Inter-Subject Analysis: A Formal Theory

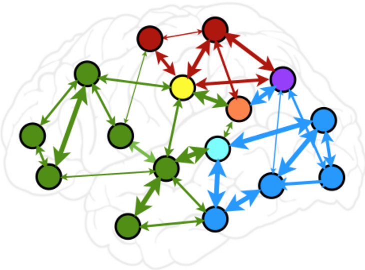
Cong Ma
Operations Research and Financial Engineering
Princeton University

Junwei Lu
Princeton ORFE

Han Liu
Princeton ORFE

From fMRI to functional connectivity

Credit: iStock



Credit: MONET

Controlled experimental settings

Each hand position was presented for 3.5 s. This is a task requiring higher order motor coordination and motor planning and in the FRB description, it was noted that this task should activate the frontal and parietal areas.

In the OM task subjects were watching an image including a central cross in the middle surrounded by 10 black boxes. Subjects were instructed to concentrate on the central cross and saccade to the surrounding box if it changed white for a moment. After this, they should have returned their gaze immediately to the central cross. In each 'on' block there were 20 fixation trials and 20 target trials. There were four fixations of each of the following durations: 800 ms, 1000 ms, 1200 ms, 1400 ms, and 1600. These were randomized and each were followed by a 200 ms target trial. This way the task was supposed to activate the visual system and the occipital lobe.

Finally, in the VG task, the images of certain objects were

Credit: Juha Pajula et al.

Naturalistic settings

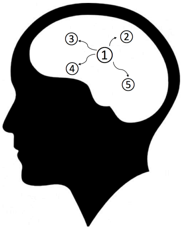
Adapted from Louise Freeman's blogpost

Naturalistic settings

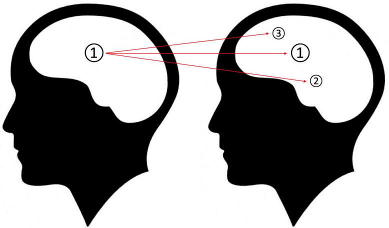
Adapted from Louise Freeman's blogpost

Challenge: noise due to intrinsic cognitive processes.

From intra- to inter-subject correlations

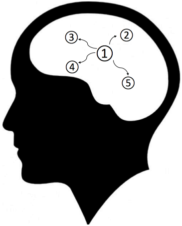


Intra-subject correlations

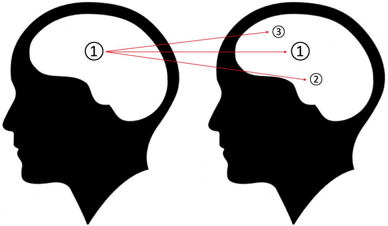


Inter-subject correlations

From intra- to inter-subject correlations



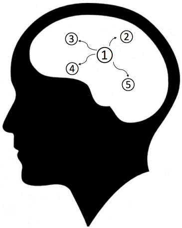
Intra-subject correlations



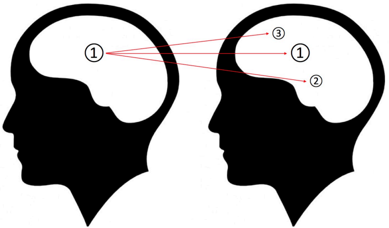
Inter-subject correlations

Modeling assumption: individual noise is uncorrelated across subjects.

From intra- to inter-subject correlations



Intra-subject correlations

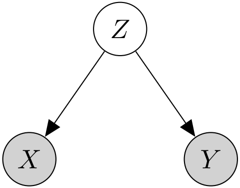


Inter-subject correlations

Modeling assumption: individual noise is uncorrelated across subjects.

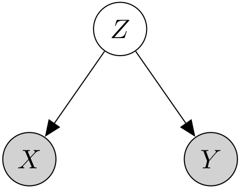
Problem: marginal dependence influenced by confounding effects!

Confounding effects



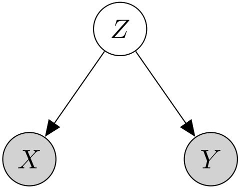
- X and Y are marginally dependent.

Confounding effects



- X and Y are marginally dependent.
- $X \perp\!\!\!\perp Y \mid Z$.

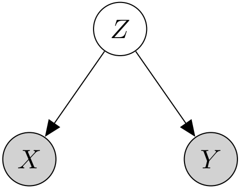
Confounding effects



- X and Y are marginally dependent.
- $X \perp\!\!\!\perp Y \mid Z$.

