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
Graphical models

Objective Bayes Model Selection

DAG-models
  Compatible Parameter Priors
  Marginal Likelihood
  Covariate-adjusted graphical models

Experimental results
Graph $\mathcal{G}$
$\mathcal{G} = (V, E)$
$V$: set of vertices
$E$: set of edges
$|V| = q$
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 $\mathbf{y}_i^\top = (y_{i1}, \ldots, y_{iq})$ which factorize according to the graph $\mathcal{G}$ is called a graphical model (wrt $\mathcal{G}$).
Factorizations

If $G = D$ is a DAG

$$f_D(y_i \mid \theta_D) = \prod_{j=1}^{q} f(y_{ij} \mid y_{i, pa_D(j)}, \theta_j)$$

$$y_{i, pa_D(j)} = \{y_{il} : l \in pa_D(j)\}; \text{pa}_D(j): \text{parents of } j$$

If $G$ is decomposable

$$f_G(y_i \mid \theta_G) = \frac{\prod_{C \in C} f(y_{i,C} \mid \theta_C)}{\prod_{S \in S} f(y_{i,S} \mid \theta_S)}$$

$C$: set of cliques; $S$: set of separators.

$y_{i,C} = \{y_{ij} : j \in C\}$. (decomposable=chordal=triangulated)

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

$S = \{\{1, 5\}, \{2, 5\}, \{3, 5\}, \{4, 5\}, \{5, 6\}\}$
$G \equiv D$: DAG

**Local Markov property**

\[ \forall u \in V \]
\[ u \perp \{ \text{nd}(u) \setminus \text{pa}(u) \} \mid \text{pa}(u) \]

$G$: UG

$A, B, S$ disjoint subsets of $V$

**Global Markov property**

If $S$ separates $A$ from $B$ in $G$, then

\[ A \perp B \mid S \]
Relationships Among Graphical Models
Typically we do **NOT** know the structure of the graph

**Aim**
Discover the graph using data

Structural learning

Graph selection
Graphical models

Objective Bayes Model Selection

DAG-models
  Compatible Parameter Priors
  Marginal Likelihood
  Covariate-adjusted graphical models

Experimental results
Bayesian model \( \mathcal{M}_k = \{ f_{\mathcal{M}_k}(Y | \theta_k), p(\theta_k) \} \)

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

\[ m_{\mathcal{M}_k}(Y) = \int f_{\mathcal{M}_k}(Y | \theta_k)p(\theta_k)d\theta_k \]

marginal likelihood of \( \mathcal{M}_k \)

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

\[ BF_{kk'}(Y) = m_{\mathcal{M}_k}(Y)/m_{\mathcal{M}_{k'}}(Y) \]
Default Improper Priors

Lack of substantive prior information
\[ p(\theta_k) = p^D(\theta_k) \]
\[ p^D(\theta_k): \text{objective default (non-informative) prior} \]
Often improper

Cannot be used naively to compute Bayes factors even when \( m_{\mathcal{M}_k}(Y) \) 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

\[ b = b(n), \ 0 < b < 1: \text{fraction of sample size } n \]

Fractional marginal likelihood of model \( \mathcal{M}_k \)

\[
m_{\mathcal{M}_k}(\mathbf{Y}; b) = \frac{\int f_{\mathcal{M}_k}(\mathbf{Y} | \theta_k) p^D(\theta_k) d\theta_k}{\int f^b_{\mathcal{M}_k}(\mathbf{Y} | \theta_k) p^D(\theta_k) d\theta_k}
\]

Can be rewritten as

\[
m_{\mathcal{M}_k}(\mathbf{Y}; b) = \int f^{1-b}_{\mathcal{M}_k}(\mathbf{Y} | \theta_k) p^F(\theta_k | b, \mathbf{Y}) d\theta_k
\]

\[ p^F(\theta_k | b, \mathbf{Y}) \propto f^b_{\mathcal{M}_k}(\mathbf{Y} | \theta_k) p^D(\theta_k): \text{fractional prior} \]

