An empirical Bayes approach to network recovery using external data

G.B. Kpogbezan (UL)
Joint work with:
Aad van der Vaart (UL), Wessel N. van Wieringen (VU & VUmc), Gwenael G.R. Leday (MRC Biostatistics Unit, Cambridge), Mark van de Wiel (VU & VUmc)

- Goal: network reconstruction from data and use of prior/external knowledge.

- Network = graph. A graph consists of a pair $(\mathcal{I}, \mathcal{E})$ where $\mathcal{I} = \{1, 2, ..., p\}$ is a set of indices representing nodes and $\mathcal{E}$ is the set of edges (relations between the nodes) in $\mathcal{I} \times \mathcal{I}$.

- Here edges reflect conditional dependencies between the nodes $\Longrightarrow$ Conditional Independence Graph.

- **Data**: $Y^j \sim^{\text{iid}} \mathrm{N}(0, \Omega_p^{-1}), \quad j \in \{1, ..., n\}$ where Ω_p^{-1} is the covariance matrix and $\Omega_p = (w_{kl})_{k,l=1,...,p}$ is the inverse covariance (or precision) matrix.

- In this setting (Gaussian Graphical Model), it holds $corr(Y_{i_1}, Y_{i_2} | \mathbf{Y}_{-i_1,-i_2}) = w_{i_1 i_2}$ (conditional dependency).

- Recontructing the network (conditional independence graph) is equivalent to determine the support of Ω_p.

- For $n \ll p$, typically for gene expression data, the problem of estimating Ω_p is not feasible.

- Some proposed solution:
  Graphical lasso : maximize the penalized log-likelihood

$$\log(det\Omega_p) - tr(S\Omega_p) - \rho||\Omega_p||_1$$

  over the space of positive definite matrices M^+ with shrinkage parameter $\rho > 0$.

SEMs

- **Simultaneous Equations Models** (SEMs): modeling of the full conditional distribution of each node and result in a system of p regression equations.
- It is:

$$Y_i = \sum_{s=1, s\neq i}^{p} \beta_{is} Y_s + \epsilon_i, \quad i \in \mathcal{I}. \tag{1}$$

- Equivalence between regression parameters and precision matrix elements, namely $\beta_{is} = w_{ii}^{-1} w_{is}$.
- Estimation of support of $\Omega_p \iff$ variables selection in p regressions.

Bayesian SEM

- Meinshausen & Buehlmann put a lasso penalty on each regression parameter to select the neighbors of each variable.
- Previously we proposed a Bayesian formulation of the SEM (BSEM) and put priors on parameters in (1). It is:

$$\begin{aligned} \epsilon_i &\sim \mathrm{N}(0_n, \sigma_i^2 \mathbf{I}_n), \\ \beta_{is} &\sim \mathrm{N}(0, \sigma_i^2 \tau_i^{-2}), \\ \tau_i^2 &\sim \Gamma(a, b), \\ \sigma_i^{-2} &\sim \Gamma(c, d) \end{aligned} \tag{2}$$

c,d: non-informative.

Variational Bayes

- Variational approximation to a distribution = closest element in a target set $\mathcal{Q}$ chosen both for computational tractability and accuracy.
- distance measured by Kullback- Leibler divergence.
- Distributions $\mathcal{Q}$ with stochastically independent marginals (i.e. product laws) are popular.
- Accuracy of approximation naturally restricted to the marginal distributions.

Why is Bayesian SEM attractive?

- Double shrinkage; Amount of shrinkage is node-specific.
- Fast posterior approximation by Variational Bayes.
- Approximate joint posterior under product laws assumption:

$$\pi(\beta_i, \tau_i^2, \sigma_i^{-2}) \approx q(\beta_i, \tau_i^2, \sigma_i^{-2}) = q_1(\beta_i) q_2(\tau_i^2) q_3(\sigma_i^{-2})$$

- Appealing EB procedure for hyperparameters estimation.
- Efficient EM-type algorithms for minimization.

Why is Variational Bayes attractive?

- It is FAST (remember: p penalized regressions...).
- It is ACCURATE (verified by Gibbs sampling).
- Analytical lower-bound for marginal likelihood: $M_i(a, b)$
  - Summation: $M(a, b) = \sum_{i=1}^{p} M_i(a, b)$
  - Maximize $M(a, b)$: EB estimate of prior parameters.
- $M_i(a, b)$ facilitates posterior edge selection

Extension of BSEM to BSEMed

- Prior knowledge on the to-be-reconstructed network topology usually available for instance
  - from pathway repositories like KEGG
  - inferred from data of pilot study.
- Natural to take such information into account during network reconstruction.
- Prior knowledge assumed to be available as a prior network, which specifies which edges are present and absent.

