Inferring model structure from data: MCMC for DAGs

Radu Craiu

Department of Statistical Sciences
University of Toronto

ESOBE - Rome - Aug 27, 2026

Outline

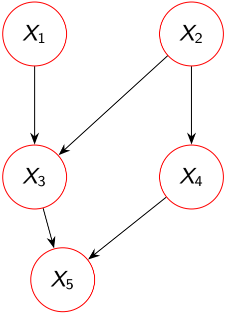
Morris Greenberg

Kieran Campbell

- Overview
- MCMC in the space of restricted orders
- Allow the restricted space to evolve
- Simulations and data analysis

Inferring a DAG from data

- Multivariate distributions can be represented through a graph plus parameters; each vertex is a unique variable.
- When it is represented by an acyclic graph, the joint distribution can be factored by a product of conditionals defined by the edges.



- DAG-type learning is of interest in fields where the causal direction is desired.
- For this graph:

$$P(X_1) \cdot P(X_2) \cdot P(X_3 \mid X_1, X_2)$$
$$P(X_4 \mid X_2) \cdot P(X_5 \mid X_3, X_4).$$

Finding model structure via MCMC

- Goal: given data D on p variables, infer a DAG that characterizes the conditional independences in the distribution of D or sample the DAG posterior $\pi(G \mid D)$.
- A typical pipeline:
  - estimate an undirected skeleton (PC);
  - explore the discrete space of DAGs by MCMC, e.g. by scoring graphs (or orders).
- Direct sampling of DAGs is prohibited by a super-exponential search space, score-equivalent plateaus, and acyclicity constraints.

Why direct sampling of DAGs is challenging

- The number of labelled directed graphs is $2^{p(p-1)}$; the fraction that are DAGs vanishes as p grows:

p	5	8	10	20
# graphs	1.0×10^{6}	7.2×10^{16}	1.2×10^{27}	2.5×10^{114}
# DAGs	2.9×10^{4}	7.8×10^{11}	4.2×10^{18}	2.3×10^{72}
# DAGs / # graphs	2.8%	1.1×10^{-5}	3.4×10^{-9}	9.5×10^{-43}

- Finding a highest-scoring DAG is NP-complete even with a bounded parent set (Chickering, 1996).
- Structure MCMC (add/delete/reverse one edge) must test acyclicity at every proposal and mixes slowly on this discrete space.
- Score-equivalent DAGs share the same likelihood, so the chain can wander inside a class for lengthy periods

Graph inference methods

- Three main methods: *constraint-based (CB)*, *score-based (SB)*, and *hybrid (H)* methods
- CB infer graph structure by estimating conditional dependencies
  - Yield consistent estimators of the true DAG under mild assumptions, e.g., faithfulness, maximum degree of the graph. (Kalisch and Bühlmann 2007)
  - Require a large number of conditional independence tests (Spirtes et al. 2000)
- SB infer structure via an objective function (*score*) that measures the relative merits of each graph
  - Score ties the graph directly to model quality
  - Scalability as sample space is exponential in p (no. of nodes)
- H combine strategically CB and SB
  - Use CB to construct a reduced space of graphs
  - Use SB to explore the space

DAG notation

- A graph is $G = (V, E_G)$ where V is set of vertices (or nodes) and set of edges is E_G
- In a directed G, edges are ordered: $i \to j$ indicates i is a parent of j and $i \leftarrow j$ implies the reverse.
- We denote the parents of node i as $pa_G(i)$ (or $pa(i)$)
- No node is both an ancestor and a descendant of another).
- Every DAG G is compatible with at least one topological order, $\prec \in \mathbb{S}^p$, i.e. a permutation of $\{1, ..., p\}$ that exhibits the ancestry of the vertices.
- If $\prec_{[i]} < \prec_{[j]}$ for all $i \to j$ in G, then $G \in \mathcal{G}_\prec$. Additionally, i is a *predecessor* of j in $\prec$, or $i \in pr_\prec(j)$ for short.

Faithfulness

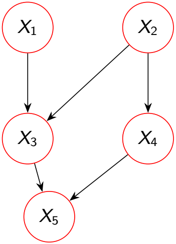
- Two DAGs are Markov equivalent if they share the same skeleton and the same v-structures (Verma and Pearl, 1990).
- When the true DAG belongs to a nontrivial equivalence class, structure learning from observational data alone becomes unidentifiable.
- Faithfulness: Every (conditional) independence in P is encoded by d-separation in G and vice versa .
- Faithfulness is needed to identify the Markov equivalence class from the conditional independence relations in the data.

