Objective Bayes Learning of Graphical Models

Guido Consonni
Università Cattolica del Sacro Cuore
Milan, Italy

Jan 20, 2017

Joint work with
Luca La Rocca (Università di Modena e Reggio Emilia)
Stefano Peluso (Università Cattolica del Sacro Cuore - Milan)

Outline

Graphical models

Objective Bayes Model Selection

DAG-models

Compatible Parameter Priors

Marginal Likelihood

Covariate-adjusted graphical models

Experimental results

Graphs

Graphs

Graph $\mathcal{G}$

$\mathcal{G} = (V, E)$

V: set of vertices

E: set of edges

$|V| = q$

Graphs

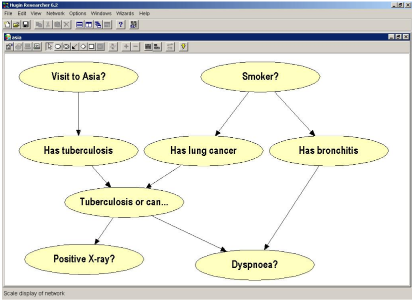
Graph $\mathcal{G}$

$\mathcal{G} = (V, E)$

V: set of vertices

E: set of edges

$|V| = q$



Graphs

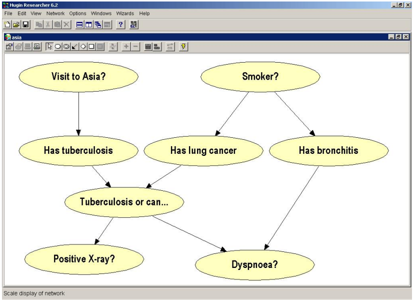
Graph $\mathcal{G}$

$\mathcal{G} = (V, E)$

V: set of vertices

E: set of edges

$|V| = q$



Graphs

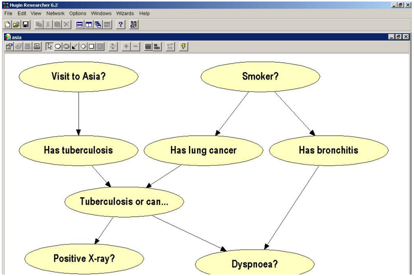
Graph $\mathcal{G}$

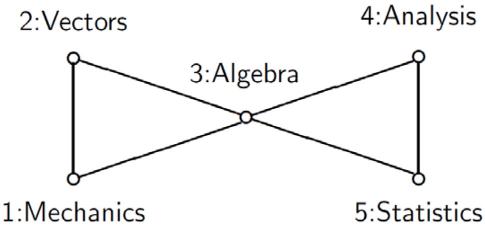
$\mathcal{G} = (V, E)$

V: set of vertices

E: set of edges

$|V| = q$





Graphical Models

When every edge in E is undirected, $\mathcal{G}$ is an undirected graph (UG).

When every edge in E is directed, $\mathcal{G}$ is a directed graph.

If a directed graph $\mathcal{G}$ has no directed cycles, then $\mathcal{G}$ is a DAG ($\mathcal{D}$).

Graphical Models

When every edge in E is undirected, $\mathcal{G}$ is an undirected graph (UG).

When every edge in E is directed, $\mathcal{G}$ is a directed graph.

If a directed graph $\mathcal{G}$ has no directed cycles, then $\mathcal{G}$ is a DAG ($\mathcal{D}$).

Given $\mathcal{G}$, a family of probability distributions for $\boldsymbol{y}_i^\top = (y_{i1}, \ldots, y_{iq})$ which factorize according to the graph $\mathcal{G}$ is called a graphical model (wrt $\mathcal{G}$).

Factorizations

If $\mathcal{G} = \mathcal{D}$ is a DAG

$$f_{\mathcal{D}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}, \boldsymbol{\theta}_j)$$

$\boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)} = \{y_{il} : l \in \mathrm{pa}_{\mathcal{D}}(j)\}$; $\mathrm{pa}_{\mathcal{D}}(j)$: parents of j

Factorizations

If $\mathcal{G} = \mathcal{D}$ is a DAG

$$f_{\mathcal{D}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}, \boldsymbol{\theta}_j)$$

$\boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)} = \{y_{il} : l \in \mathrm{pa}_{\mathcal{D}}(j)\}$; $\mathrm{pa}_{\mathcal{D}}(j)$: parents of j

Factorizations

If $\mathcal{G} = \mathcal{D}$ is a DAG

$$f_{\mathcal{D}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}, \boldsymbol{\theta}_j)$$

$\boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)} = \{y_{il} : l \in \mathrm{pa}_{\mathcal{D}}(j)\}$; $\mathrm{pa}_{\mathcal{D}}(j)$: parents of j

Factorizations

If $\mathcal{G} = \mathcal{D}$ is a DAG

$$f_{\mathcal{D}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}, \boldsymbol{\theta}_j)$$

$\boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)} = \{y_{il} : l \in \mathrm{pa}_{\mathcal{D}}(j)\}$; $\mathrm{pa}_{\mathcal{D}}(j)$: parents of j

If $\mathcal{G}$ is decomposable

$$f_{\mathcal{G}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{G}}) = \frac{\prod_{C \in \mathcal{C}} f(\boldsymbol{y}_{i,C} \mid \boldsymbol{\theta}_C)}{\prod_{S \in \mathcal{S}} f(\boldsymbol{y}_{i,S} \mid \boldsymbol{\theta}_S)}$$

$\mathcal{C}$: set of cliques; $\mathcal{S}$: set of separators.
$\boldsymbol{y}_{i,C} = \{y_{ij} : j \in C\}$.

(decomposable=chordal=triangulated)

Factorizations

If $\mathcal{G} = \mathcal{D}$ is a DAG

$$f_{\mathcal{D}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}, \boldsymbol{\theta}_j)$$

$\boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)} = \{y_{il} : l \in \mathrm{pa}_{\mathcal{D}}(j)\}$; $\mathrm{pa}_{\mathcal{D}}(j)$: parents of j

If $\mathcal{G}$ is decomposable

$$f_{\mathcal{G}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{G}}) = \frac{\prod_{C \in \mathcal{C}} f(\boldsymbol{y}_{i,C} \mid \boldsymbol{\theta}_C)}{\prod_{S \in \mathcal{S}} f(\boldsymbol{y}_{i,S} \mid \boldsymbol{\theta}_S)}$$

$\mathcal{C}$: set of cliques; $\mathcal{S}$: set of separators.
$\boldsymbol{y}_{i,C} = \{y_{ij} : j \in C\}$.

(decomposable=chordal=triangulated)

Factorizations

If $\mathcal{G} = \mathcal{D}$ is a DAG

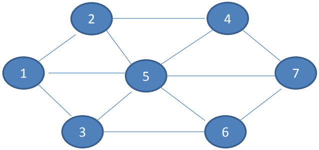
$$f_{\mathcal{D}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}, \boldsymbol{\theta}_j)$$

$\boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)} = \{y_{il} : l \in \mathrm{pa}_{\mathcal{D}}(j)\}$; $\mathrm{pa}_{\mathcal{D}}(j)$: parents of j

If $\mathcal{G}$ is decomposable

$$f_{\mathcal{G}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{G}}) = \frac{\prod_{C \in \mathcal{C}} f(\boldsymbol{y}_{i,C} \mid \boldsymbol{\theta}_C)}{\prod_{S \in \mathcal{S}} f(\boldsymbol{y}_{i,S} \mid \boldsymbol{\theta}_S)}$$

$\mathcal{C}$: set of cliques; $\mathcal{S}$: set of separators.
$\boldsymbol{y}_{i,C} = \{y_{ij} : j \in C\}$.



(decomposable=chordal=triangulated)

$\mathcal{C} = \{\{1,2,5\},\{1,3,5\},\{2,4,5\},\{3,5,6\},\{4,5,7\},\{5,6,7\}\}$

$\mathcal{S} = \{\{1,5\},\{2,5\},\{3,5\},\{4,5\},\{5,6\}\}$

Markov properties

$\mathcal{G} \equiv \mathcal{D}$: DAG

Local Markov property

$\forall u \in V$

$$u \perp\!\!\!\perp \{nd(u) \setminus pa(u)\} \mid pa(u)$$

Markov properties

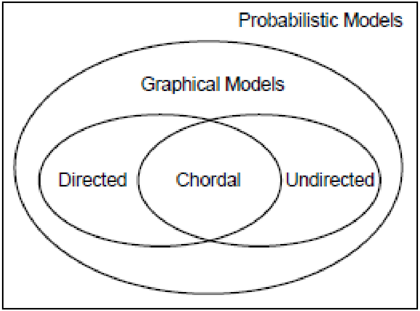
$\mathcal{G} \equiv \mathcal{D}$: DAG
Local Markov property
$\forall u \in V$

$$u \perp\!\!\!\perp \{nd(u) \setminus pa(u)\} \mid pa(u)$$

$\mathcal{G}$: UG
A, B, S disjoint subsets of V
Global Markov property
If S separates A from B in $\mathcal{G}$, then

$$A \perp\!\!\!\perp B \mid S$$

Relationships Among Graphical Models



Selection of Graphical Models

Typically we do NOT know the structure of the graph

Aim

Discover the graph using data

Selection of Graphical Models

Typically we do NOT know the structure of the graph

Aim

Discover the graph using data

Structural learning

Selection of Graphical Models

Typically we do NOT know the structure of the graph

Aim
Discover the graph using data

Structural learning

Graph selection

Bayes factor

Bayesian model $\mathcal{M}_k = \{f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k), p(\boldsymbol{\theta}_k)\}$

