Bayesian variable selection approaches for structured data

Marie Denis

December 16, 2021

Jeudis d'

This project has received funding from The European Unions's Horizon 2020 Research and innovation programme Under grant agreement No 840383.

Outline

Context

Dependence structures between variables may be induced by various factors in different applications:

- Between **observations** in the response variable:
  - Structure in time and/or space: longitudinal data, spatial data, ...
  - Structure induced by grouping factors: pedigree data, ...
  - ...

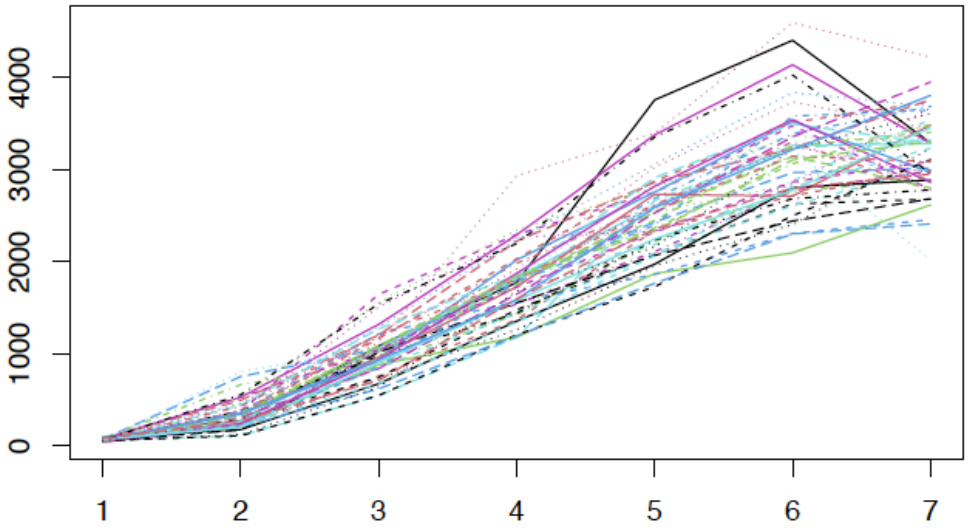


Figure 1: Evolution of the fetal weight at 6 time points representing specific gestational weeks of pregnancy and at birth (NIH)

Context

Dependence structures between variables may be induced by various factors in different applications:

- Between **predictors**:
  - Dependence structure between genes belonging to the same biological pathways or co-expressed,
  - Dependence structure between covariates collected over years,
  - ...

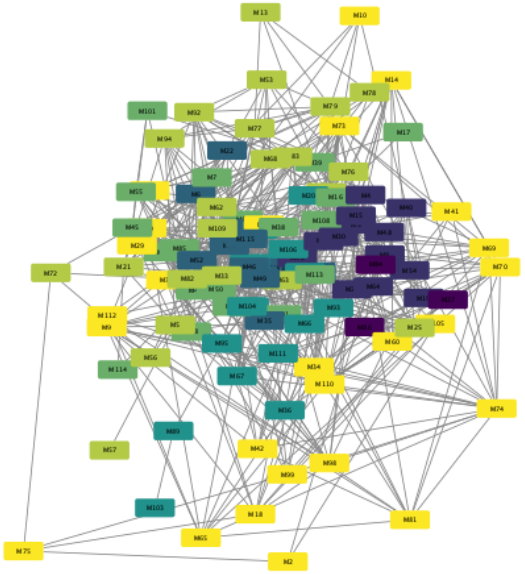


Figure 2: Gene network based on co-expressions.

Context

Dependence structures between variables may be induced by various factors in different applications:

- Dependence structures between **predictors** and **observations in the response variables**:
  - Dependence structure between miRNAs that target mRNAs,
  - Dependence structure between SNPs in genes,
  - ...

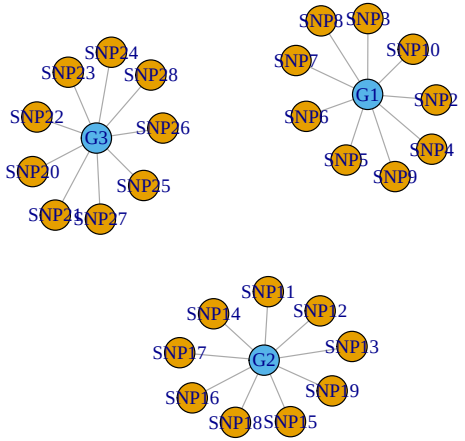


Figure 3: Network for SNPs and genes.

Context

Dependence structures:

- Between **observations in the response variable**,
- Between **predictors**,
- Between **predictors** and **observations in the response variables**,

Such structures, that are **context dependent**, need to be taken into account into statistical models

Context

In many domains high-throughput technologies have generated high-dimensional data:

- "Omic" studies: high-dimensional information at different biological levels,
- Environmental studies: regular and intense monitoring of phenotypic traits over time,
- ..

$\hookrightarrow$ The number of variables p may be greater than the number of observations n

Need to use statistical approaches preventing ill-posed problems (**non-invertible matrix**, **overfitting**) and leading to **parsimonious** models

Why classical approaches do not work?

In the linear model context:

$$Y = X\beta + \varepsilon,\ \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$$

with

- $Y = (y_1, \ldots, y_n)'$ the n-vector of outcomes,,
- X the $n \times p$ matrix of predictors which may be structured and/or of high dimension,
- $\beta = (\beta_1, \ldots, \beta_p)'$ the p-vector of coefficients,
- $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ the n-vector of residuals,
- σ^2 the residual variance.

$\hookrightarrow$ To estimate parameters β and σ^2

Why classical approaches do not work?

Ordinary Least Square (OLS) regression

To minimize the loss function $L^{OLS}(\beta) = \| Y - X\beta \|^2$:

$$\hat{\beta}^{OLS} = (X'X)^{-1}(X'Y)$$

But

- In presence of structures between predictors (as collinearity): $(X'X)^{-1}$ close to singularity and so, $\hat{\beta}^{OLS}$ not accurate
- When the number of predictors is high: $\hat{\beta}^{OLS}$ does not perform well in unseen datasets (overfitting), does not provide parsimonious models (Hadamard, 1902) and in very high dimension $(X'X)^{-1}$ not invertible

↬ Need to use **regularization methods** such as penalized regressions (Ridge (Hoerl and Kennard, 1970), Lasso (Tibshirani, 1996), ...)

