Statistical Inference with Multi-layered Graphical Models

Subhabrata Majumdar

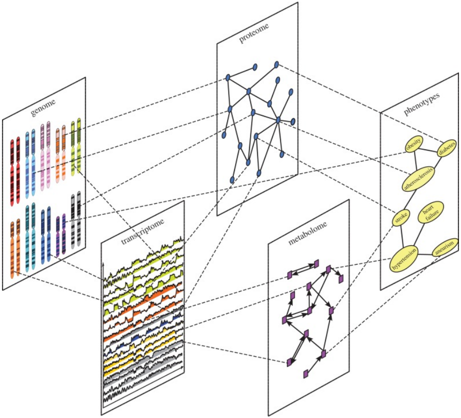
University of Florida Informatics Institute

AT&T Labs - Research, Bedminster, NJ

April 2, 2018

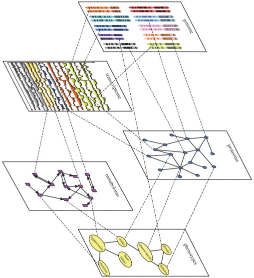
Summary

- Biological processes in the body have a natural hierarchical structure, e.g. **Gene > Protein > Metabolite**;
- There are within layer and between-layer connections in this structure.

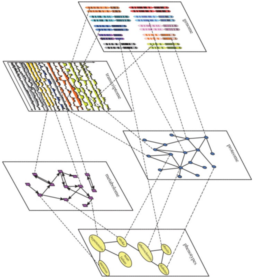


Source: Gligorijević and Pržulj (2015)

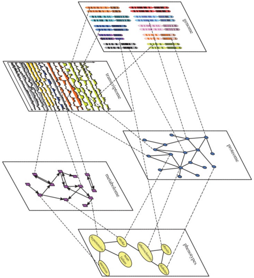
These connections can be different inside different organs, experimental conditions, or for different subtypes of the same disease;



Liver



Kidney



Lungs

A general model for data integreation.

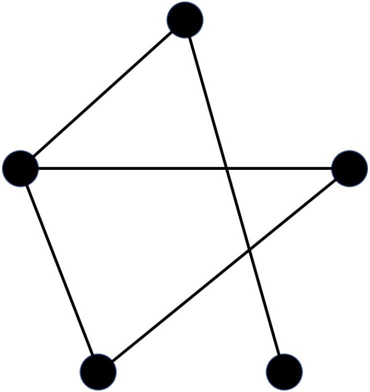
In this work we propose a general statistical framework based on graphical models for *horizontal* (i.e. across conditions or subtypes) *and vertical* (i.e. across different layers containing data on molecular compartments) *integration of information* in data from such complex biological structures.

Specifically, we perform *joint estimation and hypothesis testing* for all the connections in these structures.

Outline

Gaussian Graphical models

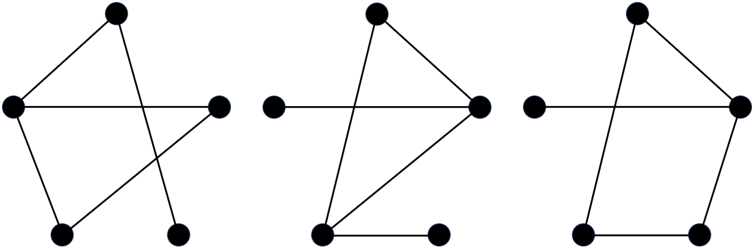
$$\mathbb{X} = (X_1, \dots, X_p)^T \sim \mathcal{N}_p(0, \Sigma_x); \quad \Omega_x = \Sigma_x^{-1}$$



Sparse estimation of Ω_x: Meinshausen and Bühlmann (2006)
Multiple testing and error control: Drton and Perlman (2007).