Bayes factor

Bayesian model $\mathcal{M}_k = \{f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k), p(\boldsymbol{\theta}_k)\}$

$\mathcal{M}_1, \ldots, \mathcal{M}_K$: K models

Bayes factor

Bayesian model $\mathcal{M}_k = \{f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k), p(\boldsymbol{\theta}_k)\}$

$\mathcal{M}_1, \dots, \mathcal{M}_K$: K models

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

Bayes factor

Bayesian model $\mathcal{M}_k = \{f_{\mathcal{M}_k}(\boldsymbol{Y} \mid \boldsymbol{\theta}_k), p(\boldsymbol{\theta}_k)\}$

$\mathcal{M}_1, \ldots, \mathcal{M}_K$: K models

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

Bayes factor for $\mathcal{M}_k$ against $\mathcal{M}_{k'}$

$$BF_{kk'}(\boldsymbol{Y}) = m_{\mathcal{M}_k}(\boldsymbol{Y}) / m_{\mathcal{M}_{k'}}(\boldsymbol{Y})$$

Default Improper Priors

Lack of substantive prior information
$p(\boldsymbol{\theta}_k) = p^D(\boldsymbol{\theta}_k)$
$p^D(\boldsymbol{\theta}_k)$: objective default (non-informative) prior
Often improper

Default Improper Priors

Lack of substantive prior information
$p(\boldsymbol{\theta}_k) = p^D(\boldsymbol{\theta}_k)$
$p^D(\boldsymbol{\theta}_k)$: objective default (non-informative) prior
Often improper

Cannot be used naively to compute Bayes factors
even when $m_{\mathcal{M}_k}(\boldsymbol{Y})$ is finite and non-zero
arbitrary constants do not cancel out in their ratios

Default Improper Priors

Lack of substantive prior information
$p(\boldsymbol{\theta}_k) = p^D(\boldsymbol{\theta}_k)$
$p^D(\boldsymbol{\theta}_k)$: objective default (non-informative) prior
Often improper

Cannot be used naively to compute Bayes factors
even when $m_{\mathcal{M}_k}(\boldsymbol{Y})$ is finite and non-zero
arbitrary constants do not cancel out in their ratios

Several methods

- Fractional Bayes Factor

Default Improper Priors

Lack of substantive prior information
$p(\boldsymbol{\theta}_k) = p^D(\boldsymbol{\theta}_k)$
$p^D(\boldsymbol{\theta}_k)$: objective default (non-informative) prior
Often improper

Cannot be used naively to compute Bayes factors
even when $m_{\mathcal{M}_k}(\boldsymbol{Y})$ is finite and non-zero
arbitrary constants do not cancel out in their ratios

Several methods

- Fractional Bayes Factor
- Intrinsic Prior

Default Improper Priors

Lack of substantive prior information
$p(\boldsymbol{\theta}_k) = p^D(\boldsymbol{\theta}_k)$
$p^D(\boldsymbol{\theta}_k)$: objective default (non-informative) prior
Often improper

Cannot be used naively to compute Bayes factors
even when $m_{\mathcal{M}_k}(\boldsymbol{Y})$ is finite and non-zero
arbitrary constants do not cancel out in their ratios

Several methods

- Fractional Bayes Factor
- Intrinsic Prior
- Expected Posterior Prior

Default Improper Priors

Lack of substantive prior information
$p(\boldsymbol{\theta}_k) = p^D(\boldsymbol{\theta}_k)$
$p^D(\boldsymbol{\theta}_k)$: objective default (non-informative) prior
Often improper

Cannot be used naively to compute Bayes factors
even when $m_{\mathcal{M}_k}(\boldsymbol{Y})$ is finite and non-zero
arbitrary constants do not cancel out in their ratios

Several methods

- Fractional Bayes Factor
- Intrinsic Prior
- Expected Posterior Prior
- ...

Fractional Bayes Factor

$b = b(n)$, $0 < b < 1$: fraction of sample size n

Fractional marginal likelihood of model $\mathcal{M}_k$

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

Fractional Bayes Factor

$b = b(n)$, $0 < b < 1$: fraction of sample size n

Fractional marginal likelihood of model $\mathcal{M}_k$

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

Can be rewritten as

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

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

Fractional Bayes Factor

$b = b(n)$, $0 < b < 1$: fraction of sample size n
Fractional marginal likelihood of model $\mathcal{M}_k$

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

Can be rewritten as

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

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

Fractional Bayes factor (FBF)

$$BF_{kk'}(\boldsymbol{Y}; b) = m_{\mathcal{M}_k}(\boldsymbol{Y}; b) / m_{\mathcal{M}_{k'}}(\boldsymbol{Y}; b)$$

Default choice: $b = n_0/n$
n_0: minimal (integer) training sample size

Outline

Difficulties with parameter priors

- parameter priors should be compatible
  related
  (e.g. to avoid paradoxes)
  Jeffreys-Lindley's paradox

Difficulties with parameter priors

- parameter priors should be compatible
  related
  (e.g. to avoid paradoxes)
  Jeffreys-Lindley's paradox
- parameter space of a graphical model is constrained

Difficulties with parameter priors

- parameter priors should be compatible
  related
  (e.g. to avoid paradoxes)
  Jeffreys-Lindley's paradox
- parameter space of a graphical model is constrained

Difficulties with parameter priors

- parameter priors should be compatible
  related
  (e.g. to avoid paradoxes)
  Jeffreys-Lindley's paradox
- parameter space of a graphical model is constrained
  Example: Gaussian graphical model

$$\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{G}} \sim N(\boldsymbol{\mu}, (\boldsymbol{\Omega}_{\mathcal{G}})^{-1})$$

$\boldsymbol{\Omega}_{\mathcal{G}}$ Markov w.r.t. UG $\mathcal{G}$

Difficulties with parameter priors

- parameter priors should be compatible
  related
  (e.g. to avoid paradoxes)
  Jeffreys-Lindley's paradox
- parameter space of a graphical model is constrained
  Example: Gaussian graphical model

$$\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{G}} \sim N(\boldsymbol{\mu}, (\boldsymbol{\Omega}_{\mathcal{G}})^{-1})$$

  $\boldsymbol{\Omega}_{\mathcal{G}}$ Markov w.r.t. UG $\mathcal{G}$
  $\boldsymbol{\Omega}_{\mathcal{G}}$ s.p.d.
  but also
  $(\boldsymbol{\Omega}_{\mathcal{G}})_{ij} = 0$ whenever there is no edge between i and j in $\mathcal{G}$
  constrained parameter space
  $p(\boldsymbol{\Omega}_{\mathcal{G}})$ must comply with this constraint

Building parameter priors under DAG-models

Collection of DAG-models
Joint sampling density under DAG-model $\mathcal{D}$

$$f_{\mathcal{D}}(\boldsymbol{y}_i \,|\, \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \,|\, \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}; \boldsymbol{\theta}_j)$$

Building parameter priors under DAG-models

Collection of DAG-models
Joint sampling density under DAG-model $\mathcal{D}$

$$f_{\mathcal{D}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}; \boldsymbol{\theta}_j)$$

Assume
[Geiger and Heckerman (2002) *Ann. Statist.*]

Building parameter priors under DAG-models

Collection of DAG-models
Joint sampling density under DAG-model $\mathcal{D}$

$$f_{\mathcal{D}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}; \boldsymbol{\theta}_j)$$

Assume
[Geiger and Heckerman (2002) *Ann. Statist.*]

- 1: *Complete model equivalence*
  Two complete DAG-models represent the same family of sampling distributions.

Building parameter priors under DAG-models

Collection of DAG-models
Joint sampling density under DAG-model $\mathcal{D}$

$$f_{\mathcal{D}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}; \boldsymbol{\theta}_j)$$

Assume
[Geiger and Heckerman (2002) *Ann. Statist.*]

- 1: *Complete model equivalence*
  Two complete DAG-models represent the same family of sampling distributions.
- 2: *Regularity*
  Smooth reparametrizations between complete models

Building parameter priors under DAG-models

Collection of DAG-models
Joint sampling density under DAG-model $\mathcal{D}$

$$f_{\mathcal{D}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}; \boldsymbol{\theta}_j)$$

Assume
[Geiger and Heckerman (2002) *Ann. Statist.*]

- 1: *Complete model equivalence*
  Two complete DAG-models represent the same family of sampling distributions.
- 2: *Regularity*
  Smooth reparametrizations between complete models
- 3: *Likelihood modularity*
  If two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are such that $\mathrm{pa}_{\mathcal{D}_1}(j) = \mathrm{pa}_{\mathcal{D}_2}(j)$, they describe the same sampling family for node j

Collection of priors under each DAG-model $\mathcal{D}$

- 4: *Prior modularity*
  If two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are such that $\mathrm{pa}_{\mathcal{D}_1}(j) = \mathrm{pa}_{\mathcal{D}_2}(j)$, prior for $\boldsymbol{\theta}_j$ under $\mathcal{D}_1$ same as prior on $\boldsymbol{\theta}_j$ under $\mathcal{D}_2$
  ["Prior compatibility"]

