Learning Markov Equivalence Classes of Directed Acyclic Graphs: an Objective Bayesian Approach

Guido Consonni

Department of Statistical Sciences
Università Cattolica del Sacro Cuore, Milan

(Joint work with Federico Castelletti, Marco Della Vedova and Stefano Peluso)

University of Helsinki
30 October 2017

Outline

Problem

- q variables

Problem

- q variables
- Aim: discover conditional independence relationships among such variables.

Problem

- q variables
- Aim: discover conditional independence relationships among such variables.
- Data: $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$

Problem

- q variables
- Aim: discover conditional independence relationships among such variables.
- Data: $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$
- $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,q})^\top \quad i = 1, \ldots, n$

Problem

- q variables
- Aim: discover conditional independence relationships among such variables.
- Data: $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$
- $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,q})^\top \quad i = 1, \ldots, n$
- $\boldsymbol{y}_i$s are i.i.d. from a parametric family of sampling distributions

Background and notation

- Graph $\mathcal{G} = (V, E)$
  - $V = \{1, \ldots, q\}$ set of nodes
  - $E \subseteq V \times V$ set of edges

Background and notation

- Graph $\mathcal{G} = (V, E)$
  - $V = \{1, \ldots, q\}$ set of nodes
  - $E \subseteq V \times V$ set of edges
- $(u, v) \in E \iff u \to v$

Background and notation

- Graph $\mathcal{G} = (V, E)$
  - $V = \{1, \ldots, q\}$ set of nodes
  - $E \subseteq V \times V$ set of edges
- $(u, v) \in E \iff u \to v$
- $\{(u, v), (v, u)\} \in E \iff u - v$

Background and notation

- Graph $\mathcal{G} = (V, E)$
  - $V = \{1, \ldots, q\}$ set of nodes
  - $E \subseteq V \times V$ set of edges
- $(u, v) \in E \iff u \to v$
- $\{(u, v), (v, u)\} \in E \iff u - v$
- Undirected graph (UG) graph with only directed edges
- Directed Acyclic Graph (DAG) graph with only directed edges (no cycles)
  - If the distributions of y_i exhibits conditional independencies determined by $\mathcal{D}$ (Markov wrt $\mathcal{D}$), then

$$f_{\mathcal{D}}(\boldsymbol{y}_i) = \prod_{j \in V} f\left(y_{i,j} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}\right)$$

  - $\mathrm{pa}_{\mathcal{D}}(j)$ *parents* of node j in $\mathcal{D}$

Background and notation

Background and notation

- DAGs encoding the same conditional independencies are called *Markov equivalent*

Background and notation

- DAGs encoding the same conditional independencies are called *Markov equivalent*

Theorem (Verma & Pearl, 1991)

Two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are Markov equivalent if and only if they have the same skeleton and the same v-structures.

Background and notation

- DAGs encoding the same conditional independencies are called *Markov equivalent*

Theorem (Verma & Pearl, 1991)

Two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are Markov equivalent if and only if they have the same skeleton and the same v-structures.

- *skeleton*: the underlying undirected graph

Background and notation

- DAGs encoding the same conditional independencies are called *Markov equivalent*

Theorem (Verma & Pearl, 1991)

Two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are Markov equivalent if and only if they have the same skeleton and the same v-structures.

- *skeleton*: the underlying undirected graph
- *v-structure*: $u \rightarrow w \leftarrow z$

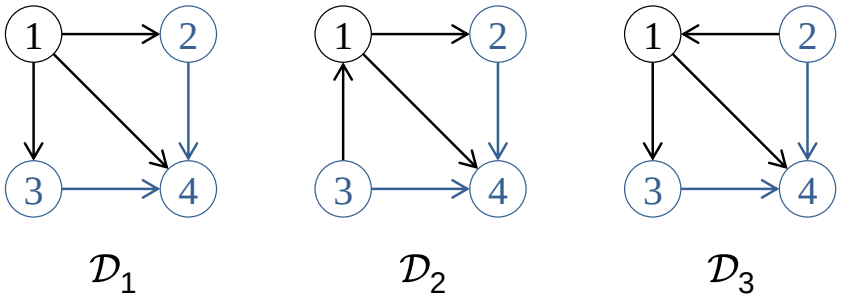
Background and notation

- DAGs encoding the same conditional independencies are called *Markov equivalent*

Theorem (Verma & Pearl, 1991)

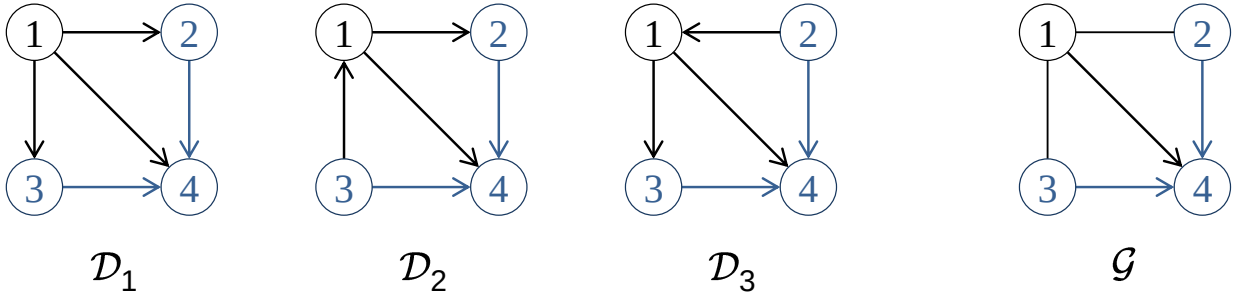
Two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are Markov equivalent if and only if they have the same skeleton and the same v-structures.

- *skeleton*: the underlying undirected graph
- *v-structure*: $u \rightarrow w \leftarrow z$
- Example: three Markov equivalent DAGs



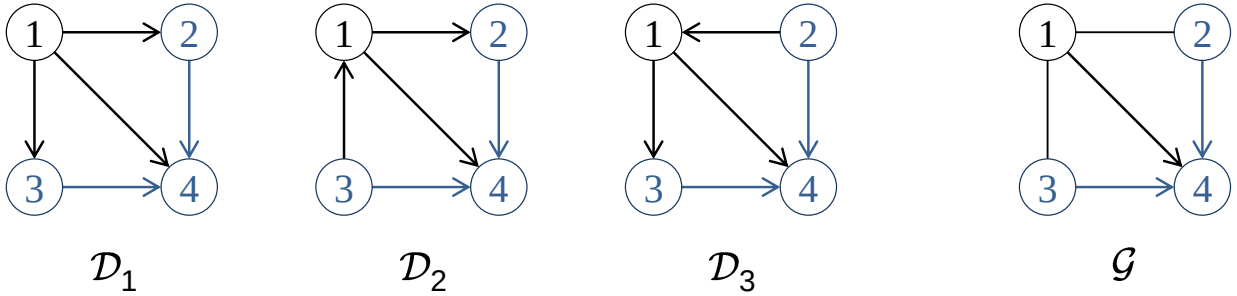
Essential graph

- *Essential Graph* (EG) $\mathcal{G}$, also called *Completed Partially Directed Graph* (CPDAG) is the union (over the edge sets) of Markov equivalent DAGs and represents a Markov equivalence class



Essential graph

- *Essential Graph* (EG) $\mathcal{G}$, also called *Completed Partially Directed Graph* (CPDAG) is the union (over the edge sets) of Markov equivalent DAGs and represents a Markov equivalence class



- $\mathcal{G}$ is a chain graph:
  - both directed and undirected edges
  - chain component: a set of nodes joined by an *undirected path*
  - $\mathcal{T}$: set of chain components of $\mathcal{G}$
  - $\mathcal{T} = \{\{1, 2, 3\}, \{4\}\}$

Essential graph (Characterization Theorem)

Theorem (Andersson et al., 1997)

A graph $\mathcal{G} = (V, E)$ is the EG for some DAG $\mathcal{D}$ with vertex set V if and only if $\mathcal{G}$ satisfies the following four conditions:

- *$\mathcal{G}$ is a CG;*
- *for each chain component $\tau \in \mathcal{T}$ the subgraph $\mathcal{G}_\tau$ is a decomposable UG;*
- *$\mathcal{G}$ has no flags (no induced subgraphs of the form $u \to v - z$);*
- *each directed edge $u \to v$ contained in $\mathcal{G}$ is strongly protected*

Bayesian model comparison

- Bayesian model $\mathcal{M}_k$

Bayesian model comparison

- Bayesian model $\mathcal{M}_k$
  - $f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)$, a family of sampling densities indexed by a model specific parameter $\boldsymbol{\theta}_k$

Bayesian model comparison

- Bayesian model $\mathcal{M}_k$
  - $f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)$, a family of sampling densities indexed by a model specific parameter $\boldsymbol{\theta}_k$
  - $p(\boldsymbol{\theta}_k)$, a prior density on $\boldsymbol{\theta}_k$, assumed to be *proper*

Bayesian model comparison

- Bayesian model $\mathcal{M}_k$
  - $f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)$, a family of sampling densities indexed by a model specific parameter $\boldsymbol{\theta}_k$
  - $p(\boldsymbol{\theta}_k)$, a prior density on $\boldsymbol{\theta}_k$, assumed to be *proper*
- $\mathcal{M}_1, \ldots, \mathcal{M}_K$ a collection of Bayesian models for the data matrix $\boldsymbol{Y}$

Bayesian model comparison

- Bayesian model $\mathcal{M}_k$
  - $f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)$, a family of sampling densities indexed by a model specific parameter $\boldsymbol{\theta}_k$
  - $p(\boldsymbol{\theta}_k)$, a prior density on $\boldsymbol{\theta}_k$, assumed to be *proper*
- $\mathcal{M}_1, \ldots, \mathcal{M}_K$ a collection of Bayesian models for the data matrix $\boldsymbol{Y}$
- Goal: comparing models

Bayesian model comparison

- Bayesian model $\mathcal{M}_k$
  - $f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)$, a family of sampling densities indexed by a model specific parameter $\boldsymbol{\theta}_k$
  - $p(\boldsymbol{\theta}_k)$, a prior density on $\boldsymbol{\theta}_k$, assumed to be *proper*
- $\mathcal{M}_1, \ldots, \mathcal{M}_K$ a collection of Bayesian models for the data matrix $\boldsymbol{Y}$
- Goal: comparing models
- Tool: marginal likelihood of $\mathcal{M}_k$ (and Bayes factor)

$$m_{\mathcal{M}_k}(\boldsymbol{Y}) = \int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k$$

Bayesian model comparison

- Bayesian model $\mathcal{M}_k$
  - $f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)$, a family of sampling densities indexed by a model specific parameter $\boldsymbol{\theta}_k$
  - $p(\boldsymbol{\theta}_k)$, a prior density on $\boldsymbol{\theta}_k$, assumed to be *proper*
- $\mathcal{M}_1, \ldots, \mathcal{M}_K$ a collection of Bayesian models for the data matrix $\boldsymbol{Y}$
- Goal: comparing models
- Tool: marginal likelihood of $\mathcal{M}_k$ (and Bayes factor)

$$m_{\mathcal{M}_k}(\boldsymbol{Y}) = \int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k$$

- If no/weak prior information set $p(\boldsymbol{\theta}_k) = p^D(\boldsymbol{\theta}_k)$ (default *objective parameter prior*)
- Problems with objective priors

Bayesian model comparison

- Bayesian model $\mathcal{M}_k$
  - $f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)$, a family of sampling densities indexed by a model specific parameter $\boldsymbol{\theta}_k$
  - $p(\boldsymbol{\theta}_k)$, a prior density on $\boldsymbol{\theta}_k$, assumed to be *proper*
- $\mathcal{M}_1, \ldots, \mathcal{M}_K$ a collection of Bayesian models for the data matrix $\boldsymbol{Y}$
- Goal: comparing models
- Tool: marginal likelihood of $\mathcal{M}_k$ (and Bayes factor)

$$m_{\mathcal{M}_k}(\boldsymbol{Y}) = \int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k$$

- If no/weak prior information set $p(\boldsymbol{\theta}_k) = p^D(\boldsymbol{\theta}_k)$ (default *objective parameter prior*)
- Problems with objective priors
  - often improper

Bayesian model comparison

- Bayesian model $\mathcal{M}_k$
  - $f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)$, a family of sampling densities indexed by a model specific parameter $\boldsymbol{\theta}_k$
  - $p(\boldsymbol{\theta}_k)$, a prior density on $\boldsymbol{\theta}_k$, assumed to be *proper*
- $\mathcal{M}_1, \ldots, \mathcal{M}_K$ a collection of Bayesian models for the data matrix $\boldsymbol{Y}$
- Goal: comparing models
- Tool: marginal likelihood of $\mathcal{M}_k$ (and Bayes factor)

$$m_{\mathcal{M}_k}(\boldsymbol{Y}) = \int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k$$

- If no/weak prior information set $p(\boldsymbol{\theta}_k) = p^D(\boldsymbol{\theta}_k)$ (default *objective parameter prior*)
- Problems with objective priors
  - often improper
  - defined up to an arbitrary constant

Bayesian model comparison

- Bayesian model $\mathcal{M}_k$
  - $f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)$, a family of sampling densities indexed by a model specific parameter $\boldsymbol{\theta}_k$
  - $p(\boldsymbol{\theta}_k)$, a prior density on $\boldsymbol{\theta}_k$, assumed to be *proper*
- $\mathcal{M}_1, \ldots, \mathcal{M}_K$ a collection of Bayesian models for the data matrix $\boldsymbol{Y}$
- Goal: comparing models
- Tool: marginal likelihood of $\mathcal{M}_k$ (and Bayes factor)

$$m_{\mathcal{M}_k}(\boldsymbol{Y}) = \int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k$$

- If no/weak prior information set $p(\boldsymbol{\theta}_k) = p^D(\boldsymbol{\theta}_k)$ (default *objective parameter prior*)
- Problems with objective priors
  - often improper
  - defined up to an arbitrary constant
  - cannot be naively used to compute marginal likelihoods

Fractional marginal likelihood

- Fractional Bayes factor (O'Hagan, 1995)

Fractional marginal likelihood

- Fractional Bayes factor (O'Hagan, 1995)
- $b = b(n)$, $0 < b < 1$, a fraction of the sample size n.