Multiple Gaussian Graphical models

$$\mathbb{X}^k = (X_1^k, \ldots, X_p^k)^T \sim \mathcal{N}_p(0, \Sigma_x^k); \quad \Omega_x^k = (\Sigma_x^k)^{-1}$$
$$k = 1, 2, \ldots, K$$



$k = 1$ $k = 2$ $k = 3$

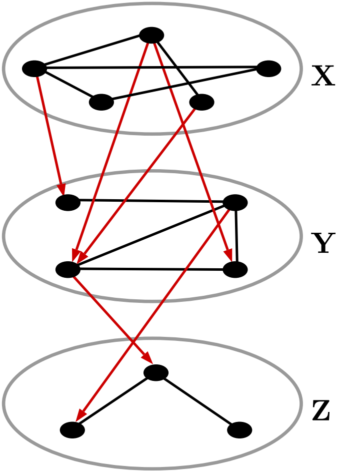
- Joint estimation of $\{\Omega_x^k\}$: Guo et al. (2011); Ma and Michailidis (2016)
- Difference and similarity testing with FDR control: Liu (2017+)

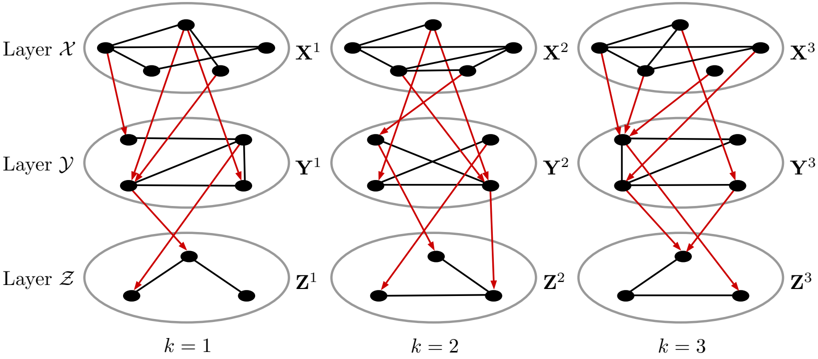
Multi-Layered Gaussian Graphical models

$$\mathbb{E} = (E_1, \ldots, E_q)^T \sim \mathcal{N}_p(0, \Sigma_y);$$
$$\mathbb{F} = (F_1, \ldots, F_r)^T \sim \mathcal{N}_r(0, \Sigma_z);$$
$$\Omega_y = (\Sigma_y)^{-1}, \Omega_z = (\Sigma_z)^{-1}$$
$$\mathbb{Y} = \mathbb{X}\mathbf{B} + \mathbb{E},$$
$$\mathbb{Z} = \mathbb{Y}\mathbf{C} + \mathbb{F}.$$

- $\Omega_x, \Omega_y, \Omega_z$ give undirected within-layer edges, while $\mathbf{B}, \mathbf{C}$ gives directed between-layer edges.

- Sparse estimation of the components: Lin et al. (2016).
- Testing: **??**





$$\mathbb{E}^k = (E_1^k, \dots, E_q^k)^T \sim \mathcal{N}_q(0, \Sigma_y^k); \quad \Omega_y^k = (\Sigma_y^k)^{-1}$$

$$\mathbb{F}^k = (F_1^k, \dots, F_r^k)^T \sim \mathcal{N}_r(0, \Sigma_z^k); \quad \Omega_y^k = (\Sigma_y^k)^{-1}$$

$$\mathbb{Y}^k = \mathbb{X}^k \mathbf{B}^k + \mathbb{E}^k,$$

$$\mathbb{Z}^k = \mathbb{Y}^k \mathbf{C}^k + \mathbb{F}^k; \quad k = 1, 2, \dots, K$$

What we do

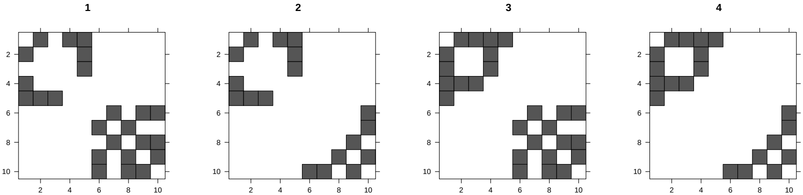
- We decompose the multi-layer problem into a series of two layer problems.
- We estimate $\{\Omega_x^k, \Omega_y^k, \mathbf{B}^k\}$ jointly for all k from a single model;
  - Incorporate structural informartion using group sparsity,
  - Propose alternating block algorithm to compute solutions,
  - Derive convergence properties of solutions.
- Devise a full pairwise testing procedure for rows of $\mathbf{B}^k$;
  - Propose a debiasing method for estimates of individual rows of $\mathbf{B}^k$, say $\mathbf{b}_i^k, i \in \{1, 2, \ldots, p\}$ and derive its asymptotic distribution;
  - For $K = 2$, propose a test for row-wise differences $\mathbf{b}_i^1 - \mathbf{b}_i^2$;
  - Perform multiple testing for elementwise differences $b_{ij}^1 = b_{ij}^2, j = 1, 2, \ldots, q$ within a row.
- Use simulations for performance evaluation.