Collection of priors under each DAG-model $\mathcal{D}$

- 4: *Prior modularity*
  If two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are such that $\mathrm{pa}_{\mathcal{D}_1}(j) = \mathrm{pa}_{\mathcal{D}_2}(j)$, prior for $\boldsymbol{\theta}_j$ under $\mathcal{D}_1$ same as prior on $\boldsymbol{\theta}_j$ under $\mathcal{D}_2$
  ["Prior compatibility"]
- 5: *Global parameter independence*

$$p_{\mathcal{D}}(\boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} p_{\mathcal{D}}(\boldsymbol{\theta}_j)$$

[local parameters $\boldsymbol{\theta}_j$'s mutually stochastically independent]

Collection of priors under each DAG-model $\mathcal{D}$

- 4: *Prior modularity*
  If two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are such that $\mathrm{pa}_{\mathcal{D}_1}(j) = \mathrm{pa}_{\mathcal{D}_2}(j)$, prior for $\boldsymbol{\theta}_j$ under $\mathcal{D}_1$ same as prior on $\boldsymbol{\theta}_j$ under $\mathcal{D}_2$
  ["Prior compatibility"]
- 5: *Global parameter independence*

$$p_{\mathcal{D}}(\boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} p_{\mathcal{D}}(\boldsymbol{\theta}_j)$$

[local parameters $\boldsymbol{\theta}_j$'s mutually stochastically independent]

Collection of priors under each DAG-model $\mathcal{D}$

- 4: *Prior modularity*
  If two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are such that $\mathrm{pa}_{\mathcal{D}_1}(j) = \mathrm{pa}_{\mathcal{D}_2}(j)$, prior for $\boldsymbol{\theta}_j$ under $\mathcal{D}_1$ same as prior on $\boldsymbol{\theta}_j$ under $\mathcal{D}_2$
  ["Prior compatibility"]
- 5: *Global parameter independence*

$$p_{\mathcal{D}}(\boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} p_{\mathcal{D}}(\boldsymbol{\theta}_j)$$

[local parameters $\boldsymbol{\theta}_j$'s mutually stochastically independent]

Assumptions 1,2,3 satisfied in the multivariate Gaussian model $N_q(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1})$

Collection of priors under each DAG-model $\mathcal{D}$

- 4: *Prior modularity*
  If two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are such that $\mathrm{pa}_{\mathcal{D}_1}(j) = \mathrm{pa}_{\mathcal{D}_2}(j)$, prior for $\boldsymbol{\theta}_j$ under $\mathcal{D}_1$ same as prior on $\boldsymbol{\theta}_j$ under $\mathcal{D}_2$
  ["Prior compatibility"]
- 5: *Global parameter independence*

$$p_{\mathcal{D}}(\boldsymbol{\theta}_{\mathcal{D}}) = \prod_{j=1}^{q} p_{\mathcal{D}}(\boldsymbol{\theta}_j)$$

[local parameters $\boldsymbol{\theta}_j$'s mutually stochastically independent]

Assumptions 1,2,3 satisfied in the multivariate Gaussian model $N_q(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1})$
Assumptions 4 and 5 satisfied with usual conjugate prior
$(\boldsymbol{\mu}, \boldsymbol{\Omega}) \sim \textit{Normal} - \textit{Wishart}$
under any complete DAG-model
and imposed under any other DAG-model to build the prior.

Outline

Marginal Lik DAG-models

If Assumptions 1-5 hold, then

$$
\begin{aligned}
m_{\mathcal{D}}(\boldsymbol{Y}) &\overset{(1)}{=} \prod_{j=1}^{q} \int p_{\mathcal{D}}(\boldsymbol{\theta}_j) \prod_{i=1}^{n} f_{\mathcal{D}}(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{D}}(j)}; \boldsymbol{\theta}_j) d\boldsymbol{\theta}_j \\
&\overset{(2)}{=} \prod_{j=1}^{q} \int p_{\mathcal{C}_j}(\boldsymbol{\theta}_j) \prod_{i=1}^{n} f_{\mathcal{C}_j}(y_{ij} \mid \boldsymbol{y}_{i,\mathrm{pa}_{\mathcal{C}_j}(j)}; \boldsymbol{\theta}_j) d\boldsymbol{\theta}_j \\
&\overset{(3)}{=} \prod_{j=1}^{q} \int p_{\mathcal{C}_j}(\boldsymbol{\theta}_j) f_{\mathcal{C}_j}(\boldsymbol{Y}_j \mid \boldsymbol{Y}_{\mathrm{pa}_{\mathcal{C}_j}(j)}; \boldsymbol{\theta}_j) d\boldsymbol{\theta}_j \\
&\overset{(4)}{=} \prod_{j=1}^{q} m_{\mathcal{C}_j}(\boldsymbol{Y}_j \mid \boldsymbol{Y}_{\mathrm{pa}_{\mathcal{C}_j}(j)}),
\end{aligned}
$$

$\mathcal{C}_j$ is any complete DAG such that $\mathrm{pa}_{\mathcal{C}_j}(j) = \mathrm{pa}_{\mathcal{D}}(j)$
(1) use global parameter independence
(2) use prior and likelihood modularity
(3) recall that $\boldsymbol{Y}_j = (\boldsymbol{y}_{ij};\ i = 1, \ldots, n)$
(4) by definition of $m_{\mathcal{C}_j}(\boldsymbol{Y}_j \mid \boldsymbol{Y}_{\mathrm{pa}_{\mathcal{C}_j}(j)})$

Marginal Lik DAG-models (ctd)

In conclusion

$$m_{\mathcal{D}}(\boldsymbol{Y}) = \prod_{j=1}^{q} m_{\mathcal{C}_j}(\boldsymbol{Y}_j \mid \boldsymbol{Y}_{\mathrm{pa}_{\mathcal{C}_j}(j)}) = \prod_{j=1}^{q} \frac{m_{\mathcal{C}_j}(\boldsymbol{Y}_{\mathrm{fa}_{\mathcal{C}_j}(j)})}{m_{\mathcal{C}_j}(\boldsymbol{Y}_{\mathrm{pa}_{\mathcal{C}_j}(j)})} = \prod_{j=1}^{q} \frac{m(\boldsymbol{Y}_{\mathrm{fa}_{\mathcal{D}}(j)})}{m(\boldsymbol{Y}_{\mathrm{pa}_{\mathcal{D}}(j)})},$$

$\mathrm{fa}_{\mathcal{D}}(j) = \mathrm{pa}_{\mathcal{D}}(j) \cup \{j\}$

$m(\cdot)$: marginal under any complete DAG-model

Marginal Lik DAG-models (ctd)

In conclusion

$$m_{\mathcal{D}}(\boldsymbol{Y}) = \prod_{j=1}^{q} m_{\mathcal{C}_j}(\boldsymbol{Y}_j \mid \boldsymbol{Y}_{\mathrm{pa}_{\mathcal{C}_j}(j)}) = \prod_{j=1}^{q} \frac{m_{\mathcal{C}_j}(\boldsymbol{Y}_{\mathrm{fa}_{\mathcal{C}_j}(j)})}{m_{\mathcal{C}_j}(\boldsymbol{Y}_{\mathrm{pa}_{\mathcal{C}_j}(j)})} = \prod_{j=1}^{q} \frac{m(\boldsymbol{Y}_{\mathrm{fa}_{\mathcal{D}}(j)})}{m(\boldsymbol{Y}_{\mathrm{pa}_{\mathcal{D}}(j)})},$$

$\mathrm{fa}_{\mathcal{D}}(j) = \mathrm{pa}_{\mathcal{D}}(j) \cup \{j\}$

$m(\cdot)$: marginal under any complete DAG-model

Bottom line
Only one single parameter prior need be elicited
under any complete (i.e. unconstrained) model
Huge simplification

Marginal Lik DAG-models (ctd)

In conclusion

$$m_{\mathcal{D}}(\boldsymbol{Y}) = \prod_{j=1}^{q} m_{\mathcal{C}_j}(\boldsymbol{Y}_j \mid \boldsymbol{Y}_{\mathrm{pa}_{\mathcal{C}_j}(j)}) = \prod_{j=1}^{q} \frac{m_{\mathcal{C}_j}(\boldsymbol{Y}_{\mathrm{fa}_{\mathcal{C}_j}(j)})}{m_{\mathcal{C}_j}(\boldsymbol{Y}_{\mathrm{pa}_{\mathcal{C}_j}(j)})} = \prod_{j=1}^{q} \frac{m(\boldsymbol{Y}_{\mathrm{fa}_{\mathcal{D}}(j)})}{m(\boldsymbol{Y}_{\mathrm{pa}_{\mathcal{D}}(j)})},$$

$\mathrm{fa}_{\mathcal{D}}(j) = \mathrm{pa}_{\mathcal{D}}(j) \cup \{j\}$

$m(\cdot)$: marginal under any complete DAG-model

Bottom line
Only one single parameter prior need be elicited
under any complete (i.e. unconstrained) model
Huge simplification

Next all we need is being able to evaluate marginals of (column) subsets of the data matrix $\boldsymbol{Y}$

Score Equivalence

Markov equivalence class

Class of DAG-models embodying same conditional independencies

Score Equivalence

Markov equivalence class
Class of DAG-models embodying same conditional independencies

Above parameter priors produce a marginal likelihood which is constant on the Markov equivalence class

Score Equivalence

Markov equivalence class

Class of DAG-models embodying same conditional independencies

Above parameter priors produce a marginal likelihood which is constant on the Markov equivalence class

Score equivalence

Score Equivalence

Markov equivalence class
Class of DAG-models embodying same conditional independencies

Above parameter priors produce a marginal likelihood which is constant on the Markov equivalence class
Score equivalence

Score equivalence is important because

- we cannot distinguish between equivalent DAGs based on (observational) data
  an inconsistency would arise if score equivalence did not hold

Score Equivalence

Markov equivalence class
Class of DAG-models embodying same conditional independencies

Above parameter priors produce a marginal likelihood which is constant on the Markov equivalence class
Score equivalence

Score equivalence is important because

- we cannot distinguish between equivalent DAGs based on (observational) data
  an inconsistency would arise if score equivalence did not hold
- we can use our method also to score a decomposable graphical model $\mathcal{G}$.