Incorporation of prior network P in form of adjacency matrix containing only zeros (no edge) and ones (edge is present):

$$
\begin{aligned}
Y_i = \sum_{s=1, s\neq i}^{p} \beta_{is} Y_s + \epsilon_i, \quad i \in \mathcal{I}, \\
\epsilon_i \sim \mathsf{N}(0_n, \sigma_i^2 \mathbf{I}_n), \\
\beta_{is} \sim \mathsf{N}(0, \sigma_i^2 \tau_{i,P_{is}}^{-2}) \\
\tau_{i,P_{is}}^2 \sim \Gamma(a_{P_{is}}, b_{P_{is}}), \\
\sigma_i^{-2} \sim \Gamma(c, d)
\end{aligned}
\tag{3}
$$

c,d: non-informative.

VB vs Gibbs sampling

- How close is the variational approximation to the true posterior distribution?
- Here we investigate this question by comparing the variational Bayes estimates of the marginal densities with the corresponding Gibbs sampling-based estimates.
- $n = 50$ independent replicates from a $\mathrm{N}(0, \Omega_p^{-1})$ with $p = 50$.
- Ω_p was chosen to be a band matrix with $b_l = b_u = 4 \Longrightarrow$ a total number of 9 band elements including the diagonal.
- Simulation study with a single regression equation (say i = 1).

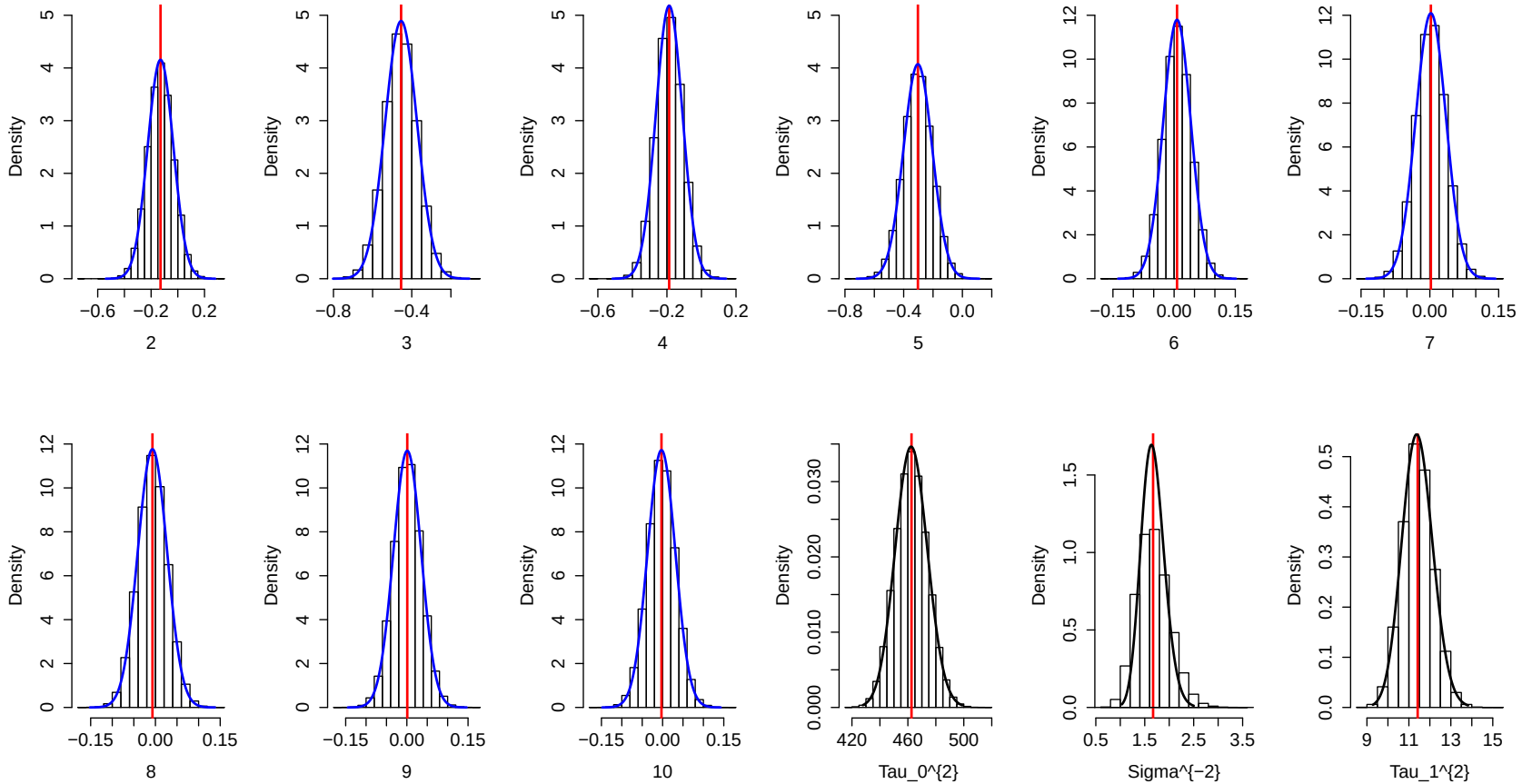
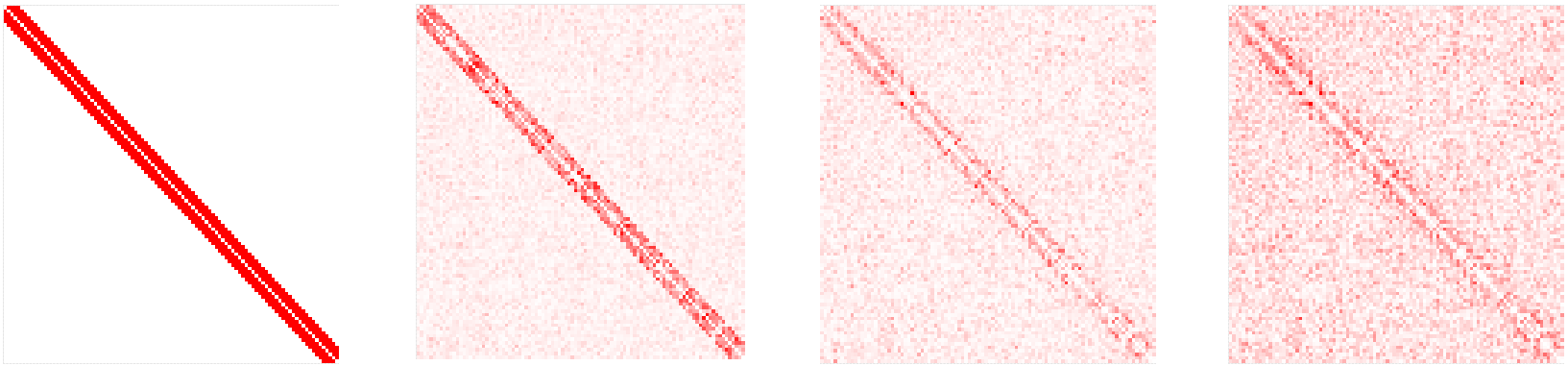