Why Bayesian approaches?

Bayesian approach is a natural framework to regularize the model and to integrate prior information:

$\looparrowright$ Regularization and integration of structure dependence are done by specifying **specific priors**

Selection

To shrink towards zero small coefficients while leaving large signals large: **Shrinkage** priors

Structure

Priors with a **variance-covariance matrix** related to structure information between variables

Shrinkage priors

Two classes of shrinkage priors:

Spike-and-slab priors

Discrete mixture of two distributions (Mitchell and Beauchamp, 1988; George and McCulloch, 1997): Introduction of a latent variable γ

$$\gamma_j = \begin{cases} 1 \text{ if variable } j \text{ is selected} \\ 0 \quad \text{otherwise} \end{cases}$$

$$\beta_j|(\gamma_j = 1) \sim p_{Slab}(\beta_j)$$

$$\beta_j|(\gamma_j = 0) \sim p_{Spike}(\beta_j)$$

Continuous shrinkage priors

Unimodal continuous distributions (Bayesian Lasso prior, Horseshoe prior, Elastic-Net prior, ...) (Kyung et al., 2010; Carvalho et al., 2008)

$$\beta_j|\tau^2, \omega_j \sim \mathcal{N}(0, \tau^2\omega_j^2) \;\; j = 1, \ldots, p$$

$$\tau^2, \omega_j^2 \sim \mathcal{F}(\tau^2; \omega_j^2)$$

where $\mathcal{F}$ a distribution to specify $\looparrowright$

Bayesian Lasso prior, **Global-local priors**, ...

Shrinkage priors

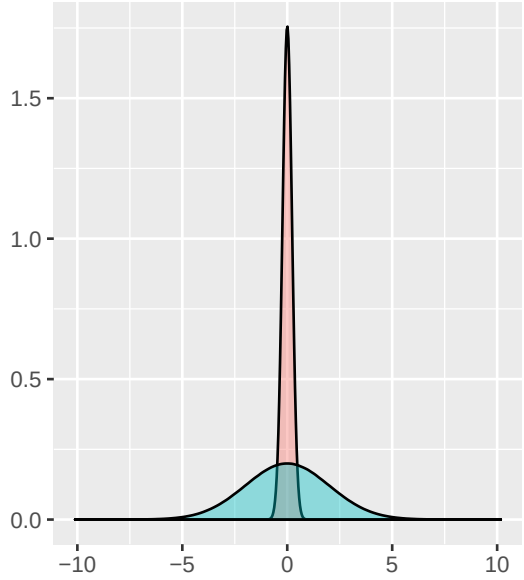


Figure 4: Spike-and-Slab prior distribution. Slab part in blue and spike part in red

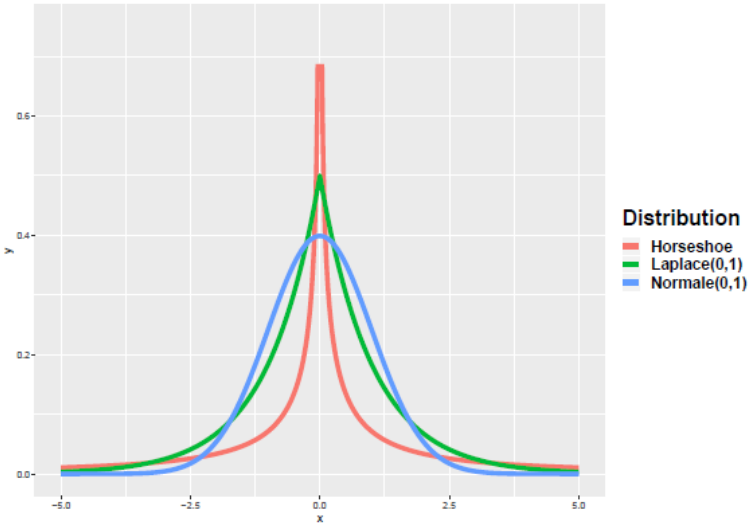


Figure 5: Continuous shrinkage prior distributions

Objective

To present Bayesian variables selection approaches for structured data

We will focus on:

- Two types of structure information: from prior knowledge (biological studies, previous analyses,..) or from sampling design (longitudinal data, spatial data, ...),
- The analysis of univariate or multivariate outcomes related to high dimensional covariate data

Univariate and multivariate linear model context

- Y a n-vector and X a $n \times p$ matrix

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon \text{ with } \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- Y a $n \times q$ matrix and X a $n \times p$ matrix

$$\begin{bmatrix} Y_{11} & & Y_{1q} \\ \vdots & & \vdots \\ Y_{n1} & & Y_{nq} \end{bmatrix} = \begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta_p} \end{bmatrix} + \boldsymbol{\varepsilon} \text{ with } \varepsilon_i \sim \mathcal{N}_q(0, \Omega)$$

with $\beta = (\beta_1, \ldots, \beta_p)'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_p')'$ the regression coefficients, σ^2 the variance parameter, and Ω the variance-covariance matrix .

Outline

Outline

Motivation: context 1

- In many domains the objective is to select a subset of predictors involved in the variability of an outcome

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon$$

↬ Variable selection approaches: penalized likelihood approaches (Lasso, Elastic-Net,...), Bayesian variable selection approaches (Shrinkage priors, ...), ...

- In many domains biological studies and previous analyses have accumulated knowledge on relationships **within** data

↬ **The objective** is to select a subset of structured predictors involved in the variability of an outcome

Motivation: context 1

- In many domains the objective is to select a subset of predictors involved in the variability of an outcome

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon$$

↬ Variable selection approaches: penalized likelihood approaches (Lasso, Elastic-Net,...), Bayesian variable selection approaches (Shrinkage priors, ...), ...

