Testing Mediation Effects using Logic of Boolean Matrices (LOGAN)

Chengchun Shi
Assistant Professor of Data Science
London School of Economics and Political Science

Our Proposal: LOGAN

Papers and Acknowledgement

- **Papers**:
  - Paper 1: **Testing Mediation Effects Using Logic of Boolean Matrices** (*JASA*, 2021, accepted)
  - Paper 2: **Sequential Pathway Inference for Multimodal Neuroimaging Analysis** (*STAT*, 2022)

- **Thanks**:
  - Lexin Li @ UC Berkeley & Helen Wills Neuroscience Institute
  - William Jagust @ UC Berkeley & Helen Wills Neuroscience Institute
  - Tengfeil Guo @ UC Berkeley & Helen Wills Neuroscience Institute

General Overview

- **Neuroimaging analysis** is a super exciting area, because
  - scientifically, understanding the inner working of human brains, and their connections with numerous neurological disorders, e.g, Alzheimer's disease, as well as normal aging, is one of the most intriguing questions
  - statistically, an array of **diverse** statistical problems, constantly calling for new models, theory, algorithms
  - large public neuroimaging databases are becoming available
  - this area is not overly crowded, yet

- A selection of neuroimaging problems:
  - imaging tensor analysis
  - brain connectivity network analysis
  - multimodal neuroimaging analysis
  - new imaging modalities: functional data analysis; ordinary differential equations; point process modeling

Scientific Motivation

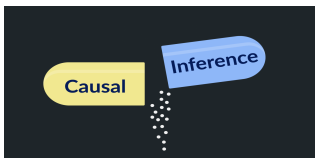
- Alzheimer's disease (AD) and normal aging:
  - AD is an irreversible neurodegenerative disorder, characterized by progressive impairment of cognitive and memory functions, then loss of independent living, and ultimately death
  - the leading form of dementia, and currently affecting 5.8 million American adults aged 65 years or older
  - prevalence continues to grow; projected to reach 13.8 million by 2050
  - there is no effective treatment
- **Scientific questions of interest**:
  - neurodegeneration measure, often captured as grey matter cortical atrophy, is a well-known biomarker associated with AD
  - amyloid-beta and tau are two hallmark pathological proteins believed to be part of the driving mechanism of AD
  - **question**: how age affects cortical thickness then cognitive outcome
  - **question**: how amyloid-beta affects tau deposition then cortical thickness then cognitive outcome

Mediation Analysis

- Widely used in a number of applications, e.g., economics, social science, medicine, neuroimaging, genetics and machine learning

(a) Neuroimaging

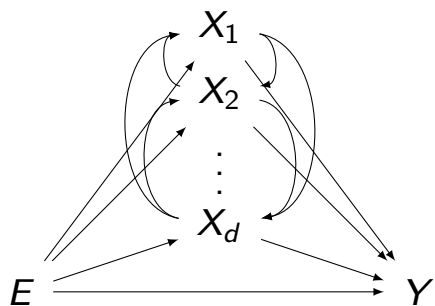
(b) Genetics



(c) Medicine

Mediation Analysis (Cont'd)

- to identify and explain the mechanism, or pathway, that underlies an observed relationship between an **exposure** and an **outcome** variable, through the inclusion of an intermediary variable, known as a **mediator**
- facilitate a better understanding of the exposure-outcome mechanism



Inference for Mediation Analysis

- **Inference** for **high-dimensional** mediation analysis:
  - question: how to infer the **significance** of individual mediators?
  - Existing solutions explicitly impose that the mediators are **conditionally independent** given the exposure, ignoring potential paths among the mediators
  - plausible in some applications, but not in others
  - in **neuroimaging**, different brain regions influence each other
  - in **genetics**, different genes interact with each other
  - challenge: the number of possible paths that go through all combinations of mediators is huge $\rightarrow$ the total number of potential paths that go through any mediator is **super-exponential** in the number of mediators