Correlation coefficient is weak criterion for measuring dependence.

Confounding effects



- X and Y are marginally dependent.
- $X \perp\!\!\!\perp Y \mid Z$.

Correlation coefficient is weak criterion for measuring dependence.

Solution: conditional dependence.

This work: a formal theory for Inter-Subject Analysis!

Modeling framework for ISA

- $X = (X_1, \ldots, X_d)^\top \sim \mathcal{N}(0, \Sigma^*)$.
  - precision matrix $\Omega^* = (\Sigma^*)^{-1} = (\omega^*_{jk})$.

- $\mathcal{G}_1, \mathcal{G}_2$: disjoint subsets of $[d]$.
  - $|\mathcal{G}_1| = d_1$.
  - $|\mathcal{G}_2| = d_2 := d - d_1$.

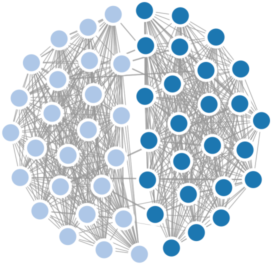
- $X_{\mathcal{G}_1}$ and $X_{\mathcal{G}_2}$: features of two different subjects.
  - $X_{\mathcal{G}_1}$: voxels in subject A.
  - $X_{\mathcal{G}_2}$: voxels in subject B.

- Data $\mathbb{X} \in \mathbb{R}^{n \times d}$.
  - each row represents a measurement of two subjects.
  - n measurements in total.

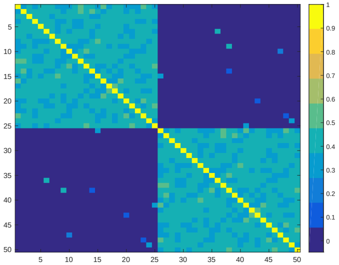
Gaussian graphical models

Fact 1

For any j and k, $X_j \perp\!\!\!\perp X_k \mid X_{\setminus\{j,k\}}$ if and only if $\omega^*_{jk} = 0$.



Graph



Ω^*

Goal of ISA

Partition covariance and precision matrices into

$$\Sigma^* = \begin{bmatrix} \Sigma_1^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\top} & \Sigma_2^* \end{bmatrix}, \quad \Omega^* = \begin{bmatrix} \Omega_1^* & \Omega_{12}^* \\ \Omega_{12}^{*\top} & \Omega_2^* \end{bmatrix}.$$

Goal of ISA

Partition covariance and precision matrices into

$$\Sigma^* = \begin{bmatrix} \Sigma_1^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\top} & \Sigma_2^* \end{bmatrix}, \quad \Omega^* = \begin{bmatrix} \Omega_1^* & \Omega_{12}^* \\ \Omega_{12}^{*\top} & \Omega_2^* \end{bmatrix}.$$

- Ω_{12}^*: driven by common stimuli.
- Ω_1^* and Ω_2^*: influenced by individual cognitive processes.
- When $n \ll d$, minimal assumption is that Ω_{12}^* is sparse.

Goal of ISA

Partition covariance and precision matrices into

$$\Sigma^* = \begin{bmatrix} \Sigma_1^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\top} & \Sigma_2^* \end{bmatrix}, \quad \Omega^* = \begin{bmatrix} \Omega_1^* & \Omega_{12}^* \\ \Omega_{12}^{*\top} & \Omega_2^* \end{bmatrix}.$$

- Ω_{12}^*: driven by common stimuli.
- Ω_1^* and Ω_2^*: influenced by individual cognitive processes.
- When $n \ll d$, minimal assumption is that Ω_{12}^* is sparse.

Goal: estimate (infer) sparse Ω_{12}^* under (possibly dense) Ω_1^* and Ω_2^* .

How to estimate Ω^*_{12}?

Candidate strategy 1: estimate Ω^*

Penalized maximum likelihood estimator for Ω^* (GLASSO):

$$\widehat{\Omega} = \text{argmin}_{\Omega} \underbrace{\mathrm{Tr}(\Omega\widehat{\Sigma}) - \log|\Omega|}_{\text{negative log-likelihood}} + \underbrace{\lambda\|\Omega\|_{1,1}}_{\text{sparsity penalty}},$$

where $\widehat{\Sigma}$ is sample covariance matrix.