Outline

Preliminaries

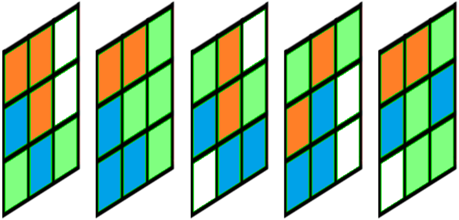
- $\mathcal{Y} = \{\mathbf{Y}^1, \ldots, \mathbf{Y}^K\}, \mathcal{X} = \{\mathbf{X}^1, \ldots, \mathbf{X}^K\}$;
- $\Omega_x = \{\Omega_x^1, \ldots, \Omega_x^K\}, \Omega_y = \{\Omega_y^1, \ldots, \Omega_y^K\}, \mathcal{B} = \{\mathbf{B}^1, \ldots, \mathbf{B}^K\}$;

Group structures in X-network is denoted by $\mathcal{G}_x = \{\mathcal{G}_{x,ii'} : i \neq i', 1 \leq i, i' \leq p\}$. Each $\mathcal{G}_{x,ii'}$ is a partition of $\{1, \ldots, K\}$ denoting grouping over k for the $(i, i')^{\text{th}}$ elements of the X-precision matrices.



Define $\mathcal{G}_y = \{\mathcal{G}_{y,jj'} : j \neq j', 1 \leq j, j' \leq q\}$ similarly.

Group structures in $\mathcal{B}$ is denoted by $\mathcal{H}$, with each $h \in \mathcal{H}$ being a collection of 3-tuples (h_i, h_j, h_k) so that $1 \leq h_i \leq p, 1 \leq h_j \leq q, 1 \leq h_k \leq K$.



1. Estimate neighborhood coefficients of each X-node, say $\zeta_i = (\zeta_i^1, \ldots, \zeta_i^K)$ using the group sparsity information $\mathcal{G}_x$:

$$\widehat{\zeta}_i = \underset{\zeta_i}{\operatorname{argmin}} \left\{ \sum_{k=1}^{K} \frac{1}{n_k} \|\mathbf{X}_i^k - \mathbf{X}_{-i}^k \zeta_i^k\|^2 + \nu_n \sum_{i' \neq i, g \in \mathcal{G}_{x,ii'}} \|\zeta_{ii'}^{[g]}\| \right\}$$

2. Estimate precision matrices as MLE over the restricted supports:

$$\widehat{E}_x^k = \{(i, i') : 1 \leq i < i' \leq p, \widehat{\zeta}_{ii'}^k \neq 0 \text{ OR } \widehat{\zeta}_{i'i}^k \neq 0\},$$
$$\widehat{\Omega}_x^k = \underset{\Omega_x^k \in \mathbb{S}_+(\widehat{E}_x^k)}{\operatorname{argmin}} \left\{ \operatorname{Tr}(\widehat{\mathbf{S}}_x^k \Omega_x^k) - \log\det(\Omega_x^k) \right\}.$$

Joint Structural Estimation Method (JSEM)
Ma and Michailidis (2016)

Estimation strategy of $\mathcal{B}, \Omega_y$

1. Estimate regression coefficient matrices using group information $\mathcal{H}$ and neighborhood coefficients of each Y-node, say $\Theta_j = (\boldsymbol{\theta}_j^1, \ldots, \boldsymbol{\theta}_j^K)$,
2. Estimate neighborhood coefficients of each Y-node, i.e. $\widehat{\Theta}_j$ using the group sparsity information $\mathcal{G}_y$ and regression matrices $\mathbf{B}^k$,
3. Estimate Y-layer precision matrices as MLE over the restricted supports.