PC skeleton

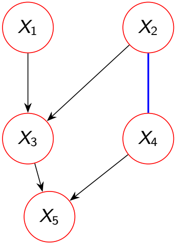
- The PC algorithm (Kalisch and Bühlmann, 2007) starts from the complete undirected graph and deletes X_i–X_j whenever $X_i \perp X_j \mid S$ for some conditioning set $S \Rightarrow$ skeleton
- Faithfulness is fragile: near path-cancellations make tests unstable (Uhler et al., 2013).
- Hybrid MCMC takes this skeleton as the initial set of allowed edges, then samples orientations/orders/subsets of edges inside it.

Faithfulness and Markov equivalence

DAG

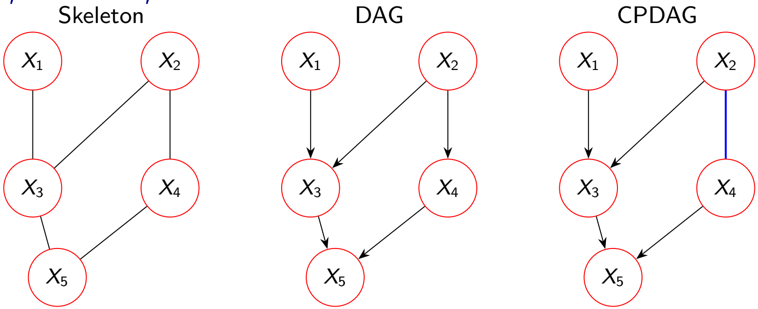


CPDAG of the same class



- Here $X_2 \rightarrow X_4$ is reversible: the DAG with $X_4 \rightarrow X_2$ is in the same class.
- The CPDAG is the unique completed representative of that class: compelled arrows stay directed; reversible edges (blue) stay undirected.
- Faithfulness means P is a perfect map of G, so the independences in the data are exactly the d-separations common to every graph in the class.
- Without faithfulness, extra independencies appear that no DAG in the class encodes, and observational data no longer identify the equivalence class.

DAGs, PDAGs, and CPDAGs



- DAG: every edge oriented.
- PDAG: a mix of directed and undirected edges (directed part acyclic). Intermediate output of orientation rules; not unique.
- CPDAG (completed PDAG / essential graph): the unique representative of a Markov equivalence class — compelled arrows stay directed; reversible edges (blue) stay undirected.
- Observational data identify a CPDAG, not a single DAG.

Ingredients: Scores

- A score $Q(G, D)$ measures how well graph G explains the data D.
- Common choices include:
  - **Information criteria:** BIC or AIC-type scores, usually based on penalized likelihood.
  - **Bayesian posterior scores:** unnormalized posterior scores such as

$$Q(G, D) \propto P(D \mid G)\pi(G).$$

  - **Bayesian network scores:** BDe for multinomial data and BGe for Gaussian data.
- Decomposable scores are especially useful:

$$Q(G, D) = \prod_{i=1}^{p} S(X_i, \mathbf{Pa}_G(i) \mid D),$$

so local graph updates only require rescoring affected nodes.

Computing the BGe scores

- Rely on closed-form marginal likelihoods $P(D \mid G)$ under conjugate priors (Heckerman and Geiger, 1995).
- BGe (Gaussian, normal–Wishart). The local score is a ratio of complete-model marginals

$$S_{\mathrm{BGe}}(X_i, \mathbf{Pa}(i) \mid D) = \frac{p(D_{i \cup \mathbf{Pa}})}{p(D_{\mathbf{Pa}})},$$

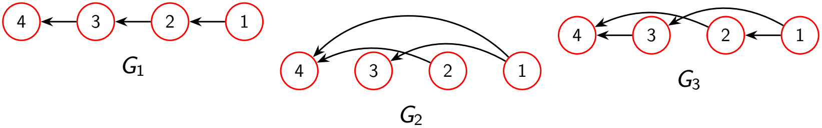
  and each $p(D_Y)$ is a product of Gamma terms and a ratio of determinants of prior/posterior scatter matrices R_0, R (Kuipers et al., 2014).
- Markov equivalent DAGs receive the same value.
- It is usually assumed that the data are approximately Gaussian, as the BGe score and other Gaussian DAG scores are fast to compute; limited work on non-Gaussian DAG scores.

Ingredients: Orders

- We move away from MCMC on the space of graphs to the space of orders.
- Using the order space:
  - Allows for "larger" moves than using the graph space.
  - Reduces the computation complexity to $O(p^K)$ (K is the max index)

Multiple DAGs for the same order

$$\prec = (1, 2, 3, 4)$$



- Each graph has arrows only from earlier to later nodes in $\prec$, that is from right to left in this layout.
- Hence $G_1, G_2, G_3 \in \mathcal{G}_\prec$ even though their edge sets differ.