Candidate strategy 1: estimate Ω^*

Penalized maximum likelihood estimator for Ω^* (GLASSO):

$$\widehat{\Omega} = \text{argmin}_{\Omega} \underbrace{\mathrm{Tr}(\Omega\widehat{\Sigma}) - \log|\Omega|}_{\text{negative log-likelihood}} + \underbrace{\lambda\|\Omega\|_{1,1}}_{\text{sparsity penalty}},$$

where $\widehat{\Sigma}$ is sample covariance matrix.

Problem: Ω^* is not sparse.

Candidate strategy 2: estimate Ω^*_{*j}

Column-wise estimator for Ω^*_{*j}:

- Neighborhood selection: regress X_j on $X_{\setminus\{j\}}$.

- CLIME:

$$\begin{aligned} \widehat{\Omega}_{*j} \;=\; & \text{argmin}_\beta && \|\beta\|_1 \\ & \text{subject to} && \|\widehat{\Sigma}\beta - e_j\|_\infty \le \lambda. \end{aligned}$$

Candidate strategy 2: estimate Ω^*_{*j}

Column-wise estimator for Ω^*_{*j}:

- Neighborhood selection: regress X_j on $X_{\setminus\{j\}}$.

- CLIME:

$$\widehat{\Omega}_{*j} = \text{argmin}_\beta \quad \|\beta\|_1 \quad \text{subject to} \quad \|\widehat{\Sigma}\beta - e_j\|_\infty \leq \lambda.$$

Problem: Ω^*_{*j} is not sparse.

Candidate strategy 3: estimate Ω_{12}^*

A naive decomposition $\Omega = \Omega_D + \Omega_O$:

$$\Omega = \begin{bmatrix} \Omega_1 & \Omega_{12} \\ \Omega_{12}^\top & \Omega_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \Omega_1 & 0 \\ 0 & \Omega_2 \end{bmatrix}}_{\Omega_D} + \underbrace{\begin{bmatrix} 0 & \Omega_{12} \\ \Omega_{12}^\top & 0 \end{bmatrix}}_{\Omega_O}.$$

Candidate strategy 3: estimate Ω_{12}^*

A naive decomposition $\Omega = \Omega_D + \Omega_O$:

$$\Omega = \begin{bmatrix} \Omega_1 & \Omega_{12} \\ \Omega_{12}^\top & \Omega_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \Omega_1 & 0 \\ 0 & \Omega_2 \end{bmatrix}}_{\Omega_D} + \underbrace{\begin{bmatrix} 0 & \Omega_{12} \\ \Omega_{12}^\top & 0 \end{bmatrix}}_{\Omega_O}.$$

$$\Longrightarrow \quad \mathcal{L}_n(\Omega_O, \Omega_D) = \mathrm{Tr}(\Omega_O \widehat{\Sigma}) + \underbrace{\mathrm{Tr}(\Omega_D \widehat{\Sigma})}_{\text{independent of } \Omega_O} - \log |\Omega_O + \Omega_D|.$$

Candidate strategy 3: estimate Ω_{12}^*

A naive decomposition $\Omega = \Omega_D + \Omega_O$:

$$\Omega = \begin{bmatrix} \Omega_1 & \Omega_{12} \\ \Omega_{12}^\top & \Omega_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \Omega_1 & 0 \\ 0 & \Omega_2 \end{bmatrix}}_{\Omega_D} + \underbrace{\begin{bmatrix} 0 & \Omega_{12} \\ \Omega_{12}^\top & 0 \end{bmatrix}}_{\Omega_O}.$$

$$\implies \quad \mathcal{L}_n(\Omega_O, \Omega_D) = \mathrm{Tr}(\Omega_O \widehat{\Sigma}) + \underbrace{\mathrm{Tr}(\Omega_D \widehat{\Sigma})}_{\text{independent of } \Omega_O} - \log|\Omega_O + \Omega_D|.$$

$$\implies \quad \mathcal{L}_n(\Omega_O) = \mathrm{Tr}(\Omega_O \widehat{\Sigma}) - \log|\Omega_O + \Omega_D|.$$

Candidate strategy 3: estimate Ω_{12}^*

A naive decomposition $\Omega = \Omega_D + \Omega_O$:

$$\Omega = \begin{bmatrix} \Omega_1 & \Omega_{12} \\ \Omega_{12}^\top & \Omega_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \Omega_1 & 0 \\ 0 & \Omega_2 \end{bmatrix}}_{\Omega_D} + \underbrace{\begin{bmatrix} 0 & \Omega_{12} \\ \Omega_{12}^\top & 0 \end{bmatrix}}_{\Omega_O}.$$

$$\implies \quad \mathcal{L}_n(\Omega_O, \Omega_D) = \mathrm{Tr}(\Omega_O \widehat{\Sigma}) + \underbrace{\mathrm{Tr}(\Omega_D \widehat{\Sigma})}_{\text{independent of } \Omega_O} - \log|\Omega_O + \Omega_D|.$$

$$\implies \quad \mathcal{L}_n(\Omega_O) = \mathrm{Tr}(\Omega_O \widehat{\Sigma}) - \log|\Omega_O + \Omega_D|.$$

Problem: estimating Ω_D is also hard.