Fractional marginal likelihood

- Fractional Bayes factor (O'Hagan, 1995)
- $b = b(n)$, $0 < b < 1$, a fraction of the sample size n.
- *fractional marginal likelihood* of model $\mathcal{M}_k$:

$$m_{\mathcal{M}_k}(\boldsymbol{Y}; b) = \frac{\int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\int f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}$$

Fractional marginal likelihood

- Fractional Bayes factor (O'Hagan, 1995)
- $b = b(n)$, $0 < b < 1$, a fraction of the sample size n.
- *fractional marginal likelihood* of model $\mathcal{M}_k$:

$$m_{\mathcal{M}_k}(\boldsymbol{Y}; b) = \frac{\int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\int f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}$$

- $f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) = \left[f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) \right]^b$

Fractional marginal likelihood

- Fractional Bayes factor (O'Hagan, 1995)
- $b = b(n)$, $0 < b < 1$, a fraction of the sample size n.
- *fractional marginal likelihood* of model $\mathcal{M}_k$:

$$m_{\mathcal{M}_k}(\boldsymbol{Y}; b) = \frac{\int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\int f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}$$

- $f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) = \left[f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) \right]^b$

$$m_{\mathcal{M}_k}(\boldsymbol{Y}; b) = \int f^{1-b}_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^F(\boldsymbol{\theta}_k \mid b, \boldsymbol{Y}) d\boldsymbol{\theta}_k$$

Fractional marginal likelihood

- Fractional Bayes factor (O'Hagan, 1995)
- $b = b(n)$, $0 < b < 1$, a fraction of the sample size n.
- *fractional marginal likelihood* of model $\mathcal{M}_k$:

$$m_{\mathcal{M}_k}(\boldsymbol{Y}; b) = \frac{\int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\int f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}$$

- $f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) = \left[f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) \right]^b$

$$m_{\mathcal{M}_k}(\boldsymbol{Y}; b) = \int f^{1-b}_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^F(\boldsymbol{\theta}_k \mid b, \boldsymbol{Y}) d\boldsymbol{\theta}_k$$

- Implied *fractional prior*:

$$p^F(\boldsymbol{\theta}_k \mid b, \boldsymbol{Y}) \propto f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k)$$

Fractional marginal likelihood

- Fractional Bayes factor (O'Hagan, 1995)
- $b = b(n)$, $0 < b < 1$, a fraction of the sample size n.
- *fractional marginal likelihood* of model $\mathcal{M}_k$:

$$m_{\mathcal{M}_k}(\boldsymbol{Y}; b) = \frac{\int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\int f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}$$

- $f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) = \left[f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)\right]^b$

$$m_{\mathcal{M}_k}(\boldsymbol{Y}; b) = \int f^{1-b}_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^F(\boldsymbol{\theta}_k \mid b, \boldsymbol{Y}) d\boldsymbol{\theta}_k$$

- Implied *fractional prior*:

$$p^F(\boldsymbol{\theta}_k \mid b, \boldsymbol{Y}) \propto f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k)$$

- Fractional likelihood is a *discounted* full likelihood

Fractional marginal likelihood

- Fractional Bayes factor (O'Hagan, 1995)
- $b = b(n)$, $0 < b < 1$, a fraction of the sample size n.
- *fractional marginal likelihood* of model $\mathcal{M}_k$:

$$m_{\mathcal{M}_k}(\boldsymbol{Y}; b) = \frac{\int f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\int f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}$$

- $f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) = \left[f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k)\right]^b$

$$m_{\mathcal{M}_k}(\boldsymbol{Y}; b) = \int f^{1-b}_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^F(\boldsymbol{\theta}_k \mid b, \boldsymbol{Y}) d\boldsymbol{\theta}_k$$

- Implied *fractional prior*:

$$p^F(\boldsymbol{\theta}_k \mid b, \boldsymbol{Y}) \propto f^b_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) p^D(\boldsymbol{\theta}_k)$$

- Fractional likelihood is a *discounted* full likelihood
- Default choice: $b = n_0/n$, n_0 minimal (integer) training sample size which makes the induced fractional prior proper

Gaussian essential graphs

- Given $\boldsymbol{\Omega}_{\mathcal{D}}$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ i.i.d. $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$

Gaussian essential graphs

- Given $\boldsymbol{\Omega}_{\mathcal{D}}$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ i.i.d. $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
- $\boldsymbol{\Omega}_{\mathcal{D}}$ Markov with respect to a DAG $\mathcal{D}$

Gaussian essential graphs

- Given $\boldsymbol{\Omega}_{\mathcal{D}}$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ i.i.d. $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
- $\boldsymbol{\Omega}_{\mathcal{D}}$ Markov with respect to a DAG $\mathcal{D}$
- no directed edge $u \to v \implies \rho_{uv \cdot \{1,\ldots,v\} \setminus \{u,v\}} = 0$
  (assuming a well-numbering of the vertices in V)

Gaussian essential graphs

- Given $\boldsymbol{\Omega}_{\mathcal{D}}$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ i.i.d. $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
- $\boldsymbol{\Omega}_{\mathcal{D}}$ Markov with respect to a DAG $\mathcal{D}$
- no directed edge $u \to v \implies \rho_{uv \cdot \{1,\ldots,v\} \setminus \{u,v\}} = 0$
  (assuming a well-numbering of the vertices in V)

- equivalence class of $\mathcal{D} \to$ EG $\mathcal{G}$ with chain components $\tau \in \mathcal{T}$

Gaussian essential graphs

- Given $\boldsymbol{\Omega}_{\mathcal{D}}$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ i.i.d. $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
- $\boldsymbol{\Omega}_{\mathcal{D}}$ Markov with respect to a DAG $\mathcal{D}$
- no directed edge $u \to v \implies \rho_{uv \cdot \{1,\ldots,v\} \setminus \{u,v\}} = 0$
  (assuming a well-numbering of the vertices in V)

- equivalence class of $\mathcal{D} \to$ EG $\mathcal{G}$ with chain components $\tau \in \mathcal{T}$
- Conditional density of variables in chain component τ

$$f_{\mathcal{G}_\tau}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\theta}_{\mathcal{G}_\tau}) = \mathcal{N}_{|\tau|}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{\mu}_\tau + \boldsymbol{\Gamma}_\tau \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

Gaussian essential graphs

- Given $\boldsymbol{\Omega}_{\mathcal{D}}$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ i.i.d. $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
- $\boldsymbol{\Omega}_{\mathcal{D}}$ Markov with respect to a DAG $\mathcal{D}$
- no directed edge $u \to v \implies \rho_{uv \cdot \{1,\ldots,v\} \setminus \{u,v\}} = 0$
  (assuming a well-numbering of the vertices in V)

- equivalence class of $\mathcal{D} \to$ EG $\mathcal{G}$ with chain components $\tau \in \mathcal{T}$
- Conditional density of variables in chain component τ

$$f_{\mathcal{G}_\tau}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\theta}_{\mathcal{G}_\tau}) = \mathcal{N}_{|\tau|}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{\mu}_\tau + \boldsymbol{\Gamma}_\tau \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

- $\boldsymbol{\mu}_\tau = \mathbb{E}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}})$

Gaussian essential graphs

- Given $\boldsymbol{\Omega}_{\mathcal{D}}$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ i.i.d. $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
- $\boldsymbol{\Omega}_{\mathcal{D}}$ Markov with respect to a DAG $\mathcal{D}$
- no directed edge $u \to v \implies \rho_{uv \cdot \{1,\ldots,v\} \setminus \{u,v\}} = 0$
  (assuming a well-numbering of the vertices in V)

- equivalence class of $\mathcal{D} \to$ EG $\mathcal{G}$ with chain components $\tau \in \mathcal{T}$
- Conditional density of variables in chain component τ

$$f_{\mathcal{G}_\tau}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\theta}_{\mathcal{G}_\tau}) = \mathcal{N}_{|\tau|}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{\mu}_\tau + \boldsymbol{\Gamma}_\tau \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

- $\boldsymbol{\mu}_\tau = \mathbb{E}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}})$
- $\boldsymbol{\Gamma}_\tau$ matrix of regression parameters

Gaussian essential graphs

- Given $\boldsymbol{\Omega}_{\mathcal{D}}$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ i.i.d. $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
- $\boldsymbol{\Omega}_{\mathcal{D}}$ Markov with respect to a DAG $\mathcal{D}$
- no directed edge $u \to v \implies \rho_{uv \cdot \{1,\ldots,v\} \setminus \{u,v\}} = 0$
  (assuming a well-numbering of the vertices in V)

- equivalence class of $\mathcal{D} \to$ EG $\mathcal{G}$ with chain components $\tau \in \mathcal{T}$
- Conditional density of variables in chain component τ

$$f_{\mathcal{G}_\tau}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\theta}_{\mathcal{G}_\tau}) = \mathcal{N}_{|\tau|}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{\mu}_\tau + \boldsymbol{\Gamma}_\tau \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

- $\boldsymbol{\mu}_\tau = \mathbb{E}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}})$
- $\boldsymbol{\Gamma}_\tau$ matrix of regression parameters
- $\boldsymbol{\Omega}_{\mathcal{G}_\tau}$ (conditional) precision matrix
  Recall (Theorem 1) that $\boldsymbol{\Omega}_{\mathcal{G}_\tau}$ is Markov with respect to the decomposable graph $\mathcal{G}_\tau$.

Likelihood and prior factorization for an EG

$$f_{\mathcal{G}_\tau}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\theta}_{\mathcal{G}_\tau}) = \mathcal{N}_{|\tau|}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{B}_\tau^\top \boldsymbol{x}_{i,\tau}, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

$$\boldsymbol{x}_{i,\tau} = \begin{bmatrix} 1 \\ \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)} \end{bmatrix}; \quad \boldsymbol{B}_\tau = \begin{bmatrix} \boldsymbol{\mu}_\tau^\top \\ \boldsymbol{\Gamma}_\tau^\top \end{bmatrix}$$

Likelihood and prior factorization for an EG

$$f_{\mathcal{G}_\tau}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\theta}_{\mathcal{G}_\tau}) = \mathcal{N}_{|\tau|}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{B}_\tau^\top \boldsymbol{x}_{i,\tau}, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

$$\boldsymbol{x}_{i,\tau} = \begin{bmatrix} 1 \\ \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)} \end{bmatrix}; \quad \boldsymbol{B}_\tau = \begin{bmatrix} \boldsymbol{\mu}_\tau^\top \\ \boldsymbol{\Gamma}_\tau^\top \end{bmatrix}$$

- $\boldsymbol{B}_\tau$ unconstrained ($\mathcal{G}$ has no flags)

Likelihood and prior factorization for an EG

$$f_{\mathcal{G}_\tau}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\theta}_{\mathcal{G}_\tau}) = \mathcal{N}_{|\tau|}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{B}_\tau^\top \boldsymbol{x}_{i,\tau}, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

$$\boldsymbol{x}_{i,\tau} = \begin{bmatrix} 1 \\ \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)} \end{bmatrix}; \quad \boldsymbol{B}_\tau = \begin{bmatrix} \boldsymbol{\mu}_\tau^\top \\ \boldsymbol{\Gamma}_\tau^\top \end{bmatrix}$$

- $\boldsymbol{B}_\tau$ unconstrained ($\mathcal{G}$ has no flags)

Likelihood and prior factorization for an EG

$$f_{\mathcal{G}_\tau}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\theta}_{\mathcal{G}_\tau}) = \mathcal{N}_{|\tau|}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{B}_\tau^\top \boldsymbol{x}_{i,\tau}, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

$$\boldsymbol{x}_{i,\tau} = \begin{bmatrix} 1 \\ \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)} \end{bmatrix}; \quad \boldsymbol{B}_\tau = \begin{bmatrix} \boldsymbol{\mu}_\tau^\top \\ \boldsymbol{\Gamma}_\tau^\top \end{bmatrix}$$

- $\boldsymbol{B}_\tau$ unconstrained ($\mathcal{G}$ has no flags)

$$\boldsymbol{Y}_\tau \,|\, \boldsymbol{X}_\tau, \boldsymbol{B}_\tau, \boldsymbol{\Omega}_{\mathcal{G}_\tau} \sim \mathcal{N}_{n,|\tau|}(\boldsymbol{X}_\tau \boldsymbol{B}_\tau, \boldsymbol{I}_n, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

Lik factorization under a chain graph

Likelihood and prior factorization for an EG

$$f_{\mathcal{G}_\tau}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\theta}_{\mathcal{G}_\tau}) = \mathcal{N}_{|\tau|}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{B}_\tau^\top \boldsymbol{x}_{i,\tau}, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

$$\boldsymbol{x}_{i,\tau} = \begin{bmatrix} 1 \\ \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)} \end{bmatrix}; \quad \boldsymbol{B}_\tau = \begin{bmatrix} \boldsymbol{\mu}_\tau^\top \\ \boldsymbol{\Gamma}_\tau^\top \end{bmatrix}$$

- $\boldsymbol{B}_\tau$ unconstrained ($\mathcal{G}$ has no flags)

$$\boldsymbol{Y}_\tau \,|\, \boldsymbol{X}_\tau, \boldsymbol{B}_\tau, \boldsymbol{\Omega}_{\mathcal{G}_\tau} \sim \mathcal{N}_{n,|\tau|}(\boldsymbol{X}_\tau \boldsymbol{B}_\tau, \boldsymbol{I}_n, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

Lik factorization under a chain graph

$$f_{\mathcal{G}}(\boldsymbol{Y} \,|\, \boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{G}}) = \prod_{\tau \in \mathcal{T}} \mathcal{N}_{n,|\tau|}(\boldsymbol{Y}_\tau \,|\, \boldsymbol{X}_\tau \boldsymbol{B}_\tau, \boldsymbol{I}_n, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

Likelihood and prior factorization for an EG

$$f_{\mathcal{G}_\tau}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)}, \boldsymbol{\theta}_{\mathcal{G}_\tau}) = \mathcal{N}_{|\tau|}(\boldsymbol{y}_{i,\tau} \,|\, \boldsymbol{B}_\tau^\top \boldsymbol{x}_{i,\tau}, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

$$\boldsymbol{x}_{i,\tau} = \begin{bmatrix} 1 \\ \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{G}}(\tau)} \end{bmatrix}; \quad \boldsymbol{B}_\tau = \begin{bmatrix} \boldsymbol{\mu}_\tau^\top \\ \boldsymbol{\Gamma}_\tau^\top \end{bmatrix}$$

- $\boldsymbol{B}_\tau$ unconstrained ($\mathcal{G}$ has no flags)

$$\boldsymbol{Y}_\tau \,|\, \boldsymbol{X}_\tau, \boldsymbol{B}_\tau, \boldsymbol{\Omega}_{\mathcal{G}_\tau} \sim \mathcal{N}_{n,|\tau|}(\boldsymbol{X}_\tau \boldsymbol{B}_\tau, \boldsymbol{I}_n, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