Score Equivalence

Markov equivalence class
Class of DAG-models embodying same conditional independencies

Above parameter priors produce a marginal likelihood which is constant on the Markov equivalence class
Score equivalence

Score equivalence is important because

- we cannot distinguish between equivalent DAGs based on (observational) data
  an inconsistency would arise if score equivalence did not hold
- we can use our method also to score a decomposable graphical model $\mathcal{G}$.

Score Equivalence

Markov equivalence class
Class of DAG-models embodying same conditional independencies

Above parameter priors produce a marginal likelihood which is constant on the Markov equivalence class
Score equivalence

Score equivalence is important because

- we cannot distinguish between equivalent DAGs based on (observational) data
  an inconsistency would arise if score equivalence did not hold
- we can use our method also to score a decomposable graphical model $\mathcal{G}$.
  [A decomposable graph $\mathcal{G}$ always admits a DAG version $\mathcal{G}^{<}$]

Marginal Lik Decomposable Graphical Models

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

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

$m(\cdot)$: marginal data distribution under any *complete* graph

Marginal Likelihood of a Gaussian DAG-model

$\boldsymbol{y}_i \,|\, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathcal{D} \overset{iid}{\sim} N_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$; $i = 1, \ldots, n$
$\boldsymbol{\Omega}_{\mathcal{D}}$: constrained precision matrix, Markov with respect to $\mathcal{D}$

Marginal Likelihood of a Gaussian DAG-model

$\boldsymbol{y}_i \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathcal{D} \overset{iid}{\sim} N_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1}); i = 1, \ldots, n$
$\boldsymbol{\Omega}_{\mathcal{D}}$: constrained precision matrix, Markov with respect to $\mathcal{D}$
Assumptions 1-3 satisfied

Marginal Likelihood of a Gaussian DAG-model

$\boldsymbol{y}_i \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathcal{D} \stackrel{iid}{\sim} N_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$; $i = 1, \ldots, n$
$\boldsymbol{\Omega}_{\mathcal{D}}$: constrained precision matrix, Markov with respect to $\mathcal{D}$
Assumptions 1-3 satisfied
Take any complete Gaussian DAG model $\mathcal{D} = \mathcal{C}$ ($\boldsymbol{\Omega}_{\mathcal{C}} = \boldsymbol{\Omega}$ *unconstrained*)

Marginal Likelihood of a Gaussian DAG-model

$\boldsymbol{y}_i \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathcal{D} \overset{iid}{\sim} N_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$; $i = 1, \ldots, n$
$\boldsymbol{\Omega}_{\mathcal{D}}$: constrained precision matrix, Markov with respect to $\mathcal{D}$
Assumptions 1-3 satisfied
Take any complete Gaussian DAG model $\mathcal{D} = \mathcal{C}$ ($\boldsymbol{\Omega}_{\mathcal{C}} = \boldsymbol{\Omega}$ *unconstrained*)
If $(\boldsymbol{\mu}, \boldsymbol{\Omega}) \sim \textit{Normal} - \textit{Wishart}$
$p(\boldsymbol{\mu} \mid \boldsymbol{\Omega}) = N_q(\boldsymbol{\mu} \mid \boldsymbol{m}, (c\boldsymbol{\Omega})^{-1})$
$p(\boldsymbol{\Omega}) = W_q(a, R)$
Then Assumptions 4-5 are satisfied

Marginal Likelihood of a Gaussian DAG-model

$\boldsymbol{y}_i \,|\, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathcal{D} \overset{iid}{\sim} N_q(\boldsymbol{\mu}, \boldsymbol{\Omega}_{\mathcal{D}}^{-1})$; $i = 1, \ldots, n$
$\boldsymbol{\Omega}_{\mathcal{D}}$: constrained precision matrix, Markov with respect to $\mathcal{D}$
Assumptions 1-3 satisfied
Take any complete Gaussian DAG model $\mathcal{D} = \mathcal{C}$ ($\boldsymbol{\Omega}_{\mathcal{C}} = \boldsymbol{\Omega}$ *unconstrained*)
If $(\boldsymbol{\mu}, \boldsymbol{\Omega}) \sim Normal - Wishart$
$p(\boldsymbol{\mu} \,|\, \boldsymbol{\Omega}) = N_q(\boldsymbol{\mu} \,|\, \boldsymbol{m}, (c\boldsymbol{\Omega})^{-1})$
$p(\boldsymbol{\Omega}) = W_q(a, R)$
Then Assumptions 4-5 are satisfied
Closed-form expressions for $m(\boldsymbol{Y}_J)$ are available ($J \subset \{1, \ldots, q\}$)
Geiger & Heckerman (2002)
[corrections in Kuipers, Moffa & Heckerman (2014, *Ann. Statist.*)]
C. and La Rocca (2012, *Scand. J. Statist.*)

Objective Bayes Marginal Likelihood of a Gaussian DAG-model

An objective Bayes version is available ($\boldsymbol{\mu} = 0$) starting from an improper prior

Objective Bayes Marginal Likelihood of a Gaussian DAG-model

An objective Bayes version is available ($\mu = 0$)
starting from an improper prior

C. and La Rocca (2012) for DAGs
start with an improper standard prior on the unconstrained $\mathbf{\Omega}$
and then use the FBF

Objective Bayes Marginal Likelihood of a Gaussian DAG-model

An objective Bayes version is available ($\boldsymbol{\mu} = 0$)
starting from an improper prior

C. and La Rocca (2012) for DAGs
start with an improper standard prior on the unconstrained $\boldsymbol{\Omega}$
and then use the FBF

Carvalho and Scott (2009, *Biometrika*)
for decomposable UGs
start with an improper Hyper Inverse Wishart on the
constrained $\boldsymbol{\Sigma}_{\mathcal{G}} = (\boldsymbol{\Omega})_{\mathcal{G}}^{-1}$
and then use the FBF

Outline

Marginal Likelihood of a Multivariate Gaussian Regression DAG Model

Extension of the previous results to the regression setting
(applied motivation follows)

Marginal Likelihood of a Multivariate Gaussian Regression DAG Model

Extension of the previous results to the regression setting (applied motivation follows)

$\boldsymbol{X}$: known design matrix

Marginal Likelihood of a Multivariate Gaussian Regression DAG Model

Extension of the previous results to the regression setting (applied motivation follows)

$\boldsymbol{X}$: known design matrix

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

Marginal Likelihood of a Multivariate Gaussian Regression DAG Model

Extension of the previous results to the regression setting (applied motivation follows)
$\boldsymbol{X}$: known design matrix

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

Assume ($\boldsymbol{B}$) and Ω unrestricted. Prior

$$\begin{aligned} \boldsymbol{B} \mid \Omega &\sim \mathcal{N}_{p+1,q}(\underline{\boldsymbol{B}}, \boldsymbol{C}^{-1}, \Omega^{-1}), \\ \Omega &\sim \mathcal{W}_q(a, \boldsymbol{R}) \end{aligned}$$

Assumptions 1-5 for the construction of compatible priors hold
C., La Rocca and Peluso (2016, *Scand. J. Statist*, to appear)

Marginal Likelihood of a Multivariate Gaussian Regression DAG Model

Extension of the previous results to the regression setting (applied motivation follows)
$\boldsymbol{X}$: known design matrix

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

Assume ($\boldsymbol{B}$) and Ω unrestricted. Prior

$$\begin{aligned} \boldsymbol{B} \mid \Omega &\sim \mathcal{N}_{p+1,q}(\underline{\boldsymbol{B}}, \boldsymbol{C}^{-1}, \Omega^{-1}), \\ \Omega &\sim \mathcal{W}_q(a, \boldsymbol{R}) \end{aligned}$$

Assumptions 1-5 for the construction of compatible priors hold
C., La Rocca and Peluso (2016, *Scand. J. Statist*, to appear)

Closed form expression for the marginal of any set of variables $\boldsymbol{Y}_J$ given $\boldsymbol{X}$ is available

Marginal Likelihood of a Multivariate Gaussian Regression DAG Model

Extension of the previous results to the regression setting (applied motivation follows)

$\boldsymbol{X}$: known design matrix

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

Assume ($\boldsymbol{B}$) and Ω unrestricted. Prior

$$\begin{aligned} \boldsymbol{B} \mid \Omega &\sim \mathcal{N}_{p+1,q}(\underline{\boldsymbol{B}}, \boldsymbol{C}^{-1}, \Omega^{-1}), \\ \Omega &\sim \mathcal{W}_q(a, \boldsymbol{R}) \end{aligned}$$

Assumptions 1-5 for the construction of compatible priors hold C., La Rocca and Peluso (2016, *Scand. J. Statist*, to appear)

Closed form expression for the marginal of any set of variables $\boldsymbol{Y}_J$ given $\boldsymbol{X}$ is available

Marginal likelihood of any regression DAG-model can be evaluated

Genetical Genomics Experiments

- measure both genetic variants and gene expression data on the same subjects

Genetical Genomics Experiments

- measure both genetic variants and gene expression data on the same subjects
- genome-wide eQTL analysis (expression quantitative trait loci)

Genetical Genomics Experiments

- measure both genetic variants and gene expression data on the same subjects
- genome-wide eQTL analysis (expression quantitative trait loci)

Genetical Genomics Experiments

- measure both genetic variants and gene expression data on the same subjects
- genome-wide eQTL analysis
  (expression quantitative trait loci)

Aim: study conditional independence structures of gene expressions
after the confounding genetic effects are taken into account.