Our strategy: an alternative parameter

Key parameter: $\Theta^* = \Omega^* - (\Sigma_D^*)^{-1}$, where $\Sigma_D^* = \text{diag}(\Sigma_1^*, \Sigma_2^*)$.

Our strategy: an alternative parameter

Key parameter: $\Theta^* = \Omega^* - (\Sigma_D^*)^{-1}$, where $\Sigma_D^* = \mathrm{diag}(\Sigma_1^*, \Sigma_2^*)$.

More concretely,

$$\Theta^* = \begin{bmatrix} \Sigma_1^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\top} & \Sigma_2^* \end{bmatrix}^{-1} - \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & \Sigma_2^* \end{bmatrix}^{-1} = \begin{bmatrix} \Theta_1^* & \Theta_{12}^* \\ \Theta_{12}^{*\top} & \Theta_2^* \end{bmatrix}.$$

Our strategy: an alternative parameter

Key parameter: $\Theta^* = \Omega^* - (\Sigma_D^*)^{-1}$, where $\Sigma_D^* = \text{diag}(\Sigma_1^*, \Sigma_2^*)$.

More concretely,

$$\Theta^* = \begin{bmatrix} \Sigma_1^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\top} & \Sigma_2^* \end{bmatrix}^{-1} - \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & \Sigma_2^* \end{bmatrix}^{-1} = \begin{bmatrix} \Theta_1^* & \Theta_{12}^* \\ \Theta_{12}^{*\top} & \Theta_2^* \end{bmatrix}.$$

Why Θ^*?

Our strategy: an alternative parameter

Key parameter: $\Theta^* = \Omega^* - (\Sigma_D^*)^{-1}$, where $\Sigma_D^* = \text{diag}(\Sigma_1^*, \Sigma_2^*)$.

More concretely,

$$\Theta^* = \begin{bmatrix} \Sigma_1^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\top} & \Sigma_2^* \end{bmatrix}^{-1} - \begin{bmatrix} \Sigma_1^* & 0 \\ 0 & \Sigma_2^* \end{bmatrix}^{-1} = \begin{bmatrix} \Theta_1^* & \Theta_{12}^* \\ \Theta_{12}^{*\top} & \Theta_2^* \end{bmatrix}.$$

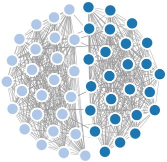
Why Θ^*?

$$\Theta_{12}^* = \Omega_{12}^*.$$

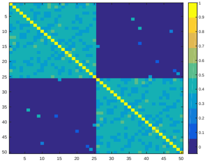
Θ^* is also sparse

Fact 2

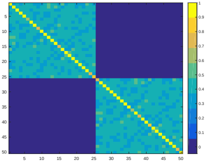
If Ω^*_{12} is s-sparse, then Θ^* is at most $(2s^2 + 2s)$-sparse.



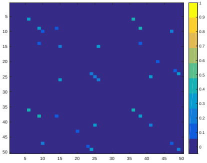
Graph



Ω^*



$(\Sigma^*_D)^{-1}$



Θ^*

STRINGS estimator

Under new decomposition $\Omega = \Theta + \Sigma_D^{-1}$, we have

$$\mathcal{L}_n(\Theta, \Sigma_D^{-1}) = \mathrm{Tr}[(\Theta + \Sigma_D^{-1})\widehat{\Sigma}] - \log|\Theta + \Sigma_D^{-1}|.$$

STRINGS estimator

Under new decomposition $\Omega = \Theta + \Sigma_D^{-1}$, we have

$$\mathcal{L}_n(\Theta, \Sigma_D^{-1}) = \mathrm{Tr}[(\Theta + \Sigma_D^{-1})\widehat{\Sigma}] - \log|\Theta + \Sigma_D^{-1}|.$$