Lik factorization under a chain graph

$$f_{\mathcal{G}}(\boldsymbol{Y} \,|\, \boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{G}}) = \prod_{\tau \in \mathcal{T}} \mathcal{N}_{n,|\tau|}(\boldsymbol{Y}_\tau \,|\, \boldsymbol{X}_\tau \boldsymbol{B}_\tau, \boldsymbol{I}_n, \boldsymbol{\Omega}_{\mathcal{G}_\tau}^{-1})$$

If $\boldsymbol{\theta}_{\mathcal{G}_\tau}$ *a priori* independent (global parameter independence)

$$m_{\mathcal{G}}(\boldsymbol{Y}) = \prod_{\tau \in \mathcal{T}} m_{\mathcal{G}_\tau}(\boldsymbol{Y}_\tau \,|\, \boldsymbol{X}_\tau)$$

Marginal likelihood of the essential graph

- Use theory developed in Consonni et al (2017) to perform objective Bayes learning for Gaussian multivariate regression DAG models

Marginal likelihood of the essential graph

- Use theory developed in Consonni et al (2017) to perform objective Bayes learning for Gaussian multivariate regression DAG models
- Default prior

$$p^D(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \propto |\boldsymbol{\Omega}_\tau|^{\frac{a_D - |\tau| - 1}{2}}$$

Marginal likelihood of the essential graph

- Use theory developed in Consonni et al (2017) to perform objective Bayes learning for Gaussian multivariate regression DAG models
- Default prior

$$p^D(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \propto |\boldsymbol{\Omega}_\tau|^{\frac{a_D - |\tau| - 1}{2}}$$

- $\mathcal{G}_\tau$ is decomposable

Marginal likelihood of the essential graph

- Use theory developed in Consonni et al (2017) to perform objective Bayes learning for Gaussian multivariate regression DAG models
- Default prior

$$p^D(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \propto |\boldsymbol{\Omega}_\tau|^{\frac{a_D - |\tau| - 1}{2}}$$

- $\mathcal{G}_\tau$ is decomposable
- $\mathcal{C}_\tau$ set of (maximal) cliques
- $\mathcal{S}_\tau$ set of separators

Marginal likelihood of the essential graph

- Use theory developed in Consonni et al (2017) to perform objective Bayes learning for Gaussian multivariate regression DAG models
- Default prior

$$p^D(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \propto |\boldsymbol{\Omega}_\tau|^{\frac{a_D - |\tau| - 1}{2}}$$

- $\mathcal{G}_\tau$ is decomposable
- $\mathcal{C}_\tau$ set of (maximal) cliques
- $\mathcal{S}_\tau$ set of separators
- Then

$$m_{\mathcal{G}_\tau}(\boldsymbol{Y}_\tau \,|\, \boldsymbol{X}_\tau) = \frac{\prod_{C \in \mathcal{C}_\tau} m_\tau(\boldsymbol{Y}_{C,\tau} \,|\, \boldsymbol{X}_\tau)}{\prod_{S \in \mathcal{S}_\tau} m_\tau(\boldsymbol{Y}_{S,\tau} \,|\, \boldsymbol{X}_\tau)}$$

Marginal likelihood of the essential graph

- Use theory developed in Consonni et al (2017) to perform objective Bayes learning for Gaussian multivariate regression DAG models
- Default prior

$$p^D(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \propto |\boldsymbol{\Omega}_\tau|^{\frac{a_D - |\tau| - 1}{2}}$$

- $\mathcal{G}_\tau$ is decomposable
- $\mathcal{C}_\tau$ set of (maximal) cliques
- $\mathcal{S}_\tau$ set of separators
- Then

$$m_{\mathcal{G}_\tau}(\boldsymbol{Y}_\tau \,|\, \boldsymbol{X}_\tau) = \frac{\prod_{C \in \mathcal{C}_\tau} m_\tau(\boldsymbol{Y}_{C,\tau} \,|\, \boldsymbol{X}_\tau)}{\prod_{S \in \mathcal{S}_\tau} m_\tau(\boldsymbol{Y}_{S,\tau} \,|\, \boldsymbol{X}_\tau)}$$

- $m_\tau(\boldsymbol{Y}_{C,\tau} \,|\, \boldsymbol{X}_\tau)$ and $m_\tau(\boldsymbol{Y}_{S,\tau} \,|\, \boldsymbol{X}_\tau)$ computed under the *complete* graph using standard priors
  (no need for hyper-Inverse Wishart of the like)

Marginal likelihood of the essential graph

- Use theory developed in Consonni et al (2017) to perform objective Bayes learning for Gaussian multivariate regression DAG models
- Default prior

$$p^D(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \propto |\boldsymbol{\Omega}_\tau|^{\frac{a_D - |\tau| - 1}{2}}$$

- $\mathcal{G}_\tau$ is decomposable
- $\mathcal{C}_\tau$ set of (maximal) cliques
- $\mathcal{S}_\tau$ set of separators
- Then

$$m_{\mathcal{G}_\tau}(\boldsymbol{Y}_\tau \,|\, \boldsymbol{X}_\tau) = \frac{\prod_{C \in \mathcal{C}_\tau} m_\tau(\boldsymbol{Y}_{C,\tau} \,|\, \boldsymbol{X}_\tau)}{\prod_{S \in \mathcal{S}_\tau} m_\tau(\boldsymbol{Y}_{S,\tau} \,|\, \boldsymbol{X}_\tau)}$$

- $m_\tau(\boldsymbol{Y}_{C,\tau} \,|\, \boldsymbol{X}_\tau)$ and $m_\tau(\boldsymbol{Y}_{S,\tau} \,|\, \boldsymbol{X}_\tau)$ computed under the *complete* graph using standard priors
  (no need for hyper-Inverse Wishart of the like)
- Finally obtain the marginal likelihood of the essential graph $\mathcal{G}$ is

$$m_{\mathcal{G}}(\boldsymbol{Y}) = \prod_{\tau \in \mathcal{T}} m_{\mathcal{G}_\tau}(\boldsymbol{Y}_\tau \,|\, \boldsymbol{X}_\tau)$$

A Markov chain on EG space

- $\mathcal{S}_q$, the set of all EGs on q nodes

A Markov chain on EG space

- $\mathcal{S}_q$, the set of all EGs on q nodes
- $\mathcal{S} \subset \mathcal{S}_q$, e.g. the set of all EGs on q nodes having fewer edges than a threshold M (sparsity constraint)

A Markov chain on EG space

- $\mathcal{S}_q$, the set of all EGs on q nodes
- $\mathcal{S} \subset \mathcal{S}_q$, e.g. the set of all EGs on q nodes having fewer edges than a threshold M (sparsity constraint)

- Define a set of operators (based on insertion or removal of directed edges, undirected edges and v-structures) that determine the transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ (He et al., 2013)

A Markov chain on EG space

- $\mathcal{S}_q$, the set of all EGs on q nodes
- $\mathcal{S} \subset \mathcal{S}_q$, e.g. the set of all EGs on q nodes having fewer edges than a threshold M (sparsity constraint)

- Define a set of operators (based on insertion or removal of directed edges, undirected edges and v-structures) that determine the transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ (He et al., 2013)
- For each $\mathcal{G} \in \mathcal{S}$ obtain a *perfect* set of operators $\mathcal{O}_{\mathcal{G}}$ inducing a Markov chain $\{\mathcal{G}_t\}$ on $\mathcal{S}$ satisfying *optimal* properties: *validity*; *distinguishability*; *irreducibility*; *reversibility*

A Markov chain on EG space

- $\mathcal{S}_q$, the set of all EGs on q nodes
- $\mathcal{S} \subset \mathcal{S}_q$, e.g. the set of all EGs on q nodes having fewer edges than a threshold M (sparsity constraint)

- Define a set of operators (based on insertion or removal of directed edges, undirected edges and v-structures) that determine the transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ (He et al., 2013)
- For each $\mathcal{G} \in \mathcal{S}$ obtain a *perfect* set of operators $\mathcal{O}_{\mathcal{G}}$ inducing a Markov chain $\{\mathcal{G}_t\}$ on $\mathcal{S}$ satisfying *optimal* properties: *validity*; *distinguishability*; *irreducibility*; *reversibility*
- Then the probability of transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ is

$$p_{\mathcal{G},\mathcal{G}'} = \begin{cases} 1/|\mathcal{O}_{\mathcal{G}}| & \text{if } \mathcal{G}' \in \mathcal{S}, \\ 0 & \text{otherwise.} \end{cases}$$

MCMC algorithm on Markov equivalence classes of DAGs

- The transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ is accepted with probability

$$\alpha = \min \left\{ 1; \frac{m_{\mathcal{G}'}(\boldsymbol{Y})}{m_{\mathcal{G}}(\boldsymbol{Y})} \cdot \frac{p(\mathcal{G}')}{p(\mathcal{G})} \cdot \frac{q(\mathcal{G} \mid \mathcal{G}')}{q(\mathcal{G}' \mid \mathcal{G})} \right\}$$

MCMC algorithm on Markov equivalence classes of DAGs

- The transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ is accepted with probability

$$\alpha = \min \left\{ 1; \frac{m_{\mathcal{G}'}(\boldsymbol{Y})}{m_{\mathcal{G}}(\boldsymbol{Y})} \cdot \frac{p(\mathcal{G}')}{p(\mathcal{G})} \cdot \frac{q(\mathcal{G} \mid \mathcal{G}')}{q(\mathcal{G}' \mid \mathcal{G})} \right\}$$

- $m_{\mathcal{G}}(\boldsymbol{Y})$ marginal likelihood of EG $\mathcal{G}$

MCMC algorithm on Markov equivalence classes of DAGs

- The transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ is accepted with probability

$$\alpha = \min \left\{ 1; \frac{m_{\mathcal{G}'}(\boldsymbol{Y})}{m_{\mathcal{G}}(\boldsymbol{Y})} \cdot \frac{p(\mathcal{G}')}{p(\mathcal{G})} \cdot \frac{q(\mathcal{G} \mid \mathcal{G}')}{q(\mathcal{G}' \mid \mathcal{G})} \right\}$$

- $m_{\mathcal{G}}(\boldsymbol{Y})$ marginal likelihood of EG $\mathcal{G}$
- $q(\mathcal{G}' \mid \mathcal{G}) = p_{\mathcal{G},\mathcal{G}'}$ proposal distribution

MCMC algorithm on Markov equivalence classes of DAGs

- The transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ is accepted with probability

$$\alpha = \min \left\{ 1; \frac{m_{\mathcal{G}'}(\boldsymbol{Y})}{m_{\mathcal{G}}(\boldsymbol{Y})} \cdot \frac{p(\mathcal{G}')}{p(\mathcal{G})} \cdot \frac{q(\mathcal{G} \mid \mathcal{G}')}{q(\mathcal{G}' \mid \mathcal{G})} \right\}$$

- $m_{\mathcal{G}}(\boldsymbol{Y})$ marginal likelihood of EG $\mathcal{G}$
- $q(\mathcal{G}' \mid \mathcal{G}) = p_{\mathcal{G},\mathcal{G}'}$ proposal distribution
- $p(\mathcal{G})$ prior on $\mathcal{G}$

MCMC algorithm on Markov equivalence classes of DAGs

- The transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ is accepted with probability

$$\alpha = \min \left\{ 1; \frac{m_{\mathcal{G}'}(\boldsymbol{Y})}{m_{\mathcal{G}}(\boldsymbol{Y})} \cdot \frac{p(\mathcal{G}')}{p(\mathcal{G})} \cdot \frac{q(\mathcal{G} \mid \mathcal{G}')}{q(\mathcal{G}' \mid \mathcal{G})} \right\}$$

- $m_{\mathcal{G}}(\boldsymbol{Y})$ marginal likelihood of EG $\mathcal{G}$
- $q(\mathcal{G}' \mid \mathcal{G}) = p_{\mathcal{G},\mathcal{G}'}$ proposal distribution
- $p(\mathcal{G})$ prior on $\mathcal{G}$
  - Assume a beta-binomial prior on the skeleton $\mathcal{G}^u$ of $\mathcal{G}$

MCMC algorithm on Markov equivalence classes of DAGs

- The transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ is accepted with probability

$$\alpha = \min \left\{ 1; \frac{m_{\mathcal{G}'}(\boldsymbol{Y})}{m_{\mathcal{G}}(\boldsymbol{Y})} \cdot \frac{p(\mathcal{G}')}{p(\mathcal{G})} \cdot \frac{q(\mathcal{G} \mid \mathcal{G}')}{q(\mathcal{G}' \mid \mathcal{G})} \right\}$$

- $m_{\mathcal{G}}(\boldsymbol{Y})$ marginal likelihood of EG $\mathcal{G}$
- $q(\mathcal{G}' \mid \mathcal{G}) = p_{\mathcal{G},\mathcal{G}'}$ proposal distribution
- $p(\mathcal{G})$ prior on $\mathcal{G}$
  - Assume a beta-binomial prior on the skeleton $\mathcal{G}^u$ of $\mathcal{G}$
  - $\mathcal{G}_{(j)}$ j-th element of the vectorized lower triangular part of the adjacency matrix of $\mathcal{G}^u$

MCMC algorithm on Markov equivalence classes of DAGs

- The transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ is accepted with probability

$$\alpha = \min\left\{1; \frac{m_{\mathcal{G}'}(\boldsymbol{Y})}{m_{\mathcal{G}}(\boldsymbol{Y})} \cdot \frac{p(\mathcal{G}')}{p(\mathcal{G})} \cdot \frac{q(\mathcal{G} \mid \mathcal{G}')}{q(\mathcal{G}' \mid \mathcal{G})}\right\}$$

- $m_{\mathcal{G}}(\boldsymbol{Y})$ marginal likelihood of EG $\mathcal{G}$
- $q(\mathcal{G}' \mid \mathcal{G}) = p_{\mathcal{G},\mathcal{G}'}$ proposal distribution
- $p(\mathcal{G})$ prior on $\mathcal{G}$
  - Assume a beta-binomial prior on the skeleton $\mathcal{G}^u$ of $\mathcal{G}$
  - $\mathcal{G}_{(j)}$ j-th element of the vectorized lower triangular part of the adjacency matrix of $\mathcal{G}^u$
  - $\mathcal{G}_{(j)} \mid \pi \sim \text{Ber}(\pi), \quad j = 1, \ldots, q(q-1)/2,$ independently

MCMC algorithm on Markov equivalence classes of DAGs

- The transition from $\mathcal{G} \in \mathcal{S}$ to $\mathcal{G}' \in \mathcal{S}$ is accepted with probability

$$\alpha = \min\left\{1; \frac{m_{\mathcal{G}'}(\boldsymbol{Y})}{m_{\mathcal{G}}(\boldsymbol{Y})} \cdot \frac{p(\mathcal{G}')}{p(\mathcal{G})} \cdot \frac{q(\mathcal{G} \mid \mathcal{G}')}{q(\mathcal{G}' \mid \mathcal{G})}\right\}$$