- In many domains biological studies and previous analyses have accumulated knowledge on relationships **within** data

↬ **The objective** is to select a subset of structured predictors involved in the variability of an outcome

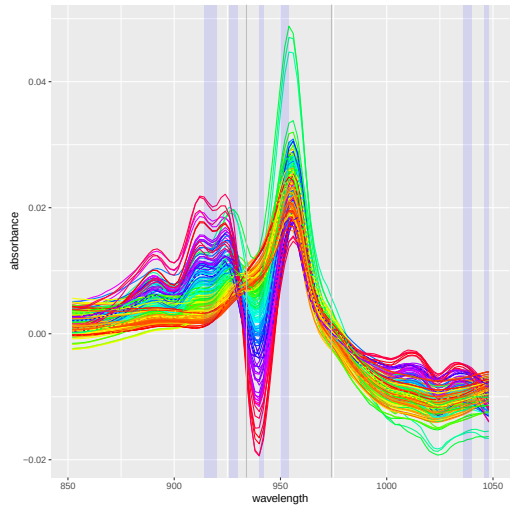
Example

In many domains the objective is to select a subset of structured predictors involved in the variability of an outcome:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon$$

↬ Spectrometric data:

$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}}_{\text{Fat content}} \sim \underbrace{\begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix}}_{\text{Spectra sampled at p wavelengths}}$$



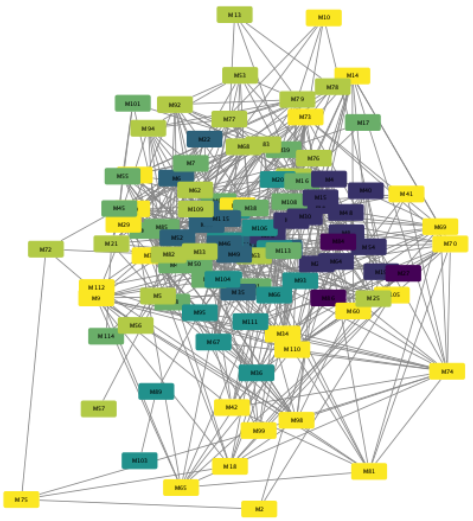
Example

In many domains the objective is to select a subset of structured predictors involved in the variability of an outcome:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon$$

↬ Genomic data:

$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}}_{\text{Production}} \sim \underbrace{\begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix}}_{\text{Expression values for p genes}}$$



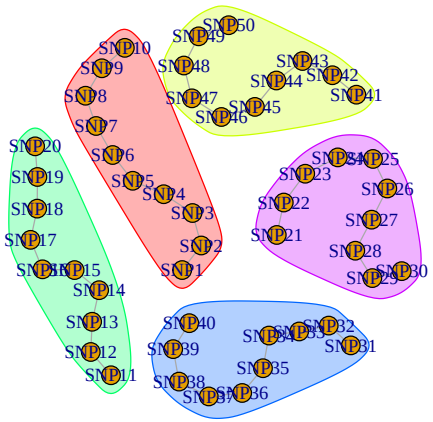
Example

In many domains the objective is to select a subset of structured predictors involved in the variability of an outcome:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon$$

↬ <u>Genetic data</u>:

$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}}_{\text{Production}} \sim \underbrace{\begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix}}_{\text{Genotypes at p SNPs}}$$



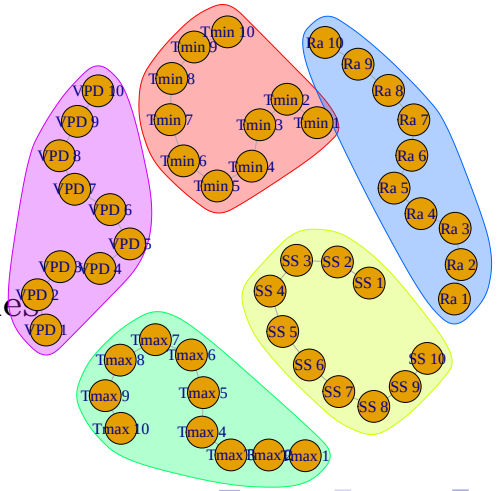
Example

In many domains the objective is to select a subset of structured predictors involved in the variability of an outcome:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon$$

↬ <u>Environmental data</u>:

$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}}_{\text{Natural Abscission}} \sim \underbrace{\begin{bmatrix} X_{11} & & X_{1(Txl)} \\ \vdots & & \vdots \\ X_{n1} & & X_{n(Txl)} \end{bmatrix}}_{l \text{ environmental variables over T times}}$$



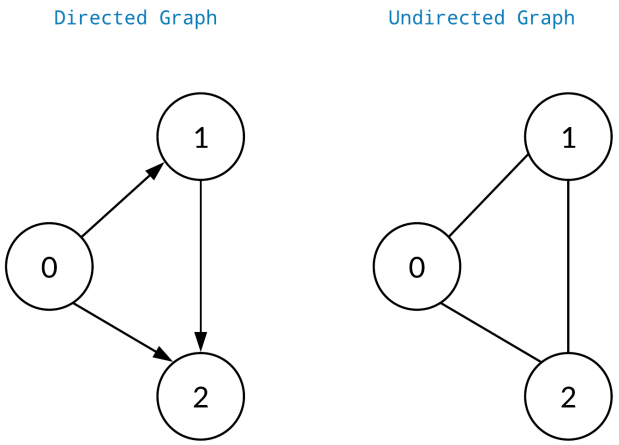
Motivations

The dependence structure described previously may be encoded by an undirected graph:

Undirected graph

A **graph** is a pair $\mathcal{G} = (V, E)$ where $V = \{1, \ldots, p\}$ is a finite set of vertices (nodes), and the set of edges E is a subset of the set $V \times V$. Two types of commonly used graphs:

- **directed graph** where edges are denoted by ordered pairs $(i, j) \in E$
- **undirected graph** where edges are denoted by unordered pairs $(i, j) \in E \iff (j, i) \in E$.



Proposed model

We propose to extend the approach proposed by Faulkner and Minin (2018); Faulkner (2019) to the more general context of graph-structured variables by combining:

1. The **efficiency and flexibility** of a shrinkage prior, the horseshoe (HS) prior, in terms of selection and estimation,
2. With a Gaussian Markov random field (GMRF) for its **appealing connection** with undirected graphs (Rue and Held, 2005):

HS-GMRF model (Denis and Tadesse, 2021, submitted to Annals of Applied Statistics)

HS-GMRF model

We assume that $\mathcal{G} = \bigcup_{i=1}^{I} \mathcal{G}_i = \bigcup_{i=1}^{I} (V_i, E_i)$ a disjoint union of I subgraphs and $\mathcal{S}$ the set of indices associated to one representative of each of the I subgraphs.