- **Mediation estimation through sparse regularization**:
  - can in effect identify important mediators
  - but estimation does not explicitly quantify the significance (p-value), and does not control the false discovery

Inference for Mediation Analysis (Cont'd)

- Chakrabortty et al. (2018):
  - allowed **mediator-by-mediator interactions**
  - formulated the **directed acyclic graph** (DAG) structure
  - defined the **individual mediation effect** of a given mediator as the **summation** of all the effects of the exposure on the outcome that can be attributed to that mediator
  - established the convergence and confidence interval for their estimator
- Our proposal:
  - similarly adopt the DAG formulation and allow **mediator-by-mediator interactions**
  - defined the **individual mediation effect** of a given mediator as the **maximum** absolute value of all effects of the exposure on the outcome attributed to that mediator

What we propose (in a nutshell)

- a new testing procedure to evaluate the individual mediation effect
- **logic of Boolean matrices** $\rightarrow$ establish the proper limiting distribution under $\mathcal{H}_0$

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- **data splitting** $\rightarrow$ type-I error control

- **screening** $\rightarrow$ power enhancement

What we propose (Cont'd)

- **decorrelated estimator** $\rightarrow$ reduce bias induced by high-dimensional mediators

- **multiplier bootstrap** $\rightarrow$ obtain critical values
- **multiple testing** $\rightarrow$ FDR control
- establish the **asymptotic size, power, and FDR control**, while allowing the number of mediators to **diverge** to ∞

Gaussian Graphical Model

- **setup**: exposure E/X_0; multivariate mediators $X_1, \ldots, X_d$; outcome Y/X_{d+1}; write $\boldsymbol{X} = (E, X_1, \ldots, X_d, Y)^\top \in \mathbb{R}^{d+2}$
- **Gaussian graphical model**:

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{W}(\boldsymbol{X} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon},$$

  - $\boldsymbol{\mu} = E(\boldsymbol{X})$; $\boldsymbol{W} \in \mathbb{R}^{(d+2)\times(d+2)}$; $\boldsymbol{\varepsilon} = (\varepsilon_0, \ldots, \varepsilon_{d+1})^\top$
  - $\boldsymbol{W}$ specifies the directional relationships among the variables in $\boldsymbol{X}$, which can be encoded by a **DAG**
  - $X_i \to X_j$: X_i is called a parent of X_j, and X_j a child of X_i
  - $X_i \to X_{i_1} \to \ldots \to X_{i_{k-1}} \to X_j$ for some $\{i_k\}_{1 \le l < k}$: X_i is called an ancestor of X_j, and X_j a descendant of X_i.
  - X_0 is not the child of any mediator $X_1, \ldots, X_d$;
    X_{d+1} is not the parent of X_0 nor any mediator $X_1, \ldots, X_d$
  - the errors ε_i, $i = 0, \ldots, d+1$, are jointly normally distributed and independent, and the error variances $\sigma_i^2 = \mathrm{Var}(\varepsilon_i)$, $i = 0, \ldots, d+1$, are **constant** (Peters and Bühlmann, 2014, Yuan et al., 2019)

Hypotheses

- **Total effect**: for a directed **path** $\zeta : X_0 \to X_{i_1} \to \ldots \to X_{i_k} \to X_{d+1}$ for some $\{i_t\}_{1 \le t \le k} \subseteq \{1, \ldots, d\}$, the total effect of X_0 on X_{d+1} attributed to this path as

$$\omega_\zeta = W_{i_1,0} \left(\prod_{t=0}^{k-1} W_{i_{t+1},i_t} \right) W_{d+1,i_k},$$

  where $W_{i,j}$ is the (i,j)th entry of $\boldsymbol{W}$. If such a path does not exist, we have $\omega_\zeta = 0$.