Ingredients: Restricted spaces

- Let $\mathcal{H} = (V, E_{\mathcal{H}})$ denote the restricted space, with corresponding adjacency matrix H (where $H_{ji} = 1$ if $j \to i \in \mathcal{H}$) and $\mathbf{h}^i = \{X_j : H_{ji} = 1\}$.
- Index (number of parents) is restricted to be less than K
- The admissible DAGs are $\mathcal{G}_{\mathcal{H}} := \{G : G = (V, E_G) \in \mathbb{G}_p,\ E_G \subseteq E_{\mathcal{H}}\}$.
- Kuipers et al. 2021 use a *restricted search space* estimated via constraint-based methods in a hybrid setup. The restricted space is allowed to grow until a stopping rule is met. Samples are collected using the final space.

Hybrid sampler in the order space

- The posterior for order $\prec$ is proportional to the score $R(\prec | D)$

$$\begin{aligned}\pi_{Ord}(\prec | D) &\propto R(\prec | D) \\ &\propto \sum_{G \in \mathcal{G}_{\prec}} \pi_{DAG}(G|D) \propto \prod_{i=1}^{p} \sum_{\mathbf{Pa} \in \mathbf{Pa}_{\prec}(i)} S(X_i, \mathbf{Pa}|D)\end{aligned}$$

- When using a restricted space $\mathcal{H}$ the posterior becomes

$$R_{\mathcal{H}}(\prec | D) \propto \prod_{i=1}^{p} \sum_{\substack{\mathbf{Pa} \subseteq \mathbf{h}^i \\ \mathbf{Pa} \in \mathbf{Pa}_{\prec}(i)}} S(X_i, \mathbf{Pa}|D)$$

- Constraining scores to be on the restricted space with graph index at most K reduces the cost of MCMC steps from $O(p^K)$ to $O(Kp)$.
- Caveat: $R_{\mathcal{H}}$ will not be the same target as π_{Ord}.

Why restricting cuts the cost: $O(p^K) \to O(Kp)$

- Every order-score step recomputes, for each node i, the local sum over admissible parent sets of size at most K:

$$\sum_{\mathbf{Pa} \in \mathbf{Pa}_{\prec}(i)} S(X_i, \mathbf{Pa} \mid D).$$

- Unrestricted: node i may draw parents from any of its $\leq p-1$ predecessors, so the number of candidate parent sets is

$$\sum_{k=0}^{K} \binom{p-1}{k} = O(p^K),$$

and this must be re-evaluated as the order changes.

- Restricted space $\mathcal{H}$: parents of i are confined to the allowed neighborhood $\mathbf{h}^i$ with $|\mathbf{h}^i| \leq K$. These subset-scores are precomputed once; a move only changes which elements of $\mathbf{h}^i$ precede i in $\prec$, an $O(K)$ update per node.

- Over the p nodes,

$$\underbrace{O(p^K)}_{\text{search over all } p \text{ variables}} \longrightarrow \underbrace{O(Kp)}_{\text{search within } \mathbf{h}^i,\ |\mathbf{h}^i| \leq K}.$$

Theoretical results

- Cao et al. (2019): DAG-Wishart priors yield strong graph-selection consistency and posterior contraction for the precision matrix, with p allowed to grow, given a known topological order.
- Lee et al. (2019): an empirical sparse Cholesky prior attains minimax (or nearly minimax) posterior rates for Cholesky factors, under weaker conditions on dimension, sparsity and signal; model-selection consistency still assumes a known order.
- Castelletti et al. (2018): score essential graphs (CPDAGs) directly with an objective fractional Bayes factor — closed-form marginal likelihood — and explore sparse EG space by MCMC, so the target is the Markov equivalence class rather than a single DAG.
- Zhou and Chang (2023): random-walk Metropolis–Hastings on sparse equivalence classes mixes in time polynomial in n and p, and an empirical Bayes model is strongly consistent. Mixing of order MCMC is largely open.

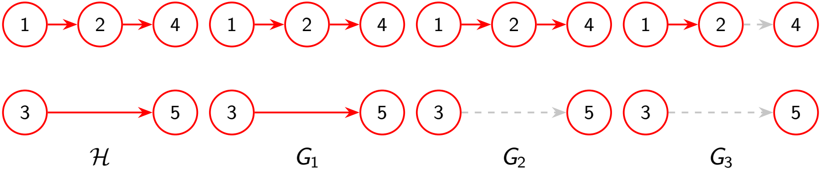
MCMC Sampling of DAGs via **B**irth-death processes **R**estricted **O**ver **O**rder **D**istributions (BROOD)

- MCMC algorithms for graphs can vary in 3 key ways
- The score function used
  - We use decomposable BGe scores
- The state space used
  - The state space is restricted ($\mathcal{H}$) but allowed to vary
  - We work in the order space and the graph space alternating
    - Sample $\prec \sim \pi_{Ord}(\prec |D, \mathcal{H})$
    - Sample $G \sim \pi_{DAG}(G| \prec, D, \mathcal{H})$
- The move proposal setup is to alternate between moves in:
  - the order space given a restricted space
  - the set of restricted spaces