- $m_{\mathcal{G}}(\boldsymbol{Y})$ marginal likelihood of EG $\mathcal{G}$
- $q(\mathcal{G}' \mid \mathcal{G}) = p_{\mathcal{G},\mathcal{G}'}$ proposal distribution
- $p(\mathcal{G})$ prior on $\mathcal{G}$
  - Assume a beta-binomial prior on the skeleton $\mathcal{G}^u$ of $\mathcal{G}$
  - $\mathcal{G}_{(j)}$ j-th element of the vectorized lower triangular part of the adjacency matrix of $\mathcal{G}^u$
  - $\mathcal{G}_{(j)} \mid \pi \sim \text{Ber}(\pi), \quad j = 1, \ldots, q(q-1)/2,$ independently
  - $\pi \sim \text{Beta}(a, b)$

Simulation study

- We compare our Objective Bayes Essential graph Search (OBES) with three benchmark methods:

Simulation study

- We compare our Objective Bayes Essential graph Search (OBES) with three benchmark methods:
  - Greedy DAG Search algorithm
    (Chickering, 2002)
  - Greedy Equivalence Search with tuning parameter $\gamma \in \{0, 0.5, 1\}$
    (Chickering, 2002; Hauser and Bühlmann, 2012)
  - PC algorithm with confidence level $\alpha \in \{0.01, 0.05, 0.1\}$
    (Spirtes et al., 2000; Colombo and Maathuis, 2014)

Simulation study

- We compare our Objective Bayes Essential graph Search (OBES) with three benchmark methods:
  - Greedy DAG Search algorithm (Chickering, 2002)
  - Greedy Equivalence Search with tuning parameter $\gamma \in \{0, 0.5, 1\}$ (Chickering, 2002; Hauser and Bühlmann, 2012)
  - PC algorithm with confidence level $\alpha \in \{0.01, 0.05, 0.1\}$ (Spirtes et al., 2000; Colombo and Maathuis, 2014)
- OBES is *fully* Bayes, and produces a posterior distribution on $\mathcal{S}$.
  To obtain a summary of this distribution we define first the marginal probability of edge inclusion

$$p_{u \to v}(\boldsymbol{Y}) = \sum_{\mathcal{G} \in \mathcal{S}_{u \to v}} p(\mathcal{G} \mid \boldsymbol{Y})$$

Simulation study

- We compare our Objective Bayes Essential graph Search (OBES) with three benchmark methods:
  - Greedy DAG Search algorithm (Chickering, 2002)
  - Greedy Equivalence Search with tuning parameter $\gamma \in \{0, 0.5, 1\}$ (Chickering, 2002; Hauser and Bühlmann, 2012)
  - PC algorithm with confidence level $\alpha \in \{0.01, 0.05, 0.1\}$ (Spirtes et al., 2000; Colombo and Maathuis, 2014)
- OBES is *fully* Bayes, and produces a posterior distribution on $\mathcal{S}$.
  To obtain a summary of this distribution we define first the marginal probability of edge inclusion
$$p_{u \to v}(\boldsymbol{Y}) = \sum_{\mathcal{G} \in \mathcal{S}_{u \to v}} p(\mathcal{G} \mid \boldsymbol{Y})$$
- Next we take the *median probability (graph) model*, namely the graph containing only those directed edges $u \to v$ such that $p_{u \to v}(\boldsymbol{Y}) \geq 0.5$ (not an EG but a Partial DAG)

Simulation study

- We compare our Objective Bayes Essential graph Search (OBES) with three benchmark methods:
  - Greedy DAG Search algorithm (Chickering, 2002)
  - Greedy Equivalence Search with tuning parameter $\gamma \in \{0, 0.5, 1\}$ (Chickering, 2002; Hauser and Bühlmann, 2012)
  - PC algorithm with confidence level $\alpha \in \{0.01, 0.05, 0.1\}$ (Spirtes et al., 2000; Colombo and Maathuis, 2014)
- OBES is *fully* Bayes, and produces a posterior distribution on $\mathcal{S}$.
  To obtain a summary of this distribution we define first the marginal probability of edge inclusion

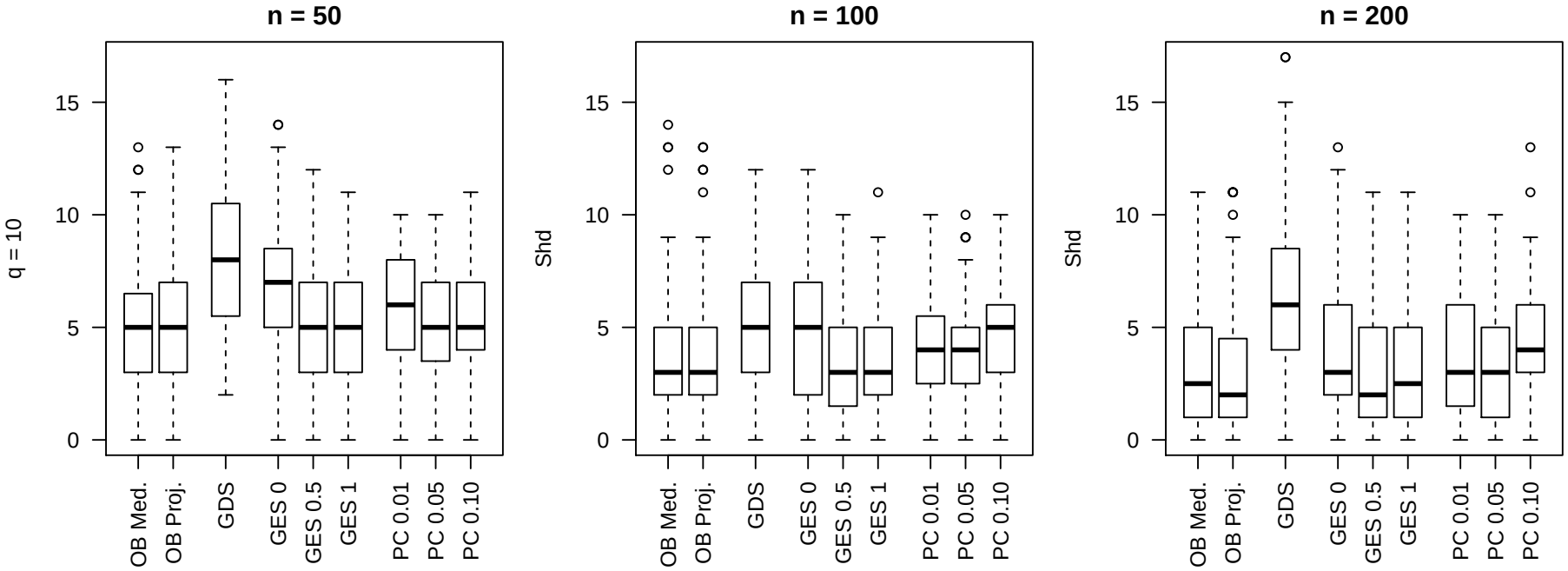
$$p_{u \to v}(\boldsymbol{Y}) = \sum_{\mathcal{G} \in \mathcal{S}_{u \to v}} p(\mathcal{G} \mid \boldsymbol{Y})$$

- Next we take the *median probability (graph) model*, namely the graph containing only those directed edges $u \to v$ such that $p_{u \to v}(\boldsymbol{Y}) \geq 0.5$ (not an EG but a Partial DAG)
- We make a consistent extension of the PDAG and finally we obtain its EG named *projected median probability model*

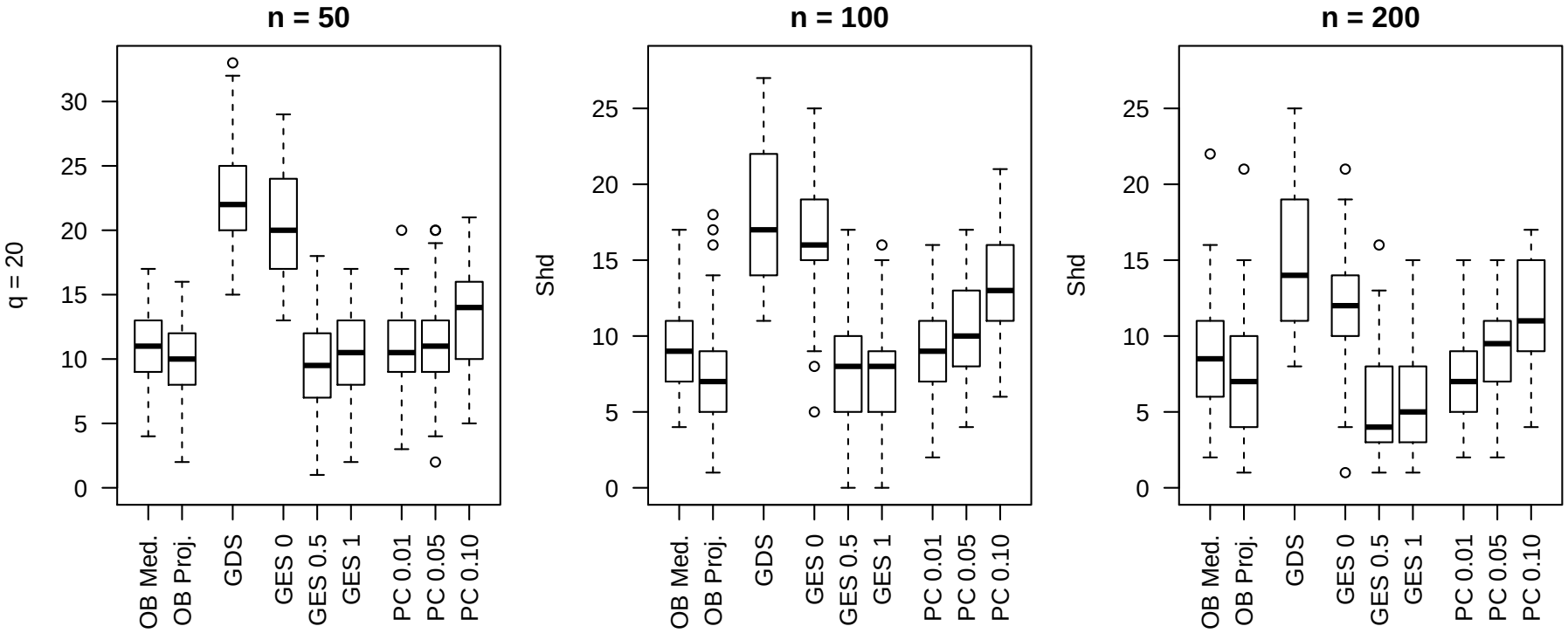
Simulation setting

- Consider $q \in \{10, 20\}$, $n \in \{50, 100, 200\}$
- For each scenario (n, q) generate 50 datasets
- To obtain each dataset, randomly generate a DAG $\mathcal{D}$ on q nodes and simulate n i.i.d observations from $\mathcal{N}(\mathbf{0}, \mathbf{\Omega}_{\mathcal{D}}^{-1})$
- For each method (OBES and benchmarks) compute the Structural Hamming Distances between estimates and true EG over the 50 simulations

Results


n = 50
n = 100
n = 200
q = 10
Shd
Shd
0
5
10
15
OB Med.
OB Proj.
GDS
GES 0
GES 0.5
GES 1
PC 0.01
PC 0.05
PC 0.10

Results


n = 50
n = 100
n = 200
q = 20
Shd
Shd
0
5
10
15
20
25
30
OB Med.
OB Proj.
GDS
GES 0
GES 0.5
GES 1
PC 0.01
PC 0.05
PC 0.10

Protein-signaling data

- Nine datasets from Sachs et al. (2005)

Protein-signaling data

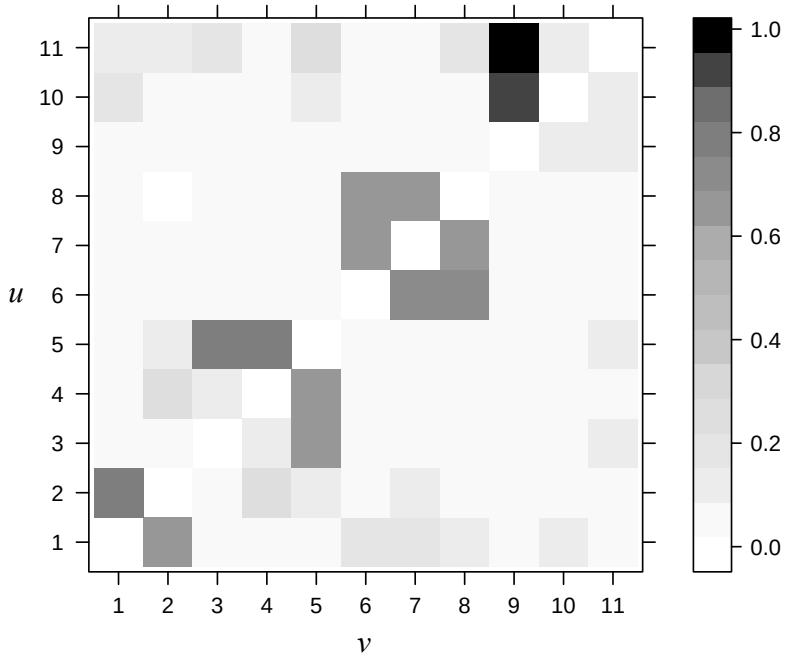
- Nine datasets from Sachs et al. (2005)
- Data include the levels of $q = 11$ phosphorylated proteins and phospholipids quantified using flow cytometry under different experimental conditions, each with sample size in the range 700-1000

Protein-signaling data

- Nine datasets from Sachs et al. (2005)
- Data include the levels of $q = 11$ phosphorylated proteins and phospholipids quantified using flow cytometry under different experimental conditions, each with sample size in the range 700-1000
- We implement OBES on each dataset setting the maximum number of edges at 22) and running 10^5 MCMC iterations

