

# Bayesian inference on high-dimensional Seemingly Unrelated Regressions

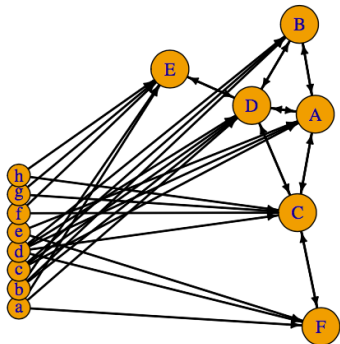
Alex Lewin and Marco Banterle

London School of Hygiene and Tropical Medicine

June 2018

# Motivation

- Genetic studies aiming at identifying association between point mutations (SNPs) and **multivariate** phenotypes:
  - gene expression measurements
  - metabolomics data
  - protein concentrations
  - ...
- Looking for sparse variable selection
- Take into account data correlations (possibly sparse)



# Bayesian Setting

## High dimensional data

- $p \approx 10^4$  to  $10^6$  variables in  $X$
- $q$  ranging from 1 to  $10^4$  variables in  $Y$
- Around  $n = 5000$  observations

## Focus on **Sparse Bayesian Variable Selection** (sparse BVS)

- Estimate using MCMC
- Provides the **posterior probability of association** for each predictor and each response (model averaging).

Frame the problem as a multivariate linear regression model:

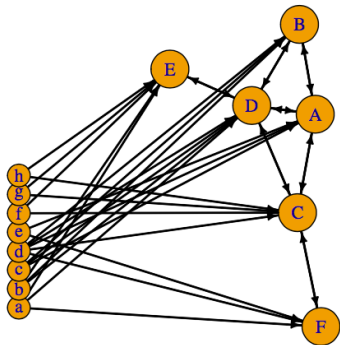
$$Y = X B + E$$

$n \times q$      $n \times p$   $p \times q$      $n \times p$

or equivalently:

$$Y \sim \mathcal{MN}(XB, \mathbb{I}_n, C)$$

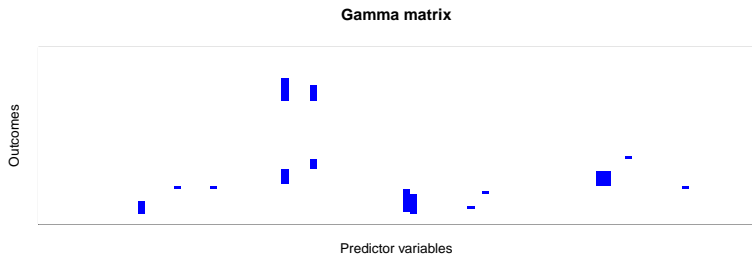
- Sparse variable selection on associations ( $B$ )
- Sparse covariance selection ( $C$ )



Variable selection performed through binary matrix  $\Gamma$  ( $p \times q$ )

$$\gamma_{jk} = \begin{cases} 1 & \implies B_{jk} \neq 0 \\ 0 & \implies B_{jk} = 0 \end{cases}$$

Sparsity prior  $\gamma_{jk} \sim \text{Bern}(\omega_{jk})$ ,  $\omega_{jk} \sim \text{Beta}()$

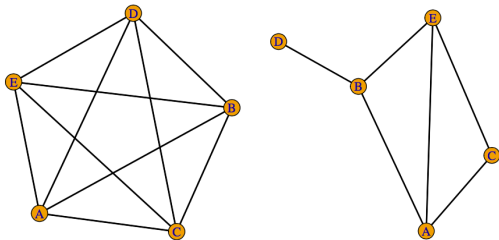


Predictor  $X_j$  only appears in a regression if  $\gamma_{jk}$  is 1.

## Two options for Covariance matrix:

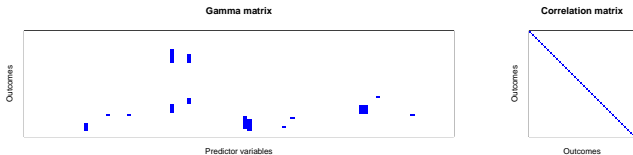
- Dense covariance matrix  $C$  with Inverse Wishart prior
- Covariance selection performed using Gaussian Graphical models:

$(C^{-1})_{ij} \leftrightarrow$  outcomes  $y_i$  and  $y_j$  are conditionally independent.