Genetical Genomics Experiments

- measure both genetic variants and gene expression data on the same subjects
- genome-wide eQTL analysis (expression quantitative trait loci)

Aim: study conditional independence structures of gene expressions
after the confounding genetic effects are taken into account.

This greatly improves inference on the network of genes
Cai *et. al.* (2013, *Biometrika*)

Genetical Genomics Experiments

- measure both genetic variants and gene expression data on the same subjects
- genome-wide eQTL analysis
(expression quantitative trait loci)

Aim: study conditional independence structures of gene expressions
after the confounding genetic effects are taken into account.

This greatly improves inference on the network of genes
Cai *et. al.* (2013, *Biometrika*)

Typical Scenarios

$p_\star$: number of potential predictors

Typical Scenarios

$p_\star$: number of potential predictors
p: number of truly effective predictors

Typical Scenarios

$p_\star$: number of potential predictors
p: number of truly effective predictors
sparsity
$p << p_\star$

Typical Scenarios

$p_\star$: number of potential predictors
p: number of truly effective predictors
sparsity
$p << p_\star$

- $p_\star \approx 100 - 500$; $q \approx 50 - 300$; $n \approx 50 - 200$

Typical Scenarios

$p_\star$: number of potential predictors
p: number of truly effective predictors
sparsity
$p << p_\star$

- $p_\star \approx 100 - 500$; $q \approx 50 - 300$; $n \approx 50 - 200$
- $p_\star \approx 3,000$; $q \approx 100$; $n \approx 100$

Fractional Marginal Likelihood

Default prior on complete DAG

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

Fractional Marginal Likelihood

Default prior on complete DAG

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

$\mathcal{G}$ Undirected decomposable graph

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

$$m(\boldsymbol{Y}_C) = \pi^{-\frac{(n-n_0)|C|}{2}} \frac{\Gamma_{|C|}\left(\frac{a_D + n - p - 1 - |\bar{C}|}{2}\right)}{\Gamma_{|C|}\left(\frac{a_D + n_0 - p - 1 - |\bar{C}|}{2}\right)} \left(\frac{n_0}{n}\right)^{\frac{|C|(a_D + n_0 - |\bar{C}|)}{2}} |\hat{\boldsymbol{E}}_C^\top \hat{\boldsymbol{E}}_C|^{-\frac{n-n_0}{2}}$$

$\Gamma_{|C|}(\cdot)$: multivariate gamma function; $\hat{\boldsymbol{E}}_C$: residuals clique C

Fractional Marginal Likelihood

Default prior on complete DAG

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

$\mathcal{G}$ Undirected decomposable graph

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

$$m(\boldsymbol{Y}_C) = \pi^{-\frac{(n-n_0)|C|}{2}} \frac{\Gamma_{|C|}\left(\frac{a_D + n - p - 1 - |\bar{C}|}{2}\right)}{\Gamma_{|C|}\left(\frac{a_D + n_0 - p - 1 - |\bar{C}|}{2}\right)} \left(\frac{n_0}{n}\right)^{\frac{|C|(a_D + n_0 - |\bar{C}|)}{2}} |\hat{\boldsymbol{E}}_C^\top \hat{\boldsymbol{E}}_C|^{-\frac{n-n_0}{2}}$$

$\Gamma_{|C|}(\cdot)$: multivariate gamma function; $\hat{\boldsymbol{E}}_C$: residuals clique C

Fractional Marginal Likelihood

Default prior on complete DAG

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

$\mathcal{G}$ Undirected decomposable graph

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

$$m(\boldsymbol{Y}_C) = \pi^{-\frac{(n-n_0)|C|}{2}} \frac{\Gamma_{|C|}\left(\frac{a_D+n-p-1-|\overline{C}|}{2}\right)}{\Gamma_{|C|}\left(\frac{a_D+n_0-p-1-|\overline{C}|}{2}\right)} \left(\frac{n_0}{n}\right)^{\frac{|C|(a_D+n_0-|\overline{C}|)}{2}} |\hat{\boldsymbol{E}}_C^\top \hat{\boldsymbol{E}}_C|^{-\frac{n-n_0}{2}}$$

$\Gamma_{|C|}(\cdot)$: multivariate gamma function; $\hat{\boldsymbol{E}}_C$: residuals clique C

Formula for $m_{\mathcal{G}}(\boldsymbol{Y})$ holds provided $n > p + |C|$, $C \in \mathcal{C}$
sparsity condition on regression and graphical structure

Recommended settings $a_D = q - 1$; $n_0 = p + 2$

Variable selection

We do not know which are the truly effective p predictors

Variable selection

We do not know which are the truly effective p predictors

Need to perform variable selection

Variable selection

We do not know which are the truly effective p predictors

Need to perform variable selection

Variable selection

We do not know which are the truly effective p predictors

Need to perform variable selection

variable indicators
$\gamma_1 = 1$: intercept
$\gamma_i = 1$ if covariate $i - 1$ is in the model; otherwise $\gamma_i = 0$;
$i = 2, \ldots p_\star + 1$

Variable selection

We do not know which are the truly effective p predictors

Need to perform variable selection

variable indicators
$\gamma_1 = 1$: intercept
$\gamma_i = 1$ if covariate $i - 1$ is in the model; otherwise $\gamma_i = 0$;
$i = 2, \ldots p_\star + 1$

$\boldsymbol{\gamma}^\top = (\gamma_1, \ldots, \gamma_{p_\star+1})$:
regression model

Variable selection

We do not know which are the truly effective p predictors

Need to perform variable selection

variable indicators
$\gamma_1 = 1$: intercept
$\gamma_i = 1$ if covariate $i - 1$ is in the model; otherwise $\gamma_i = 0$;
$i = 2, \ldots p_\star + 1$

$\boldsymbol{\gamma}^\top = (\gamma_1, \ldots, \gamma_{p_\star+1})$:
regression model

$p_\gamma = \sum_{j=2}^{p_\star+1} \gamma_j$
number of predictors in model γ

Variable selection

We do not know which are the truly effective p predictors

Need to perform variable selection

variable indicators
$\gamma_1 = 1$: intercept
$\gamma_i = 1$ if covariate $i - 1$ is in the model; otherwise $\gamma_i = 0$; $i = 2, \ldots p_\star + 1$

$\boldsymbol{\gamma}^\top = (\gamma_1, \ldots, \gamma_{p_\star+1})$:
regression model

$p_\gamma = \sum_{j=2}^{p_\star+1} \gamma_j$
number of predictors in model γ

Under sparse regression
$p_\gamma << p_\star$

Joint variable and graph selection

γ: regression structure
set of predictors to include in the linear model

Joint variable and graph selection

γ: regression structure
set of predictors to include in the linear model

$\mathcal{G}$: covariance structure
decomposable graph for the precision matrix

Joint variable and graph selection

γ: regression structure
set of predictors to include in the linear model

$\mathcal{G}$: covariance structure
decomposable graph for the precision matrix

Graphical Gaussian multivariate regression model

$$\boldsymbol{Y} \mid \boldsymbol{B}_\gamma, \Omega_\mathcal{G}, \gamma, \mathcal{G} \sim \mathcal{N}_{n,q}(\boldsymbol{X}_\gamma \boldsymbol{B}_\gamma, \boldsymbol{I}_n, \Omega_\mathcal{G}^{-1})$$

Joint variable and graph selection

γ: regression structure
set of predictors to include in the linear model

$\mathcal{G}$: covariance structure
decomposable graph for the precision matrix

Graphical Gaussian multivariate regression model

$$\boldsymbol{Y} \mid \boldsymbol{B}_\gamma, \Omega_\mathcal{G}, \gamma, \mathcal{G} \sim \mathcal{N}_{n,q}(\boldsymbol{X}_\gamma \boldsymbol{B}_\gamma, \boldsymbol{I}_n, \Omega_\mathcal{G}^{-1})$$

