Amir Bashir and Beatrix Jones
A.Bashir@massey.ac.nz M.B.Jones@massey.ac.nz
Institute of Natural and Mathematical Sciences
Massey University, Auckland New Zealand

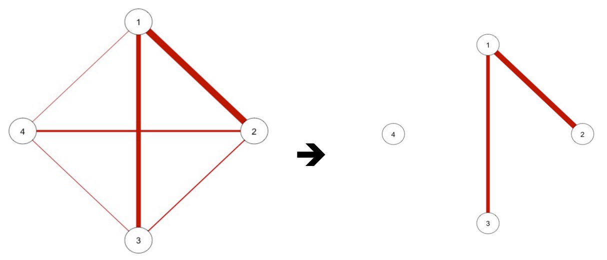
Sparse estimates from dense precision matrix posteriors

Introduction

Model selection is a key statistical process because it highlights important variables and relationships. In addition, by reducing the number of parameters to be estimated it improves estimation for the remaining parameters. For a precision matrix, this looks like:

$$\begin{pmatrix} 16.7 & -3.9 & 8.5 & -0.2 \\ -3.9 & 2.7 & -0.5 & -1.4 \\ -8.5 & -0.5 & 30.0 & -0.6 \\ -0.2 & -1.4 & -0.6 & 34.6 \end{pmatrix} \Rightarrow \begin{pmatrix} 17.9 & -4.2 & -8.9 & 0 \\ -4.2 & 2.5 & 0 & 0 \\ -8.9 & 0 & 29.3 & 0 \\ 0 & 0 & 0 & 32.2 \end{pmatrix}$$

Or, visualize the matrix as a graph, with edges representing non –zero elements.



Searching or sampling over space of sparse matrices is computationally intensive. For precision matrices, even after the effort to sample a posterior over the space of sparse matrices, the resulting Bayes estimate is typically not sparse. The fit to future data (based on log likelihood) is:

$$fit(\Gamma) = \log\det\Gamma - \mathrm{tr}\left(\frac{X^{*}X^{*T}\Gamma}{n^{*}}\right)$$

Which has its expected value maximized at

$$\Gamma = \bar{\Sigma}^{-1}$$

- Actual fit of this estimate is a random variable governed by the posterior predictive distribution [1].
- Consider top 95% of this distribution, seeking a sparser choice for Γ with fit still in this range.
- This methodology is applicable to *any* posterior over precision matrices, even if the sampled matrices have no zero elements.
- Can also be used to consider *differences* in precision matrices.

Example Data:

Fecal Volatilome: 174 compounds measured by mass spectrometry of fecal samples.

- Control group: 49 8-year-old children born at term.
- 42 Cases, children of the same age born pre-term.

Organic Acids: 7 organic acids measured in 3 groups of roughly 30 people each [2]:

- HIV positive (with treatment) and obese,
- HIV positive (with treatment) with normal weight,
- HIV negative and obese.

Methods of "sparsifying" the posterior mean

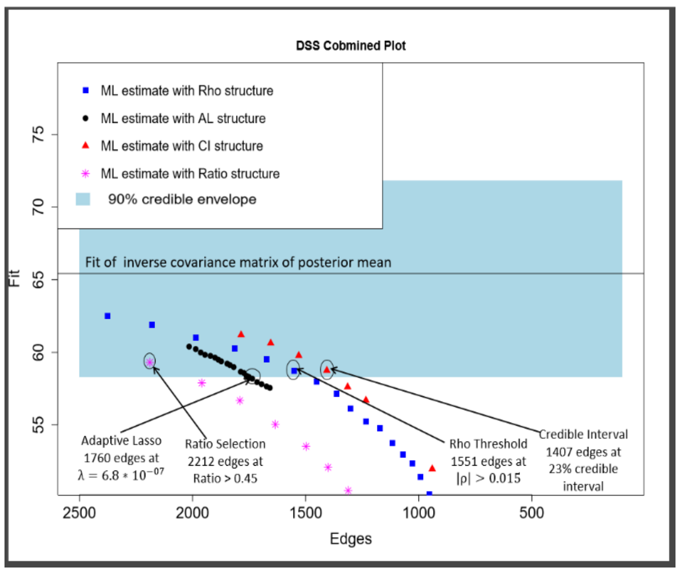
Set elements to zero if

- Partial correlation (ρ) less than X
- X% credible interval includes zero
- Ratio with estimate based on conjugate Wishart prior is less than X
- Adaptive Lasso with shrinkage parameter X

In each case, find best Γ with specified zero structure. Vary X to see how the fit is impacted. Select the sparsest model in the blue envelope.

Example: Fecal Volatilome control data:

- Posterior generated using Bayesian Adaptive Lasso [3,4]
- Samples all dense precision matrices.
- Credible interval method produces sparsest graphs (red triangles).



Finding sparse matrix differences

Interested in how (if) precision matrix differs across *C* conditions (eg case/control).