HS-GMRF model

$$
\begin{aligned}
\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2 &\sim \mathcal{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \mathrm{I}_n) \\
\boldsymbol{\beta}|\boldsymbol{\tau}^2, \lambda^2 &\sim \mathcal{N}_p(0, \lambda^2 \boldsymbol{Q}^{-1}) \\
\tau_{jj'} &\sim \mathcal{C}^+(0,1) \text{ for } (j,j') \in \bigcup_{i=1}^{I} E_i; \tau_j \sim \mathcal{C}^+(0,1) \text{ for } j \in \mathcal{S} \\
\lambda|\sigma &\sim \mathcal{C}^+(0,\sigma);\ \sigma^2 \sim \mathcal{IG}(a_0, b_0)
\end{aligned}
$$

with $\boldsymbol{Q}$ the precision matrix related to the graph structure, λ the global shrinkage hyperparameter, and $\boldsymbol{\tau}$ the local shrinkage hyperparameters

Inference done with a Gibbs sampling.

Application

Objective

To identify gene expressions involved in the variability of riboflavin production using data on 71 samples

- A total of 142 gene expressions considered
- Estimation of an undirected graph with 157 edges
- 5-fold cross-validation procedure

Methods	CV-MSPE	Selected genes
HS-GMRF	**0.29**	4 (90% HPD) 8 (80% HPD)
HS-GMRF-nosign	0.31	4(90% HPD) 6 (80% HPD)
HS	0.33	0(90% HPD) 0 (80% HPD)
SS-Ising	0.37	21 ($PPI > 0.5$) 16 ($PPI > 0.8$)
Lasso	0.41	16

↬ HS-GMRF yields the smallest CV-MSPE

Application

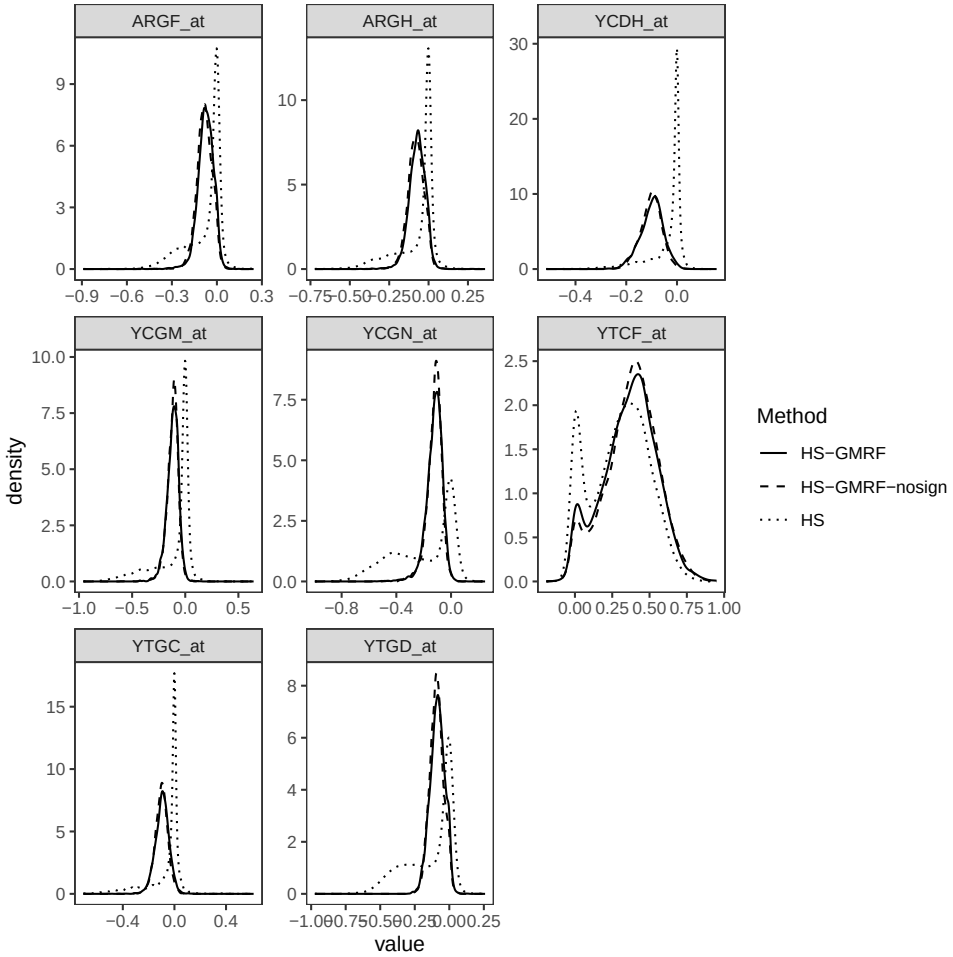


Figure 6: Posterior densities for the selected genes

For moderate non-zero effects:

- HS estimates densities concentrated around 0 with long tails or bimodal densities with one of the modes around 0,
- HS-GMRF-based methods estimate unimodal densities or bimodal densities with the mode around 0 less than with HS.

HS-GMRF-based approaches select groups of genes involved in the same biological pathway.