Fractional Bayes factor (FBF)

\[
\text{BF}_{kk'}(\mathbf{Y}; b) = m_{\mathcal{M}_k}(\mathbf{Y}; b) / m_{\mathcal{M}_{k'}}(\mathbf{Y}; b)
\]

Default choice: \( b = n_0/n \)

\( n_0 \): minimal (integer) training sample size
Graphical models

Objective Bayes Model Selection

**DAG-models**
- Compatible Parameter Priors
- Marginal Likelihood
- Covariate-adjusted graphical models

Experimental results
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

\[ y \mid \mu, \Omega_\mathcal{G} \sim N(\mu, (\Omega_\mathcal{G})^{-1}) \]

\( \Omega_\mathcal{G} \) Markov w.r.t. UG \( \mathcal{G} \)
\( \Omega_\mathcal{G} \) s.p.d.

but also
\( (\Omega_\mathcal{G})_{ij} = 0 \) whenever there is no edge between \( i \) and \( j \) in \( \mathcal{G} \)

Constrained parameter space
\( p(\Omega_\mathcal{G}) \) must comply with this constraint
Collection of DAG-models
Joint sampling density under DAG-model $\mathcal{D}$

\[
f_{\mathcal{D}}(y_i \mid \theta_{\mathcal{D}}) = \prod_{j=1}^{q} f(y_{ij} \mid y_{i,pa_{\mathcal{D}}(j)}; \theta_j)
\]

Assume

- 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 $pa_{\mathcal{D}_1}(j) = pa_{\mathcal{D}_2}(j)$, 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 $\text{pa}_{\mathcal{D}_1}(j) = \text{pa}_{\mathcal{D}_2}(j)$, prior for $\theta_j$ under $\mathcal{D}_1$ same as prior on $\theta_j$ under $\mathcal{D}_2$ [“Prior compatibility”]

- **5: Global parameter independence**

  $$ p_\mathcal{D}(\theta_\mathcal{D}) = \prod_{j=1}^{q} p_\mathcal{D}(\theta_j) $$

  [local parameters $\theta_j$’s mutually stochastically independent]

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

If Assumptions 1-5 hold, then

\[ m_D(Y) \stackrel{(1)}{=} \prod_{j=1}^{q} \int p_D(\theta_j) \prod_{i=1}^{n} f_D(y_{ij} \mid y_i, \text{pa}_D(j); \theta_j) d\theta_j \]

\[ = \prod_{j=1}^{q} \int p_{C_j}(\theta_j) \prod_{i=1}^{n} f_{C_j}(y_{ij} \mid y_i, \text{pa}_{C_j}(j); \theta_j) d\theta_j \]

\[ = \prod_{j=1}^{q} \int p_{C_j}(\theta_j) f_{C_j}(Y_j \mid Y_{\text{pa}_{C_j}(j)}; \theta_j) d\theta_j \]

\[ = \prod_{j=1}^{q} m_{C_j}(Y_j \mid Y_{\text{pa}_{C_j}(j)}) \]

\( C_j \) is any complete DAG such that \( \text{pa}_{C_j}(j) = \text{pa}_D(j) \)

(1) use global parameter independence
(2) use prior and likelihood modularity
(3) recall that \( Y_j = (y_{ij}; i = 1, \ldots, n) \)
(4) by definition of \( m_{C_j}(Y_j \mid Y_{\text{pa}_{C_j}(j)}) \)
In conclusion

\[
m_D(Y) = \prod_{j=1}^{q} m_{C_j}(Y_j \mid Y_{pa_{C_j}(j)}) = \prod_{j=1}^{q} \frac{m_{C_j}(Y_{fa_{C_j}(j)})}{m_{C_j}(Y_{pa_{C_j}(j)})} = \prod_{j=1}^{q} \frac{m(Y_{fa_D(j)})}{m(Y_{pa_D(j)})},
\]

\[fa_D(j) = pa_D(j) \cup \{j\}\]

\[m(\cdot)\]: 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 \(Y\)
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 $\mathcal{G}^<$]
$C$: set of cliques
$S$: set of separators