- **Hypotheses**: for an individual **mediator** X_q, $q = 1, \ldots, d$,

$$\begin{aligned} \mathcal{H}_0(q) &: \omega_\zeta = 0, \quad \text{for all } \zeta \text{ that passes through } X_q, \\ \mathcal{H}_1(q) &: \omega_\zeta \neq 0, \quad \text{for some } \zeta \text{ that passes through } X_q. \end{aligned}$$

  when $\mathcal{H}_1(q)$ holds, we say X_q is a **significant mediator**

Hypotheses (Cont'd)

- **Equivalent hypotheses**:

$$\mathcal{H}_0(q) : 0 \notin \mathrm{ACT}(q, \boldsymbol{W}) \quad \text{or} \quad q \notin \mathrm{ACT}(d+1, \boldsymbol{W}),$$
$$\mathcal{H}_1(q) : 0 \in \mathrm{ACT}(q, \boldsymbol{W}) \quad \text{and} \quad q \in \mathrm{ACT}(d+1, \boldsymbol{W}).$$

  where $\mathrm{ACT}(j, \boldsymbol{W})$ denotes the set of true **ancestors** of X_j

- **Hypotheses we target**: for $q_1 = 0, \ldots, d$, $q_2 = 1, \ldots, d+1$,

$$\mathcal{H}_0(q_1, q_2) : q_1 \notin \mathrm{ACT}(q_2, \boldsymbol{W}),$$
$$\mathcal{H}_1(q_1, q_2) : q_1 \in \mathrm{ACT}(q_2, \boldsymbol{W}).$$

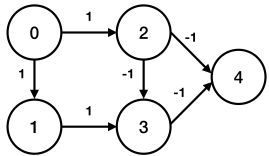
  - the null hypothesis $\mathcal{H}_0(q)$ can be **decomposed** into a **union** of the two null hypotheses $\mathcal{H}_0(0, q)$ and $\mathcal{H}_0(q, d+1)$
  - by the union-intersection principle, $\max\left\{p(0, q), p(q, d+1)\right\}$ is a valid p-value for testing $\mathcal{H}_0(q)$

Hypotheses (Cont'd)

- **Alternative definition** of a significant mediator (Chakrabortty et al., 2018):

$$\mathcal{H}_0^*(q): \sum \omega_\zeta = 0, \quad \text{versus} \quad \mathcal{H}_1^*(q): \sum \omega_\zeta \neq 0,$$

where the summation is taken for all ζ that pass through X_q



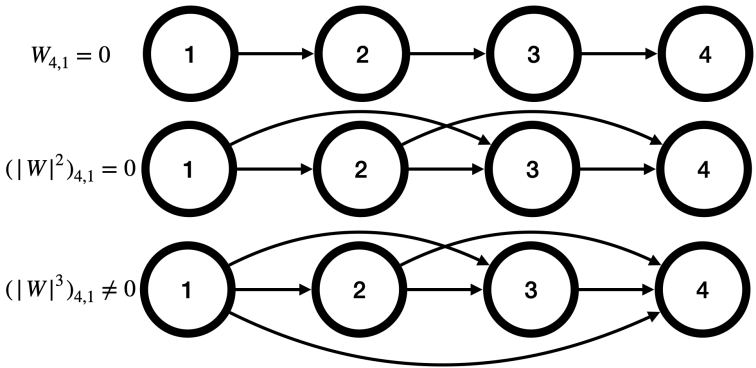
  - the effects along the path ζ may **cancel out** with each other, resulting in a zero sum, even though there are significant positive and negative mediation effects along ζ
  - e.g., for X_2, two paths, $X_0 \to X_2 \to X_4$ and $X_0 \to X_2 \to X_3 \to X_4$, both pass through X_2, while the aggregated total effect is $\sum_\zeta \omega_\zeta = 1 \times \{-1 + (-1) \times (-1)\} = 0$

Key Observation

- **Power of matrices**:

$$\mathcal{H}_0(q_1, q_2) \text{ holds if and only if } (|\boldsymbol{W}|^k)_{q_2,q_1} = 0, \text{ for any } k = 1, \ldots, d.$$