Compatible graphs for $\mathcal{H}$ and $\prec$

$$\prec = (1, 2, 3, 4, 5), \qquad E_{\mathcal{H}} = \{1 \to 2,\ 2 \to 4,\ 3 \to 5\}.$$

- Think of $\mathcal{H}$ as a small graph of allowed edges.
- A compatible DAG $G \in \mathcal{G}_{\mathcal{H},\prec}$ keeps a subset of the edges in $E_{\mathcal{H}}$: it may drop some allowed edges but never adds new ones.

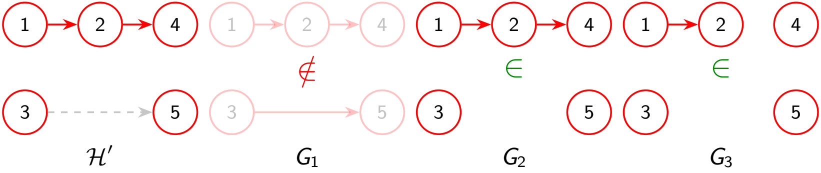


Solid red edges are kept; gray dashed edges are dropped. Each G keeps a subset of the edges of $\mathcal{H}$.

Changing the space: Removing one edge restricts the compatible set

$$E_{\mathcal{H}'} = E_{\mathcal{H}} \setminus \{3 \to 5\}.$$

- The edge $3 \to 5$ is no longer admissible.
- Every graph that used $3 \to 5$ now drops out of the compatible set; the rest survive.

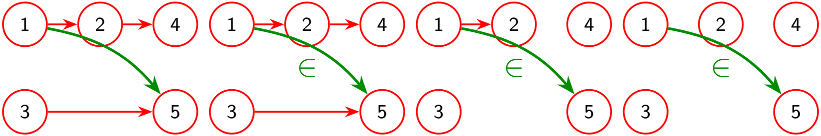


The move $\mathcal{H} \mapsto \mathcal{H}'$ changes the sum from $\sum_{G \in \mathcal{G}_{\mathcal{H}}}$ to a smaller $\sum_{G \in \mathcal{G}_{\mathcal{H}'}}$.

Adding one edge enlarges the compatible set

$$E_{\mathcal{H}''} = E_{\mathcal{H}} \cup \{1 \to 5\}.$$

- The new edge $1 \to 5$ (in green) becomes admissible.
- Graphs that use $1 \to 5$ were impossible before; they now enter the compatible set.



$\mathcal{H}''$ G_4 (new) G_5 (new) G_6 (new)

The move $\mathcal{H} \mapsto \mathcal{H}''$ changes the sum from $\sum_{G \in \mathcal{G}_{\mathcal{H}}}$ to a larger $\sum_{G \in \mathcal{G}_{\mathcal{H}''}}$.

Cost of restricting: Order space

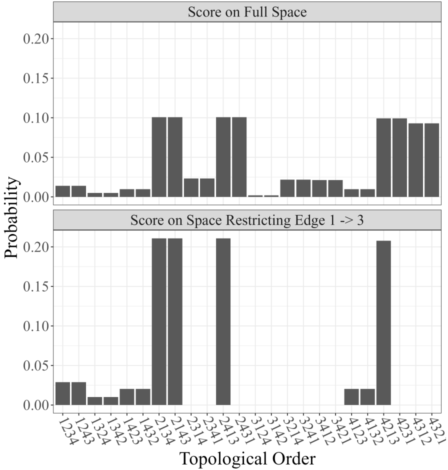


Figure: 4-node example: order posterior for the full search space (top) vs. the restricted space (bottom), where the edge $1 \rightarrow 3$ is excluded.

MCMC on restricted spaces

- **Theorem 1** Let $\mathcal{H} \in \mathbb{D}_p$ be a restricted space with corresponding edge set $E_{\mathcal{H}}$ and admissible DAG set $\mathcal{G}_{\mathcal{H}}$. Let π_{DAG}, π_{Ord} denote the posterior distributions over DAGs and orders, respectively, using the full DAG space $\mathbb{G}_p$, and let $\pi_{Ord}^{(\mathcal{H})}$ be the posterior over orders restricted to $\mathcal{G}_{\mathcal{H}}$. Define $\varepsilon_{\mathcal{H}} := 1 - \pi_{DAG}(\mathcal{H}|D)$ as the posterior mass assigned to DAGs outside of $\mathcal{H}$. Then, for some constant $c \in [0, 1]$,

$$1 - \sqrt{1 - c\varepsilon_{\mathcal{H}}} \leq D_{TV}(\pi_{Ord}^{(\mathcal{H})}, \pi_{Ord}) \leq \min\left(\sqrt{2 - 2\sqrt{1 - c\varepsilon_{\mathcal{H}}}}, 1\right)$$

  where D_{TV} denotes the total variation distance between distributions.
- Shows errors are incurred but is of limited use in algorithmic design.