1. Estimate regression coefficient matrices using group information $\mathcal{H}$ and neighborhood coefficients of each Y-node, say $\Theta_j = (\boldsymbol{\theta}_j^1, \ldots, \boldsymbol{\theta}_j^K)$,
2. Estimate neighborhood coefficients of each Y-node, i.e. $\widehat{\Theta}_j$ using the group sparsity information $\mathcal{G}_y$ and regression coefficient matrices $\mathbf{B}^k$,
3. Estimate Y-layer precision matrices as MLE over the restricted supports.

Solution: Iterate!

A note on testing

Our testing methodology only requires *generic* estimates that satisfy the following convergence rates:

$$\|\widehat{\zeta}^k - \zeta_0^k\|_1 \leq O\left(\sqrt{\frac{\log p}{n}}\right)$$

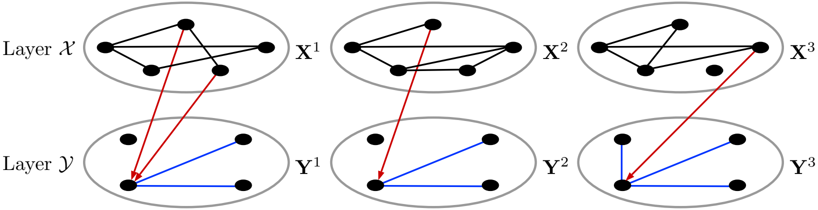
$$\|\widehat{\Omega}_y^k - \Omega_{y0}^k\|_\infty \leq O\left(\sqrt{\frac{\log(pq)}{n}}\right)$$

$$\|\widehat{\mathbf{B}}^k - \mathbf{B}_0^k\|_1 \leq O\left(\sqrt{\frac{\log(pq)}{n}}\right)$$

The estimation procedure gives one suitable set of estimates.

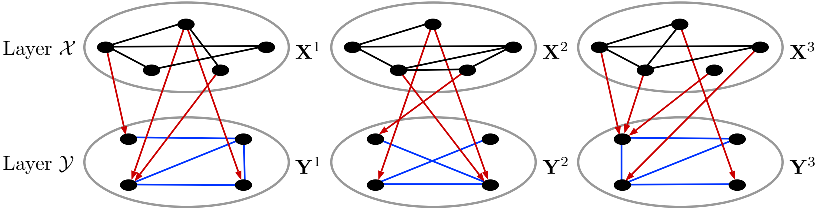
Outline

The objective function



$$\mathbf{Y}_j^k \quad - \quad \mathbf{E}_{-j}^k \boldsymbol{\theta}_j^k \quad - \quad \mathbf{X}^k \mathbf{B}_j^k$$

The objective function



$$\sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{q} \left\| \mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \mathbf{B}_{-j}^k) \boldsymbol{\theta}_j^k - \mathbf{X}^k \mathbf{B}_j^k \right\|^2$$

$$+ \lambda_n \sum_{h \in \mathcal{H}} \| \mathbf{B}^{[h]} \| + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{jj'}} \| \boldsymbol{\theta}_{jj'}^{[g]} \|$$

1. Estimate regression coefficient matrices using group information $\mathcal{H}$:

$$\widehat{\mathcal{B}} = \underset{\mathcal{B}}{\text{argmin}} \left\{ \sum_{k=1}^{K} \frac{1}{n_k} \| (\mathbf{Y}^k - \mathbf{X}^k \mathbf{B}^k) \widehat{\mathbf{T}}^k \|_F^2 + \lambda_n \sum_{h \in \mathcal{H}} \| \mathbf{B}^{[h]} \| \right\}$$

where $\widehat{t}_{jj}^k = 1, \widehat{t}_{jj'}^k = -\widehat{\theta}_{jj'}^k$.