Figure: Comparison of variational marginal densities of $\beta_{1,2}, \ldots, \beta_{1,10}$ (blue curves) and $\tau_{1,0}^2$, $\tau_{1,1}^2$ and σ_1^{-2} (black curves) with corresponding Gibbs sampling-based histograms. The red vertical lines display the variational marginal means.

VB vs Gibbs sampling: Time comparison

	BSEMed	Gibbs sampling
time in seconds	40	$2542 \times 50 = 127,100$

Computing times for an R-implementation of the variational Bayes method and the Gibbs sampling method with $n = p = 50$.



(a) True graph (b) BSEMed: true (c) BSEMed: 50 % (d) BSEM

Figure: Visualization of BSEMed estimate using perfect prior (b), BSEMed estimate using 50% true edges information (c), BSEM estimate (d) and the true graph (a) in case $n = 50$ and $p = 100$.

Data: Gene expression data from GEO

Lung: 49 Normals, 58 Cancers
Apoptosis pathway: p = 84 genes

Pancreas: 39 Normals, 39 Cancers
p53 pathway: p = 68 genes.

Idea: Use network fitted on Normals to inform network for Cancers.

EB estimate of prior mean $\tau^2_{i,0}$ and $\tau^2_{i,1}$

	Not in Normal Network	In Normal Nerwork	ratio
Lung	27.32	1.71	20.13
Pancreas	20.03	1.21	12.97

- Prior networks are clearly of use:
  - the mean prior precision for regression parameters corresponding to the edges absent in the prior network is relatively large
  - stronger shrinkage towards zero compared to mean prior precision corresponding to edges present in the prior network.

Reproducibility

100 random splits of the data: What proportion of top 50 edges reproduces, on average?

# edges	BSEM	SEMLasso	glasso	BSEMed
50	9.12%	2.64%	6.84%	59.16%

Lung data, average percentage edges in overlap.

# edges	BSEM	SEMLasso	glasso	BSEMed
50	14.84%	5.6%	9.04%	55.64%

Pancreas data, average percentage edges in overlap.

- BSEMed and BSEM are attractive framework for network inference and computationally very fast.
- Performance of BSEMed increase when the external data is relevant.
- BSEMed performs **as good as** BSEM when external data are **not relevant at all**.
- In case of multiple sources of external data BSEMed can be easily used: one at a time.

THANK YOU!!!