Separate Θ from Σ_D^{-1},

$$\mathcal{L}_n(\Theta, \Sigma_D^{-1}) = \mathrm{Tr}(\Theta\widehat{\Sigma}) + \underbrace{\mathrm{Tr}(\Sigma_D^{-1}\widehat{\Sigma}) - 2\log|\Sigma_D^{-1}|}_{\text{independent of } \Theta} - \log\underbrace{|\Sigma_D\Theta\Sigma_D + \Sigma_D|}_{\text{plug-in estimator}}.$$

STRINGS estimator

Under new decomposition $\Omega = \Theta + \Sigma_D^{-1}$, we have

$$\mathcal{L}_n(\Theta, \Sigma_D^{-1}) = \mathrm{Tr}[(\Theta + \Sigma_D^{-1})\widehat{\Sigma}] - \log|\Theta + \Sigma_D^{-1}|.$$

Separate Θ from Σ_D^{-1},

$$\mathcal{L}_n(\Theta, \Sigma_D^{-1}) = \mathrm{Tr}(\Theta\widehat{\Sigma}) + \underbrace{\mathrm{Tr}(\Sigma_D^{-1}\widehat{\Sigma}) - 2\log|\Sigma_D^{-1}|}_{\text{independent of } \Theta} - \log\underbrace{|\Sigma_D\Theta\Sigma_D + \Sigma_D|}_{\text{plug-in estimator}}.$$

Sparse edge es**T**imato**R** for **I**ntense **N**uisance **G**raph**S**:

$$\widehat{\Theta} = \mathrm{argmin}_{\Theta}\, \mathrm{Tr}(\Theta\widehat{\Sigma}) - \log|\widehat{\Sigma}_D\Theta\widehat{\Sigma}_D + \widehat{\Sigma}_D| + \lambda\|\Theta\|_{1,1}.$$

Estimation consistency

Theorem 3

Suppose $\|\Omega^*_{12}\|_0 \le s$. *Under sample size condition* $s^4\sqrt{\log d/n} = \mathcal{O}(1)$ *and some regularity conditions, if* $\lambda \asymp \sqrt{\log d/n}$, *with high probability*

$$\|\widehat{\Theta} - \Theta^*\|_F = \mathcal{O}\left(\sqrt{\frac{s^2 \log d}{n}}\right) \text{ and } \|\widehat{\Theta} - \Theta^*\|_{1,1} = \mathcal{O}\left(s^2\sqrt{\frac{\log d}{n}}\right).$$

Estimation consistency

Theorem 3

Suppose $\|\Omega^*_{12}\|_0 \leq s$. *Under sample size condition* $s^4\sqrt{\log d/n} = \mathcal{O}(1)$ *and some regularity conditions, if* $\lambda \asymp \sqrt{\log d/n}$, *with high probability*

$$\|\widehat{\Theta} - \Theta^*\|_F = \mathcal{O}\left(\sqrt{\frac{s^2 \log d}{n}}\right) \text{ and } \|\widehat{\Theta} - \Theta^*\|_{1,1} = \mathcal{O}\left(s^2\sqrt{\frac{\log d}{n}}\right).$$

- Rate of convergence: intrinsic to this method.
  - recall $\|\Theta^*\|_0 \lesssim s^2$.

What about inference?

A general strategy: de-biasing estimators

Consider a general regularized estimator

$$\widehat{\beta} = \text{argmin}_\beta \underbrace{\mathcal{L}_n(\beta)}_{\text{negative log-likelihood}} + \underbrace{\lambda\|\beta\|}_{\text{regularizer}} .$$

A general strategy: de-biasing estimators

Consider a general regularized estimator

$$\widehat{\beta} = \mathsf{argmin}_\beta \underbrace{\mathcal{L}_n(\beta)}_{\text{negative log-likelihood}} + \underbrace{\lambda\|\beta\|}_{\text{regularizer}} .$$

A de-biased estimator takes following form:

$$\widehat{\beta}^u = \widehat{\beta} - \underbrace{M}_{\text{bias correction matrix}} \nabla\mathcal{L}_n(\widehat{\beta}).$$

Rationale behind general strategy

De-biased estimator: $\widehat{\beta}^u = \widehat{\beta} - M\nabla\mathcal{L}_n(\widehat{\beta})$.

Taylor expansion:

$$\sqrt{n}\cdot(\widehat{\beta}^u - \beta^*) \approx - \underbrace{\sqrt{n}M\nabla\mathcal{L}_n(\beta^*)}_{\text{asymptotically normal}} - \underbrace{\sqrt{n}[M\nabla^2\mathcal{L}_n(\beta^*) - I](\widehat{\beta} - \beta^*)}_{o_p(1)}.$$

How to de-bias $\widehat{\Theta}$

De-biased estimator:

$$\widehat{\Theta}^u = \widehat{\Theta} - \mathcal{M} \underbrace{\left(\widehat{\Sigma}\widehat{\Theta}\widehat{\Sigma}_D + \widehat{\Sigma} - \widehat{\Sigma}_D\right)}_{\text{counterpart of } \nabla \mathcal{L}_n(\widehat{\Theta})},$$

where $\mathcal{M}$ is bias correction operator.