2. Estimate neighborhood coefficients of each Y-node, say $\Theta_j = (\boldsymbol{\theta}_j^1, \ldots, \boldsymbol{\theta}_j^K)$ using the group sparsity information $\mathcal{G}_y$:

$$\widehat{\Theta}_j = \underset{\Theta_j}{\text{argmin}} \left\{ \sum_{k=1}^{K} \frac{1}{n_k} \| \widehat{\mathbf{E}}_j^k - \widehat{\mathbf{E}}_{-j}^k \boldsymbol{\theta}_j^k \|^2 + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{y,jj'}} \| \boldsymbol{\theta}_{jj'}^{[g]} \| \right\}$$

3. Estimate precision matrices as MLE over the restricted supports:

$$\widehat{E}_y^k = \{(j, j') : 1 \leq j < j' \leq q, \widehat{\theta}_{jj'}^k \neq 0 \text{ OR } \widehat{\theta}_{j'j}^k \neq 0\},$$

$$\widehat{\Omega}_y^k = \underset{\Omega_y^k \in \mathbb{S}_+(\widehat{E}_y^k)}{\text{argmin}} \left\{ \text{Tr}(\widehat{\mathbf{S}}_y^k \Omega_y^k) - \log\det(\Omega_y^k) \right\}.$$

Joint Multiple Multi-Level Estimation (JMMLE)

1. Solve for $\{\mathcal{B}, \Theta\}$:

$$\{\widehat{\mathcal{B}}, \widehat{\Theta}\} = \underset{\mathcal{B},\Theta}{\operatorname{argmin}} \left\{ \sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{q} \left\| \mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \mathbf{B}_{-j}^k) \boldsymbol{\theta}_j^k - \mathbf{X}^k \mathbf{B}_j^k \right\|^2 \right.$$
$$\left. + \lambda_n \sum_{h \in \mathcal{H}} \| \mathbf{B}^{[h]} \| + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{jj'}} \| \boldsymbol{\theta}_{jj'}^{[g]} \| \right\}$$

2. Estimate precision matrices as MLE over the restricted supports:

$$\widehat{E}_y^k = \{(j, j') : 1 \leq j < j' \leq q, \widehat{\theta}_{jj'}^k \neq 0 \text{ OR } \widehat{\theta}_{j'j}^k \neq 0\},$$
$$\widehat{\Omega}_y^k = \underset{\Omega_y^k \in \mathbb{S}_+(\widehat{E}_y^k)}{\operatorname{argmin}} \left\{ \operatorname{Tr}(\widehat{\mathbf{S}}_y^k \Omega_y^k) - \log\det(\Omega_y^k) \right\}.$$

Computational algorithm

$$\{\widehat{\mathcal{B}}, \widehat{\Theta}\} = \underset{\mathcal{B},\Theta}{\operatorname{argmin}} \{f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta) + P(\mathcal{B}) + Q(\Theta)\}$$

The objective function is biconvex, so we solve the above by the following alternating iterative algorithm:

1. Start with initial estimates of $\mathcal{B}$ and Θ, say $\mathcal{B}^{(0)}, \Theta^{(0)}$.
2. Iterate:

$$\mathcal{B}^{(t+1)} = \underset{\mathcal{B}}{\operatorname{argmin}} \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta^{(t)}) + Q(\mathcal{B}) \right\}$$
$$\Theta^{(t+1)} = \underset{\Theta}{\operatorname{argmin}} \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}^{(t+1)}, \Theta) + P(\Theta) \right\}$$

3. Continue till convergence.

The two subproblems

$$\widehat{\mathcal{B}} = \underset{\mathcal{B}}{\arg\min} \left\{ \sum_{k=K}^{q} \frac{1}{n_k} \sum_{j=1}^{q} \left\| \mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \mathbf{B}_{-j}^k) \widehat{\boldsymbol{\theta}}_j^k - \mathbf{X}^k \mathbf{B}_j^k \right\|^2 + \lambda_n \sum_{h \in \mathcal{H}} \|\mathbf{B}^{[h]}\| \right\}$$

$$= \underset{\mathcal{B}}{\arg\min} \left\{ \sum_{k=1}^{K} \frac{1}{n_k} \|(\mathbf{Y}^k - \mathbf{X}^k \mathbf{B}^k) \widehat{\mathbf{T}}^k\|_F^2 + + \lambda_n \sum_{h \in \mathcal{H}} \|\mathbf{B}^{[h]}\| \right\}$$

where $\widehat{t}_{jj}^k = 1, \widehat{t}_{jj'}^k = -\widehat{\theta}_{jj'}^k$.

