Estimation and Inference for Brain Connectivity Analysis

Lexin Li

Division of Biostatistics
University of California, Berkeley

Outline

- talk outline:
  - motivation
  - estimation: multiple partial correlation matrices
  - inference: comparison of two partial correlation matrices
  - association modeling: symmetric tensor predictor regression
- collaboration:
  - William Jagust Lab @ Helen Wills Neuroscience Institute
  - Yunzhang Zhu @ Ohio State
  - Yin Xia @ UNC, Chapel Hill
  - Hua Zhou @ UCLA, Weixin Cai @ UC Berkeley
- thanks:
  - NSF DMS-1310319
  - Hernando Ombao, Daniel Rowe

Motivation

- scientific background:
  - Alzheimer's disease (AD) and normal aging
  - amyloid beta (Aβ) is a form of protein that is toxic to neurons in the brain, and it accumulates outside neurons and forms sticky buildup called Aβ plaques
  - Aβ plaques destroy synapses, i.e., contact points via which nerve cells relay signals to one another, and eventually lead to nerve cell death
  - Aβ plaques are the hallmark neuropathology markers of Alzheimer's disease (AD), and are also commonly found in elderly normal controls
  - previous studies have demonstrated that brain networks degrade among AD subjects

Motivation

- Berkeley Aging Cohort (BAC):
  - Aβ deposition was measured using Pittsburgh compound-B positron emission tomography (PIB-PET) imaging
  - $n = 140$ cognitively normal elder subjects
    - a continuous measure for each subject (Box-Cox transformation)
    - a binary measure: dichotomized into two groups, Aβ negative (111), Aβ positive (29)
  - brain connectivity network was measured by resting-state functional magnetic resonance imaging (rs-fMRI)
    - **preprocessed**
    - Freesurfer Desikan-Killany atlas: $p = 80$ regions-of-interest
    - TR = 1.89 sec, temporal dimension $q = 256$ time points
      TR = 2.20 sec, temporal dimension $q = 187$ time points
  - additional covariates: age, gender, education

Motivation

- broad question of interest:
  how Aβ deposition are related to brain connectivity patterns in cognitively normal elder subjects

Motivation

- broad question of interest:
  how Aβ deposition are related to brain connectivity patterns in cognitively normal elder subjects
  - how brain networks differ between the Aβ negative group and Aβ positive group
  - how to quantify the statistical significance of such difference
  - how brain networks relate to Aβ deposition in cognitively normal elder subjects

Motivation

- broad question of interest:
  how Aβ deposition are related to brain connectivity patterns in cognitively normal elder subjects
  - how brain networks differ between the Aβ negative group and Aβ positive group
  - how to quantify the statistical significance of such difference
  - how brain networks relate to Aβ deposition in cognitively normal elder subjects
- some possible formulations to tackle the problem
  - **estimation**: estimate multiple connectivity networks for the Aβ negative and positive groups
  - **inference**: quantify the statistical significance of comparing two connectivity networks
  - **association modeling**: regress the (binary / continuous) Aβ deposition measure on the connectivity network plus additional covariates

Symmetric tensor predictor regression

- association modeling:
  - extends from **tensor predictor regression** (Zhou et al., 2013)
  - fits a regression with Aβ deposition as the response (binary or continuous), the **symmetric, connectivity matrix** that describes the brain connectivity network as the predictor
  - has easy interpretation of the effect of individual links between brain regions on the phenotype
  - works with binary or continuous connectivity network (e.g., correlation or thresholded correlation matrix)
  - permits individual variation of functional connectivity
  - permits inference at the individual level, so potentially useful clinically
  - takes any connectivity matrix as input, both in time domain and frequency domain: **correlation, partial correlation, mutual information, partial mutual information**