How to de-bias $\widehat{\Theta}$

De-biased estimator:

$$\widehat{\Theta}^u = \widehat{\Theta} - \mathcal{M} \underbrace{\left(\widehat{\Sigma}\widehat{\Theta}\widehat{\Sigma}_D + \widehat{\Sigma} - \widehat{\Sigma}_D\right)}_{\text{counterpart of } \nabla \mathcal{L}_n(\widehat{\Theta})},$$

where $\mathcal{M}$ is bias correction operator.

De-biased estimator:

$$\widehat{\Theta}^u = \widehat{\Theta} - M(\widehat{\Sigma}\widehat{\Theta}\widehat{\Sigma}_D + \widehat{\Sigma} - \widehat{\Sigma}_D)P^\top.$$

Decomposition of $\widehat{\Theta}^u$

$\widehat{\Theta}^u - \Theta^* =$ Leading $+$ Remainder:

$$\text{Leading} = -M[(\widehat{\Sigma} - \Sigma^*)(I + \Theta^*\Sigma_D^*) - (I - \Sigma^*\Theta^*)(\widehat{\Sigma}_D - \Sigma_D^*)]P^\top.$$

Decomposition of $\widehat{\Theta}^u$

$\widehat{\Theta}^u - \Theta^* =$ Leading + Remainder:

$$\begin{aligned}\text{Remainder} = & - M(\widehat{\Sigma} - \Sigma^*)\Theta^*(\widehat{\Sigma}_D - \Sigma^*_D)P^\top \\ & - (M\widehat{\Sigma} - I)(\widehat{\Theta} - \Theta^*)(\widehat{\Sigma}_D P^\top - I) \\ & - (M\widehat{\Sigma} - I)(\widehat{\Theta} - \Theta^*) - (\widehat{\Theta} - \Theta^*)(\widehat{\Sigma}_D P^\top - I).\end{aligned}$$

Decomposition of $\widehat{\Theta}^u$

$\widehat{\Theta}^u - \Theta^* =$ Leading + Remainder:

$$\begin{aligned}\text{Remainder} = &- M(\widehat{\Sigma} - \Sigma^*)\Theta^*(\widehat{\Sigma}_D - \Sigma_D^*)P^\top \\ &- (M\widehat{\Sigma} - I)(\widehat{\Theta} - \Theta^*)(\widehat{\Sigma}_D P^\top - I) \\ &- (M\widehat{\Sigma} - I)(\widehat{\Theta} - \Theta^*) - (\widehat{\Theta} - \Theta^*)(\widehat{\Sigma}_D P^\top - I).\end{aligned}$$

Problem: cannot estimate Ω^* and $(\Sigma_D^*)^{-1}$ due to lack of sparsity.

Constrained optimization is sufficient

$$\begin{aligned}\text{Remainder} = & -M(\widehat{\Sigma} - \Sigma^*)\Theta^*(\widehat{\Sigma}_D - \Sigma_D^*)P^\top \\ & -(M\widehat{\Sigma} - I)(\widehat{\Theta} - \Theta^*)(\widehat{\Sigma}_D P^\top - I) \\ & -(M\widehat{\Sigma} - I)(\widehat{\Theta} - \Theta^*) - (\widehat{\Theta} - \Theta^*)(\widehat{\Sigma}_D P^\top - I).\end{aligned}$$

Key: M and P need not to be estimates of Ω^* and $(\Sigma_D^*)^{-1}$.

Constrained optimization is sufficient

$$\begin{aligned}\text{Remainder} = &- M(\widehat{\Sigma} - \Sigma^*)\Theta^*(\widehat{\Sigma}_D - \Sigma_D^*)P^\top \\ &- (M\widehat{\Sigma} - I)(\widehat{\Theta} - \Theta^*)(\widehat{\Sigma}_D P^\top - I) \\ &- (M\widehat{\Sigma} - I)(\widehat{\Theta} - \Theta^*) - (\widehat{\Theta} - \Theta^*)(\widehat{\Sigma}_D P^\top - I).\end{aligned}$$

Key: M and P need not to be estimates of Ω^* and $(\Sigma_D^*)^{-1}$.