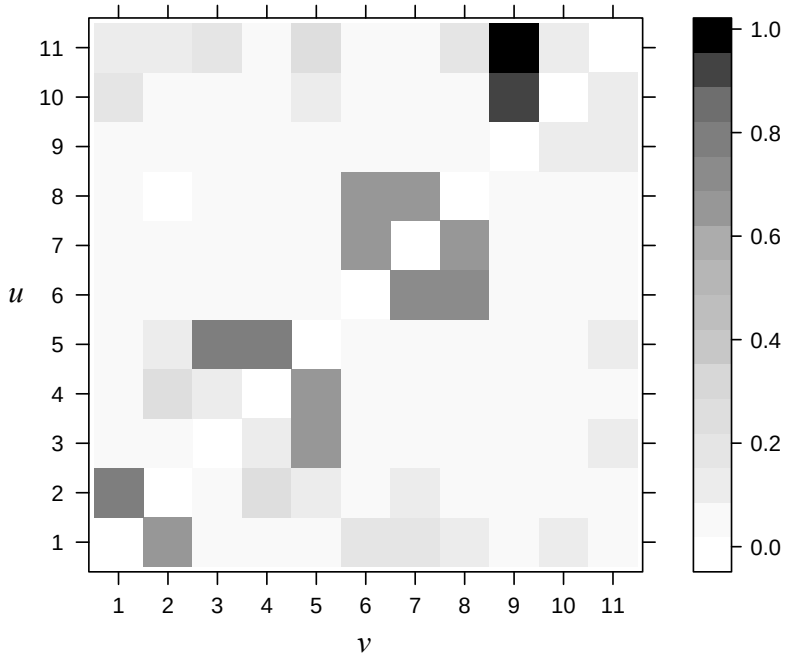
Results (Dataset I)

- Heat map with posterior probabilities of edge inclusion $p_{u \to v}(\boldsymbol{Y})$



Results (Dataset I)

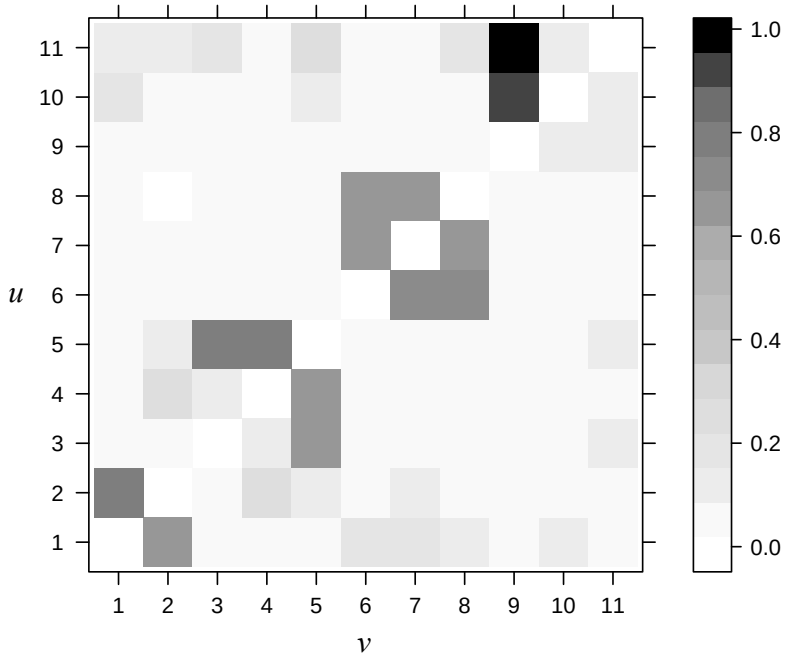
- Heat map with posterior probabilities of edge inclusion $p_{u \to v}(\boldsymbol{Y})$



- We observe $p_{u \to v}(\boldsymbol{Y}) \geq 0.5$ for

Results (Dataset I)

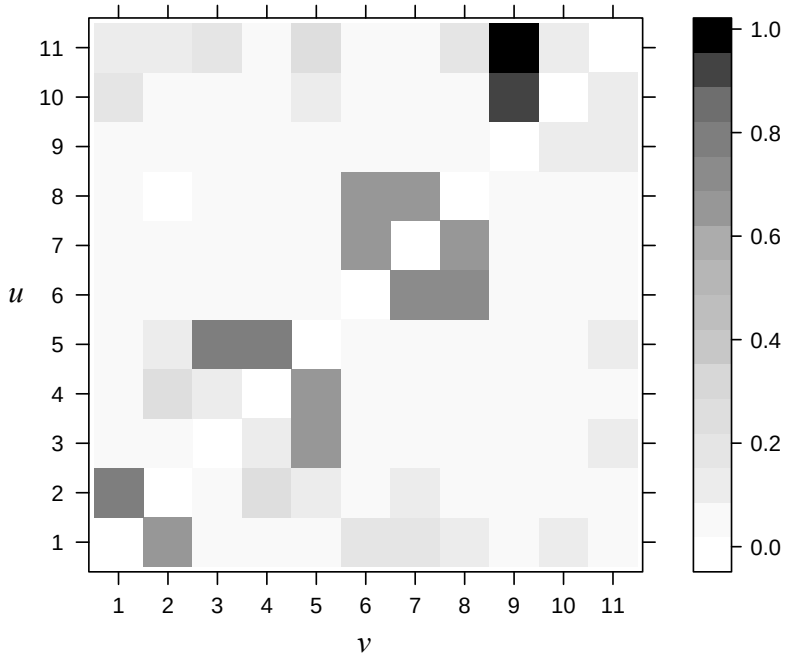
- Heat map with posterior probabilities of edge inclusion $p_{u \to v}(\boldsymbol{Y})$



- We observe $p_{u \to v}(\boldsymbol{Y}) \geq 0.5$ for
  - 6 undirected edges

Results (Dataset I)

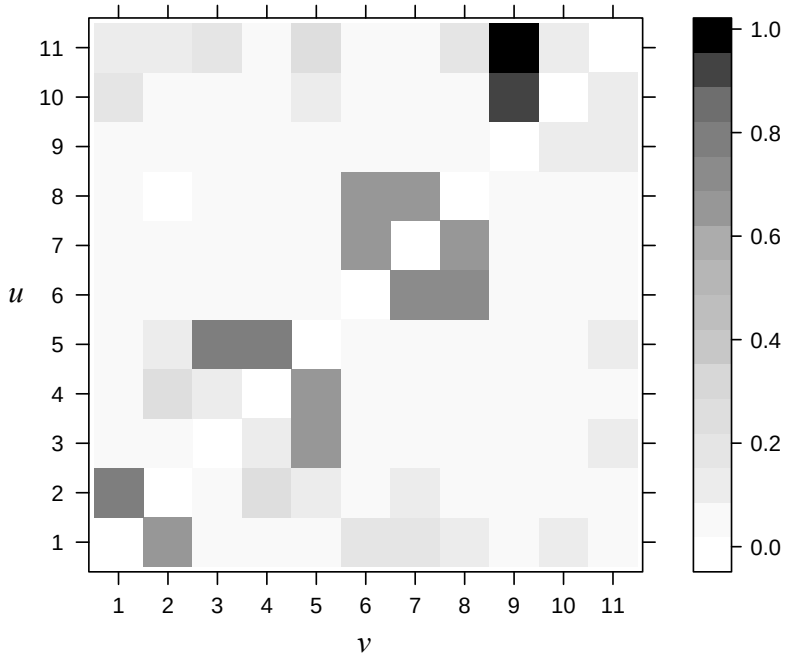
- Heat map with posterior probabilities of edge inclusion $p_{u \to v}(\boldsymbol{Y})$



- We observe $p_{u \to v}(\boldsymbol{Y}) \geq 0.5$ for
  - 6 undirected edges
  - 2 directed edges: $10 \to 9 \leftarrow 11$ (a v-structure)

Results (Dataset I)

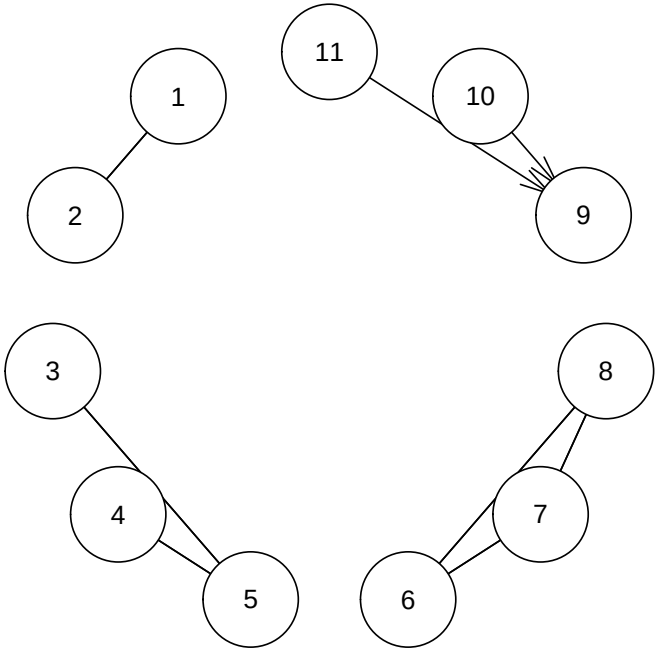
- Heat map with posterior probabilities of edge inclusion $p_{u \to v}(\boldsymbol{Y})$



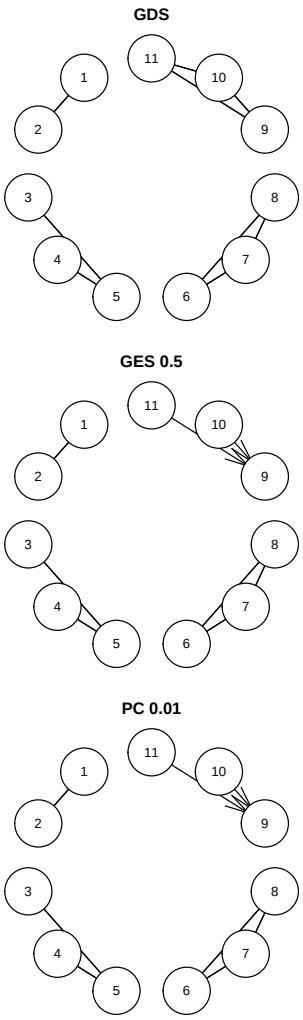
- We observe $p_{u \to v}(\boldsymbol{Y}) \geq 0.5$ for
  - 6 undirected edges
  - 2 directed edges: $10 \to 9 \leftarrow 11$ (a v-structure)

- For comparison with benchmarks we construct the median probability graph model including all edges $u \to v$ such that $p_{u \to v} \geq 0.5$

Results (Dataset I)

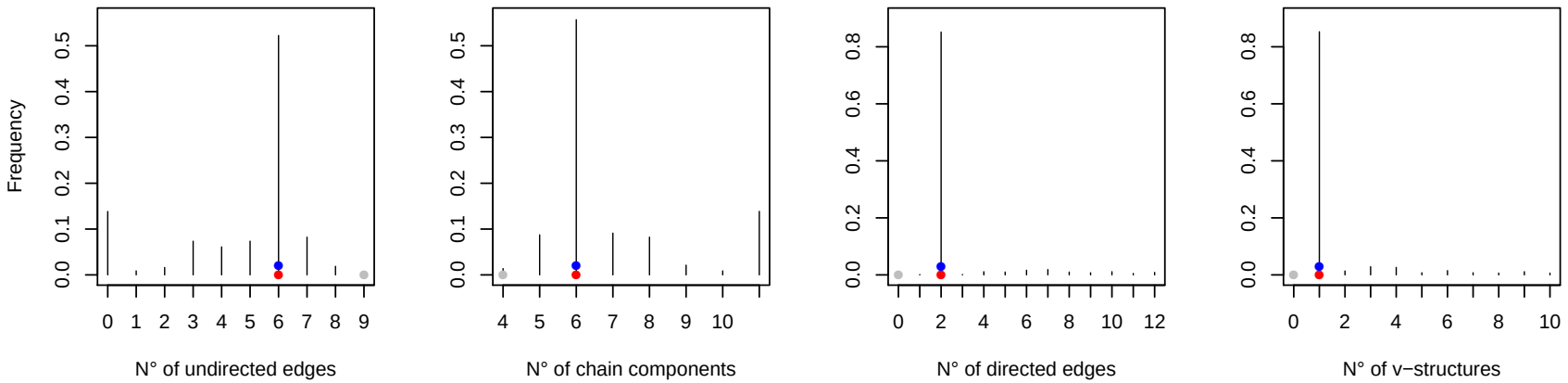


OBES median probability model



Results (Dataset I)

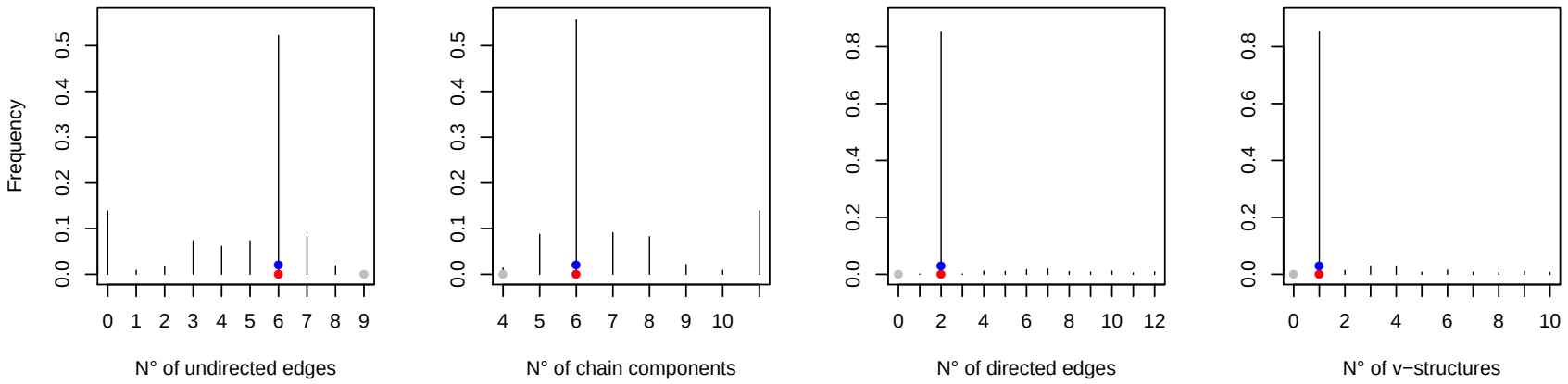
- Define the *quantile probability graph model* of order p^* as the graph including all edges $u \to v$ such that $p_{u \to v}(\boldsymbol{Y}) \geq p^*$



GDS (●), GES 0.5 (●), PC 0.01 (●)

Results (Dataset I)