- Generate independent posteriors for each condition.
- Fit = sum of fits across conditions, weighted by sample size.
- No ready algorithm to find best fitting, positive definite matrices obeying a particular set of constraints (eg, constrain elements corresponding to overlapping credible intervals to be the same).
- Joint Graphical Lasso (JGL) [5] used instead to sparsify difference. L1 penalty on matrix differences (λ_2) and size of off diagonal elements(λ_1). We found use of an adaptive penalty crucial:

$$\max_{\Gamma}\left[\sum_{c=1}^{C} n_c\left(\log\det\Gamma_c - \mathrm{tr}(\bar{\Sigma}_c\Gamma_c)\right) + \lambda_1\sum_{c=1}^{C}\sum_{i\neq j}\frac{|\gamma_{cij}|}{\sqrt{\gamma^{*}_{cij}}} + \lambda_2\sum_{c<c'}\sum_{i,j}\frac{|\gamma_{cij}-\gamma_{c'ij}|}{\sqrt{d_{ij}}}\right],$$

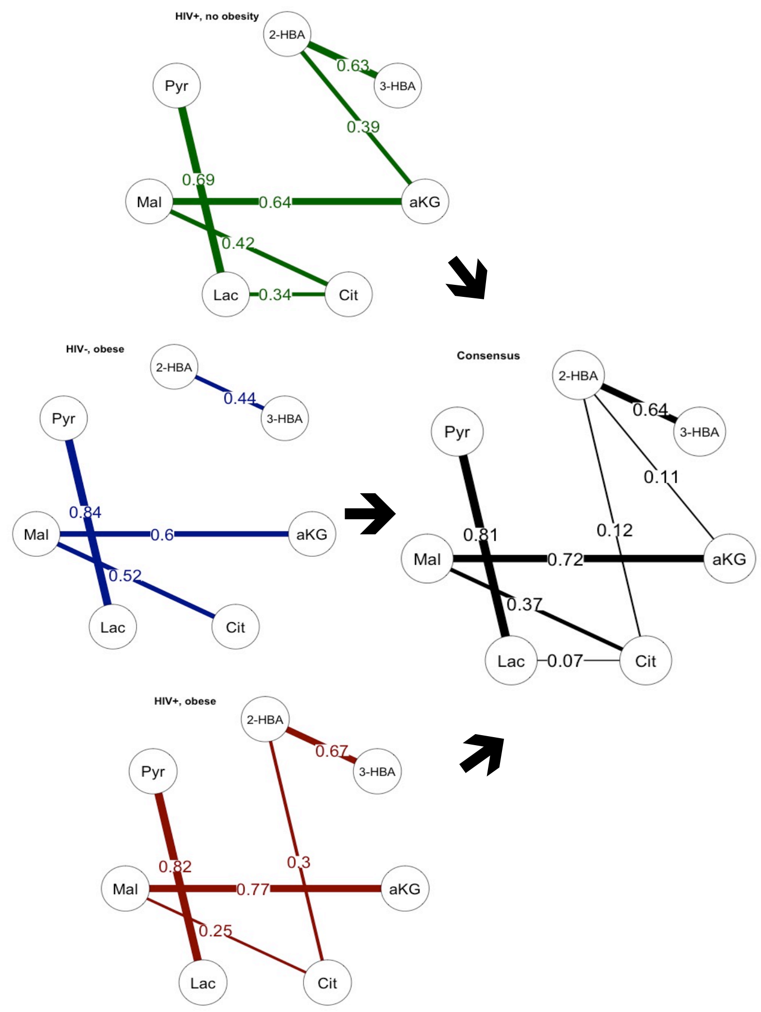
where

$$\begin{aligned}\gamma{*}_{cij} &= \max(|\gamma^{(1)}_{ij}|, 0.00001) \text{ and} \\ d_{ij} &= \max(\sum_{c=1}^{C}|\gamma^{(1)}_{cij} - \bar{\gamma}^{(1)}_{ij}|, 0.00001).\end{aligned}$$

- Benefit of starting with sparsified matrices (share some zero elements). Starting graphs for each condition chosen via credible interval method, but in top 60% of fit.
- λ_1 ensures common zeros maintained as other parts of the matrix are modified.

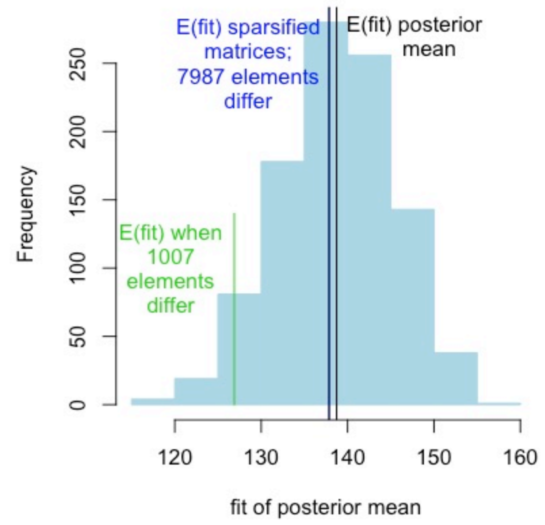
Application to Organic Acids data

Edges annotated with corresponding partial correlations. JGL produced the consensus graph on the right, which is still in the 95% fit window.



Application to fecal volatilome data

Same strategy applied to higher dimensional dataset.



Conclusion

- Sensible sparse estimates can be produced from (potentially more tractable) posteriors over dense matrices.
- Readily extended to the case of differences in the precision matrix across conditions.

References

1. Hahn, P. R. and Carvalho, C. M. (2015). Decoupling shrinkage and selection in Bayesian linear models: a posterior summary perspective. JASA, 110: 435-448.
2. Koethe, J. R., Jenkins, C. A., Petucci, C., et al. (2015). Superior Glucose Tolerance and Metabolomic Profiles, independent of adiposity, in HIV infected women compared with men on anti-retroviral therapy. Medicine, 95(19): e3634.
3. Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. Bayesian Analysis, 7(867-886).
4. Peterson, C. B., Vannucci, M., Karakas et al (2013). Inferring metabolic networks using the Bayesian adaptive graphical lasso with informative priors. Statistics and its Interface, 6(4): 437-558
5. Danaher, P., Wang, P., and Witten, D. M. (2013). The joint graphical lasso for inverse covariance estimation across multiple classes. JRSSB 76(2): 373-397.