Symmetric tensor predictor regression

- association modeling:
  - extends from **tensor predictor regression** (Zhou et al., 2013)
  - fits a regression with Aβ deposition as the response (binary or continuous), the **symmetric, connectivity matrix** that describes the brain connectivity network as the predictor
  - has easy interpretation of the effect of individual links between brain regions on the phenotype
  - works with binary or continuous connectivity network (e.g., correlation or thresholded correlation matrix)
  - permits individual variation of functional connectivity
  - permits inference at the individual level, so potentially useful clinically
  - takes any connectivity matrix as input, both in time domain and frequency domain: **correlation, partial correlation, mutual information, partial mutual information**
  - is applicable to applications beyond neuroimaging; e.g., in genetic epistasis studies, where D-way gene interactions can be formulated as an order-D symmetric tensor

Model

- notations:
  - Y = univariate response; e.g., continuous or binary Aβ deposition
  - $\boldsymbol{Z} \in \mathbb{R}^q$ = additional covariate vector containing age, gender, education
  - $\boldsymbol{X} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$ = order-D tensor-valued predictor; e.g., $D = 2$ for connectivity matrix, $D = 2, 3$ for two-way, or three-way interactions

- consider a generalized linear model (GLM) with a link function:

$$g(\mu) = \alpha + \gamma^\mathsf{T} \boldsymbol{Z} + \langle \mathcal{B}, \boldsymbol{X} \rangle$$

  - $\mu = E(Y | \boldsymbol{X}, \boldsymbol{Z})$
  - the inner product $\langle \mathcal{B}, \boldsymbol{X} \rangle = \langle \mathrm{vec}\mathcal{B}, \mathrm{vec}\boldsymbol{X} \rangle$
  - this model is prohibitive, **if no further constraint**, as the number of parameters is $1 + p_0 + \prod_{d=1}^{D} p_d$; e.g., $p = 80 \rightarrow 6,400$; $p = 1,000 \rightarrow 10^6$ for 2-way interactions

Model

- key idea: impose a low rank decomposition of $\mathcal{B}$
  - an array $\mathcal{B} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$ admits a **rank-R CP decomposition** if

$$\mathcal{B} = \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \cdots \circ \boldsymbol{\beta}_D^{(r)} = [\![\boldsymbol{B}_1, \ldots, \boldsymbol{B}_D]\!]$$

where $\boldsymbol{\beta}_d^{(r)} \in \mathbb{R}^{p_d}, d = 1, \ldots, D, r = 1, \ldots, R$, are all column vectors, $\circ$ denotes an outer product, and $\boldsymbol{B}_d = [\boldsymbol{\beta}_d^{(1)} \ldots \boldsymbol{\beta}_d^{(R)}] \in \mathbb{R}^{p_d \times R}$

Model

- key idea: impose a low rank decomposition of $\mathcal{B}$
  - an array $\mathcal{B} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$ admits a rank-R CP decomposition if

$$\mathcal{B} = \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \cdots \circ \boldsymbol{\beta}_D^{(r)} = [\![\boldsymbol{B}_1, \ldots, \boldsymbol{B}_D]\!]$$

where $\boldsymbol{\beta}_d^{(r)} \in \mathbb{R}^{p_d}, d = 1, \ldots, D, r = 1, \ldots, R$, are all column vectors, $\circ$ denotes an outer product, and $\boldsymbol{B}_d = [\boldsymbol{\beta}_d^{(1)} \ldots \boldsymbol{\beta}_d^{(R)}] \in \mathbb{R}^{p_d \times R}$

  - for $D = 2,\ R = 1,\ \mathcal{B} = [\![\boldsymbol{B}_1, \boldsymbol{B}_2]\!], \boldsymbol{B}_1 = \boldsymbol{\beta}_1, \boldsymbol{B}_2 = \boldsymbol{\beta}_2,$

$$\mathcal{B} = \boldsymbol{\beta}_1 \circ \boldsymbol{\beta}_2$$

  - for $D = 2,\ R = 2,\ \mathcal{B} = [\![\boldsymbol{B}_1, \boldsymbol{B}_2]\!], \boldsymbol{B}_1 = [\boldsymbol{\beta}_1^{(1)}, \boldsymbol{\beta}_1^{(2)}], \boldsymbol{B}_2 = [\boldsymbol{\beta}_2^{(1)}, \boldsymbol{\beta}_2^{(2)}],$

$$\mathcal{B} = \boldsymbol{\beta}_1^{(1)} \circ \boldsymbol{\beta}_2^{(1)} + \boldsymbol{\beta}_1^{(2)} \circ \boldsymbol{\beta}_2^{(2)}$$