- **Example**:



Test Statistics

- **Key observation**:

$$\mathcal{H}_0(q_1, q_2) \text{ holds if and only if } (|\boldsymbol{W}|^k)_{q_2,q_1} = 0, \text{ for any } k = 1, \ldots, d.$$

- a natural test statistic is $\{(|\widehat{\boldsymbol{W}}|^k)_{q_2,q_1}\}_{1 \le k \le d}$, where $\widehat{\boldsymbol{W}}$ is some consistent estimator for $\boldsymbol{W}$
- however, it is difficult to obtain the limiting distribution of $(|\widehat{\boldsymbol{W}}|^k)_{q_2,q_1}$ under $\mathcal{H}_0(q_1, q_2)$

Logic of Boolean Matrices

- for two real-valued matrices $\boldsymbol{A}_1 = \{a_{1,i,j}\}_{ij} \in \mathbb{R}^{q_1 \times q_2}$, $\boldsymbol{A}_2 = \{a_{2,i,j}\}_{ij} \in \mathbb{R}^{q_2 \times q_3}$
- define a **new matrix multiplication operator** and **a new matrix addition operator** to replace the usual matrix multiplication and addition
- define $\boldsymbol{A}_1 \otimes \boldsymbol{A}_2$ to be a $q_1 \times q_3$ matrix whose (i,j)th entry equals $\max_{k \in \{1,\cdots,q_2\}} \min(a_{1,i,k}, a_{2,k,j}) \rightarrow$ replace the multiplication operation in the usual matrix multiplication with the minimum operator, and replace the addition operation with the maximum operator
- define $\boldsymbol{A}_1 \oplus \boldsymbol{A}_2$ to be a $q_1 \times q_2$ matrix whose (i,j)th entry equals $\max(a_{1,i,j}, a_{2,i,j})$
- when $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ are binary matrices, the minimum and maximum operators are equivalent to the logic operators "and" and "or" in Boolean algebra
- when $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ are binary matrices, "$\otimes$" operator is equivalent to the Boolean matrix multiplication operator
- when $\boldsymbol{A}_1, \boldsymbol{A}_2$ are binary matrices, "$\oplus$" operator is equivalent to the Boolean matrix addition operator

Logic of Boolean Matrices (Cont'd)

- **Key observation**:

$$\mathcal{H}_0(q_1, q_2) \text{ holds if and only if } (|\boldsymbol{W}|^{(k)})_{q_2,q_1} = 0, \text{ for any } k = 1, \ldots, d.$$

- **Aggregating** $|\boldsymbol{W}|^{(k)}$ for all k-**step paths**, $k = 1, \ldots, d$,

$$\boldsymbol{W}^* = |\boldsymbol{W}| \oplus |\boldsymbol{W}|^{(2)} \oplus \cdots \oplus |\boldsymbol{W}|^{(d)}.$$

  $\mathcal{H}_0(q_1, q_2)$ holds if and only if $(\boldsymbol{W}_0^*)_{q_2,q_1} = 0$

- Test statistic: $\widehat{\boldsymbol{W}}^*_{q_2,q_1}$ for $\mathcal{H}_0(q_1, q_2)$, where $\widehat{\boldsymbol{W}}$ is some consistent estimator for $\boldsymbol{W}$
- Its limiting distribution under $\mathcal{H}_0$ is stochastically smaller than the maximum of certain Gaussian variables whose distribution can be well-approximated via **high-dimensional Gaussian bootstrap** (Chernozhukov et al., 2014)

Step 1: Data Splitting

- split the data into two equal halves $\{\boldsymbol{x}_i\}_{i\in\mathcal{I}_1} \cup \{\boldsymbol{x}_i\}_{i\in\mathcal{I}_2}$, where $\mathcal{I}_\ell$ is the set of indices of subsamples, $\ell = 1, 2$
- ensure the resulting test achieves a valid **type-I error rate** under minimal conditions
- commonly used in statistical testing (Romano and DiCiccio, 2019)
- construct two test statistics based on both halves of data, then combine them
- can also do **multiple** splits, at the cost of heavier computations