Instead, we can solve following optimization problems:

$$\begin{aligned}&\text{find} \quad M \\ &\text{s.t.} \quad \|M\widehat{\Sigma} - I\|_{\max} \le \lambda'.\end{aligned} \qquad \begin{aligned}&\text{find} \quad P \\ &\text{s.t.} \quad \|P\widehat{\Sigma}_D - I\|_{\max} \le \lambda'.\end{aligned}$$

Sample splitting

$$\text{Leading} = -\underbrace{M}_{M(\widehat{\Sigma})}[(\widehat{\Sigma}-\Sigma^*)(I+\Theta^*\Sigma_D^*)-(I-\Sigma^*\Theta^*)(\widehat{\Sigma}_D-\Sigma_D^*)]\underbrace{P^\top}_{P(\widehat{\Sigma}_D)} .$$

Sample splitting

$$\text{Leading} = -\underbrace{M}_{M(\widehat{\Sigma})}[(\widehat{\Sigma}-\Sigma^*)(I+\Theta^*\Sigma_D^*)-(I-\Sigma^*\Theta^*)(\widehat{\Sigma}_D-\Sigma_D^*)]\underbrace{P^\top}_{P(\widehat{\Sigma}_D)}.$$

Not asymptotically normal since M and P are dependent on $\widehat{\Sigma}$.

Sample splitting

$$\text{Leading} = -\underbrace{M}_{M(\widehat{\Sigma})}[(\widehat{\Sigma}-\Sigma^*)(I+\Theta^*\Sigma_D^*)-(I-\Sigma^*\Theta^*)(\widehat{\Sigma}_D-\Sigma_D^*)]\underbrace{P^\top}_{P(\widehat{\Sigma}_D)}.$$

Not asymptotically normal since M and P are dependent on $\widehat{\Sigma}$.

Trick: sample splitting.

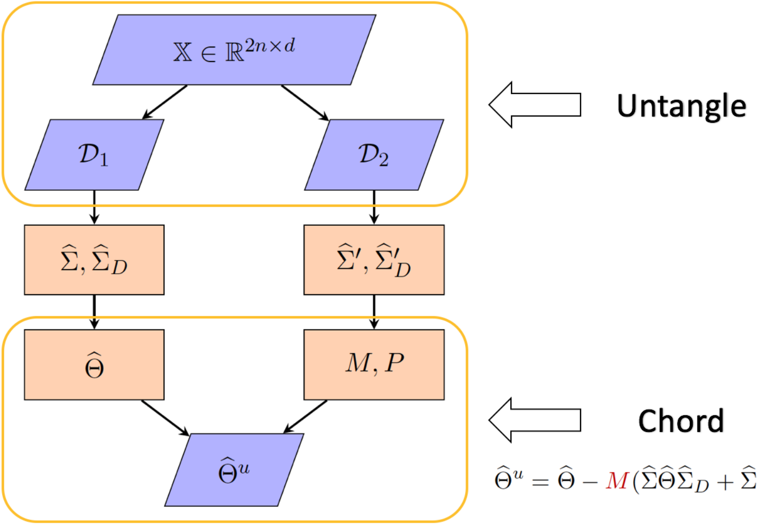
Sample splitting

$$\text{Leading} = -\underbrace{M}_{M(\widehat{\Sigma}')}[(\widehat{\Sigma}-\Sigma^*)(I+\Theta^*\Sigma_D^*)-(I-\Sigma^*\Theta^*)(\widehat{\Sigma}_D-\Sigma_D^*)]\underbrace{P^\top}_{P(\widehat{\Sigma}'_D)}.$$

Not asymptotically normal since M and P are dependent on $\widehat{\Sigma}$.

Trick: sample splitting.

"Untangle and Chord"



$$\widehat{\Theta}^u = \widehat{\Theta} - M(\widehat{\Sigma}\widehat{\Theta}\widehat{\Sigma}_D + \widehat{\Sigma} - \widehat{\Sigma}_D)P^\top$$

Asymptotic normality

Theorem 4

Suppose conditions in Theorem 3 hold. Further assume $\|\Omega^*\|_1 = \mathcal{O}(1)$*. Let* $\widehat{\Theta}^u = (\widehat{\theta}^u_{jk})$ *be de-biased estimator with* $\lambda' \asymp \sqrt{\log d/n}$*. Under scaling condition* $s^2 \log d/\sqrt{n} = o(1)$*,*

$$\sqrt{n} \cdot (\widehat{\theta}^u_{jk} - \widehat{\theta}^*_{jk})/\widehat{\xi}_{jk} \rightsquigarrow \mathcal{N}(0, 1).$$