Model

- CP tensor predictor regression:
  - the link function:

$$g(\mu) = \alpha + \boldsymbol{\gamma}^{\mathsf{T}} \boldsymbol{Z} + \langle \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \cdots \circ \boldsymbol{\beta}_D^{(r)}, \boldsymbol{X} \rangle$$

  - reduces the dimensionality from the order of $p_1 \times \ldots \times p_D$ to $R \times (p_1 + \ldots + p_D)$
  - **estimation**: a block-relaxation algorithm
    alternatively update $\boldsymbol{B}_d$, and each update is simply a standard GLM, because although $g(\mu)$ is not linear in $(\boldsymbol{B}_1, \ldots, \boldsymbol{B}_D)$ jointly, it is linear in $\boldsymbol{B}_d$ individually
  - **regularized estimation**: another block-relaxation algorithm
    each update is a penalized GLM

Model

- symmetric tensor predictor regression:
  - if $\boldsymbol{X}$ is a symmetric tensor, then $\boldsymbol{B}$ should be symmetric too, i.e.,

$$\boldsymbol{B} = \sum_{r=1}^{R} \lambda_r \, \boldsymbol{\beta}^{(r)} \circ \cdots \circ \boldsymbol{\beta}^{(r)} = [\![\boldsymbol{\lambda}; \boldsymbol{B}, \ldots, \boldsymbol{B}]\!]$$

    where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_R)^{\mathsf{T}}, \boldsymbol{B} \in \mathbb{R}^{p \times R}$
  - the link function:

$$g(\mu) = \alpha + \boldsymbol{\gamma}^{\mathsf{T}} \boldsymbol{Z} + \langle \sum_{r=1}^{R} \lambda_r \, \boldsymbol{\beta}^{(r)} \circ \cdots \circ \boldsymbol{\beta}^{(r)}, \boldsymbol{X} \rangle$$

  - reduces the dimensionality further from the order of RDp to $R(p+1)$
  - estimation: can **not** apply the block-relaxation algorithm!
  - in addition, plan to add sparsity regularization

Optimization

- solve the sparsity regularized estimation:

$$\min \ell(\gamma, \boldsymbol{\lambda}, \boldsymbol{B}) + \rho \|\mathrm{vec}\boldsymbol{B}\|_1$$

  - update of γ given $\boldsymbol{\lambda}$ and $\boldsymbol{B}$: a classical GLM with offset $\langle \mathcal{B}, \boldsymbol{X}_i \rangle$
  - update $\boldsymbol{\lambda}$ given γ and $\boldsymbol{B}$: a GLM with R-dimensional covariates $(\mathrm{vec}\, \boldsymbol{X}_i)^{\mathsf{T}}(\boldsymbol{B} \odot \cdots \odot \boldsymbol{B})$ and offset $\gamma^{\mathsf{T}} \boldsymbol{Z}_i$, because

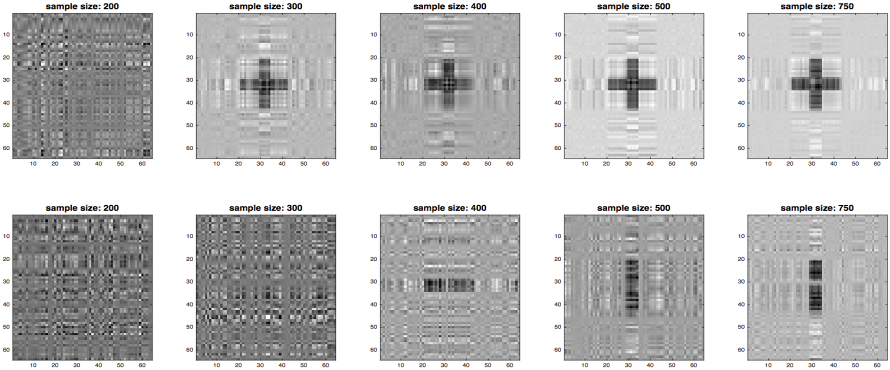
$$\langle \mathcal{B}, \boldsymbol{X}_i \rangle = \langle \mathrm{vec}\, \mathcal{B}, \mathrm{vec}\, \boldsymbol{X}_i \rangle = (\mathrm{vec}\, \boldsymbol{X}_i)^T (\boldsymbol{B} \odot \cdots \odot \boldsymbol{B}) \boldsymbol{\lambda}$$