Step 2: Initial DAG Estimation

- compute an initial estimator $\widetilde{\boldsymbol{W}}^{(\ell)}$ for $\boldsymbol{W}_0$, given each half of the data $\{\boldsymbol{x}_i\}_{i\in\mathcal{I}_\ell}$, $\ell = 1, 2$
- several choices: **Zheng et al. (2018)**; Yuan et al. (2019)
- a novel characterization of the acyclic constraint:

$$\widetilde{\boldsymbol{W}}^{(\ell)} = \operatorname{argmin}_{\boldsymbol{W}\in\mathbb{R}^{(d+2)\times(d+2)}} \sum_{i\in\mathcal{I}_\ell} \|\widetilde{\boldsymbol{x}}_i - \boldsymbol{W}\widetilde{\boldsymbol{x}}_i\|_2^2 + \lambda|\mathcal{I}_\ell| \sum_{i,j} |W_{i,j}|$$
$$\text{subject to } \operatorname{trace}\{\exp(\boldsymbol{W}\circ\boldsymbol{W})\} = d+2.$$

- only require $\widetilde{\boldsymbol{W}}^{(\ell)}$ to be **consistent** to $\boldsymbol{W}_0$; considerably weaker than requiring $\widetilde{\boldsymbol{W}}^{(\ell)}$ to be **selection consistent**; i.e., $\mathbb{I}(\widetilde{W}^{(l)}_{i,j} = 0) = \mathbb{I}(W_{0,i,j} = 0)$ for any $i, j = 0, \ldots, d+1$
- we establish the properties of $\widetilde{\boldsymbol{W}}^{(\ell)}$, which is not available in Zheng et al. (2018)

Steps 3 & 4

- **Step 3: screening**
  - compute the binary matrix $\widehat{\boldsymbol{B}}^{(\ell)}$ given the initial estimator $\widetilde{\boldsymbol{W}}^{(\ell)}$
  - use the nonzero entries of $\widehat{\boldsymbol{B}}^{(\ell)}$ to determine the support of the subsequent decorrelated estimation step
  - bring down the number of potential paths to a moderate level $\rightarrow$ reduce the variance of the test statistic $\rightarrow$ enhance the power of the test

- **Step 4: decorrelated estimation of $\boldsymbol{W}$ using cross-fitting**
  - use one set of samples $\mathcal{I}_\ell$ to obtain the initial estimator $\widetilde{\boldsymbol{W}}^{(\ell)}$ and $\widehat{\boldsymbol{B}}^{(\ell)}$, then use the other set of samples $\mathcal{I}_\ell^c$ to compute the entries of the decorrelated estimator $\widehat{\boldsymbol{W}}^{(\ell)}$
  - reduce the bias of $\widetilde{\boldsymbol{W}}^{(\ell)}$ under the setting of high-dimensional mediators
  - guarantee the entry of $\widehat{\boldsymbol{W}}^{(\ell)}$ is $\sqrt{n}$-consistent and asymptotically normal

Step 5: Multiplier Bootstrap

- for the test statistic:

$$\sqrt{|\mathcal{I}_\ell^c|}(\widehat{\boldsymbol{W}}^{*(\ell)})_{q_1,q_2} \leq \max_{(i,j)\in\mathcal{S}} \sqrt{|\mathcal{I}_\ell^c|}\,|\widehat{W}_{i,j}^{(\ell)} - W_{0,i,j}|,$$

  where $\mathcal{S}$ denotes the screening set.

- The right-hand-side is to converge to a **maximum of Gaussian**, whose limiting distribution can be obtained via high-dimensional Gaussian multiplier (Chernozhukov et al., 2014).

Finally...

