaussian: unfaithful parameters have Lebesgue measure zero in $\mathbb{R}^m$.
Meek 1995 [5]	Discrete: same — Lebesgue measure zero in $\mathbb{R}^m$.
Lin et al. 2020 [4]	Discrete: faithful parameters are also dense and open.

With no canonical Lebesgue measure on a nonparametric space of Bayesian networks, we use topological notion of 'typical'.

Our approach: topological typicality

In the example above, the unfaithful parameters form a 2-dimensional manifold $\beta_{AC} - \beta_{AB}\beta_{BC} = 0$ in the 3-dimensional parameter space $(\beta_{AB}, \beta_{BC}, \beta_{AC})$:



We show that the **faithful models** are **open and dense** for two categories:

- In **Euclidean parameter spaces** (linear-Gaussian, discrete, regular conditional exponential families).
- In **nonparametric classes** (unconstrained, and regular conditional density models) the faithful Bayesian networks $m = (\mathbb{P}_m(X_v \mid X_{\mathrm{pa}(v)}))_{v\in V}$ are *open and dense* in the metric

$$d^\circ_{TV}(m, m') := \sum_{v\in V} \sup_{x_{\mathrm{pa}(v)}} d_{TV}\big(\mathbb{P}_m(X_v \mid x_{\mathrm{pa}(v)}),\ \mathbb{P}_{m'}(X_v \mid x_{\mathrm{pa}(v)})\big).$$

Nonparametric results in two topologies

Conditional independence is dense in total variation and weak topology. For it to be closed:

- **Total variation**: conditional independence is closed [3].
- **Weak topology**: conditional independence is closed **under regularity conditions**. **Weakly closed sets are consistently testable** [2, 1].

Under regularity they coincide, and are metrised by d_{TV} (distributions) and d°_{TV} (Markov kernels).

Main results

For each model class we ask three questions: are the faithful *parameters* typical? Are the faithful *distributions* typical?

Model class	Faithful parameters (tuples of Markov kernels)	Faithful distributions (space of $\mathbb{P}(X_V)$)
1. Unconstrained any standard Borel $\mathcal{X}_v$	Open + dense in d°_{TV}	Open + dense in d_{TV}
2. Conditional exponential families sample spaces $\subseteq \mathbb{R}$	Open + dense in $\mathbb{R}^m$; full Lebesgue measure	Open + dense in d_{TV} = weak topology
3. Nonparametric conditional densities any standard Borel $\mathcal{X}_v$	Open + dense in d°_{TV}	Open + dense in d_{TV} = weak topology

Consequence: consistent causal discovery

In model classes **2** and **3** of the table:

(i) Faithfulness is open and dense in the weak topology

(ii) Conditional independence is closed in the weak topology, hence there exists a consistent CI test

Combined: any sound constraint-based causal discovery algorithm (e.g. PC, FCI) is consistent on an open and dense — hence typical — set of Bayesian networks.

Proof sketch

Step 1 — finite reduction. Given DAG G, faithfulness is the *finite intersection*

$$F_G = \bigcap_{A \not\perp^d_G B \mid C} \{\mathbb{P} \text{ Markov w.r.t. } G : X_A \not\perp\!\!\!\perp_{\mathbb{P}} X_B \mid X_C\}.$$

A finite intersection of dense open sets is dense and open — so it suffices to show, *for one d-connection at a time*, that the set of distributions exhibiting the corresponding dependence is dense and open.

Step 2 — open. By Lauritzen 2024 [3], $\{\mathbb{P} : X_A \perp\!\!\!\perp_{\mathbb{P}} X_B \mid X_C\}$ is closed in total variation. So its complement in M_G is open.

Step 3 — dense via kernel interpolation. Given $\mathbb{P}_0$ with $X_A \perp\!\!\!\perp_{\mathbb{P}_0} X_B \mid X_C$, pick $\mathbb{P}_1$ with $X_A \not\perp\!\!\!\perp_{\mathbb{P}_1} X_B \mid X_C$ and interpolate the kernels:

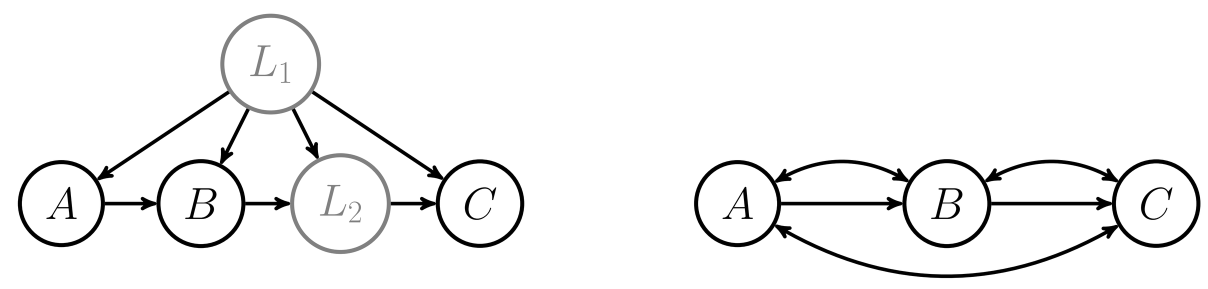
$$\mathbb{P}_\lambda := \bigotimes_{v\in V} \big((1-\lambda)\mathbb{P}_0(X_v \mid X_{\mathrm{pa}(v)}) + \lambda \mathbb{P}_1(X_v \mid X_{\mathrm{pa}(v)})\big).$$

Then $\mathbb{P}_\lambda \xrightarrow{tv} \mathbb{P}_0$ and $X_A \not\perp\!\!\!\perp_{\mathbb{P}_\lambda} X_B \mid X_C$ in the tail. So **conditional dependence is dense**.

Latent variables — a weaker requirement

Our results yield that faithfulness w.r.t. the full DAG G (over all variables, observed O and latent L) is typical. Faithfulness w.r.t. the *latent projection* G_O involves only the observed variables and is therefore a *strictly weaker* requirement — so it is typical too.

DAG G Latent projection $G_{A,B,C}$



faithful w.r.t. $\{A, B, C, L_1, L_2\}$ in G

$\Downarrow$

faithful w.r.t. $\{A, B, C\}$ in G $\iff$ faithful w.r.t. $G_{A,B,C}$

[1] Philip Boeken, Eduardo Skapinakis, Konstantin Genin, and Joris M. Mooij. Topological Criteria for Hypothesis Testing with Finite-Precision Measurements, 2026.

[2] Konstantin Genin and Kevin T. Kelly. The Topology of Statistical Verifiability. *Electronic Proceedings in Theoretical Computer Science*, 251:236–250, 2017.

[3] Steffen Lauritzen. Total variation convergence preserves conditional independence. *Statistics & Probability Letters*, 214:110200, 2024.

[4] Hanti Lin and Jiji Zhang. On Learning Causal Structures from Non-Experimental Data without Any Faithfulness Assumption. In *Proceedings of the 31st International Conference on Algorithmic Learning Theory*, pages 554–582. PMLR, 2020.

[5] 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.

[6] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*, volume 81 of *Lecture Notes in Statistics*. Springer, New York, NY, 1993.