# Previous work in Bayesian multivariate regression

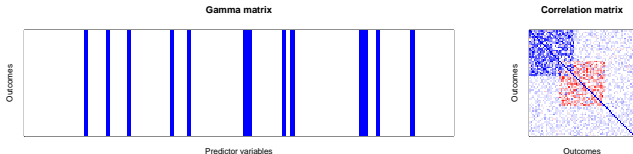
- Either assume diagonal covariance matrix



*Bottolo L, Chadeau-Hyam, M et al. (2013)*

*Lewin A et al. (2015)*

- Or assume all responses related to the same set of predictors

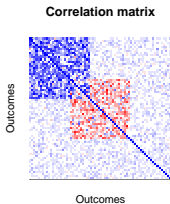
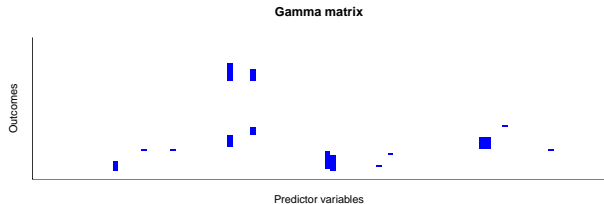


*Bottolo L, Petretto, E et al. (2011)*

*Bhadra A and Mallick BK (2013)*

## Our work on SUR model

- Full selection matrix  $\Gamma$ ; Full covariance matrix  $R$



Formulate as a Seemingly Unrelated Regressions (SUR) model:

$$\mathbf{y}_k = X_{\gamma_k} \beta_{\gamma_k} + \epsilon_k \quad \text{for } k = 1, \dots, q$$

$n \times 1$        $n \times d_k$     $d_k \times 1$        $n \times 1$

$Cov[\epsilon_k \epsilon_l] = C_{kl} \neq 0 \implies$  Outcomes do not naturally separate as in previous hierarchical model.



In both “previous” cases, models are conjugate in  $B$  and  $C$   
→ only  $\Gamma$  (variable selection) are updated.

- In the SUR model, Standard priors (Normal, Inverse Wishart) → Not Conjugate in  $B$  or  $C$
- Can calculate posterior full conditionals for  $\beta_k$  and  $C$  → Gibbs sampler for  $\gamma_k, \beta_k$  and  $C$ .
- However, computationally intensive if use naive updates.

## A Factorisation of the Covariance Matrix

From Zellner and Ando (2010): decompose the Likelihood:

$$\begin{cases} \mathbf{y}_1 = X_{\gamma_1} \boldsymbol{\beta}_{\gamma_1} + \boldsymbol{\varepsilon}_1 \\ \mathbf{y}_2 = X_{\gamma_2} \boldsymbol{\beta}_{\gamma_2} + \rho_{21}(\mathbf{y}_1 - X_{\gamma_1} \boldsymbol{\beta}_{\gamma_1}) + \boldsymbol{\varepsilon}_2 \\ \vdots \\ \mathbf{y}_k = X_{\gamma_k} \boldsymbol{\beta}_{\gamma_k} + \sum_{l < k} \rho_{kl}(\mathbf{y}_l - X_{\gamma_l} \boldsymbol{\beta}_{\gamma_l}) + \boldsymbol{\varepsilon}_k \end{cases}$$

$$\text{with } \mathbb{E}[\boldsymbol{\varepsilon}_k, \boldsymbol{\varepsilon}_l] = \begin{cases} 0 & k \neq l \\ \sigma_k^2 \mathbb{I}_n & k = l \end{cases}$$

So Likelihood separates across separate responses.

Reparametrisation is  $C \longleftrightarrow \{\sigma_k^2, \rho_{kl}\}$

We reformulate the reparametrisation as a factorisation of the Covariance matrix:

$$C_{(j)} = \begin{pmatrix} C_{(j-1)} & \mathbf{c}_j \\ \mathbf{c}_j^t & c_j \end{pmatrix}$$

for  $j = 2, \dots, q$

$$\sigma_j^2 \equiv c_j - \mathbf{c}_j^t C_{(j-1)}^{-1} \mathbf{c}_j$$

$$\boldsymbol{\rho}_j \equiv C_{(j-1)}^{-1} \mathbf{c}_j$$

Complete factorisation of the Covariance matrix (or equivalently the Precision matrix).