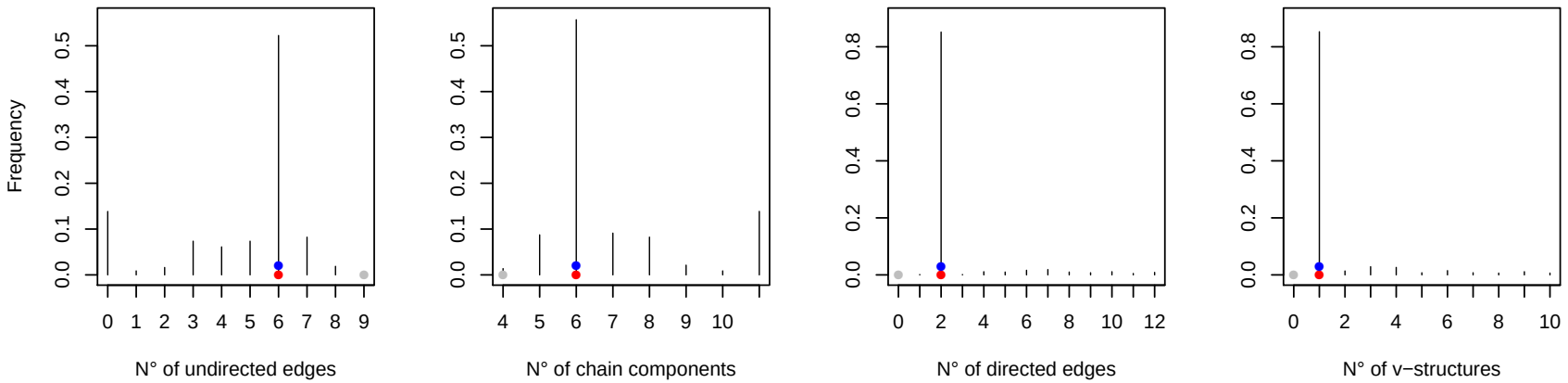
- Define the *quantile probability graph model* of order p^* as the graph including all edges $u \to v$ such that $p_{u \to v}(\boldsymbol{Y}) \geq p^*$
- For a grid of thresholds $p^* \in [0.1, 0.9]$ obtain a collection of quantile probability graph models



GDS (●), GES 0.5 (●), PC 0.01 (●)

Results (Dataset I)

- Define the *quantile probability graph model* of order p^* as the graph including all edges $u \to v$ such that $p_{u \to v}(\boldsymbol{Y}) \geq p^*$
- For a grid of thresholds $p^* \in [0.1, 0.9]$ obtain a collection of quantile probability graph models
- For each graph choose a feature of interest (e.g number of directed edges) and obtain its frequency distribution by varying p^*



GDS (●), GES 0.5 (●), PC 0.01 (●)

Summary

Summary

Summary

- We presented an objective Bayes method to obtain the marginal likelihood of an EG

Summary

- We presented an objective Bayes method to obtain the marginal likelihood of an EG
- We constructed an MCMC sampler to explore the space of EGs under sparsity constraints

Summary

- We presented an objective Bayes method to obtain the marginal likelihood of an EG
- We constructed an MCMC sampler to explore the space of EGs under sparsity constraints
- We compared our Objective Bayes Essential graph Search (OBES) method with current benchmarks

Summary

- We presented an objective Bayes method to obtain the marginal likelihood of an EG
- We constructed an MCMC sampler to explore the space of EGs under sparsity constraints
- We compared our Objective Bayes Essential graph Search (OBES) method with current benchmarks
- OBES is fully Bayes, and thus can provide an uncertainty evaluation of any feature of interest (e.g. the probability of inclusion of a particular edge) and not only a single estimate of the EG

Summary

- We presented an objective Bayes method to obtain the marginal likelihood of an EG
- We constructed an MCMC sampler to explore the space of EGs under sparsity constraints
- We compared our Objective Bayes Essential graph Search (OBES) method with current benchmarks
- OBES is fully Bayes, and thus can provide an uncertainty evaluation of any feature of interest (e.g. the probability of inclusion of a particular edge) and not only a single estimate of the EG
- Being objective, it is virtually free from prior specifications

On going research

- The Sachs data were collected under different experimental conditions

On going research

- The Sachs data were collected under different experimental conditions
- We analyzed them separately
  Extension to structural learning of *multiple* essential graphs with features potentially shared across graphs (Peterson et al., 2015)

On going research

- The Sachs data were collected under different experimental conditions
- We analyzed them separately
  Extension to structural learning of *multiple* essential graphs with features potentially shared across graphs (Peterson et al., 2015)
- Extension of the methodology to jointly model *observational* and *interventional* data
  (produced under different exogenous perturbations of variables or by randomized intervention experiments) (Hauser & Bülmann, 2015)

Main references

ANDERSSON, S. A., MADIGAN, D. & PERLMAN, M. D. (1997).
A characterization of Markov equivalence classes for acyclic digraphs.
Ann. Statist. **25**, 505–541.

BHADRA, A. & MALLICK, B. K. (2013).
Joint high-dimensional Bayesian variable and covariance selection with an application to eQTL analysis.
Biometrics **69**, 447–457.

CARVALHO, C. & SCOTT, J. (2012).
Objective Bayesian model selection in Gaussian graphical models.
Biometrika **96**, 497–512.

CASTELO, R. & PERLMAN, M. D. (2004).
Learning essential graph Markov models from data.
In *Advances in Bayesian networks*, vol. 146 of *Stud. Fuzziness Soft Comput.* Springer, Berlin, 255–269.

CHICKERING, D. M. (2002).
Learning equivalence classes of Bayesian-network structures.
Journal of Machine Learning Research **2 (3)**, 445–498.

Main references

CONSONNI, G. & LA ROCCA, L. (2012).
Objective Bayes factors for Gaussian directed acyclic graphical models.
Scandinavian Journal of Statistics **39**, 743–756.

CONSONNI, G., La Rocca, L. & Peluso, S. (2017).
Objective Bayes Covariate-Adjusted Sparse Graphical Model Selection.
Scandinavian Journal of Statistics, **44**, 741–764.

DAWID, A. P. & LAURITZEN, S. L. (1993).
Hyper Markov laws in the statistical analysis of decomposable graphical models.
The Annals of Statistics **21**, 1272–1317.

GEIGER, D. & HECKERMAN, D. (2002).
Parameter priors for directed acyclic graphical models and the characterization of several probability distributions.
The Annals of Statistics **30**, pp. 1412–1440.

HAUSER, A. & BÜHLMANN, P. (2015).
Jointly interventional and observational data: estimation of interventional Markov equivalence classes of directed acyclic graphs.
Journal of the Royal Statistical Society B **77**, 291–318.

Main references

HE, Y., JIA, J. & YU, B. (2013).
Reversible MCMC on Markov equivalence classes of sparse directed acyclic graphs.
Ann. Statist. **41**, 1742–1779.

O'HAGAN, A. (1995).
Fractional Bayes factors for model comparison.
Journal of the Royal Statistical Society B **57**, 99–138.

PETERSON, C., STINGO, F. C. & Vannucci, M. (2015).
Bayesian inference of multiple Gaussian graphical models.
Journal of the American Statistical Association, **110 (509)**, 159–174.

SACHS, K., Perez, O., Pe'er, D., Lauffenburger, D. A. & Nolan, G. P. (2005).
Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data.
Science, **308**, 523-529.

SPIRTES, P. AND GLYMOUR, C. AND SCHEINES, R. (2000).
Causation, Prediction and Search (2nd edition).
Cambridge, MA: The MIT Press., 1-16.

Supplementary slides

Gaussian multivariate regression DAGs

- Use theory developed in Consonni et al (2017) to perform objective Bayes learning for Gaussian multivariate regression DAG models
- They extend the methodology of Geiger and Heckerman (2002) for constructing compatible priors across DAG models
  - to the regression setting with covariate $\boldsymbol{X}$
  - to the objective Bayes setting
    (using the fractional Bayes factor)
  - Key ideas
    - Assign a prior on the *unconstrained* parameters for the *complete* DAG model
      (easy; can use standard conjugate matrix normal Wishart)
    - Show that the fractional prior belongs to the above conjugate family
    - Deduce the marginal distribution of Y_J with $J \subseteq V$ under the *complete* model
      (easy because of conjugacy)
    - If $\mathcal{G}$ is decomposable, marginal likelihood is

$$m_{\mathcal{G}}(\boldsymbol{Y} \mid \boldsymbol{X}) = \frac{\prod_{C \in \mathcal{C}} m(\boldsymbol{Y}_C \mid \boldsymbol{X})}{\prod_{S \in \mathcal{S}} m(\boldsymbol{Y}_S \mid \boldsymbol{X})}$$

    - Important: both $m(\boldsymbol{Y}_C)$ and $m(\boldsymbol{Y}_S)$ are in closed form and computed under the *complete* graph using standard priors
      (no need for hyper-Inverse Wishart of the like)

Properties of the Markov chain on EG space

- *validity*: the transition probability from $\mathcal{G}$ to $\mathcal{G}'$ that is not an EG is zero and the transition probability of an EG to itself is null

Properties of the Markov chain on EG space

- *validity*: the transition probability from $\mathcal{G}$ to $\mathcal{G}'$ that is not an EG is zero and the transition probability of an EG to itself is null
- *distinguishability*: for any $\mathcal{G}'$ *direct successor* of $\mathcal{G}$, there is a unique operator that transforms $\mathcal{G}$ in $\mathcal{G}'$

Properties of the Markov chain on EG space

- *validity*: the transition probability from $\mathcal{G}$ to $\mathcal{G}'$ that is not an EG is zero and the transition probability of an EG to itself is null
- *distinguishability*: for any $\mathcal{G}'$ *direct successor* of $\mathcal{G}$, there is a unique operator that transforms $\mathcal{G}$ in $\mathcal{G}'$
- *irreducibility*: starting from $\mathcal{G}$, there is a positive probability of reaching any other EG in $\mathcal{S}$ via a sequence of operators

Properties of the Markov chain on EG space

- *validity*: the transition probability from $\mathcal{G}$ to $\mathcal{G}'$ that is not an EG is zero and the transition probability of an EG to itself is null
- *distinguishability*: for any $\mathcal{G}'$ *direct successor* of $\mathcal{G}$, there is a unique operator that transforms $\mathcal{G}$ in $\mathcal{G}'$
- *irreducibility*: starting from $\mathcal{G}$, there is a positive probability of reaching any other EG in $\mathcal{S}$ via a sequence of operators
- *reversibility*: if $\mathcal{G}'$ is a direct successor of $\mathcal{G}$, then $\mathcal{G}$ is also a direct successor of $\mathcal{G}'$

MCMC algorithm

- Start from an arbitrary $\mathcal{G}_0$ (e.g. the null graph)
  For $t = 1, \ldots, T$

MCMC algorithm

- Start from an arbitrary $\mathcal{G}_0$ (e.g. the null graph)
  For $t = 1, \ldots, T$
  - set $\mathcal{G} = \mathcal{G}_{t-1}$

MCMC algorithm

- Start from an arbitrary $\mathcal{G}_0$ (e.g. the null graph)
  For $t = 1, \ldots, T$
  - set $\mathcal{G} = \mathcal{G}_{t-1}$
  - generate $\mathcal{G}'$ from the proposal $q(\mathcal{G}' \mid \mathcal{G})$

MCMC algorithm

- Start from an arbitrary $\mathcal{G}_0$ (e.g. the null graph)
  For $t = 1, \ldots, T$
  - set $\mathcal{G} = \mathcal{G}_{t-1}$
  - generate $\mathcal{G}'$ from the proposal $q(\mathcal{G}' \mid \mathcal{G})$
  - compute the probability of acceptance α

MCMC algorithm

- Start from an arbitrary $\mathcal{G}_0$ (e.g. the null graph)
  For $t = 1, \ldots, T$
  - set $\mathcal{G} = \mathcal{G}_{t-1}$
  - generate $\mathcal{G}'$ from the proposal $q(\mathcal{G}' \mid \mathcal{G})$
  - compute the probability of acceptance α
  - update $\mathcal{G}_t = \mathcal{G}'$ with probability α, $\mathcal{G}_t = \mathcal{G}_{t-1}$ with probability $1 - \alpha$

Data generation for simulations

- Fix q, number of nodes

Data generation for simulations

- Fix q, number of nodes
- Randomly generate a topologically ordered DAG with probability of edge inclusion $p_{edge} = 3/(2q-2)$

Data generation for simulations

- Fix q, number of nodes
- Randomly generate a topologically ordered DAG with probability of edge inclusion $p_{edge} = 3/(2q-2)$
- The DAG thus obtained implies the set of linear equations

$$Y_{i,j} = \mu_j + \sum_{k \in pa(j)} \beta_{k,j} Y_{i,k} + \varepsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, q$$

Data generation for simulations

- Fix q, number of nodes
- Randomly generate a topologically ordered DAG with probability of edge inclusion $p_{edge} = 3/(2q-2)$
- The DAG thus obtained implies the set of linear equations

$$Y_{i,j} = \mu_j + \sum_{k \in pa(j)} \beta_{k,j} Y_{i,k} + \varepsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, q$$

- with $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_j^2)$ independently

Data generation for simulations

- Fix q, number of nodes
- Randomly generate a topologically ordered DAG with probability of edge inclusion $p_{edge} = 3/(2q-2)$
- The DAG thus obtained implies the set of linear equations

$$Y_{i,j} = \mu_j + \sum_{k \in pa(j)} \beta_{k,j} Y_{i,k} + \varepsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, q$$

- with $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_j^2)$ independently
- $\mu_j <- 0$

Data generation for simulations

- Fix q, number of nodes
- Randomly generate a topologically ordered DAG with probability of edge inclusion $p_{edge} = 3/(2q-2)$
- The DAG thus obtained implies the set of linear equations

$$Y_{i,j} = \mu_j + \sum_{k \in pa(j)} \beta_{k,j} Y_{i,k} + \varepsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, q$$

- with $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_j^2)$ independently
- μ_j <- 0
- σ_j^2 <- `runif` $(0, 2)$

Data generation for simulations

- Fix q, number of nodes
- Randomly generate a topologically ordered DAG with probability of edge inclusion $p_{edge} = 3/(2q-2)$
- The DAG thus obtained implies the set of linear equations

$$Y_{i,j} = \mu_j + \sum_{k \in pa(j)} \beta_{k,j} Y_{i,k} + \varepsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, q$$

- with $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_j^2)$ independently
- μ_j <- 0
- σ_j^2 <- `runif` $(0, 2)$
- $\beta_{k,j}$ <- `runif` $[-1, -0.1] \cup [0.1, 1]$.