$$\widehat{\Theta} = \underset{\Theta}{\arg\min} \left\{ \sum_{k=K}^{q} \frac{1}{n_k} \sum_{j=1}^{q} \left\| \mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \widehat{\mathbf{B}}_{-j}^k) \boldsymbol{\theta}_j^k - \mathbf{X}^k \widehat{\mathbf{B}}_j^k \right\|^2 + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{jj'}} \|\boldsymbol{\theta}_{jj'}^{[g]}\| \right\}$$

$$= \underset{\Theta}{\arg\min} \left\{ \sum_{k=K}^{q} \frac{1}{n_k} \sum_{j=1}^{q} \|\widehat{\mathbf{E}}_j^k - \widehat{\mathbf{E}}_{-j}^k \boldsymbol{\theta}_j^k\|^2 + \gamma_n \sum_{j' \neq j, g \in \mathcal{G}_{jj'}} \|\boldsymbol{\theta}_{jj'}^{[g]}\| \right\}$$

where $\widehat{\mathbf{E}}^k = \mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k$.

For $\lambda_n \geq 4\sqrt{|h_{\max}|}\mathbb{R}_0\sqrt{\frac{\log(pq)}{n}}$, the following hold with probability approaching 1 as $n \to \infty$,

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1 \leq \frac{48\sqrt{|h_{\max}|}s_\beta \lambda_n}{\psi^*}$$

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \leq \frac{12\sqrt{s_\beta}\lambda_n}{\psi^*}$$

$$\sum_{h \in \mathcal{H}} \|\boldsymbol{\beta}^{[h]} - \boldsymbol{\beta}_0^{[h]}\| \leq \frac{48 s_\beta \lambda_n}{\psi^*}$$

with $\psi^*, \mathbb{R}_0$ being constants, and $\boldsymbol{\beta} = (\text{vec}(\mathbf{B}^1)^T, \ldots, \text{vec}(\mathbf{B}^K)^T)^T$, $|h_{\max}|$ the maximum group size in $\boldsymbol{\beta}_0$ (the true $\boldsymbol{\beta}$) and s_β the sparsity of $\boldsymbol{\beta}_0$.

Error bounds for $\widehat{\Theta}, \widehat{\Omega}$

For $\gamma_n = 4\sqrt{|g_{\max}|}\mathbb{Q}_0\sqrt{\frac{\log(pq)}{n}}$, the following hold with probability approaching 1 as $n \to \infty$,

$$\|\widehat{\Theta}_j - \Theta_{0,j}\|_F \leq \frac{12\sqrt{s_j}\gamma_n}{\psi}$$

$$\sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\widehat{\boldsymbol{\theta}}_{jj'}^{[g]} - \boldsymbol{\theta}_{0,jj'}^{[g]}\| \leq \frac{48 s_j \gamma_n}{\psi}$$

$$\left|\text{supp}(\widehat{\Theta}_j)\right| \leq \frac{128 s_j}{\psi}$$

$$\frac{1}{K}\sum_{k=1}^K \|\widehat{\Omega}_y^k - \Omega_y^k\|_F \leq O\left(\frac{\sqrt{S}\gamma_n}{\sqrt{K}}\right)$$

with $\psi, \mathbb{Q}_0$ being constants, $|g_{\max}|$ the maximum group size in Θ_0, s_j the sparsity of Θ_j and $S = \sum_j s_j$.

Outline

- Consider the case $K = 2$, and suppose we are interested in testing if the effect of variable i in the X-data is different across the two populations.
- For this we use the i^{th} rows of the estimates $\widehat{\mathbf{B}}^1$ and $\widehat{\mathbf{B}}^2$.
- We debias these row vectors using the neighborhood coefficients in the X-network computed previously using JSEM: In this setup, define the desparsified estimate of $\mathbf{b}_i^k$ as

$$\widehat{\mathbf{c}}_i^k = \widehat{\mathbf{b}}_i^k + \frac{1}{n t_i^k} \left(\mathbf{X}_i^k - \mathbf{X}_{-i}^k \widehat{\zeta}_i^k \right)^T (\mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k)$$

for $k = 1, 2$, where $t_i^k := (\mathbf{X}_i^k - \mathbf{X}_{-i}^k \widehat{\zeta}_i^k)^T \mathbf{X}_{-i}^k / n$.