Asymptotic normality

Theorem 4

Suppose conditions in Theorem 3 hold. Further assume $\|\Omega^*\|_1 = \mathcal{O}(1)$. *Let* $\widehat{\Theta}^u = (\widehat{\theta}^u_{jk})$ *be de-biased estimator with* $\lambda' \asymp \sqrt{\log d/n}$. *Under scaling condition* $s^2 \log d/\sqrt{n} = o(1)$*,*

$$\sqrt{n} \cdot (\widehat{\theta}^u_{jk} - \widehat{\theta}^*_{jk})/\widehat{\xi}_{jk} \rightsquigarrow \mathcal{N}(0, 1).$$

- Stronger scaling condition: $s^2 \log d/n = o(1)$ for estimation.

Asymptotic normality

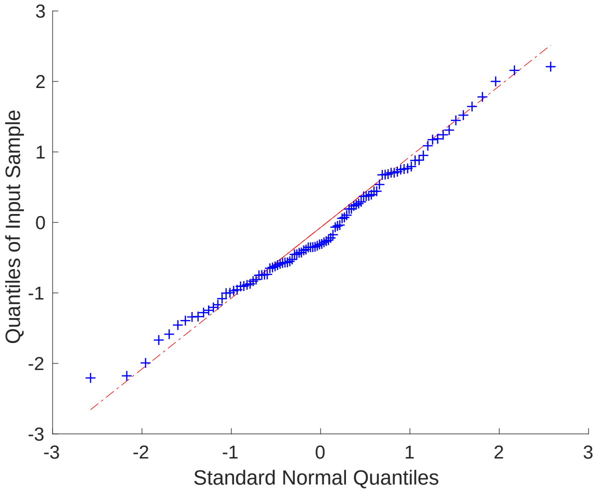
Theorem 4

Suppose conditions in Theorem 3 hold. Further assume $\|\Omega^*\|_1 = \mathcal{O}(1)$. Let $\widehat{\Theta}^u = (\widehat{\theta}^u_{jk})$ be de-biased estimator with $\lambda' \asymp \sqrt{\log d/n}$. Under scaling condition $s^2 \log d/\sqrt{n} = o(1)$,

$$\sqrt{n} \cdot (\widehat{\theta}^u_{jk} - \widehat{\theta}^*_{jk})/\widehat{\xi}_{jk} \rightsquigarrow \mathcal{N}(0,1).$$

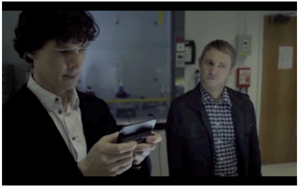
- Stronger scaling condition: $s^2 \log d/n = o(1)$ for estimation.
- Removal of sparsity condition: at the expense of $\|\Omega^*\|_1 = \mathcal{O}(1)$.

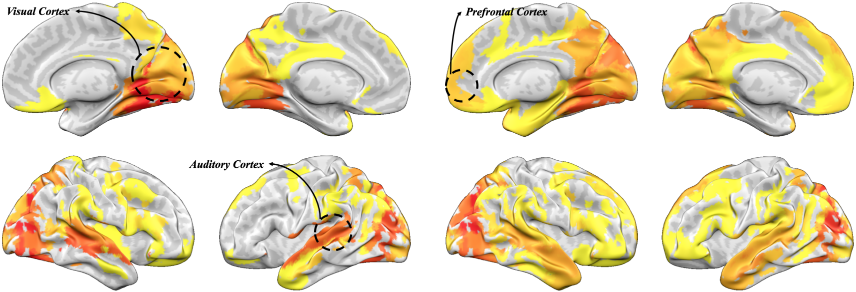
Synthetic data



$$\sqrt{n} \cdot (\widehat{\theta}^{u}_{11} - \theta^{*}_{11})/\widehat{\xi}_{11}$$

Real data experiments





Summary

- A formal theory for Inter-Subject Analysis motivated by fMRI.
    - studied estimation and inference of precision matrix in absence of sparsity.
- Future directions:
    - lower bound for the estimation error?
    - general inferential strategy: no sample splitting? no ℓ_1-norm assumption?

Paper:

"Inter-Subject Analysis: Inferring Sparse Interactions with Dense Intra-Graphs", Cong Ma, Junwei Lu, Han Liu.