Challenges

- We consider moves that grow AND shrink $\mathcal{H}_t$
- By varying $\mathcal{H}_t$ we aim to reduce the error established in Theorem 1
- The restricted spaces $\mathcal{H}_t$ are of varying dimension
- Of interest is the stationary distribution on $(\boldsymbol{H}, \boldsymbol{O}) \in \mathbb{D}_p \times \mathbb{S}^p$
- This requires theoretical validation.

A sampler with varying state spaces

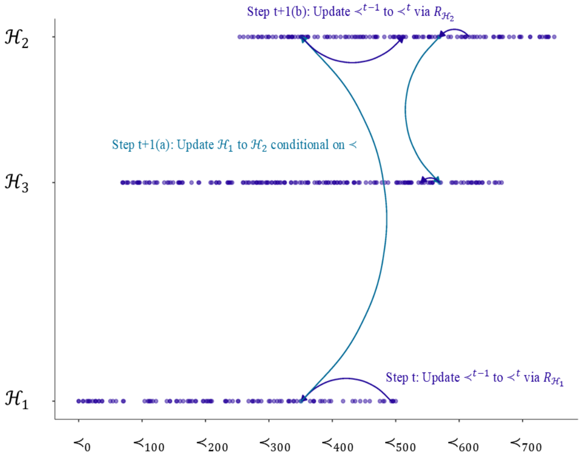


Figure: Illustration of the BROOD Algorithm

Sampling via Restricted MCMC

- We need to sample:
  - Step H: Challenging: sample the restricted space ($\mathcal{H}$) conditional on the order ($\prec$).
  - Step O: Sample order given the restricted space
  - Step G: Finally, given the restricted space and the order, sample from the set of graphs compatible with both $\mathcal{H}$ and $\prec$.
- We rely on the framework developed by Geyer and Møller (1994)
- Consider the MH algorithm sampling on $(\mathcal{H}, \prec) \in \mathbb{D}_p \times \mathbb{S}^p$ that is a mixture of 2 transition kernels:

$$Q_\ell(\mathcal{H}, \prec \mid \mathcal{H}_t, \prec_t) = (1-\ell)Q_0(\prec \mid \mathcal{H}_t, \prec_t) + \ell Q_1(\mathcal{H} \mid \mathcal{H}_t, \prec_t) \quad 0 \leq \ell \leq 1$$

- Q_0 is a standard MH procedure concentrated on $\mathcal{H}_t$ for Step 0.
- Q_1 is concentrated on $H_{m_t-1} \cup H_{m_t} \cup H_{m_t+1}$ where $\mathcal{H}_t \in H_{m_t}$ and the index m_t denotes the number of edges.

The development steps

- Given the current state $(\mathcal{H}, \prec)$
- Construct a birth and death process with rates $\beta(\mathcal{H}, \prec)$ and $\delta(\mathcal{H}, \prec)$ that has a stationary distribution.
- Define a MH algorithm with acceptance probabilities that depend on β and δ
- Ergodicity follows from the work of Geyer and Møller (1994) (see also Preston, 1977).

Sampling via Restricted MCMC

Theorem 2: Let $\mathcal{H} \in \mathbb{D}_p$, and $\prec \in \mathbb{S}^p$. Define $B_e(\mathcal{H}, \prec)$ and $D_e(\mathcal{H}, \prec)$ as follows:

$$B_e(\mathcal{H}, \prec) \propto \frac{\frac{1}{|\mathcal{G}_{\mathcal{H}^{+e}}|}\pi_{DAG}(\mathcal{H}^{+e} \mid \prec, D)}{\frac{1}{|\mathcal{G}_{\mathcal{H}}|}\pi_{DAG}(\mathcal{H} \mid \prec, D)} = \frac{1}{2}\frac{\pi_{DAG}(\mathcal{H}^{+e} \mid \prec, D)}{\pi_{DAG}(\mathcal{H} \mid \prec, D)}$$

$$D_e(\mathcal{H}, \prec) \propto \frac{\frac{c^*}{|\mathcal{G}_{\mathcal{H}^{-e}}|}\pi_{DAG}(\mathcal{H}^{-e} \mid \prec, D)}{\frac{1}{|\mathcal{G}_{\mathcal{H}}|}\pi_{DAG}(\mathcal{H} \mid \prec, D)} = 2c^*\frac{\pi_{DAG}(\mathcal{H}^{-e} \mid \prec, D)}{\pi_{DAG}(\mathcal{H} \mid \prec, D)}$$

where $c^* \in (0, 1]$. Let

$$\beta(\mathcal{H}, \prec) = \sum_{e \notin E_{\mathcal{H}}} B_e(\mathcal{H}, \prec)$$

$$\delta(\mathcal{H}, \prec) = \sum_{e \in E_{\mathcal{H}}} D_e(\mathcal{H}, \prec).$$

Then, a birth-death process defined by the rates β, δ yields a unique solution with a limit that is a tractable probability distribution.