Assume we have 'good enough' estimators:

$$\|\widehat{\zeta}^k - \zeta_0^k\|_1 = O\left(\sqrt{\frac{\log p}{n}}\right); \quad \|\widehat{\mathbf{B}}^k - \mathbf{B}_0^k\|_1 = O\left(\sqrt{\frac{\log(pq)}{n}}\right)$$

$$\left\|(\widehat{\Omega}_y^k)^{1/2} - (\Omega_y^k)^{1/2}\right\|_\infty = O\left(\sqrt{\frac{\log q}{n}}\right)$$

Also define

$$\widehat{s}_i^k := \sqrt{\|\mathbf{X}_i^k - \mathbf{X}_{-i}^k \widehat{\zeta}_i^k\|^2 / n}; \quad m_i^k := \sqrt{n} t_i^k / \widehat{s}_i^k$$

Then for sample size satisfying $n \gtrsim \log(pq), \log p = o(n^{1/2}), \log q = o(n^{1/2})$ we have

$$\begin{bmatrix} \widehat{\Omega}_y^1 & \\ & \widehat{\Omega}_y^2 \end{bmatrix}^{1/2} \begin{bmatrix} m_i^1(\widehat{\mathbf{c}}_i^1 - \mathbf{b}_i^1) & \\ & m_i^2(\widehat{\mathbf{c}}_i^2 - \mathbf{b}_i^2) \end{bmatrix} \sim \mathcal{N}_{2q}(\mathbf{0}, \mathbf{I}) + o_P(1)$$

Global test for $H_0 : \mathbf{b}_{0i}^1 = \mathbf{b}_{0i}^2$ at level $\alpha, 0 < \alpha < 1$

1. Obtain the debiased estimators $\widehat{\mathbf{c}}_i^1, \widehat{\mathbf{c}}_i^2$;
2. Calculate the test statistic

$$D_i = \left\| m_i^1 (\widehat{\Omega}_y^1)^{1/2} \widehat{\mathbf{c}}_i^1 - m_i^2 (\widehat{\Omega}_y^2)^{1/2} \widehat{\mathbf{c}}_i^2 \right\|^2$$

3. Reject H_0 if $D_i \geq \chi^2_{2q, 1-\alpha}$.

Simultaneous tests for $H_0^j : b_{0ij}^1 = b_{0ij}^2$ at level $\alpha, 0 < \alpha < 1$

1. Calculate the pairwise test statistics d_{ij} for $j = 1, \ldots, q$:

$$d_{ij} = \frac{\tau_{ij}^1 \widehat{\mathbf{c}}_{ij}^1 - \tau_{ij}^2 \widehat{\mathbf{c}}_{ij}^2}{1/m_i^1 + 1/m_i^2}$$

where τ_{ij}^k is the $(i,j)^{\text{th}}$ element of $(\widehat{\Omega}_y^k)^{1/2}, k = 1, 2$.

2. Obtain the threshold

$$\widehat{\tau} = \inf \left\{ \tau \in \mathbb{R} : 1 - \Phi(\tau) \leq \frac{\alpha}{2q} \max \left(\sum_{j \in \mathcal{I}_q} \mathbb{I}(|d_{ij}| \geq \tau), 1 \right) \right\}$$

3. For $j \in \mathcal{I}_q$, reject H_0^j if $|d_{ij}| \geq \widehat{\tau}$.

Outline

Simulation setup

- Number of categories (K) = 5;
- Structured $\{\Omega_x\}, \{\Omega_y\}, \mathcal{B}$;
- Groups in $\mathcal{B}, \Omega_x$ are non-zero with probability $5/p$, and their elements come from Unif$(-1, -0.5) \cup (0.5, 1)$;
- Groups in Ω_y are non-zero with probability $5/q$, and their elements come from Unif$(-1, -0.5) \cup (0.5, 1)$;
- We generate size-n i.i.d. samples $\mathbf{X}^k$ from $\mathcal{N}_p(0, \Sigma_x^k)$, and $\mathbf{E}^k$ from $\mathcal{N}_p(0, \Sigma_y^k)$, then obtain $\mathbf{Y}^k = \mathbf{X}^k \mathbf{B}^k + \mathbf{E}^k$;
- 100 Replications.