Aim: making inference simultaneously on $(\gamma, \mathcal{G})$

Joint variable and graph selection

γ: regression structure
set of predictors to include in the linear model

$\mathcal{G}$: covariance structure
decomposable graph for the precision matrix

Graphical Gaussian multivariate regression model

$$\boldsymbol{Y} \mid \boldsymbol{B}_\gamma, \Omega_\mathcal{G}, \gamma, \mathcal{G} \sim \mathcal{N}_{n,q}(\boldsymbol{X}_\gamma \boldsymbol{B}_\gamma, \boldsymbol{I}_n, \Omega_\mathcal{G}^{-1})$$

Aim: making inference simultaneously on $(\gamma, \mathcal{G})$
Joint graph and variable selection

Hierarchical model

$$\begin{aligned}
\boldsymbol{Y} \,|\, \gamma, \mathcal{G} &\sim m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \gamma) \\
\gamma_i \,|\, \omega_\gamma &\overset{iid}{\sim} Ber(\omega_\gamma); \quad i = 2, \ldots, p_\star + 1 \\
G_i \,|\, \omega_G &\overset{iid}{\sim} Ber(\omega_G); \quad i = 2, \ldots, q \cdot (q-1)/2 \\
\omega_\gamma &\sim U(0,1) \\
\omega_G &\sim U(0,1)
\end{aligned}$$

Hierarchical model

$$
\begin{aligned}
\boldsymbol{Y} \,|\, \gamma, \mathcal{G} &\sim m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \gamma) \\
\gamma_i \,|\, \omega_\gamma &\overset{iid}{\sim} Ber(\omega_\gamma); \quad i = 2, \ldots, p_\star + 1 \\
G_i \,|\, \omega_G &\overset{iid}{\sim} Ber(\omega_G); \quad i = 2, \ldots, q \cdot (q-1)/2 \\
\omega_\gamma &\sim U(0,1) \\
\omega_G &\sim U(0,1)
\end{aligned}
$$

$m_{\mathcal{G}}(\boldsymbol{Y} \,|\, \gamma)$: marginal likelihood
generated through our O'Bayes method based on the FBF

$\boldsymbol{G} = (G_1, \ldots, G_{q(q-1)/2})$:
vectorized *adjacency matrix* corresponding to $\mathcal{G}$

Hierarchical model

$$
\begin{aligned}
\boldsymbol{Y} \mid \gamma, \mathcal{G} &\sim m_{\mathcal{G}}(\boldsymbol{Y} \mid \gamma) \\
\gamma_i \mid \omega_\gamma &\overset{iid}{\sim} Ber(\omega_\gamma); \quad i = 2, \ldots, p_\star + 1 \\
G_i \mid \omega_G &\overset{iid}{\sim} Ber(\omega_G); \quad i = 2, \ldots, q \cdot (q-1)/2 \\
\omega_\gamma &\sim U(0,1) \\
\omega_G &\sim U(0,1)
\end{aligned}
$$

$m_{\mathcal{G}}(\boldsymbol{Y} \mid \gamma)$: marginal likelihood
generated through our O'Bayes method based on the FBF

$\boldsymbol{G} = (G_1, \ldots, G_{q(q-1)/2})$:
vectorized *adjacency matrix* corresponding to $\mathcal{G}$

Graphical models

Objective Bayes Model Selection

DAG-models

Compatible Parameter Priors

Marginal Likelihood

Covariate-adjusted graphical models

Experimental results

Two simulation settings

- *Sparse block*
  $n = 50$
  $q \in \{30, 60, 120\}$
  $p_\star = 100$

Two simulation settings

- *Sparse block*
  $n = 50$
  $q \in \{30, 60, 120\}$
  $p_\star = 100$
- *Magnified block*
  $n = 50$
  $p_\star = 100$
  $q = 150$

Two simulation settings

- *Sparse block*
  $n = 50$
  $q \in \{30, 60, 120\}$
  $p_\star = 100$
- *Magnified block*
  $n = 50$
  $p_\star = 100$
  $q = 150$

Two simulation settings

- *Sparse block*
  $n = 50$
  $q \in \{30, 60, 120\}$
  $p_\star = 100$
- *Magnified block*
  $n = 50$
  $p_\star = 100$
  $q = 150$

Parameter values generated as in the settings we compare our method to
In particular
Chen et al (2016)
Bhadra and Mallick (2013)

Sparse block

- Graph structure
  G_i=0 for $i \leq q(q-1)/2-10$ and G_i=1 otherwise for $i > q(q-1)/2-10$
  $q \times q$ adjacency matrix has a sparse bottom-right block of active edges
  sparsity of G increases with q.

Sparse block

- Graph structure

  G_i=0 for $i \leq q(q-1)/2 - 10$ and G_i=1 otherwise for $i > q(q-1)/2 - 10$

  $q \times q$ adjacency matrix has a sparse bottom-right block of active edges

  sparsity of G increases with q.
- Regression structure

  Out of $p_\star = 100$ potential covariates, true predictors are only the first and the third

  $\gamma_i = 1$ for $i \in \{1, 2, 4\}$

  $\gamma_i = 0$ otherwise

  $p_\gamma = 2$

Sparse block

- Graph structure
  G_i=0 for $i \leq q(q-1)/2-10$ and G_i=1 otherwise for $i > q(q-1)/2-10$
  $q \times q$ adjacency matrix has a sparse bottom-right block of active edges
  sparsity of G increases with q.
- Regression structure
  Out of $p_\star = 100$ potential covariates, true predictors are only the first and the third
  $\gamma_i = 1$ for $i \in \{1, 2, 4\}$
  $\gamma_i = 0$ otherwise
  $p_\gamma = 2$
- Given true G, $\Omega_\mathcal{G}$ is sampled from the G-Wishart distribution with 10 degrees of freedom and scale matrix equal to the identity

Sparse block

- Graph structure

  G_i=0 for $i \leq q(q-1)/2 - 10$ and G_i=1 otherwise for $i > q(q-1)/2 - 10$

  $q \times q$ adjacency matrix has a sparse bottom-right block of active edges

  sparsity of G increases with q.
- Regression structure

  Out of $p_\star = 100$ potential covariates, true predictors are only the first and the third

  $\gamma_i = 1$ for $i \in \{1, 2, 4\}$

  $\gamma_i = 0$ otherwise

  $p_\gamma = 2$
- Given true G, $\Omega_\mathcal{G}$ is sampled from the G-Wishart distribution with 10 degrees of freedom and scale matrix equal to the identity
- Given true γ and $\Omega_\mathcal{G}$

  $\boldsymbol{B}$ sampled from the Matrix Normal $N_{p_\star+1,q}\left(0_{p_\star+1,q}, 0.3^2 I_{p_\star+1}, \Omega_\mathcal{G}^{-1}\right)$

Sparse block

- Graph structure
  G_i=0 for $i \leq q(q-1)/2 - 10$ and G_i=1 otherwise for $i > q(q-1)/2 - 10$
  $q \times q$ adjacency matrix has a sparse bottom-right block of active edges
  sparsity of G increases with q.
- Regression structure
  Out of $p_\star = 100$ potential covariates, true predictors are only the first and the third
  $\gamma_i = 1$ for $i \in \{1, 2, 4\}$
  $\gamma_i = 0$ otherwise
  $p_\gamma = 2$
- Given true G, $\Omega_\mathcal{G}$ is sampled from the G-Wishart distribution with 10 degrees of freedom and scale matrix equal to the identity
- Given true γ and $\Omega_\mathcal{G}$
  $\boldsymbol{B}$ sampled from the Matrix Normal $N_{p_\star+1,q}\left(0_{p_\star+1,q}, 0.3^2 I_{p_\star+1}, \Omega_\mathcal{G}^{-1}\right)$
- Elements of $\boldsymbol{X}$ from the second to last column are randomly drawn from $N(10, 1)$

Magnified block

- $q = 150$

Magnified block

- $q = 150$
- Graph structure
  Fix a 50×50 adjacency matrix G^l as above
  Full G is block diagonal with G^l replicated three times

Magnified block

- $q = 150$
- Graph structure
  Fix a 50×50 adjacency matrix G^l as above
  Full G is block diagonal with G^l replicated three times

Magnified block

- $q = 150$
- Graph structure
  Fix a 50×50 adjacency matrix G^l as above
  Full G is block diagonal with G^l replicated three times
  Corresponding $\mathbf{\Omega}_{\mathcal{G}}$ is block diagonal with the three blocks $\mathbf{\Omega}^l_{\mathcal{G}}$, $5\mathbf{\Omega}^l_{\mathcal{G}}$, $10\mathbf{\Omega}^1_{\mathcal{G}}$
  $\mathbf{\Omega}_{\mathcal{G}}$ has sequentially magnified signals
- Regression structure
  True γ produced by randomly choosing each predictor with probability 0.05 among $p_\star$ potential predictors

Comparison with state-of-the-art methods for covariate-adjusted graphical model selection

- Objective Fractional Bayes Factor
  OBFBF
  (O'Bayes variable and graph selection)

Comparison with state-of-the-art methods for covariate-adjusted graphical model selection

- Objective Fractional Bayes Factor
  OBFBF
  (O'Bayes variable and graph selection)
- Two-step ANTAC (Asymptotically Normal with Thresholding after Adjusting Covariates) estimator
  Chen et al (2016, *J. Am. Statist. Asssoc.*)
  (graph selection; estimation of regression parameters intermediate step)

Comparison with state-of-the-art methods for covariate-adjusted graphical model selection