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

$m(\cdot)$: marginal data distribution under any complete graph
Marginal Likelihood of a Gaussian DAG-model

\[ \mathbf{y}_i \mid \mu, \Omega, \mathcal{D} \overset{\text{iid}}{\sim} N_q(\mu, \Omega^{-1}_D); \ i = 1, \ldots, n \]

\( \Omega_D \): constrained precision matrix, Markov with respect to \( \mathcal{D} \)
Assumptions 1-3 satisfied
Take any complete Gaussian DAG model \( \mathcal{D} = \mathcal{C} \) (\( \Omega_C = \Omega \) unconstrained)
If \( (\mu, \Omega) \sim \text{Normal} - \text{Wishart} \)
\[ p(\mu \mid \Omega) = N_q(\mu \mid m, (c\Omega)^{-1}) \]
\[ p(\Omega) = W_q(a, R) \]
Then Assumptions 4-5 are satisfied
Closed-form expressions for \( m(\mathbf{Y}_J) \) 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 ($\mu = 0$) starting from an improper prior.

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

Carvalho and Scott (2009, *Biometrika*) for decomposable UGs start with an improper Hyper Inverse Wishart on the constrained $\Sigma_G = (\Omega)^{-1}_G$ and then use the FBF.
Marginal Likelihood of a Multivariate Gaussian Regression DAG Model

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

\(X\): known design matrix

\[ Y \mid B, \Omega \sim \mathcal{N}_{n,q}(XB, I_n, \Omega^{-1}_D) \]

Assume \((B)\) and \(\Omega\) unrestricted. Prior

\[ B \mid \Omega \sim \mathcal{N}_{p+1,q}(B, C^{-1}, \Omega^{-1}), \]
\[ \Omega \sim \mathcal{W}_q(a, R) \]

Assumptions 1-5 for the construction of compatible priors hold


Closed form expression for the marginal of any set of variables
\(Y_J\) given \(X\) 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
- 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
\( p \): 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

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

\( G \) Undirected decomposable graph

\[ m_G(Y) = \frac{\prod_{C \in C} m(Y_C)}{\prod_{S \in S} m(Y_S)} \]

\[ m(Y_C) = \pi^{-\frac{(n-n_0)|C|}{2}} \frac{\Gamma_{|C|} \left( \frac{a_D + n - p - 1 - |C|}{2} \right)}{\Gamma_{|C|} \left( \frac{a_D + n_0 - p - 1 - |C|}{2} \right)} \left( \frac{n_0}{n} \right)^{\frac{|C|(a_D + n_0 - |C|)}{2}} \left| \hat{E}_C \hat{E}_C \right|^{-\frac{n-n_0}{2}} \]

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

Formula for \( m_G(Y) \) holds provided \( n > p + |C|, \ C \in C \)

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

Need to perform variable selection

variable indicators

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

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

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

Under sparse regression

$p_{\gamma} << p_\star$
Joint variable and graph selection

$\gamma$: 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

$$Y \mid B_\gamma, \Omega_\mathcal{G}, \gamma, \mathcal{G} \sim \mathcal{N}_{n,q}(X_\gamma B_\gamma, I_n, \Omega_\mathcal{G}^{-1})$$

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

\[ \textbf{Y} | \gamma, G \sim m_g(\textbf{Y} | \gamma) \]
\[ \gamma_i | \omega_\gamma \sim \text{Ber}(\omega_\gamma); \quad i = 2, \ldots, p_\star + 1 \]
\[ G_i | \omega_G \sim \text{Ber}(\omega_G); \quad i = 2, \ldots, q \cdot (q - 1)/2 \]
\[ \omega_\gamma \sim U(0, 1) \]
\[ \omega_G \sim U(0, 1) \]

\( m_g(\textbf{Y} | \gamma) \): marginal likelihood
generated through our O’Bayes method based on the FBF

\[ G = (G_1, \ldots, G_{q(q-1)/2}) \]
vectorized \textit{adjacency matrix} corresponding to \( G \)
Graphical models

Objective Bayes Model Selection

DAG-models
  Compatible Parameter Priors
  Marginal Likelihood
  Covariate-adjusted graphical models

Experimental results
Two simulation settings

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

- **Magnified block**
  - $n = 50$
  - $p^* = 100$
  - $q = 150$

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
  \[ G_i = 0 \text{ for } i \leq q(q - 1)/2 - 10 \text{ and } G_i = 1 \text{ otherwise for } i > q(q - 1)/2 - 10. \]
  
  The \( q \times q \) adjacency matrix has a sparse bottom-right block of active edges.
  