  - update $\boldsymbol{B}$ given γ and $\boldsymbol{\lambda}$: **the proximal gradient method**
    the surrogate function s to minimize is the first-order approximation to the objective function at the current point $\boldsymbol{B}^{(t)}$

$$\begin{aligned} s(\boldsymbol{B} \mid \boldsymbol{B}^{(t)}, \delta) &= \ell(\boldsymbol{B}^{(t)}) + \langle \nabla \ell(\boldsymbol{B}^{(t)}), \boldsymbol{B} - \boldsymbol{B}^{(t)} \rangle + \frac{1}{2\delta} \|\boldsymbol{B} - \boldsymbol{B}^{(t)}\|_{\mathsf{F}}^2 + \rho \|\mathrm{vec}\boldsymbol{B}\|_1 \\ &= \frac{1}{2\delta} \|\boldsymbol{B} - \{\boldsymbol{B}^{(t)} - \delta \nabla \ell(\boldsymbol{B}^{(t)})\}\|_{\mathsf{F}}^2 + \rho \|\mathrm{vec}\boldsymbol{B}\|_1 \end{aligned}$$

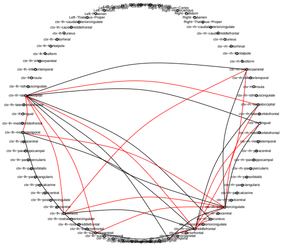
    s is minimized by soft-thresholding $\boldsymbol{B}^{(t)} - \delta \nabla \ell(\boldsymbol{B}^{(t)})$ at threshold $\rho\delta$

Simulation

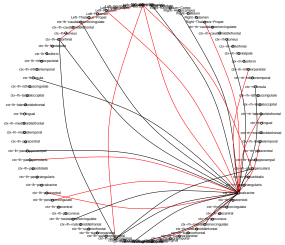

sample size: 200
sample size: 300
sample size: 400
sample size: 500
sample size: 750
sample size: 200
sample size: 300
sample size: 400
sample size: 500
sample size: 750
10 20 30 40 50 60