- Objective Fractional Bayes Factor
  OBFBF
  (O'Bayes variable and graph selection)
- Two-step ANTAC (Asymptotically Normal with Thresholding after Adjusting Covariates) estimator
  Chen et al (2016, *J. Am. Statist. Asssoc.*)
  (graph selection; estimation of regression parameters intermediate step)
- Graphical lasso
  GLASSO
  (only graph selection)
  Friedman, Hastie, Tibshirani (2008, *Biostatistics*)

Comparison with state-of-the-art methods for covariate-adjusted graphical model selection

- Objective Fractional Bayes Factor
  OBFBF
  (O'Bayes variable and graph selection)
- Two-step ANTAC (Asymptotically Normal with Thresholding after Adjusting Covariates) estimator
  Chen et al (2016, *J. Am. Statist. Asssoc.*)
  (graph selection; estimation of regression parameters intermediate step)
- Graphical lasso
  GLASSO
  (only graph selection)
  Friedman, Hastie, Tibshirani (2008, *Biostatistics*)
- Hyper-matrix t method
  HYPERT
  Bhadri and Mallick (2013, *Biometrics*)
  (Bayes variable and graph selection)

Comparison with state-of-the-art methods for covariate-adjusted graphical model selection

- Objective Fractional Bayes Factor
  OBFBF
  (O'Bayes variable and graph selection)
- Two-step ANTAC (Asymptotically Normal with Thresholding after Adjusting Covariates) estimator
  Chen et al (2016, *J. Am. Statist. Asssoc.*)
  (graph selection; estimation of regression parameters intermediate step)
- Graphical lasso
  GLASSO
  (only graph selection)
  Friedman, Hastie, Tibshirani (2008, *Biostatistics*)
- Hyper-matrix t method
  HYPERT
  Bhadri and Mallick (2013, *Biometrics*)
  (Bayes variable and graph selection)
- Sparse Gaussian Conditional method
  CONDIT
  Wytock and Kolter (2013, *J. Mach. Learn. Res.*)
  (graph selection; estimation of regression parameters intermediate step)

Comparison with state-of-the-art methods for covariate-adjusted graphical model selection

- Objective Fractional Bayes Factor
  OBFBF
  (O'Bayes variable and graph selection)
- Two-step ANTAC (Asymptotically Normal with Thresholding after Adjusting Covariates) estimator
  Chen et al (2016, *J. Am. Statist. Asssoc.*)
  (graph selection; estimation of regression parameters intermediate step)
- Graphical lasso
  GLASSO
  (only graph selection)
  Friedman, Hastie, Tibshirani (2008, *Biostatistics*)
- Hyper-matrix t method
  HYPERT
  Bhadri and Mallick (2013, *Biometrics*)
  (Bayes variable and graph selection)
- Sparse Gaussian Conditional method
  CONDIT
  Wytock and Kolter (2013, *J. Mach. Learn. Res.*)
  (graph selection; estimation of regression parameters intermediate step)
- Low Rank latent variables and sparse method
  LOWRANK
  Chandrasekaran et al. (2012, *Ann. Statist.*) (2012)
  (graph selection with unobserved latent variable)

Measures of performance for graph selection (Adjacency matrix)

- Misspecification rate

Measures of performance for graph selection (Adjacency matrix)

- Misspecification rate
- Specificity= True negative rate

Measures of performance for graph selection (Adjacency matrix)

- Misspecification rate
- Specificity= True negative rate
- Sensitivity=True positive rate

Measures of performance for graph selection (Adjacency matrix)

- Misspecification rate
- Specificity= True negative rate
- Sensitivity=True positive rate
- Matthews correlation coefficient

Measures of performance for graph selection (Adjacency matrix)

- Misspecification rate
- Specificity= True negative rate
- Sensitivity=True positive rate
- Matthews correlation coefficient

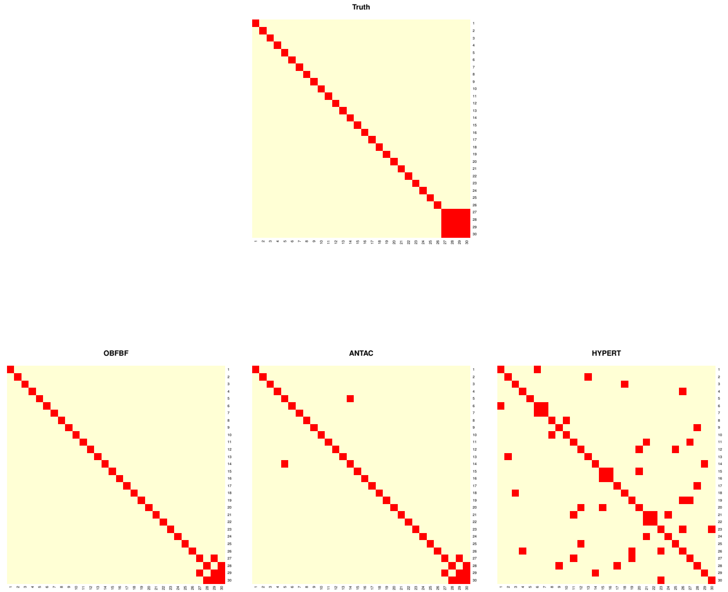
Measures of performance for graph selection (Adjacency matrix)

- Misspecification rate
- Specificity= True negative rate
- Sensitivity=True positive rate
- Matthews correlation coefficient

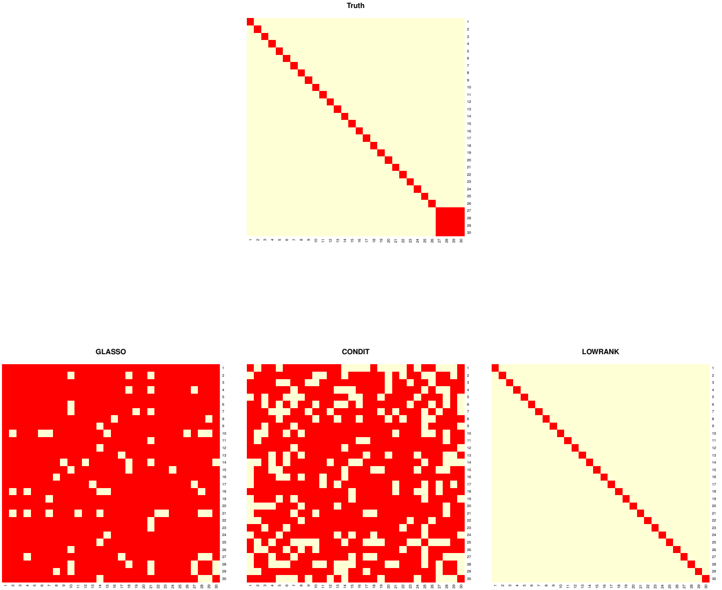
$$
\begin{aligned}
MISR &= \frac{FN + FP}{q(q-1)}, \quad SPE = \frac{TN}{TN + FP}, \\
SEN &= \frac{TP}{TP + FN} \\
MCC &= \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}
\end{aligned}
$$

Setting	(n, p^*, q)	Method	MISR	SPE	SEN	MCC	Time
Sparse	(50, 100, 30)	OBFBF	9(1)	92(1)	74(3)	47(5)	4769
		HYPERT	10(1)	91(1)	74(4)	46(2)	4270
		ANTAC	1(0)	100(0)	72(1)	84(1)	34
		GLASSO	83(5)	17(5)	86(4)	15(2)	8
		CONDIT	52(11)	48(11)	90(7)	21(4)	99
		LOWRANK	49(14)	50(14)	91(7)	22(5)	75
Sparse	(50, 100, 60)	OBFBF	3(2)	97(2)	84(1)	60(19)	5550
		HYPERT	5(0)	95(0)	84(2)	47(1)	5990
		ANTAC	0(0)	100(0)	83(1)	91(0)	109
		GLASSO	59(5)	41(5)	93(2)	12(1)	57
		CONDIT	27(19)	73(20)	89(4)	24(6)	268
		LOWRANK	81(3)	18(3)	97(3)	7(1)	236
Sparse	(50, 100, 120)	OBFBF	0(0)	100(0)	100(0)	95(5)	3745
		HYPERT	2(0)	98(0)	91(1)	54(1)	5941
		ANTAC	0(0)	100(0)	91(0)	95(0)	676
		GLASSO	36(4)	64(4)	95(1)	12(1)	547
		CONDIT	48(25)	52(25)	96(2)	11(5)	861
		LOWRANK	94(1)	6(1)	99(1)	1(0)	1002
Magnified	(50, 100, 150)	OBFBF	0(0)	100(0)	93(0)	92(12)	5498
		HYPERT	2(0)	99(0)	93(0)	54(1)	6770
		ANTAC	0(0)	100(0)	93(0)	96(0)	1971
		GLASSO	78(5)	22(5)	97(1)	5(1)	4570
		CONDIT	96(3)	4(3)	100(1)	2(1)	3517
		LOWRANK	98(0)	2(0)	100(0)	1(0)	5452