Application

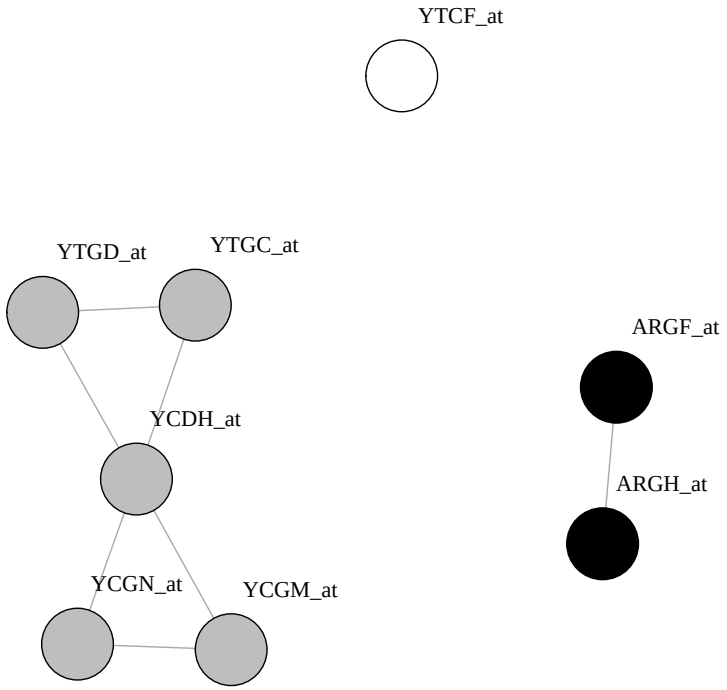


Figure 7: The estimated network

To sum up

From a statistical point of view:

- To increase power to detect associations
- To improve the predictive power
- To circumvent the problem of high collinearity

From a biological point of view:

- To encourage the identification of groups of dependent variables acting jointly on the response, especially those with subtle effects

Outline

Motivation: context 2

- In many domains the objective is to select subsets of predictors involved in the variability of subsets of outcomes

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{n1} \end{bmatrix} \begin{bmatrix} Y_{1q} \\ \vdots \\ Y_{nq} \end{bmatrix} = \begin{bmatrix} X_{11} & & & X_{1p} \\ \vdots & & & \vdots \\ X_{n1} & & & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon$$

↬ Reduction dimension approaches (Canonical Correlation Analysis (CCA), Partial Least Square regression (PLS), ...), Bayesian approaches (Stochastic partitioning method (Monni and Tadesse, 2009), Multivariate Spike-and-Slab LASSO (Deshpande et al., 2019), ...)

- In many domains biological studies have accumulated knowledge on relationships **between** different types of data (SNPs-genes, mRNAs-microRNAs (miRNAs), ...)

↬ **The objective** is to select subsets of predictors involved in the variability of subsets of outcomes while integrating prior knowledge on their relationships and to estimate Ω

Motivation: context 2

- In many domains the objective is to select subsets of predictors involved in the variability of subsets of outcomes

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{n1} \end{bmatrix} \begin{bmatrix} Y_{1q} \\ \vdots \\ Y_{nq} \end{bmatrix} = \begin{bmatrix} X_{11} & & & X_{1p} \\ \vdots & & & \vdots \\ X_{n1} & & & X_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon$$

↬ Reduction dimension approaches (Canonical Correlation Analysis (CCA), Partial Least Square regression (PLS), ...), Bayesian approaches (Stochastic partitioning method (Monni and Tadesse, 2009), Multivariate Spike-and-Slab LASSO (Deshpande et al., 2019), ...)

- In many domains biological studies have accumulated knowledge on relationships **between** different types of data (SNP-gene, mRNA-miRNA, ...)

↬ **The objective** is to select subsets of predictors involved in the variability of subsets of outcomes while integrating prior knowledge on their relationships and to estimate Ω

Motivation

Liver cancer (hepatocellular carcinoma, HCC)

miRNAs are important regulators of mRNAs in HCC Varghese et al. (2020)

- To select mRNA-miRNA pairs that are biological relevance to HCC by using prior knowledge
- To estimate a graph for mRNAs network while adjusting for miRNAs

Proposed approach

Two-step procedure combining

1. A spike-and-slab prior on the regression coefficients integrating prior knowledge at the indicator variable level (SS-int) Stingo et al. (2011)
2. With a Gaussian graphical model (Dempster, 1972)

Covariate-adjusted Gaussian graphical model (Conditional graphical model)

Proposed approach

Integration of prior information

1. Definition of scores $\boldsymbol{s}$ from biological studies relating the belief in the association between the predictors and the response
2. Modelisation of the prior inclusion probability of the variable j as follows:

$$p(\delta_j = 1|\tau) = \frac{\exp(\eta + \tau s_j)}{1 + \exp(\eta + \tau s_j)}$$

with s_j the score associated to the variable j, τ to estimate, and η fixed. Inference done with a Metropolis-within-Gibbs algorithm.

Estimation of the graphical structure accounting for covariates

- Use of graphical lasso on residuals

Simulation study

Simulated data

- Number of individuals n=250/ Number of variables p=n/2, n, 5n
- Only four variables having an effect X_1, X_2, X_3, X_4
- Four scores tested: $s_{null} = (0, 0, 0, 0)$, $s_1 = (0, 0, 1, 1)$, $s_2 = (1, 0.4, 1, 0.4)$, $s_{full} = (1, 1, 1, 1)$

Results ($p = n/2$)

- Mean of posterior inclusion probabilities over 10 replications for the four scores

Scores	X_1 (weak)	X_2 (large)	X_{90} (weak)	X_{100} (large)
s_{full}	0.75	0.99	0.79	0.99
s_1	**0.42**	1.00	0.78	1.00
s_2	0.84	0.99	0.87	0.99
s_{null}	**0.42**	0.99	**0.33**	0.99

↬ Integration of prior knowledge helps identify relevant variables with subtle effects

Application in liver cancer

Data provided by Ressom/Omics Lab (Chen et al., 2020)

- 64 patients (25 controls and 39 cases) patients recruited at MedStar Georgetown University Hospital
- 90 mRNAs and 193 miRNAs selected from mRNA-seq and miRNA-seq data
- **Scores** defined with IPA software (QIAGEN Inc.) with different levels of confidence (Experimentally observed or predicted)