Towards MCMC with varying $\mathcal{H}$

Corollary

Construct a birth-death Markov process $(\boldsymbol{H}, \boldsymbol{O}) \in \mathbb{D}_p \times \mathbb{S}^p$ with the birth and death rates specified in Theorem 2. Then, the marginal in $\mathcal{H} \in \mathbb{D}_p$ of the stationary distribution admits the following form:

$$\mu_{Space}^{(c^*)}(\mathcal{H}) \propto (2c^*)^{-|E_{\mathcal{H}}|} \sum_{\prec \in \mathbb{S}^p} \pi_{DAG}(\mathcal{H}| \prec, D), \tag{1}$$

where $\pi_{DAG}(\mathcal{H}| \prec, D) = \sum_{G \in \mathcal{G}_{\mathcal{H}}} \pi_{DAG}(G| \prec, D)$.

Putting it all together

Theorem 3: Let μ be a trans-dimensional probability measure that is compatible with a birth-death process defined by rates β, δ. Then $r(x, \xi)$, the acceptance ratio in the mixture kernel sampler of Geyer and Møller (1994), is defined by :

$$r(x, \xi) = \frac{\frac{1}{\beta(x \cup \xi) + \delta(x \cup \xi)}}{\frac{1}{\beta(x) + \delta(x)}}$$

Time reversibility is ensured with the following acceptance rates:

$$\text{death} \;\rightarrow A_1(x \mid x \cup \xi) = \begin{cases} \min\{1, 1/r(x, \xi)\} & \text{if } x \cup \xi \in H, \\ 0 & \text{otherwise} \end{cases}$$

$$\text{birth} \;\rightarrow A_1(x \cup \xi \mid x) = \begin{cases} \min\{1, r(x, \xi)\} & \text{if } x \cup \xi \in H, \\ 0 & \text{otherwise} \end{cases}$$

Dealing with the order-induced bias

- BROOD returns triples $(\mathcal{H}_t, \prec_t, G_t)$: the space comes from the birth–death kernel, and then $\prec_t \sim \pi_{\mathrm{Ord}}(\prec \mid \mathcal{H}_t, D) \propto R_{\mathcal{H}_t}(\prec \mid D)$ and $G_t \sim \pi_{\mathrm{DAG}}(G \mid \prec_t, \mathcal{H}_t, D)$.
- Each order is drawn conditionally on the space, so the sampling law of a triple is

$$q(\mathcal{H}, \prec, G) = \mu_{Space}^{(c^*)}(\mathcal{H}) \, \frac{R_{\mathcal{H}}(\prec \mid D)}{Z_{\mathcal{H}}} \, \frac{S(G \mid D) \, \mathbb{1}\{G \in \mathcal{G}_{\mathcal{H},\prec}\}}{R_{\mathcal{H}}(\prec \mid D)} \, .$$

- The score $S(G \mid D) = \prod_{i=1}^{p} S(X_i, \mathbf{Pa}_G(i) \mid D) \propto \pi_{DAG}(G \mid D)$ carries no $\prec$: the order enters only through the admissible set and the normalizers $R_{\mathcal{H}}(\prec \mid D) = \sum_{G \in \mathcal{G}_{\mathcal{H},\prec}} S(G \mid D)$ (which cancels) and $Z_{\mathcal{H}} = \sum_{\prec} R_{\mathcal{H}}(\prec \mid D)$.

Dealing with the order-induced bias

- Since $\mathcal{G}_{\mathcal{H},\prec} = \mathcal{G}_{\mathcal{H}} \cap \mathcal{G}_{\prec}$, the space does not change which orders admit a given G: $\sum_{\prec} \mathbb{1}\{G \in \mathcal{G}_{\mathcal{H},\prec}\} = n(G)$ for $G \in \mathcal{G}_{\mathcal{H}}$, so

$$q(\mathcal{H}, G) = \mu^{(c^*)}_{Space}(\mathcal{H}) \, \frac{n(G)\, S(G \mid D)}{Z_{\mathcal{H}}}, \qquad n(G) = \left|\{\prec \in \mathbb{S}^p : G \in \mathcal{G}_{\prec}\}\right|.$$