Data generation for simulations

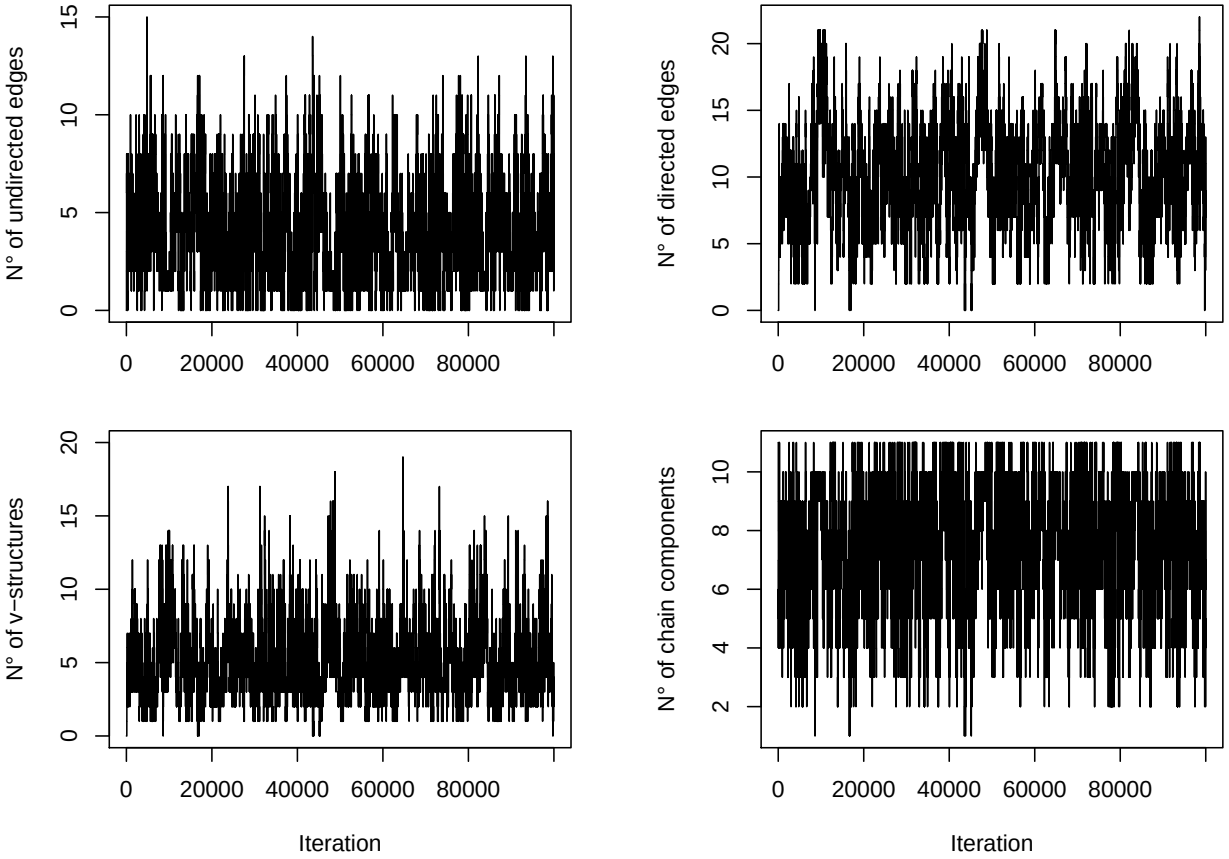
- Fix q, number of nodes
- Randomly generate a topologically ordered DAG with probability of edge inclusion $p_{edge} = 3/(2q-2)$
- The DAG thus obtained implies the set of linear equations

$$Y_{i,j} = \mu_j + \sum_{k \in pa(j)} \beta_{k,j} Y_{i,k} + \varepsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, q$$

- with $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_j^2)$ independently
- μ_j <- 0
- σ_j^2 <- `runif` $(0, 2)$
- $\beta_{k,j}$ <- `runif` $[-1, -0.1] \cup [0.1, 1]$.
- Generate datasets of size n accordingly

Diagnostics (Dataset I)

- MCMC traceplots of some graph features for the visited graphs



Further slides

Possibly add condition on n and n_0

Gaussian multivariate regression DAGs

- Use theory developed in Consonni et al (2017) to perform objective Bayes learning for Gaussian multivariate regression DAG models
- They extend the methodology of Geiger and Heckerman (2002) for constructing compatible priors across DAG models
  - to the regression setting with covariate $\boldsymbol{X}$
  - to the objective Bayes setting
    (using the fractional Bayes factor)
  - Key ideas
    - Assign a prior on the *unconstrained* parameters for the *complete* DAG model
      (easy; can use standard conjugate matrix normal Wishart)
    - Show that the fractional prior belongs to the above conjugate family
    - Deduce the marginal distribution of Y_J with $J \subseteq V$ under the *complete* model
      (easy because of conjugacy)
    - If $\mathcal{G}$ is decomposable, marginal likelihood is

$$m_{\mathcal{G}}(\boldsymbol{Y} \mid \boldsymbol{X}) = \frac{\prod_{C \in \mathcal{C}} m(\boldsymbol{Y}_C \mid \boldsymbol{X})}{\prod_{S \in \mathcal{S}} m(\boldsymbol{Y}_S \mid \boldsymbol{X})}$$

    - Important: both $m(\boldsymbol{Y}_C)$ and $m(\boldsymbol{Y}_S)$ are in closed form and computed under the *complete* graph using standard priors
      (no need for hyper-Inverse Wishart of the like)

Gaussian multivariate regression DAG models

- $\boldsymbol{Y} \,|\, \boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma} \sim \mathcal{N}_{n,q}(\boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \iff \mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_q(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$

Gaussian multivariate regression DAG models

- $\boldsymbol{Y} \mid \boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma} \sim \mathcal{N}_{n,q}(\boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \iff \mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_q(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$
  - $\boldsymbol{M}$ mean matrix

Gaussian multivariate regression DAG models

- $\boldsymbol{Y} \mid \boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma} \sim \mathcal{N}_{n,q}(\boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \iff \mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_q(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$
  - $\boldsymbol{M}$ mean matrix
  - $\boldsymbol{\Phi}$ row covariance matrix

Gaussian multivariate regression DAG models

- $\boldsymbol{Y} \,|\, \boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma} \sim \mathcal{N}_{n,q}(\boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \iff \mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_q(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$
  - $\boldsymbol{M}$ mean matrix
  - $\boldsymbol{\Phi}$ row covariance matrix
  - $\boldsymbol{\Sigma}$ column covariance matrix

Gaussian multivariate regression DAG models

- $\boldsymbol{Y} \,|\, \boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma} \sim \mathcal{N}_{n,q}(\boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \iff \mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_q(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$
  - $\boldsymbol{M}$ mean matrix
  - $\boldsymbol{\Phi}$ row covariance matrix
  - $\boldsymbol{\Sigma}$ column covariance matrix

- $\boldsymbol{Y} \,|\, \boldsymbol{B}, \boldsymbol{\Omega}_{\mathcal{D}} \sim \mathcal{N}_{n,q}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$

Gaussian multivariate regression DAG models

- $\boldsymbol{Y} \mid \boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma} \sim \mathcal{N}_{n,q}(\boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \iff \mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_q(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$
  - $\boldsymbol{M}$ mean matrix
  - $\boldsymbol{\Phi}$ row covariance matrix
  - $\boldsymbol{\Sigma}$ column covariance matrix

- $\boldsymbol{Y} \mid \boldsymbol{B}, \boldsymbol{\Omega}_{\mathcal{D}} \sim \mathcal{N}_{n,q}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
  - $\boldsymbol{Y}$ $n \times q$ matrix of observations on the responses

Gaussian multivariate regression DAG models

- $\boldsymbol{Y} \mid \boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma} \sim \mathcal{N}_{n,q}(\boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \iff \mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_q(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$
  - $\boldsymbol{M}$ mean matrix
  - $\boldsymbol{\Phi}$ row covariance matrix
  - $\boldsymbol{\Sigma}$ column covariance matrix

- $\boldsymbol{Y} \mid \boldsymbol{B}, \boldsymbol{\Omega}_{\mathcal{D}} \sim \mathcal{N}_{n,q}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
  - $\boldsymbol{Y}$ $n \times q$ matrix of observations on the responses
  - $\boldsymbol{X}$ $n \times (p+1)$ design matrix of observations on the p exogenous variables (including a column vector with all entries equal to 1 for the intercept term)

Gaussian multivariate regression DAG models

- $\boldsymbol{Y} \,|\, \boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma} \sim \mathcal{N}_{n,q}(\boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \iff \mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_q(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$
  - $\boldsymbol{M}$ mean matrix
  - $\boldsymbol{\Phi}$ row covariance matrix
  - $\boldsymbol{\Sigma}$ column covariance matrix

- $\boldsymbol{Y} \,|\, \boldsymbol{B}, \boldsymbol{\Omega}_{\mathcal{D}} \sim \mathcal{N}_{n,q}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
  - $\boldsymbol{Y}$ $n \times q$ matrix of observations on the responses
  - $\boldsymbol{X}$ $n \times (p+1)$ design matrix of observations on the p exogenous variables (including a column vector with all entries equal to 1 for the intercept term)
  - $\boldsymbol{B}$ $(p+1) \times q$ matrix of unconstrained regression parameters

Gaussian multivariate regression DAG models

- $\boldsymbol{Y} \,|\, \boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma} \sim \mathcal{N}_{n,q}(\boldsymbol{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \iff \mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{N}_q(\mathrm{vec}(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$
  - $\boldsymbol{M}$ mean matrix
  - $\boldsymbol{\Phi}$ row covariance matrix
  - $\boldsymbol{\Sigma}$ column covariance matrix

- $\boldsymbol{Y} \,|\, \boldsymbol{B}, \boldsymbol{\Omega}_{\mathcal{D}} \sim \mathcal{N}_{n,q}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_n, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$
  - $\boldsymbol{Y}$ $n \times q$ matrix of observations on the responses
  - $\boldsymbol{X}$ $n \times (p+1)$ design matrix of observations on the p exogenous variables (including a column vector with all entries equal to 1 for the intercept term)
  - $\boldsymbol{B}$ $(p+1) \times q$ matrix of unconstrained regression parameters
  - $\boldsymbol{\Omega}_{\mathcal{D}}$ $q \times q$ precision matrix, Markov with respect to a DAG $\mathcal{D}$

Case 1) $\mathcal{D}$ complete

- $\mathcal{D}$ complete $\implies$ all pairs of edges present (no conditional independencies among the q responses)

Case 1) $\mathcal{D}$ complete

- $\mathcal{D}$ complete $\implies$ all pairs of edges present (no conditional independencies among the q responses)
- $\mathbf{\Omega}$ (precision matrix) *unconstrained* (only s.p.d.)

Case 1) $\mathcal{D}$ complete

- $\mathcal{D}$ complete $\implies$ all pairs of edges present (no conditional independencies among the q responses)
- $\mathbf{\Omega}$ (precision matrix) *unconstrained* (only s.p.d.)
- Consonni et al. (2017) extended the methodology of Geiger & Heckerman (2002) for the construction of parameter priors to Gaussian regression DAG models using the fractional Bayes factor

Case 1) $\mathcal{D}$ complete

- $\mathcal{D}$ complete $\implies$ all pairs of edges present (no conditional independencies among the q responses)
- $\boldsymbol{\Omega}$ (precision matrix) *unconstrained* (only s.p.d.)
- Consonni et al. (2017) extended the methodology of Geiger & Heckerman (2002) for the construction of parameter priors to Gaussian regression DAG models using the fractional Bayes factor
- Start from the default prior $p^D(\boldsymbol{B}, \boldsymbol{\Omega}) \propto |\boldsymbol{\Omega}|^{\frac{a_D+q-1}{2}}$

Case 1) $\mathcal{D}$ complete

- $\mathcal{D}$ complete $\implies$ all pairs of edges present (no conditional independencies among the q responses)
- $\boldsymbol{\Omega}$ (precision matrix) *unconstrained* (only s.p.d.)
- Consonni et al. (2017) extended the methodology of Geiger & Heckerman (2002) for the construction of parameter priors to Gaussian regression DAG models using the fractional Bayes factor
- Start from the default prior $p^D(\boldsymbol{B}, \boldsymbol{\Omega}) \propto |\boldsymbol{\Omega}|^{\frac{a_D+q-1}{2}}$
- Obtain the fractional prior $p^F(\boldsymbol{B}, \boldsymbol{\Omega})$

Case 1) $\mathcal{D}$ complete

- $\mathcal{D}$ complete $\implies$ all pairs of edges present (no conditional independencies among the q responses)
- $\boldsymbol{\Omega}$ (precision matrix) *unconstrained* (only s.p.d.)
- Consonni et al. (2017) extended the methodology of Geiger & Heckerman (2002) for the construction of parameter priors to Gaussian regression DAG models using the fractional Bayes factor
- Start from the default prior $p^D(\boldsymbol{B}, \boldsymbol{\Omega}) \propto |\boldsymbol{\Omega}|^{\frac{a_D+q-1}{2}}$
- Obtain the fractional prior $p^F(\boldsymbol{B}, \boldsymbol{\Omega})$
- $p^F(\boldsymbol{B}, \boldsymbol{\Omega})$ is a matrix normal Wishart, which satisfies Geiger & Heckerman assumptions and produces a closed form expression for the marginal density of any (column) submatrix of $\boldsymbol{Y}$

Case 2) $\mathcal{D}$ not complete

- $\boldsymbol{\Omega}_{\mathcal{G}}$ Markov with respect to a decomposable UG $\mathcal{G}$

$$m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \boldsymbol{X}) = \frac{\prod_{C \in \mathcal{C}} m(\boldsymbol{Y}_C \,|\, \boldsymbol{X})}{\prod_{S \in \mathcal{S}} m(\boldsymbol{Y}_S \,|\, \boldsymbol{X})}$$

Case 2) $\mathcal{D}$ not complete

- $\boldsymbol{\Omega}_{\mathcal{G}}$ Markov with respect to a decomposable UG $\mathcal{G}$
- Marginal likelihood of $\mathcal{G}$

$$m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \boldsymbol{X}) = \frac{\prod_{C \in \mathcal{C}} m(\boldsymbol{Y}_C \,|\, \boldsymbol{X})}{\prod_{S \in \mathcal{S}} m(\boldsymbol{Y}_S \,|\, \boldsymbol{X})}$$

Case 2) $\mathcal{D}$ not complete

- $\boldsymbol{\Omega}_{\mathcal{G}}$ Markov with respect to a decomposable UG $\mathcal{G}$
- Marginal likelihood of $\mathcal{G}$

$$m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \boldsymbol{X}) = \frac{\prod_{C \in \mathcal{C}} m(\boldsymbol{Y}_C \,|\, \boldsymbol{X})}{\prod_{S \in \mathcal{S}} m(\boldsymbol{Y}_S \,|\, \boldsymbol{X})}$$