Sparse setting: $n = 200$, $p_\star = 100$, $q = 30$



Sparse setting: $n = 200$, $p_\star = 100$, $q = 30$

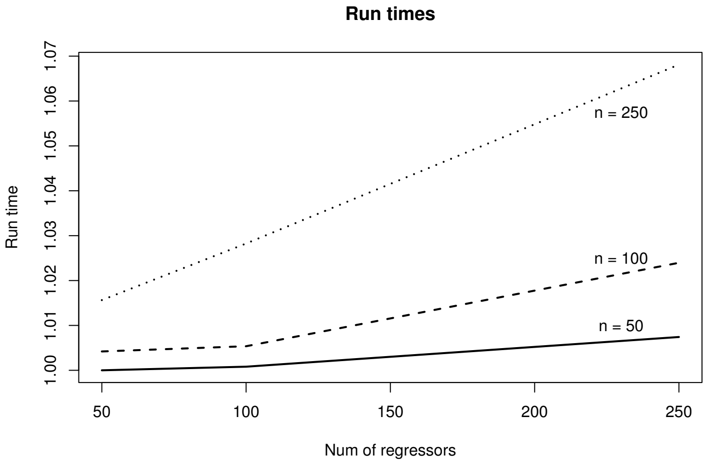


- *Computational time for MCMC based methods (OBFBF and HYPERT) higher than rest*
  However

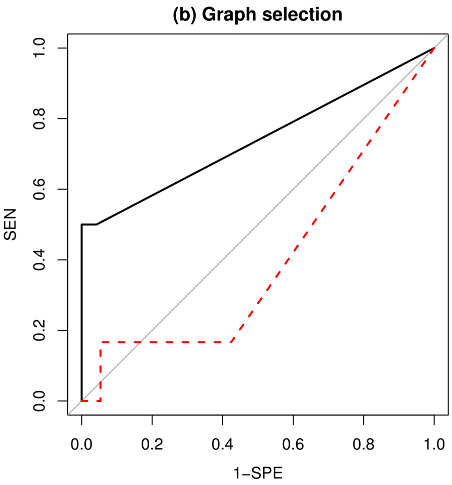
- *Computational time for MCMC based methods (OBFBF and HYPERT) higher than rest*
  However
  - *they perform also variable selection and return a richer output*

- *Computational time for MCMC based methods (OBFBF and HYPERT) higher than rest*
  However
  - *they perform also variable selection and return a richer output*
  - *Computational time for OBFBF increases only marginally (up to 7% from least to most complex setting)*

Runtimes for OBFBF

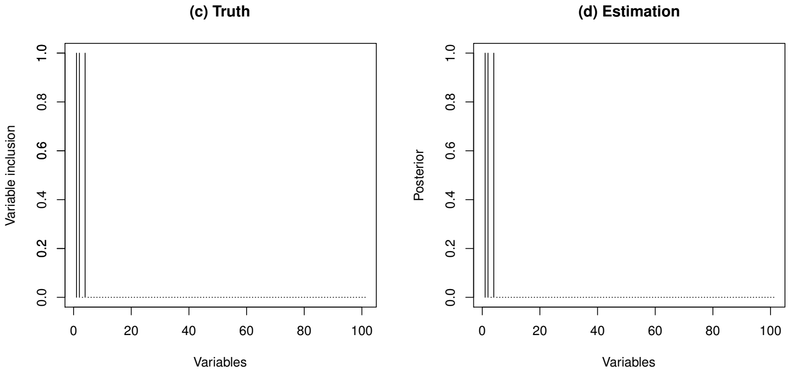


ROC curve: graph selection



Black: OBFBF
Red: HYPERT

Variable selection OBFBF



Conclusions I

Compatible parameter priors for the comparison of DAG-models can be constructed based on a single prior for the complete graph
(unconstrained parameter space)
Can use standard conjugate priors
Our contributions

- Objective Bayes (OB) method for comparing Gaussian DAG-models
  start with default prior
  and then apply the Fractional Bayes Factor

Conclusions II

- Covariate-adjusted OB method
  Joint graph and variable selection
  OBFBF comparable to ANTAC in graph selection for large and sparse networks
  although ANTAC does not perform variable selection explicitly
  OBFBF outperforms Bayesian competitor HYPERT as well remaining penalization-based methods
  OBFBF excellent performance in variable selection
  Computing time for MCMC-based methods higher but scales nicely with n, q and p

Looking ahead

- Extend the scope of covariate-adjusted graph selection beyond the regression setting and accommodate for

Looking ahead

- Extend the scope of covariate-adjusted graph selection beyond the regression setting and accommodate for

Looking ahead

- Extend the scope of covariate-adjusted graph selection beyond the regression setting and accommodate for serial dependence

Looking ahead

- Extend the scope of covariate-adjusted graph selection beyond the regression setting and accommodate for
  serial dependence
  spatial dependence

Looking ahead

- Extend the scope of covariate-adjusted graph selection beyond the regression setting and accommodate for
  serial dependence
  spatial dependence
- Explore the space of Essential DAGs (Markov Equivalence Class)
  Require calculations for Chain Graphs

Looking ahead

- Extend the scope of covariate-adjusted graph selection beyond the regression setting and accommodate for
  serial dependence
  spatial dependence
- Explore the space of Essential DAGs (Markov Equivalence Class)
  Require calculations for Chain Graphs

Looking ahead

- Extend the scope of covariate-adjusted graph selection beyond the regression setting and accommodate for
  serial dependence
  spatial dependence
- Explore the space of Essential DAGs (Markov Equivalence Class)
  Require calculations for Chain Graphs
  Use observational and interventional data

Main references

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

Main references

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

Cai, T. T. *et al.* (2013).
Covariate-adjusted precision matrix estimation with an application in genetical genomics.
Biometrika **100**, 139–156.

Main references

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

CAI, T. T. *et al.* (2013).
Covariate-adjusted precision matrix estimation with an application in genetical genomics.
Biometrika **100**, 139–156.

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

Main references

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

CAI, T. T. *et al.* (2013).
Covariate-adjusted precision matrix estimation with an application in genetical genomics.
Biometrika **100**, 139–156.

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

CHANDRASEKARAN, V., PARRILO, P. & WILLSKY, A. (2012).
Latent variable graphical model selection via convex optimization.
Ann. Statist. **40**, 1935–1967.

Main references

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

CAI, T. T. *et al.* (2013).
Covariate-adjusted precision matrix estimation with an application in genetical genomics.
Biometrika **100**, 139–156.

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

CHANDRASEKARAN, V., PARRILO, P. & WILLSKY, A. (2012).
Latent variable graphical model selection via convex optimization.
Ann. Statist. **40**, 1935–1967.

Chen, M. *et al.* (2016).
Asymptotically normal and efficient estimation of covariate-adjusted Gaussian graphical model.
Journal of the American Statistical Association **111**, 394–406.

Main references

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

CAI, T. T. *et al.* (2013).
Covariate-adjusted precision matrix estimation with an application in genetical genomics.
Biometrika **100**, 139–156.

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

CHANDRASEKARAN, V., PARRILO, P. & WILLSKY, A. (2012).
Latent variable graphical model selection via convex optimization.
Ann. Statist. **40**, 1935–1967.

Chen, M. *et al.* (2016).
Asymptotically normal and efficient estimation of covariate-adjusted Gaussian graphical model.
Journal of the American Statistical Association **111**, 394–406.

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

Main references

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

CAI, T. T. *et al.* (2013).
Covariate-adjusted precision matrix estimation with an application in genetical genomics.
Biometrika **100**, 139–156.

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

CHANDRASEKARAN, V., PARRILO, P. & WILLSKY, A. (2012).
Latent variable graphical model selection via convex optimization.
Ann. Statist. **40**, 1935–1967.

Chen, M. *et al.* (2016).
Asymptotically normal and efficient estimation of covariate-adjusted Gaussian graphical model.
Journal of the American Statistical Association **111**, 394–406.

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. (2016).
Objective Bayes covariate-adjusted sparse graphical model selection.
Scandinavian Journal of Statistics. To appear.

Main references

FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. (2008).
Sparse inverse covariance estimation with the graphical lasso.
Biostatistics **9**, 432–441.

Main references

FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. (2008).
Sparse inverse covariance estimation with the graphical lasso.
Biostatistics **9**, 432–441.

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.

Main references

FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. (2008).
Sparse inverse covariance estimation with the graphical lasso.
Biostatistics **9**, 432–441.

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.

KUIPERS, J., MOFFA, G. & HECKERMAN, D. (2014).
Addendum on the scoring of Gaussian directed acyclic graphical models.
Ann. Statist. **42**, 1689–1691.

Main references

FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. (2008).
Sparse inverse covariance estimation with the graphical lasso.
Biostatistics **9**, 432–441.

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.

KUIPERS, J., MOFFA, G. & HECKERMAN, D. (2014).
Addendum on the scoring of Gaussian directed acyclic graphical models.
Ann. Statist. **42**, 1689–1691.

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

Main references

FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. (2008).
Sparse inverse covariance estimation with the graphical lasso.
Biostatistics **9**, 432–441.

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.

KUIPERS, J., MOFFA, G. & HECKERMAN, D. (2014).
Addendum on the scoring of Gaussian directed acyclic graphical models.
Ann. Statist. **42**, 1689–1691.

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

WYTOCK, M. & KOLTER, J. Z. (2013).
Sparse Gaussian conditional random fields: Algorithms, theory, and application to energy forecasting.
In *Proceedings of the 30th International Conference on Machine Learning (ICML)*. pp. 1265–1273. JMLR: W&CP Volume 28.