Evaluation metrics

1. True positives-

$$\text{TP}(\widehat{\mathcal{B}}) = \frac{\sum_k |\,\text{supp}(\widehat{\mathbf{B}}^k) \cup \text{supp}(\mathbf{B}_0^k)|}{\sum_k |\,\text{supp}(\mathbf{B}_0^k)|}$$

2. True negatives-

$$\text{TN}(\widehat{\mathcal{B}}) = \frac{\sum_k |\,\text{supp}^c(\widehat{\mathbf{B}}^k) \cup \text{supp}^c(\mathbf{B}_0^k)|}{\sum_k |\,\text{supp}^c(\mathbf{B}_0^k)|}$$

3. Relative error in Frobenius norm-

$$\text{rel.Frob}(\widehat{\mathcal{B}}) = \sum_{k=1}^{K} \frac{\|\widehat{\mathbf{B}}^k - \mathbf{B}_0^k\|_F}{\|\mathbf{B}_0^k\|_F}$$

Same metrics are used for $\widehat{\Theta}$.

Results

Method	TP($\widehat{\mathcal{B}}$)	TN($\widehat{\mathcal{B}}$)	rel.Frob($\widehat{\mathcal{B}}$)	TP($\widehat{\Theta}$)	TN($\widehat{\Theta}$)	rel.Frob($\widehat{\Theta}$)
Setting 1: $n = 100$, $p = 60$, $q = 30$, $K = 5$						
Joint (JMMLE)	0.999 (2e-3)	0.99 (0.01)	0.19 (0.02)	0.66 (0.06)	0.95 (0.01)	0.33 (0.02)
Separate	0.95 (0.02)	0.99 (2e-3)	0.27 (0.03)	0.89 (0.02)	0.63 (0.01)	0.77 (0.04)
Setting 2: $n = 100$, $p = 30$, $q = 60$, $K = 5$						
Joint (JMMLE)	0.996 (4e-3)	0.99 (6e-3)	0.21 (0.01)	0.58 (0.04)	0.98 (3e-3)	0.32 (8e-3)
Separate	0.66 (0.04)	0.994 (1e-3)	0.59 (0.03)	0.62 (0.03)	0.81 (7e-3)	0.43 (0.01)
Setting 3: $n = 150$, $p = 200$, $q = 200$, $K = 5$						
Joint (JMMLE)	1.00 (0)	1.00 (0)	0.12 (5e-3)	0.39 (0.04)	0.996 (2e-3)	0.30 (7e-3)
Separate	0.42 (0.03)	0.99 (1e-3)	0.48 (0.02)	0.41 (0.02)	0.73 (0.01)	0.44 (0.02)

Future work

- Larger simulation: effect of signal strength and sparsity;
- Real data application, application to personalized medicine;
- Mediation analysis.

References

M. Drton and M. D. Perlman. Multiple Testing and Error Control in Gaussian Graphical Model Selection. *Statist. Sci.*, 22(3):430–449, 2007.

V. Gligorijević and N. Pržulj. Methods for biological data integration: perspectives and challenges. *J. R. Soc. Interface*, 12(112):20150571, 2015.

J. Guo, E. Levina, G. Michailidis, and J. Zhu. Joint estimation of multiple graphical models. *Biometrika*, 98(1): 1–15, 2011.

J. Lin, S. Basu, M. Banerjee, and G. Michailidis. Penalized Maximum Likelihood Estimation of Multi-layered Gaussian Graphical Models. *J. Mach. Learn. Res.*, 17:5097–5147, 2016.

W. Liu. Structural similarity and difference testing on multiple sparse Gaussian graphical models. *Ann. Statist.*, To appear, 2017+.

J. Ma and G. Michailidis. Joint Structural Estimation of Multiple Graphical Models. *J. Mach. Learn. Res.*, 17: 5777–5824, 2016.

N. Meinshausen and P. Bühlmann. High dimensional graphs and variable selection with the Â¨ lasso. *Ann. Statist.*, 34(3):1436–1462, 2006.

THANK YOU!

Questions?