- Both multiplicities favour sparsity: a sparse G has more compatible orders and sits in more restricted spaces.
- When using the form of $\mu^{(c^*)}_{Space}(\mathcal{H})$, each draw is reweighted by

$$w_t = \frac{(2c^*)^{|E_{\mathcal{H}_t}|}}{n(G_t)\, M(G_t)}, \qquad \widehat{\mathbb{E}}[f(G)] = \frac{\sum_t w_t\, f(G_t)}{\sum_t w_t},$$

where $M(G) = \left|\{\mathcal{H} \in \mathbb{D}_p : E_G \subseteq E_{\mathcal{H}}\}\right| = \prod_{i=1}^{p} \sum_{k=d_i}^{K} \binom{p-1-d_i}{k-d_i}$ counts the spaces admitting G, with $d_i = |\mathbf{Pa}_G(i)|$.

Oil returns and geopolitical risk

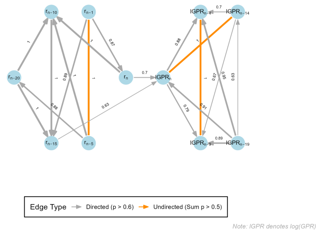
- Daily WTI oil prices and the geopolitical risk (GPR) index of Caldara and Iacoviello (2022), recorded from 1986/01/02 to 2026/04/20.
- Let $r_n = \log(\mathrm{Oil}_{n+1}) - \log(\mathrm{Oil}_n)$. The object of interest is the transition

  $$p\Big(r_n \mid r_{n-1}, r_{n-5}, r_{n-10}, r_{n-15}, r_{n-20}, \ell\mathrm{GPR}_n, \ell\mathrm{GPR}_{n-4}, \ldots, \ell\mathrm{GPR}_{n-19}\Big),$$

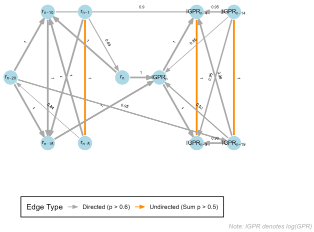
  with $\ell\mathrm{GPR} = \log(\mathrm{GPR})$.
- Median graphs contain only those edges $u \to v$ such that $p_{u\to v} + p_{v\to u} \geq 0.5$.
- If one of those probabilities is greater than 0.6 the edge is directed.

Oil returns and geopolitical risk

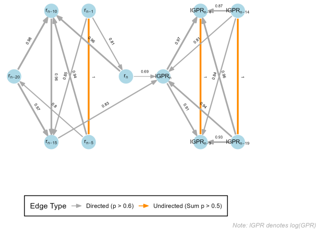
Full Space Sampler



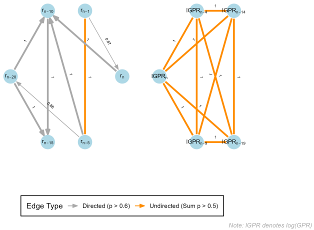
BiDAG Hybrid Sampler



BiDAG with Ad-Hoc Plus-1 Graph Sampling



BROOD with B-D Kernels



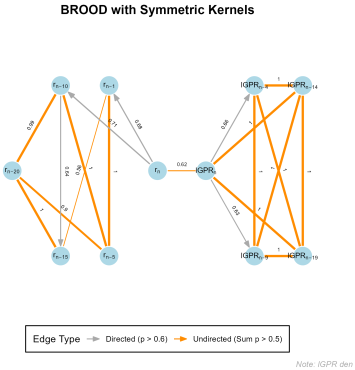
Improving BROOD

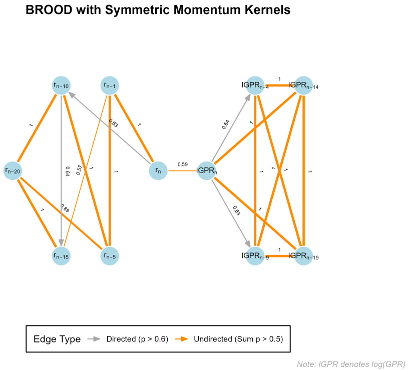
- We explore a symmetric kernel in which birth/death are proposed with prob 1/2.
- The acceptance ratio for a proposed birth is

$$\alpha_{\mathrm{sym}}(H \to H^{+e}) = \min\big\{1,\ \beta(H)/\delta(H^{+e})\big\}.$$