We define this for any covariance matrix (dense or sparse).

- Factorise priors across the  $q$  response variables:  $C \sim \mathcal{IW}(\nu, M)$  becomes  $\prod_{j=1}^q \mathcal{N}(\boldsymbol{\rho}_j | \sigma_j^2, M) \times \mathcal{IG}(\sigma_j^2 | \nu, M)$

So posterior conditionals factorise also:

$$\prod_{j=1}^q \mathcal{N}(\boldsymbol{\rho}_j | \sigma_j^2, M, X, Y, B, \Gamma) \times \mathcal{IG}(\sigma_j^2 | \nu, M, X, Y, B, \Gamma)$$

So MCMC updates for  $C$  parameters factorise over responses.

- Regression coefficients prior:  $\prod_{j=1}^q \mathcal{N}(\beta_j | \gamma_j, W)$

MCMC for  $B$  not so straightforward: Zellner and Ando used simplified factorisation + Gibbs resampling

We have calculated correct factorised full conditionals:

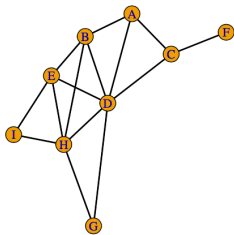
$$\prod_{j=1}^q \beta_{\gamma_j} | (B \setminus \beta_j), W, X, Y, C, \Gamma$$

- Update for  $\gamma_j$  parameters also factorised over response variables (using the ESS (evolutionary stochastic search) algorithm developed by Bottolo et al. )

## Sparse covariance selection

Replace IW prior by Hyper-IW prior conditional on a graph.

Decomposable (chordal or triangulated) graph:  $C \sim \mathcal{HIW}_G(\nu, M)$



HIW factorises over connected components of graph  $G$ :

$$p(C) = p(C_{P_1}) \prod_{k=2}^K p(C_{P_k} \mid C_{S_k})$$

- $S_k$  are separators in the graph
- $P_k$  are cliques

Remaining elements of  $C$  are updated using a “completion operation”.

For decomposable graphs, there is a nice connection between  $\rho_j$  and Precision matrix:

$$\rho_{ji} = 0 \iff (C^{-1})_{ij} = 0$$

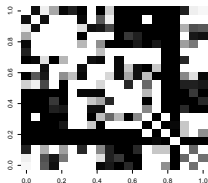
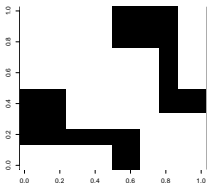
MCMC sampler:

- Update graph structure (single edge or junction tree moves).
- Retain simple Normal and Inverse Wishart priors on  $\rho_j$  and  $\sigma_j^2$ .
- Given graph, only need to update the non-zero  $\rho_{ji}$ .

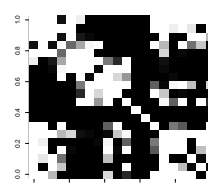
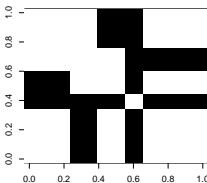
Sparsity leads to another computational gain (only non-zero  $\rho_{ji}$ ).

Bayesian "model averaging" means non-decomposable graphs can be approximated by decomposable graphs.

Decomposable:



Non-decomposable:



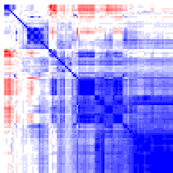


# Case study: mQTL discovery in the North Finland Birth Cohort study (NFBC)

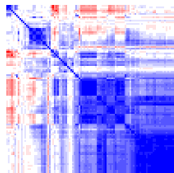
- The NFBC66 is a cohort of 12000 adults followed since 1966
- Question of interest is the **discovery of genetic markers associated with metabolite regulation of lipids**
- These responses are highly structured, with strong correlations
- After quality control,
  - n = 4023 people
  - q = 103 metabolites
  - p = 9172 SNPs on chromosome 16