Results

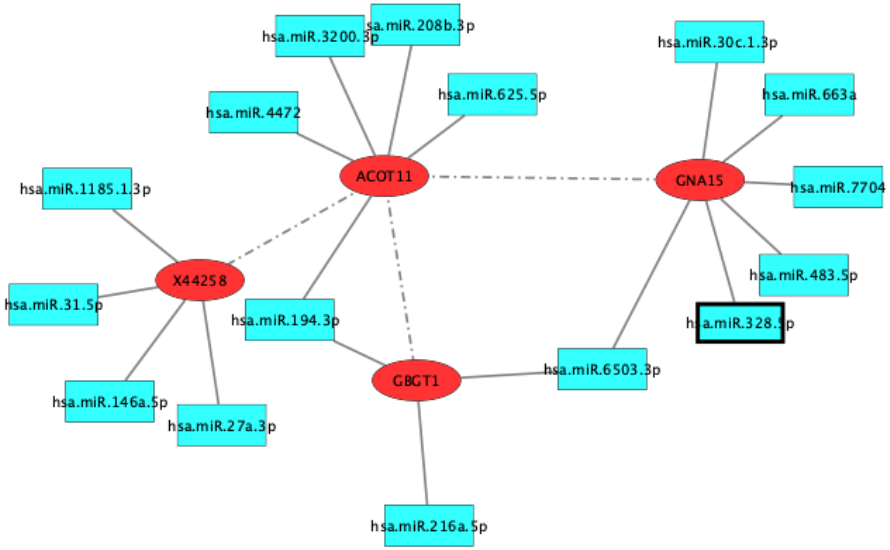
Among the identified mRNA-miRNAs pairs:

- 15% exact matched pairs recovered with SS-int/2% with EN in control group
- 7% exact matched pairs recovered with SS-int/ 4% with EN in case group

Integration of biological information helps recover verified pairs involved in the pathogenesis of HCC but biological information may be more or less relevant

Results

Estimated undirected graph for a subset of features in case group:



- Relations between the four genes disappear when adjusting for miRNAs $\rightsquigarrow$ mRNAs coexpressed via miRNAs
- Different results in control group: regulation heterogeneity between the two groups

To sum up

From a statistical point of view

- Integration of prior knowledge helps identify variables with subtle effects
- Use of weighted scores is important to balance the biological information
- Conditional graphical models help disentangle relations between variables

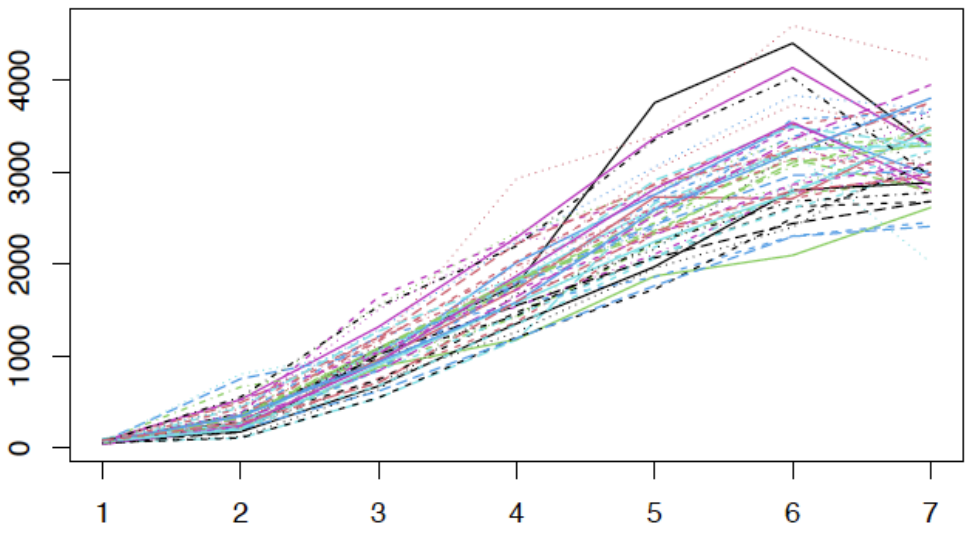
From a biological point of view

- Finer understanding of the biological mechanisms

Outline

Context: longitudinal data

For example the evolution of the fetal weight at 6 time points representing specific gestational weeks of pregnancy and at birth (data provided by NIH)

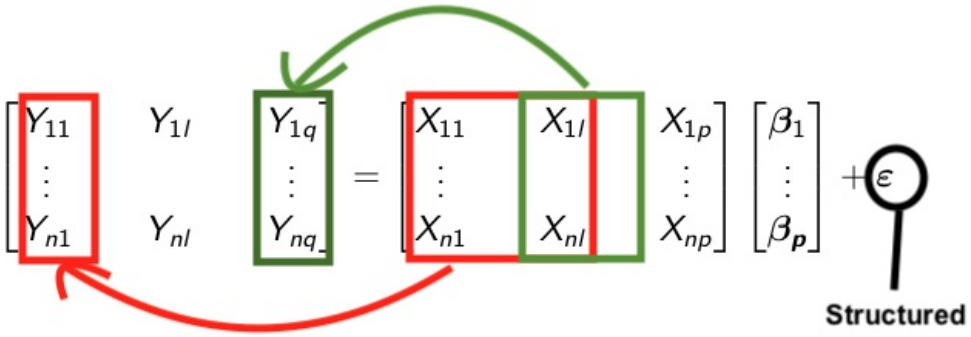


Two objectives:

1. To study the dynamic genetic architecture across the pregnancy (functional mapping),
2. To identify cluster specific covariates (Genetic variables, Biomarkers, Socio-demographic variables,...)

Dynamic genetic architecture

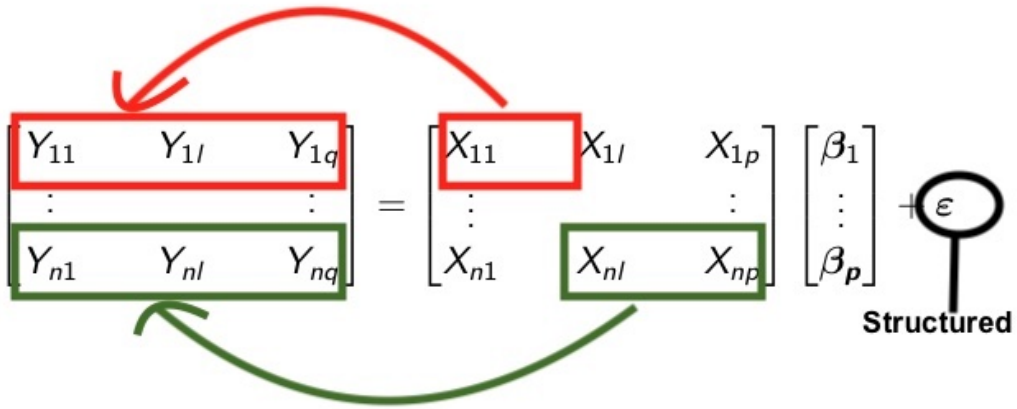
Bayesian variable selection approach for selecting subsets of variables involved in the variability of outcomes for specific time periods (Heuclin et al., 2021) (R package VCGSS)