- **Decision making**:
  - reject the null $\mathcal{H}_0(q)$ when $\mathcal{H}_0(0, q)$ and $\mathcal{H}_0(q, d+1)$ are both rejected
  - for each half of the data $\ell = 1, 2$, we have made a decision $\mathcal{D}^{(\ell)}$ regarding $\mathcal{H}_0(q) \rightarrow$ we reject $\mathcal{H}_0(q)$ when either $\mathcal{D}^{(1)}$ or $\mathcal{D}^{(2)}$ decides to reject $\rightarrow$ by Bonferroni's inequality, this yields a valid α-level test

- **Multiple testing**:
  - adopt the **ScreenMin** procedure of Djordjilović et al. (2019) for multiple testing and false discovery control

Theory

- **Asymptotic size**:

$$\mathbb{P}\big\{\boldsymbol{\mathcal{H}}_0(q) \text{ is rejected} \mid \boldsymbol{\mathcal{H}}_0(q) \text{ holds}\big\} \leq \alpha + o(1).$$

- **Asymptotic power**:

$$\mathbb{P}\big\{\boldsymbol{\mathcal{H}}_0(q) \text{ is rejected} \mid \boldsymbol{\mathcal{H}}_1(q) \text{ holds}\big\} \to 1, \quad \text{as} \quad n \to \infty.$$

- **Asymptotic FDR control**:

$$\mathrm{FDR} \leq \alpha + o(1)$$

- **Consistency of the initial DAG estimator**:
  - the convergence rate of the initial DAG estimator $\widetilde{\boldsymbol{W}}^{(\ell)}$ obtained from Zheng et al. (2018) is the same as that of the oracle estimator

Simulation I

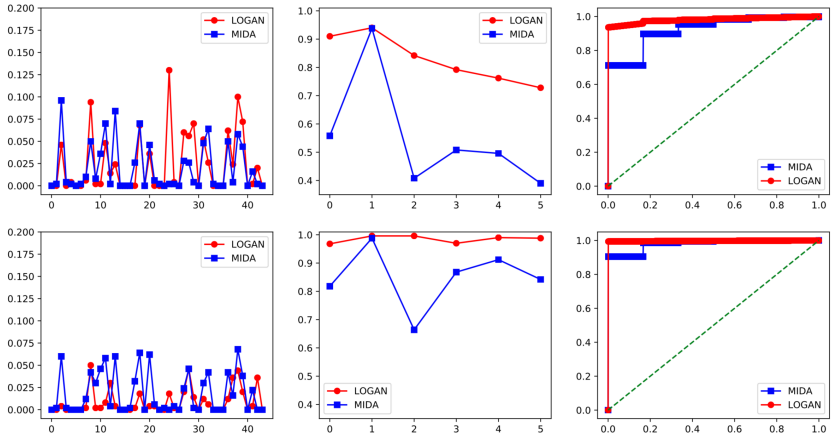


Figure: Empirical rejection rate and ROC curve of the proposed test, LOGAN, and the test of Chakrabortty et al. (2018), MIDA, when $d = 50$. The upper panels: $n = 100$, and the bottom panels: $n = 200$. The left panels: under $\mathcal{H}_0$, the middles panels: under $\mathcal{H}_1$, where the horizontal axis is the mediator index, and the right panels: the average ROC curve.

AD Case Study I

- Mediation inference:
  - **exposure**: age; **outcome**: PACC score; **mediators**: gray matter cortical thickness of $d = 68$ brain regions-of-interest (ROIs)
  - $n = 389$ subjects
  - set FDR level at 10%

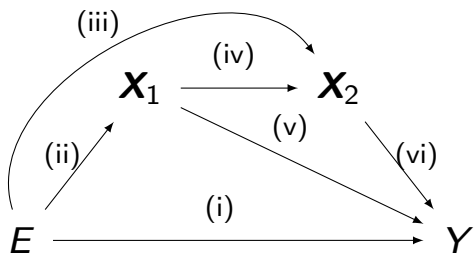
- Findings:

amyloid negative group	
l-entorhinal	l-precuneus
l-superiortemporal	r-inferiorparietal
r-superiorfrontal	r-superiortemporal