BAC data analysis: continuous response

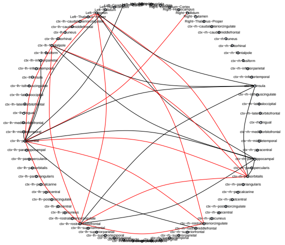
Pearson correlation



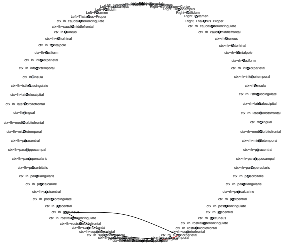
Partial correlation



Mutual information

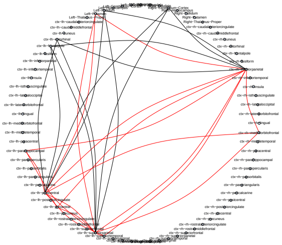


Partial mutual information

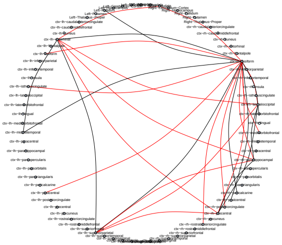


BAC data analysis: binary response

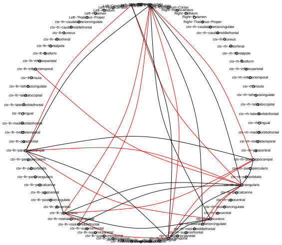
Pearson correlation



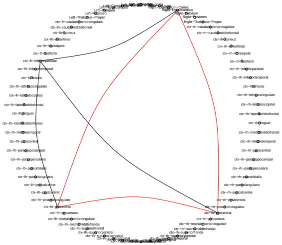
Partial correlation



Mutual information



Partial mutual information



BAC data analysis

- some observations:
  - negative links (red) suggests that, having this link decreases the chance to be Aβ positive, or lower Aβ value — another way to look at this is, it is more likely that this link would disappear in the Aβ positive group compared to the Aβ negative group
  - the difference of connectivity patterns of cognitive normal elder subjects between Aβ positive group and Aβ negative group are similar to that between AD and normal control
  - the four connectivity measures have overlapping findings and do not contradict to each other
  - the findings from a continuous response overlap with those from a binary response

BAC data analysis: continuous response

Links	Pearson correlation	Partial correlation	Mutual information	Partial mutual information	Findings
Negative	precuneus — posteriorcingulate	pericalcarine — amygdala, posteriorcingulate, middletemporal			‡ Decrease in connection between posterior cingulate cortex/precuneus and medial prefrontal cortex, hippocampus (Bluhm et al., 2008)
	middletemporal — posteriorcingulate				‡ Decrease in connectivity inside posterior cingulate cortex/precuneus (Bluhm et al., 2008)
		supramarginal — amygdala		supramarginal — superiorparietal	‡ AD affected superior occipital, supramarginal, superior temporal, inferior parietal, angular and inferior frontal gyri, putamen, thalamus and posterior cerebellum (Sidlauskaite et al., 2015); ‡ Decrease between the auditory network and temporal gyrus, supramarginal gyrus, and post-central gyrus. (Hafkemeijer et al., 2015)
			rostralanterior cingulate — paracentral		‡ AD group showed lower proportion of fibers in the rostral anterior cingulate (Daianu, 2013)
Positive	middletemporal — precuneus		parahippocampal — paracentral	precuneus — supramarginal, superiorparietal	‡ Unknown

BAC data analysis: binary response

Links	Pearson correlation	Partial correlation	Mutual information	Partial mutual information	Findings
Negative	inferior parietal — superior parietal, pre-/post-central, parahippocamal, medial orbital frontal		precuneus — superior-temporal, amygdala		‡ Clinically normal older adults harboring amyloid burden show disruption of functional connectivity in default network (posterior cingulate, lateral parietal, and medial prefrontal cortices) that cannot be accounted for by increased age or structural atrophy. (Hedden, 2009); ‡ Decrease in connection between posterior cingulate cortex/precuneus and medial prefrontal cortex, hippocampus (Bluhm et al., 2008)
		precentral — superior-parietal	parstriangularis — parahippocampal		‡ Decrease in connection between back of brain and frontal region in general (Meunier et al., 2009)
		fusiform — posterior-/anterior cingulate	parahippocampal — superior-temporal, amygdala, precuneus	hippocampus — pre-central, left & right	‡ RSFC between the hippocampus and the posterior cingulate cortex was found to be positively correlated with performance on a memory task (Wang et al., 2010)
	inferior parietal — putamen				Unknown
Positive	frontal pole — interior & superior parietal, post-central	middle temporal — entorhinal		inferiorparietal — pre-central, hippocampus,	Unknown

Estimation of brain networks

- **estimation** of multiple partial correlation matrices:
  - adopts **matrix normal** distribution and Gaussian graphical modeling characterizes brain connectivity network through **partial correlation matrix**
  - imposes both a sparsity penalty and a group sparsity principle, under the belief that the connectivity network is **sparse**, and the **difference** of connectivity networks across groups is **sparse**
  - employs a **non-convex** sparsity penalty and a **non-convex** group sparsity penalty
  - develops a highly scalable optimization algorithm: sequential convex relaxation via MM + solve each relaxation via ADMM
  - establishes **sparsistency** for the actual optimizer of our estimation algorithm, under a relatively weak set of regularity conditions
  - shows numerical advantage compared to a convex alternative