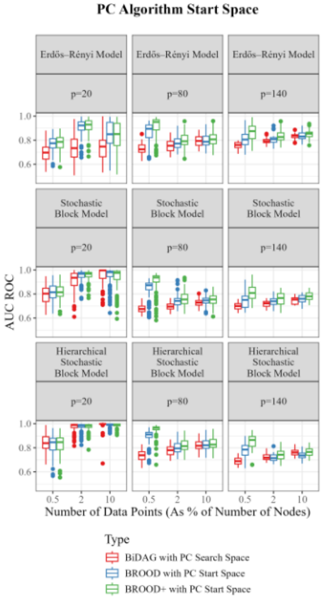
- We also consider a symmetric momentum kernel that acts on (H, v) with $v \in \{+1, -1\}$ and targets $\tilde{\mu}(H, v) = \frac{1}{2}\mu(H)(\mathbf{1}_{\{v=-1\}} + \mathbf{1}_{\{v=1\}})$.
  If $v_t = +1$: propose birth only and accept with same probability as above. Accept $\Rightarrow v_{t+1} = +1$; Reject $\Rightarrow v_{t+1} = -1$ and state becomes $(H, -1)$.
- These yield cleaner median graphs and more decisive edge probabilities.
- Within-series dependence in the r lags and in the $\ell\mathrm{GPR}$ lags is recovered more sharply; the contemporaneous r_n–$\ell\mathrm{GPR}_n$ link is weaker and often undirected.

Improving BROOD

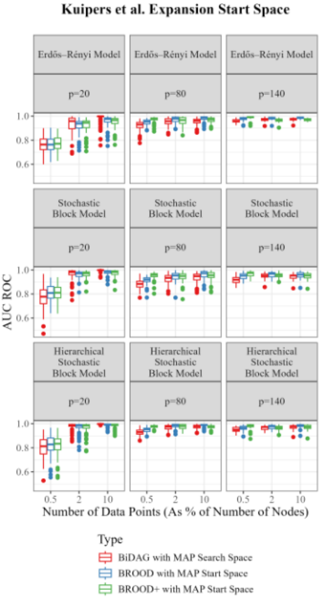




Simulation results



Simulation results - fixed sparsity



- What did we learn?
  - We lose the intuition given in Euclidian spaces by distances
  - Space is discrete but enormous and missing "parts" of it can be consequential
  - Theory for such samplers is challenging, but badly needed.
- What do we want to know?
  - In most examples we must impose restrictions on directionality
  - Incorporate auxiliary information about the nodes and their interrelationships
  - Design an adaptive version so tuning parameters, say c^*, are optimized?
  - Integrate everything into downstream data analysis...
- Paper available here: `raducraiu.com/publications/`

References

CALDARA, D. and IACOVIELLO, M. (2022). Measuring geopolitical risk. *American Economic Review* **112** 1194–1225.

CAO, X., KHARE, K. and GHOSH, M. (2019). Posterior graph selection and estimation consistency for high-dimensional Bayesian DAG models. *The Annals of Statistics* **47** 319–348.

CASTELLETTI, F., CONSONNI, G., VEDOVA, M. L. D. and PELUSO, S. (2018). Learning Markov Equivalence Classes of Directed Acyclic Graphs: An Objective Bayes Approach. *Bayesian Analysis* **13** 1235–1260.

CHICKERING, D. M. (1996). Learning Bayesian Networks is NP-Complete. In *Learning from Data: Artificial Intelligence and Statistics V* (D. Fisher and H.-J. Lenz, eds.). Springer, New York, NY, 121–130.

GEYER, C. J. and MØLLER, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics* **21** 359–373.

HECKERMAN, D. and GEIGER, D. (1995). Learning Bayesian networks: a unification for discrete and Gaussian domains. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. UAI'95, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

KALISCH, M. and BÜHLMANN, P. (2007). Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm. *Journal of Machine Learning Research* **8** 613–636.

KUIPERS, J., MOFFA, G. and HECKERMAN, D. (2014). Addendum on the scoring of Gaussian directed acyclic graphical models. *The Annals of Statistics* **42** 1689–1691.

KUIPERS, J., SUTER, P. and MOFFA, G. (2021). Efficient Sampling and Structure Learning of Bayesian Networks. *Journal of Computational and Graphical Statistics* **0** 1–12.

LEE, K., LEE, J. and LIN, L. (2019). Minimax posterior convergence rates and model selection consistency in high-dimensional DAG models based on sparse Cholesky factors. *The Annals of Statistics* **47** 3413–3437.

PRESTON, C. (1977). Spatial birth-and-death processes. *Bulletin of the International Statistical Institute* **46** 371–391.

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). Causation, Prediction, and Search.

UHLER, C., RASKUTTI, G., BÜHLMANN, P. and YU, B. (2013). Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics* **41** 436–463.

VERMA, T. and PEARL, J. (1990). Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*. UAI '90, Elsevier Science Inc., USA.

ZHOU, Q. and CHANG, H. (2023). Complexity analysis of Bayesian learning of high-dimensional DAG models and their equivalence classes. *The Annals of Statistics* **51** 1058–1085.