  - entorhinal cortex functions as a hub in a widespread network for memory, navigation and the perception of time; one of the most heavily damaged cortices in AD
  - precuneus is involved with episodic memory, visuospatial processing, reflections upon self, and aspects of consciousness, and is found to be an AD-signature region

Sequential Mediation Analysis

- **Question**: how amyloid-beta affects tau deposition then cortical thickness then cognitive outcome
- **Challenge**: **multiple sets** of mediators are **sequentially** ordered on the potential pathways following certain biological constraints



Simulation II

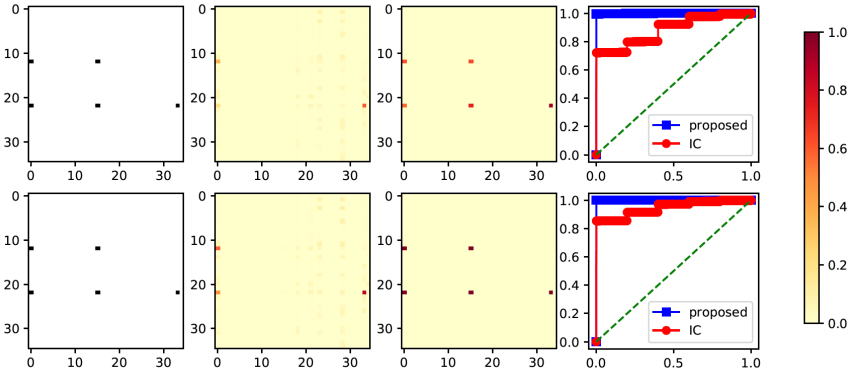


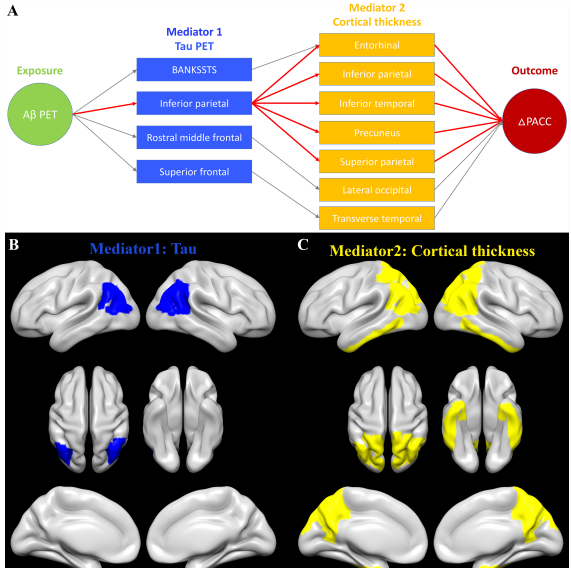
Figure: Empirical size and power when $d_1 = d_2 = 35$. First column: the vertical axis denotes the indices of the mediators in the first set, and the horizontal axis the second set. The black dots indicate the true significant mediator pairs. Second and third columns: the empirical rejection rate by the method of Chakrabortty et al. (2018), and our sequential test, respectively. Fourth column: the average ROC curve with a varying significance level. First row: $n = 200$, and second row: $n = 400$.

AD Case Study II

- mediation inference:
  - **exposure**: amyloid-beta;
    outcome: change of PACC score of two consecutive visits;
    mediator set 1: tau deposition of $d_1 = 35$ brain ROIs;
    mediator set 2: gray matter cortical thickness of $d_2 = 34$ brain ROIs
  - $n = 184$ subjects
  - set FDR level at 10%

AD Case Study II (Cont'd)

- findings:



Thank You!

☺Papers and softwares can be found on my personal website

`callmespring.github.io`