  Sparsity of \( G \) increases with \( q \).

- Regression structure
  Out of \( p_* = 100 \) potential covariates, true predictors are only the first and the third:
  \[ \gamma_i = 1 \text{ for } i \in \{1, 2, 4\} \]
  \[ \gamma_i = 0 \text{ otherwise} \]
  
  \( p_\gamma = 2 \)

- Given true \( G \), \( \Omega_G \) is sampled from the G-Wishart distribution with 10 degrees of freedom and scale matrix equal to the identity.

- Given true \( \gamma \) and \( \Omega_G \)
  \( B \) sampled from the Matrix Normal \( N_{p_* + 1, q} \left( \begin{array}{c} 0_{p_* + 1, q} \end{array} \right, 0.3^2 I_{p_* + 1}, \Omega_G^{-1} \)\)

- Elements of \( X \) from the second to last column are randomly drawn from \( N(10, 1) \)
• \( q = 150 \)

• Graph structure
  Fix a \( 50 \times 50 \) adjacency matrix \( G' \) as above
  Full \( G \) is block diagonal with \( G' \) replicated three times
  Corresponding \( \Omega_G \) is block diagonal with the three blocks \( \Omega'_{G}, 5\Omega'_{G}, 10\Omega^1_{G} \)
  \( \Omega_G \) has sequentially magnified signals

• Regression structure
  True \( \gamma \) produced by randomly choosing each predictor with probability 0.05 among \( p_* \) 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)
- Two-step ANTAC (Asymptotically Normal with Thresholding after Adjusting Covariates) estimator
  (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*)
- Sparse Gaussian Conditional method (CONDIT)
  (graph selection; estimation of regression parameters intermediate step)
- Low Rank latent variables and sparse method (LOWRANK)
  (graph selection with unobserved latent variable)
Measures of performance for graph selection (Adjacency matrix)

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

\[
\text{MISR} = \frac{FN + FP}{q(q-1)}, \quad \text{SPE} = \frac{TN}{TN + FP},
\]

\[
\text{SEN} = \frac{TP}{TP + FN}
\]