# Precision Graph mostly sparse!

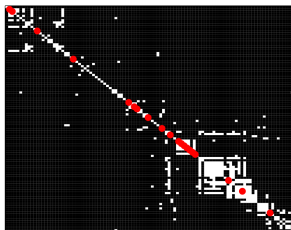
**Data Correlation**



**Residual Correlation**

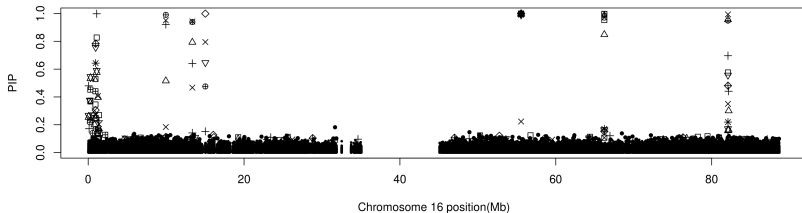


**MAP (adj) Graph**

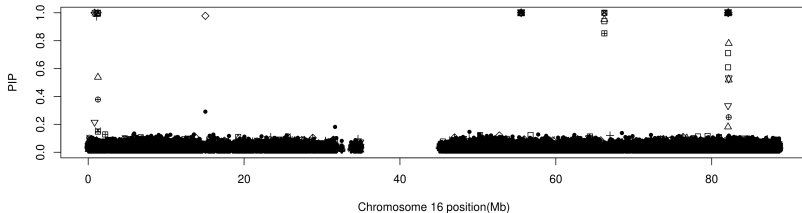


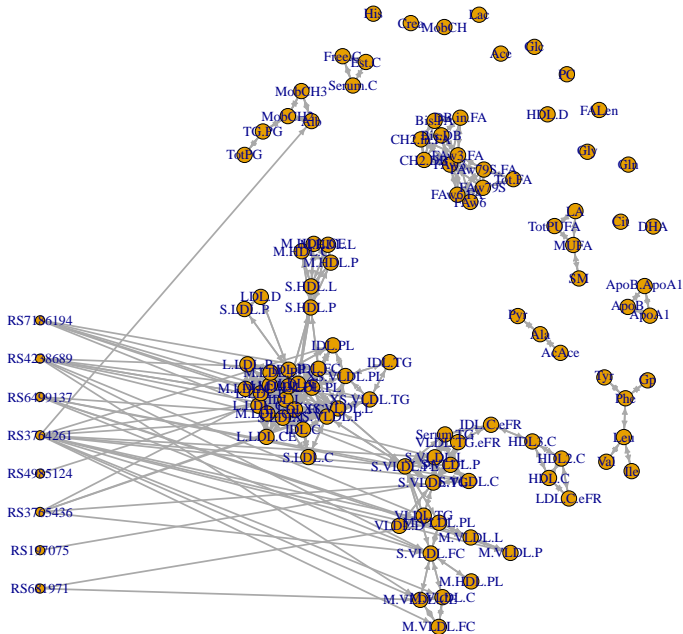
# Evidence of enhanced linkage for Chromosome 16

Manhattan plot – HESS



Manhattan plot – SSUR





# Summary

- Bayesian SUR model with sparsity prior to perform variable selection for multiple responses.
- Estimating the residual covariance matrix increases the accuracy of the variable selection
- Modelling sparsity in the residual covariance matrix aids computations
- Computational speed-up → model can be used on large genomic data sets.

# Thank you!

- Sylvia Richardson
- Leonardo Bottolo
- Marjo-Riitta Jarvelin
- Habib Saadi
- Marc Chadeau-Hyam

*Lewin A et al. (2015)*

MT-HESS: an efficient Bayesian approach for simultaneous association detection in OMICS datasets, with application to eQTL mapping in multiple tissues, **Bioinformatics** **10.1093**

*Bottolo L, Chadeau-Hyam, M et al. (2013)*

Guess-ing polygenic associations with multiple phenotypes using a GPU-based Evolutionary Stochastic Search algorithm, **PLoS Genetics** **10.1371**.

*Bhadra A and Mallick BK (2013)*

Joint high-dimensional Bayesian variable and covariance selection with an application to eQTL analysis, **Biometrics** **10.1111**.

*Bottolo L, Petretto, E et al. (2011)*

Bayesian detection of expression quantitative trait loci hot spots, **Genetics** **189**:1449–1459.