- To identify molecular markers which control the outcome over time
- To estimate the dynamic effect of the selected markers over time

Cluster specific covariates

Bayesian variable selection approach for selecting subsets of variables involved in the variability of clustered outcome profiles



Existing approaches propose mixture regression models for longitudinal data to partition a collection of individuals into homogeneous subsets but do not select (De la Cruz-Mesía et al., 2008; Xu et al., 2018)

- ↬ Inclusion of all predictors may be detrimental to recover the true cluster structure (Tadesse et al., 2005): Need to select the relevant predictors providing information about the group structure of the observations

Proposed approach

Stochastic partitioning approach

Extension of the stochastic partitioning approach developed by Monni and Tadesse (2009) to cluster longitudinal data and to select variables that discriminate them

The objective is to partition the data into an unknown number K of components (X_I, Y_J) with $J \subset \{1, \ldots, p\}$ and $I \subset \{1, \ldots, n\}$ where

- Individuals are assigned to one and only one component
- Predictors may belong to several components

$\hookrightarrow$ Association of subsets of X to clusters of Y profiles

Work in progress

Preliminary results

- Simulated data: n=50, p=100, T=10 with K=2 components
- Two components recovered along with variables related to each component

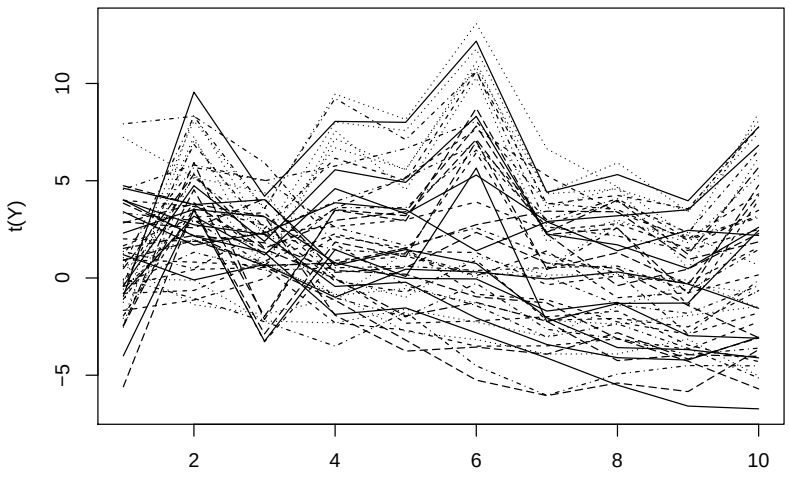


Figure 8: Outcome profiles

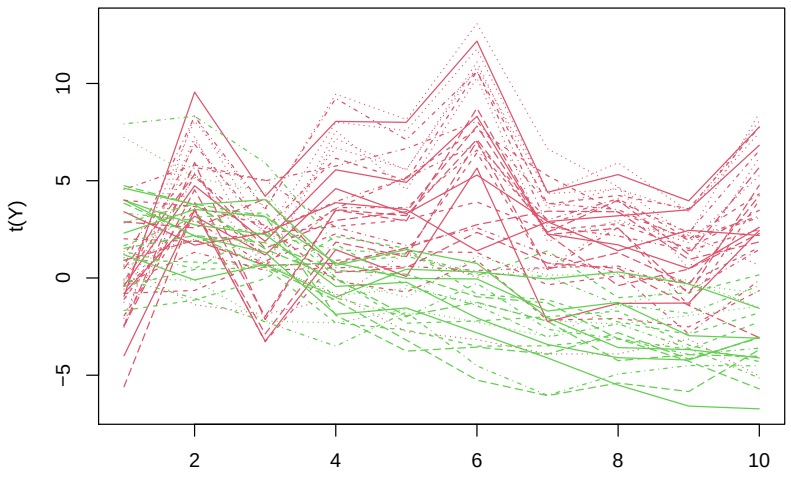


Figure 9: Clustered outcome profiles

Work in progress

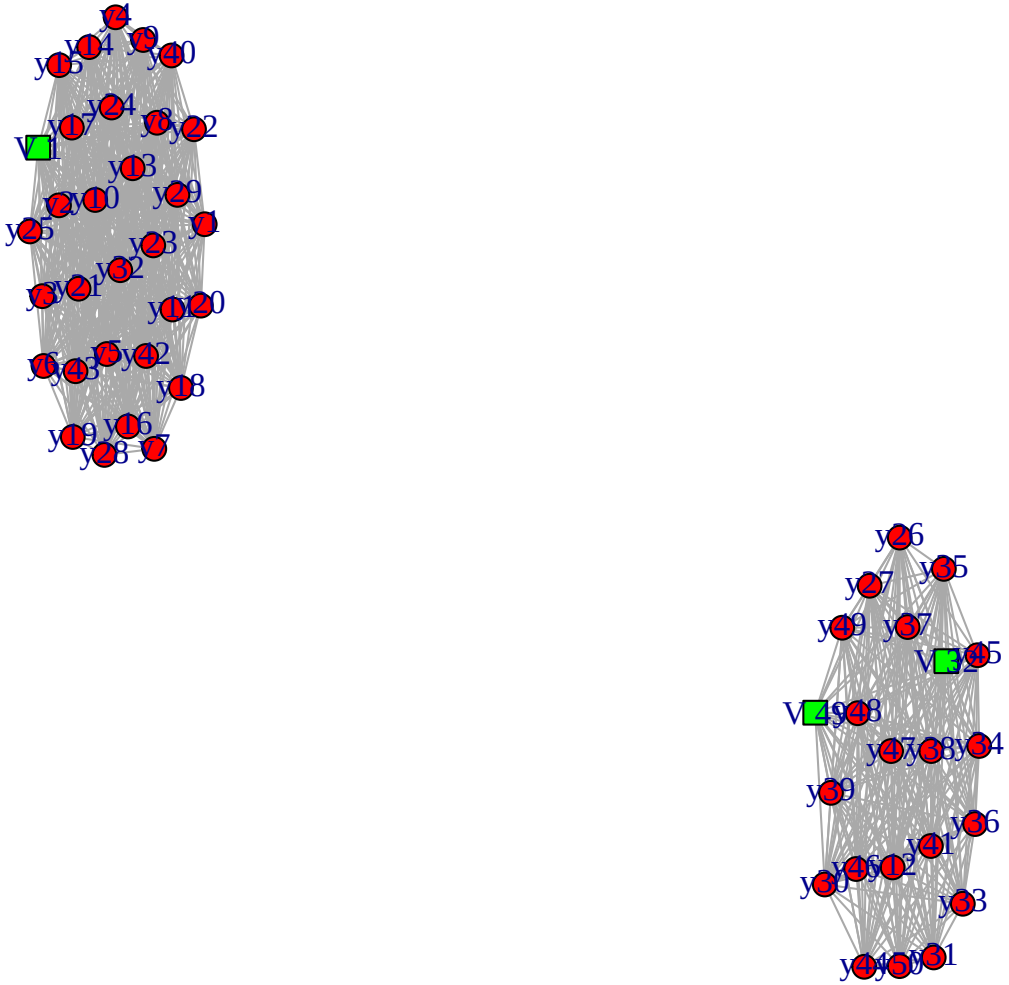


Figure 10: Network for mRNAs and miRNAs

Outline

Conclusion

We proposed **Bayesian variables selection approaches** to analyze univariate or multivariate outcomes while accounting for the **dependence structure** between variables

From a statistical point of view

Such approaches help:

- The model building process by reducing model complexity and by circumventing high collinearity problem through identifiability constraints,
- Identify the relevant variables even those with subtle effects
- Improve the predictive power

Conclusion

From a biological point of view

Such approaches help:

- **Gain insight** into biological mechanisms between different biological levels by integrating prior knowledge or by discovering new ones
- Encompass a broad type of dependence structures: **applicable in plant context**
  - HS-GMRF approach: to explain natural abscission according to structured environmental data (Oil palm/...) / to predict yield according to gene network (Oil palm/ Hevea/ Rice/ ...)
  - SS-int/ Conditional graphical model/Multivariate SS Lasso: to estimate gene networks while accounting for genetic data (Oil palm/Hevea/Eucalyptus/...)
  - Two last approaches: to select genetic markers involved in the variation of longitudinal outcomes and to estimate their functional effects/ to cluster individuals wrt their profiles over time and to select genetic markers (Oil palm/Hevea/Eucalyptus/...)
  - ...

Conclusion

Perspectives

- To extend the HS-GMRF prior by integrating prior knowledge on strengths of connections
- To simultaneously select predictors, estimate covariance and integrate prior knowledge (Co-supervision of a post-doctoral researcher (Zhen Liu) at Georgetown University with Prof. Mahlet Tadesse)