\[
\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}
\]
<table>
<thead>
<tr>
<th>Setting</th>
<th>$(n, p^*, q)$</th>
<th>Method</th>
<th>MISR</th>
<th>SPE</th>
<th>SEN</th>
<th>MCC</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse</td>
<td>$(50, 100, 30)$</td>
<td>OBFBF</td>
<td>9(1)</td>
<td>92(1)</td>
<td>74(3)</td>
<td>47(5)</td>
<td>4769</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HYPERT</td>
<td>10(1)</td>
<td>91(1)</td>
<td>74(4)</td>
<td>46(2)</td>
<td>4270</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANTAC</td>
<td>1(0)</td>
<td>100(0)</td>
<td>72(1)</td>
<td>84(1)</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GLASSO</td>
<td>83(5)</td>
<td>17(5)</td>
<td>86(4)</td>
<td>15(2)</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CONDIT</td>
<td>52(11)</td>
<td>48(11)</td>
<td>90(7)</td>
<td>21(4)</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LOWRANK</td>
<td>49(14)</td>
<td>50(14)</td>
<td>91(7)</td>
<td>22(5)</td>
<td>75</td>
</tr>
<tr>
<td>Sparse</td>
<td>$(50, 100, 60)$</td>
<td>OBFBF</td>
<td>3(2)</td>
<td>97(2)</td>
<td>84(1)</td>
<td>60(19)</td>
<td>5550</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HYPERT</td>
<td>5(0)</td>
<td>95(0)</td>
<td>84(2)</td>
<td>47(1)</td>
<td>5990</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANTAC</td>
<td>0(0)</td>
<td>100(0)</td>
<td>83(1)</td>
<td>91(0)</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GLASSO</td>
<td>59(5)</td>
<td>41(5)</td>
<td>93(2)</td>
<td>12(1)</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CONDIT</td>
<td>27(19)</td>
<td>73(20)</td>
<td>89(4)</td>
<td>24(6)</td>
<td>268</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LOWRANK</td>
<td>81(3)</td>
<td>18(3)</td>
<td>97(3)</td>
<td>7(1)</td>
<td>236</td>
</tr>
<tr>
<td>Sparse</td>
<td>$(50, 100, 120)$</td>
<td>OBFBF</td>
<td>0(0)</td>
<td>100(0)</td>
<td>100(0)</td>
<td>95(5)</td>
<td>3745</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HYPERT</td>
<td>2(0)</td>
<td>98(0)</td>
<td>91(1)</td>
<td>54(1)</td>
<td>5941</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANTAC</td>
<td>0(0)</td>
<td>100(0)</td>
<td>91(0)</td>
<td>95(0)</td>
<td>676</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GLASSO</td>
<td>36(4)</td>
<td>64(4)</td>
<td>95(1)</td>
<td>12(1)</td>
<td>547</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CONDIT</td>
<td>48(25)</td>
<td>52(25)</td>
<td>96(2)</td>
<td>11(5)</td>
<td>861</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LOWRANK</td>
<td>94(1)</td>
<td>6(1)</td>
<td>99(1)</td>
<td>1(0)</td>
<td>1002</td>
</tr>
<tr>
<td>Magnified</td>
<td>$(50, 100, 150)$</td>
<td>OBFBF</td>
<td>0(0)</td>
<td>100(0)</td>
<td>93(0)</td>
<td>92(12)</td>
<td>5498</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HYPERT</td>
<td>2(0)</td>
<td>99(0)</td>
<td>93(0)</td>
<td>54(1)</td>
<td>6770</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANTAC</td>
<td>0(0)</td>
<td>100(0)</td>
<td>93(0)</td>
<td>96(0)</td>
<td>1971</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GLASSO</td>
<td>78(5)</td>
<td>22(5)</td>
<td>97(1)</td>
<td>5(1)</td>
<td>4570</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CONDIT</td>
<td>96(3)</td>
<td>4(3)</td>
<td>100(1)</td>
<td>2(1)</td>
<td>3517</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LOWRANK</td>
<td>98(0)</td>
<td>2(0)</td>
<td>100(0)</td>
<td>1(0)</td>
<td>5452</td>
</tr>
</tbody>
</table>
Sparse setting: $n = 200$, $p_{\star} = 100$, $q = 30$
Sparse setting: $n = 200, p_* = 100, q = 30$
• 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)
ROC curve: graph selection

(b) Graph selection

1−SPE

SEN

0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0

Black: OBFBF
Red: HYPERT
Variable selection OBFBF

(c) Truth

(d) Estimation
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 serial dependence, spatial dependence.
- Explore the space of Essential DAGs (Markov Equivalence Class).
  Require calculations for Chain Graphs.
  Use observational and interventional data.
Joint high-dimensional Bayesian variable and covariance selection with an application to eQTL analysis. *Biometrics* 69, 447–457.

Covariate-adjusted precision matrix estimation with an application in genetical genomics. *Biometrika* 100, 139–156.

Objective Bayesian model selection in Gaussian graphical models. *Biometrika* 96, 497–512.


Sparse inverse covariance estimation with the graphical lasso.
*Biostatistics* 9, 432–441.

Parameter priors for directed acyclic graphical models and the characterization of several probability distributions.

Addendum on the scoring of Gaussian directed acyclic graphical models.

Fractional Bayes factors for model comparison.

Sparse Gaussian conditional random fields: Algorithms, theory, and application to energy forecasting.