- $\mathcal{C}$ set of cliques

Case 2) $\mathcal{D}$ not complete

- $\boldsymbol{\Omega}_{\mathcal{G}}$ Markov with respect to a decomposable UG $\mathcal{G}$
- Marginal likelihood of $\mathcal{G}$

$$m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \boldsymbol{X}) = \frac{\prod_{C \in \mathcal{C}} m(\boldsymbol{Y}_C \,|\, \boldsymbol{X})}{\prod_{S \in \mathcal{S}} m(\boldsymbol{Y}_S \,|\, \boldsymbol{X})}$$

- $\mathcal{C}$ set of cliques
- $\mathcal{S}$ set of separators

Case 2) $\mathcal{D}$ not complete

- $\boldsymbol{\Omega}_{\mathcal{G}}$ Markov with respect to a decomposable UG $\mathcal{G}$
- Marginal likelihood of $\mathcal{G}$

$$m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \boldsymbol{X}) = \frac{\prod_{C \in \mathcal{C}} m(\boldsymbol{Y}_C \,|\, \boldsymbol{X})}{\prod_{S \in \mathcal{S}} m(\boldsymbol{Y}_S \,|\, \boldsymbol{X})}$$

- $\mathcal{C}$ set of cliques
- $\mathcal{S}$ set of separators
- $\boldsymbol{Y}_C$ submatrix of responses belonging to $C \in \mathcal{C}$

Case 2) $\mathcal{D}$ not complete

- $\boldsymbol{\Omega}_{\mathcal{G}}$ Markov with respect to a decomposable UG $\mathcal{G}$
- Marginal likelihood of $\mathcal{G}$

$$m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \boldsymbol{X}) = \frac{\prod_{C \in \mathcal{C}} m(\boldsymbol{Y}_C \,|\, \boldsymbol{X})}{\prod_{S \in \mathcal{S}} m(\boldsymbol{Y}_S \,|\, \boldsymbol{X})}$$

- $\mathcal{C}$ set of cliques
- $\mathcal{S}$ set of separators
- $\boldsymbol{Y}_C$ submatrix of responses belonging to $C \in \mathcal{C}$
- $\boldsymbol{Y}_S$ submatrix of responses belonging to $S \in \mathcal{S}$

Case 2) $\mathcal{D}$ not complete

- $\boldsymbol{\Omega}_{\mathcal{G}}$ Markov with respect to a decomposable UG $\mathcal{G}$
- Marginal likelihood of $\mathcal{G}$

$$m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \boldsymbol{X}) = \frac{\prod_{C \in \mathcal{C}} m(\boldsymbol{Y}_C \,|\, \boldsymbol{X})}{\prod_{S \in \mathcal{S}} m(\boldsymbol{Y}_S \,|\, \boldsymbol{X})}$$

- $\mathcal{C}$ set of cliques
- $\mathcal{S}$ set of separators
- $\boldsymbol{Y}_C$ submatrix of responses belonging to $C \in \mathcal{C}$
- $\boldsymbol{Y}_S$ submatrix of responses belonging to $S \in \mathcal{S}$

- Important: $m(\boldsymbol{Y}_C)$ is computed under any *complete* graph using the matrix normal Wishart fractional prior (and similarly for $m(\boldsymbol{Y}_S)$)

Case 2) $\mathcal{D}$ not complete

- $\boldsymbol{\Omega}_{\mathcal{G}}$ Markov with respect to a decomposable UG $\mathcal{G}$
- Marginal likelihood of $\mathcal{G}$

$$m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \boldsymbol{X}) = \frac{\prod_{C \in \mathcal{C}} m(\boldsymbol{Y}_C \,|\, \boldsymbol{X})}{\prod_{S \in \mathcal{S}} m(\boldsymbol{Y}_S \,|\, \boldsymbol{X})}$$

- $\mathcal{C}$ set of cliques
- $\mathcal{S}$ set of separators
- $\boldsymbol{Y}_C$ submatrix of responses belonging to $C \in \mathcal{C}$
- $\boldsymbol{Y}_S$ submatrix of responses belonging to $S \in \mathcal{S}$

- Important: $m(\boldsymbol{Y}_C)$ is computed under any *complete* graph using the matrix normal Wishart fractional prior (and similarly for $m(\boldsymbol{Y}_S)$)
- Use (1) to obtain the marginal likelihood of an EG.

EG Chain component τ

Fractional prior

$$
\begin{aligned}
p^F(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \quad &\propto \quad |\boldsymbol{\Omega}_\tau|^{\frac{a_D + n_0 - |\mathrm{pa}_{\mathcal{G}}(\tau)| - |\tau| - 2}{2}} \\
&\cdot \quad e^{-\frac{n_0}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_\tau \left\{ (\boldsymbol{B}_\tau - \hat{\boldsymbol{B}}_\tau)^\top \tilde{\boldsymbol{C}}_\tau \, (\boldsymbol{B}_\tau - \hat{\boldsymbol{B}}_\tau) + \tilde{\boldsymbol{R}}_\tau \right\}\right)},
\end{aligned}
$$

EG Chain component τ

Fractional prior

$$
\begin{aligned}
p^F(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \quad &\propto \quad |\boldsymbol{\Omega}_\tau|^{\frac{a_D + n_0 - |\mathrm{pa}_{\mathcal{G}}(\tau)| - |\tau| - 2}{2}} \\
&\cdot \quad e^{-\frac{n_0}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_\tau \left\{ (\boldsymbol{B}_\tau - \hat{\boldsymbol{B}}_\tau)^\top \tilde{\boldsymbol{C}}_\tau \, (\boldsymbol{B}_\tau - \hat{\boldsymbol{B}}_\tau) + \tilde{\boldsymbol{R}}_\tau \right\}\right)},
\end{aligned}
$$

- $\hat{\boldsymbol{B}}_\tau = (\boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau)^{-1} \boldsymbol{X}_\tau^\top \boldsymbol{Y}_\tau$

EG Chain component τ

Fractional prior

$$
\begin{aligned}
p^F(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \quad &\propto \quad |\boldsymbol{\Omega}_\tau|^{\frac{a_D + n_0 - |\mathrm{pa}_{\mathcal{G}}(\tau)| - |\tau| - 2}{2}} \\
&\cdot \quad e^{-\frac{n_0}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_\tau \left\{ (\boldsymbol{B}_\tau - \hat{\boldsymbol{B}}_\tau)^\top \tilde{\boldsymbol{C}}_\tau \, (\boldsymbol{B}_\tau - \hat{\boldsymbol{B}}_\tau) + \tilde{\boldsymbol{R}}_\tau \right\}\right)},
\end{aligned}
$$

- $\hat{\boldsymbol{B}}_\tau = (\boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau)^{-1} \boldsymbol{X}_\tau^\top \boldsymbol{Y}_\tau$
- $\hat{\boldsymbol{E}}_\tau = (\boldsymbol{Y}_\tau - \boldsymbol{X}_\tau \hat{\boldsymbol{B}}_\tau)$

EG Chain component τ

Fractional prior

$$
\begin{aligned}
p^F(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \quad &\propto \quad |\boldsymbol{\Omega}_\tau|^{\frac{a_D + n_0 - |\mathrm{pa}_{\mathcal{G}}(\tau)| - |\tau| - 2}{2}} \\
&\cdot \quad e^{-\frac{n_0}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_\tau \left\{ (\boldsymbol{B}_\tau - \hat{\boldsymbol{B}}_\tau)^\top \tilde{\boldsymbol{C}}_\tau \, (\boldsymbol{B}_\tau - \hat{\boldsymbol{B}}_\tau) + \tilde{\boldsymbol{R}}_\tau \right\}\right)},
\end{aligned}
$$

- $\hat{\boldsymbol{B}}_\tau = (\boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau)^{-1} \boldsymbol{X}_\tau^\top \boldsymbol{Y}_\tau$
- $\hat{\boldsymbol{E}}_\tau = (\boldsymbol{Y}_\tau - \boldsymbol{X}_\tau \hat{\boldsymbol{B}}_\tau)$
- $\tilde{\boldsymbol{C}}_\tau = n^{-1} \boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau$

EG Chain component τ

Fractional prior

$$
\begin{aligned}
p^F(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \quad &\propto \quad |\boldsymbol{\Omega}_\tau|^{\frac{a_D + n_0 - |\mathrm{pa}_{\mathcal{G}}(\tau)| - |\tau| - 2}{2}} \\
&\cdot \quad e^{-\frac{n_0}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_\tau \left\{ (\boldsymbol{B}_\tau - \hat{\boldsymbol{B}}_\tau)^\top \tilde{\boldsymbol{C}}_\tau \, (\boldsymbol{B}_\tau - \hat{\boldsymbol{B}}_\tau) + \tilde{\boldsymbol{R}}_\tau \right\}\right)},
\end{aligned}
$$

- $\hat{\boldsymbol{B}}_\tau = (\boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau)^{-1} \boldsymbol{X}_\tau^\top \boldsymbol{Y}_\tau$
- $\hat{\boldsymbol{E}}_\tau = (\boldsymbol{Y}_\tau - \boldsymbol{X}_\tau \hat{\boldsymbol{B}}_\tau)$
- $\tilde{\boldsymbol{C}}_\tau = n^{-1} \boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau$
- $\tilde{\boldsymbol{R}}_\tau = n^{-1} \hat{\boldsymbol{E}}_\tau^\top \hat{\boldsymbol{E}}_\tau$

EG Chain component τ

Fractional prior

$$p^F(\boldsymbol{B}_\tau, \boldsymbol{\Omega}_\tau) \propto |\boldsymbol{\Omega}_\tau|^{\frac{a_D+n_0-|\mathrm{pa}_{\mathcal{G}}(\tau)|-|\tau|-2}{2}} \cdot e^{-\frac{n_0}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_\tau\left\{(\boldsymbol{B}_\tau-\hat{\boldsymbol{B}}_\tau)^\top \tilde{\boldsymbol{C}}_\tau (\boldsymbol{B}_\tau-\hat{\boldsymbol{B}}_\tau)+\tilde{\boldsymbol{R}}_\tau\right\}\right)},$$

- $\hat{\boldsymbol{B}}_\tau = (\boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau)^{-1}\boldsymbol{X}_\tau^\top \boldsymbol{Y}_\tau$
- $\hat{\boldsymbol{E}}_\tau = (\boldsymbol{Y}_\tau - \boldsymbol{X}_\tau \hat{\boldsymbol{B}}_\tau)$
- $\tilde{\boldsymbol{C}}_\tau = n^{-1}\boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau$
- $\tilde{\boldsymbol{R}}_\tau = n^{-1}\hat{\boldsymbol{E}}_\tau^\top \hat{\boldsymbol{E}}_\tau$

- Matrix normal Wishart
  conjugate to the conditional sampling model in chain component τ

Marginal density of data submatrix in τ

$$
\begin{aligned}
m_\tau(\boldsymbol{Y}_{J,\tau} \mid \boldsymbol{X}_\tau) \quad &= \quad \pi^{-\frac{(n-n_0)|J|}{2}} \frac{\Gamma_{|J|}\left(\frac{a_D+n-|\mathrm{pa}_{\mathcal{G}}(\tau)|-1-|\bar{J}|}{2}\right)}{\Gamma_{|J|}\left(\frac{a_D+n_0-|\mathrm{pa}_{\mathcal{G}}(\tau)|-1-|\bar{J}|}{2}\right)} \\
&\cdot \quad \left(\frac{n_0}{n}\right)^{\frac{|J|(a_D+n_0-|\bar{J}|)}{2}} |\hat{\boldsymbol{E}}_{J,\tau}^\top \hat{\boldsymbol{E}}_{J,\tau}|^{-\frac{n-n_0}{2}}
\end{aligned}
$$

- $\bar{J} = \tau \setminus J$, $|\bar{J}| = |\tau| - |J|$

Marginal density of data submatrix in τ

$$
\begin{aligned}
m_\tau(\boldsymbol{Y}_{J,\tau} \mid \boldsymbol{X}_\tau) \quad &= \quad \pi^{-\frac{(n-n_0)|J|}{2}} \frac{\Gamma_{|J|}\left(\frac{a_D+n-|\mathrm{pa}_{\mathcal{G}}(\tau)|-1-|\bar{J}|}{2}\right)}{\Gamma_{|J|}\left(\frac{a_D+n_0-|\mathrm{pa}_{\mathcal{G}}(\tau)|-1-|\bar{J}|}{2}\right)} \\
&\cdot \quad \left(\frac{n_0}{n}\right)^{\frac{|J|(a_D+n_0-|\bar{J}|)}{2}} |\hat{\boldsymbol{E}}_{J,\tau}^\top \hat{\boldsymbol{E}}_{J,\tau}|^{-\frac{n-n_0}{2}}
\end{aligned}
$$

- $\bar{J} = \tau \setminus J$, $|\bar{J}| = |\tau| - |J|$
- $\hat{\boldsymbol{E}}_{J,\tau} = (\boldsymbol{Y}_{J,\tau} - \boldsymbol{X}_\tau \hat{\boldsymbol{B}}_{J,\tau})$

Marginal density of data submatrix in τ

$$
\begin{aligned}
m_\tau(\boldsymbol{Y}_{J,\tau} \mid \boldsymbol{X}_\tau) \quad &= \quad \pi^{-\frac{(n-n_0)|J|}{2}} \frac{\Gamma_{|J|}\left(\frac{a_D+n-|\mathrm{pa}_{\mathcal{G}}(\tau)|-1-|\bar{J}|}{2}\right)}{\Gamma_{|J|}\left(\frac{a_D+n_0-|\mathrm{pa}_{\mathcal{G}}(\tau)|-1-|\bar{J}|}{2}\right)} \\
&\cdot \quad \left(\frac{n_0}{n}\right)^{\frac{|J|(a_D+n_0-|\bar{J}|)}{2}} |\hat{\boldsymbol{E}}_{J,\tau}^\top \hat{\boldsymbol{E}}_{J,\tau}|^{-\frac{n-n_0}{2}}
\end{aligned}
$$

- $\bar{J} = \tau \setminus J$, $|\bar{J}| = |\tau| - |J|$
- $\hat{\boldsymbol{E}}_{J,\tau} = (\boldsymbol{Y}_{J,\tau} - \boldsymbol{X}_\tau \hat{\boldsymbol{B}}_{J,\tau})$
- $\hat{\boldsymbol{B}}_{J,\tau} = (\boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau)^{-1} \boldsymbol{X}_\tau^\top \boldsymbol{Y}_{J,\tau}$