Also

- Estimation of networks based on different correlations (Pearson correlation, partial correlation) or different approaches (Graphical Lasso, Graphical horseshoe,...)
- Differential network analysis between two or more conditions: to compare networks for plants under different treatments
- Estimation of dynamic networks (ex: Fused graphical Lasso)

Thanks for your attention !

Bénédicte Favreau
Jean-Marc Gion
Benjamin Heuclin
Frédéric Mortier
Sébastien Tisné

Department of mathematics and statistics
Mahlet G. Tadesse
Zhen Liu

Ressom/Omics Lab
Habtom Ressom
Rency Varghese
Megan Barefoot

NICHD
Fasil Ayele

Outline

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2008). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Chen, Y., Barefoot, M. E., Varghese, R. S., Wang, K., Di Poto, C., and Ressom, H. W. (2020). Integrative analysis to identify race-associated metabolite biomarkers for hepatocellular carcinoma. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 5300–5303. IEEE.

De la Cruz-Mesía, R., Quintana, F. A., and Marshall, G. (2008). Model-based clustering for longitudinal data. *Computational Statistics & Data Analysis*, 52(3):1441–1457.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, pages 157–175.

Deshpande, S. K., Ročková, V., and George, E. I. (2019). Simultaneous variable and covariance selection with the multivariate spike-and-slab lasso. *Journal of Computational and Graphical Statistics*, 28(4):921–931.

Faulkner, J. R. (2019). *Adaptive Bayesian Nonparametric Smoothing with Markov Random Fields and Shrinkage Priors*. PhD thesis.

Faulkner, J. R. and Minin, V. N. (2018). Locally adaptive smoothing with markov random fields and shrinkage priors. *Bayesian analysis*, 13(1):225.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica sinica*, pages 339–373.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55.

Kyung, M., Gill, J., Ghosh, M., Casella, G., et al. (2010). Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis*, 5(2):369–411.

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the american statistical association*, 83(404):1023–1032.

Monni, S. and Tadesse, M. G. (2009). A stochastic partitioning method to associate high-dimensional responses and covariates. *Bayesian Analysis*, 4(3):413–436.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

Stingo, F. C., Chen, Y. A., Tadesse, M. G., and Vannucci, M. (2011). Incorporating biological information into linear models: A Bayesian approach to the selection of pathways and genes. *The annals of applied statistics*, 5(3).

Tadesse, M. G., Sha, N., and Vannucci, M. (2005). Bayesian variable selection in clustering high-dimensional data. *Journal of the American Statistical Association*, 100(470):602–617.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Varghese, R. S., Zhou, Y., Barefoot, M., Chen, Y., Di Poto, C., Balla, A. K., Oliver, E., Sherif, Z. A., Kumar, D., Kroemer, A. H., et al. (2020). Identification of mirna-mrna associations in hepatocellular carcinoma using hierarchical integrative model. *BMC medical genomics*, 13(1):1–14.

Xu, P., Peng, H., and Huang, T. (2018). Unsupervised learning of mixture regression models for longitudinal data. *Computational Statistics & Data Analysis*, 125:44–56.