Estimation of brain networks

- estimation in a nutshell:
  - matrix normal distribution:

$$\boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{kn_k} \overset{\text{i.i.d.}}{\sim} N(\boldsymbol{M}_k, \boldsymbol{\Sigma}_{L_k} \otimes \boldsymbol{\Sigma}_{T_k}), \quad k = 1, \cdots, K$$

  - of interest: spatial partial correlation matrix

$$\boldsymbol{\Omega}_k = \mathrm{Diag}(\boldsymbol{\Sigma}_{L_k}^{-1/2}) \boldsymbol{\Sigma}_{L_k}^{-1} \mathrm{Diag}(\boldsymbol{\Sigma}_{L_k}^{-1/2})$$

  - penalized optimization:

$$\sum_{k=1}^{K} n_k \left\{ \mathrm{trace}(\boldsymbol{\Omega}_k \widehat{\boldsymbol{\Gamma}}_k) - \log\det(\boldsymbol{\Omega}_k) \right\} + \sum_{k=1}^{K} n_k \sum_{i \neq j} p_{\lambda_1}\left(|\omega_{kij}|\right) + n \sum_{i \neq j} p_{\lambda_2}\left(\sqrt{\omega_{1ij}^2 + \cdots \omega_{Kij}^2}\right)$$

    - $p_\lambda(x)$ is nondecreasing and differentiable on $\mathbb{R}^+$ and $p_\lambda(0) = 0$
    - $\lim_{x \to 0^+} p'_\lambda(x) = \lambda$
    - $p_\lambda(x) + \frac{1}{b}x^2$ is convex for some constant $b > 0$
    - $p'_\lambda(x) = 0$ for $|x| > a\lambda$ for some constant $a \geq \frac{b}{2}$

Inference of brain networks

- **hypothesis testing** of comparing two partial correlation matrices:
  - from **estimation** to **inference**: statistical significance quantification
  - adopts **matrix normal** distribution and Gaussian graphical modeling characterizes brain connectivity network through **partial correlation matrix**
  - builds the test statistics based on a regression representation of partial correlations
  - proposes a **global testing** procedure and an **entry(link)-wise testing** procedure with FDR control
  - for the global test, derives the asymptotic distribution of the test statistic under the null distribution, obtains the power of the test under the sparse alternative, and shows that it is minimax rate optimal
  - for the multiple testing procedure, shows that it controls the false discovery proportion and false discovery rate at a pre-specified level asymptotically

Inference of brain networks

- inference in a nutshell:
  - matrix normal distribution:

$$\begin{aligned} \mathrm{cov}\{\mathrm{vec}(\boldsymbol{X}_{ki})\} &= \boldsymbol{\Sigma}_{L_k} \otimes \boldsymbol{\Sigma}_{T_k}, \quad k = 1, 2, i = 1, \ldots, n_i \\ \mathrm{cov}^{-1}\{\mathrm{vec}(\boldsymbol{X}_{ki})\} &= \boldsymbol{\Sigma}_{L_k}^{-1} \otimes \boldsymbol{\Sigma}_{T_k}^{-1}, \quad k = 1, 2, i = 1, \ldots, n_i \end{aligned}$$

  - of interest: spatial partial correlation matrix

$$\boldsymbol{\Omega}_{L_k} = \mathrm{Diag}(\boldsymbol{\Sigma}_{L_k}^{-1/2})\boldsymbol{\Sigma}_{L_k}^{-1}\mathrm{Diag}(\boldsymbol{\Sigma}_{L_k}^{-1/2})$$

  - hypotheses:

$$\begin{aligned} \text{global test:} &\quad \boldsymbol{\Omega}_{L_1} = \boldsymbol{\Omega}_{L_2} \ \text{ versus } \ \boldsymbol{\Omega}_{L_1} \neq \boldsymbol{\Omega}_{L_2} \\ \text{entry-wise test:} &\quad \omega_{L_1,i,j} = \omega_{L_2,i,j} \ \text{ versus } \ \omega_{L_1,i,j} \neq \omega_{L_2,i,j} \end{aligned}$$

  - treat the temporal precision matrix as a nuisance: known and estimated

Thank You!