Covariate Assisted Principal (CAP) Regression for Matrix Outcomes

Xi (Rossi) **LUO**

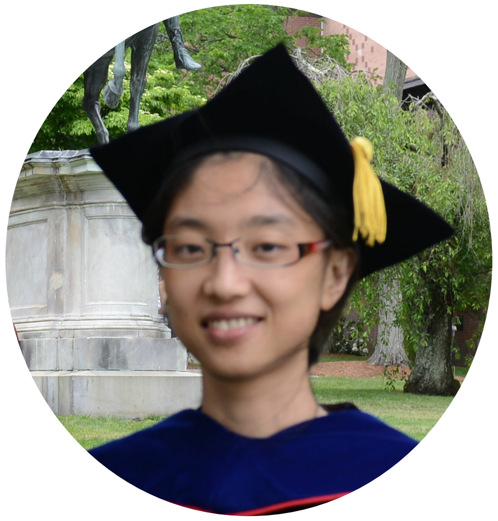
The University of Texas
Health Science Center
School of Public Health
Dept of Biostatistics
and Data Science
ABCD Research Group

BigComplexDATA.com

ICSA, Hangzhou, CHINA
December 20, 2019

Funding: NIH R01EB022911; NSF/DMS (BD2K) 1557467

Co-Authors



Yi Zhao
Indiana Univ

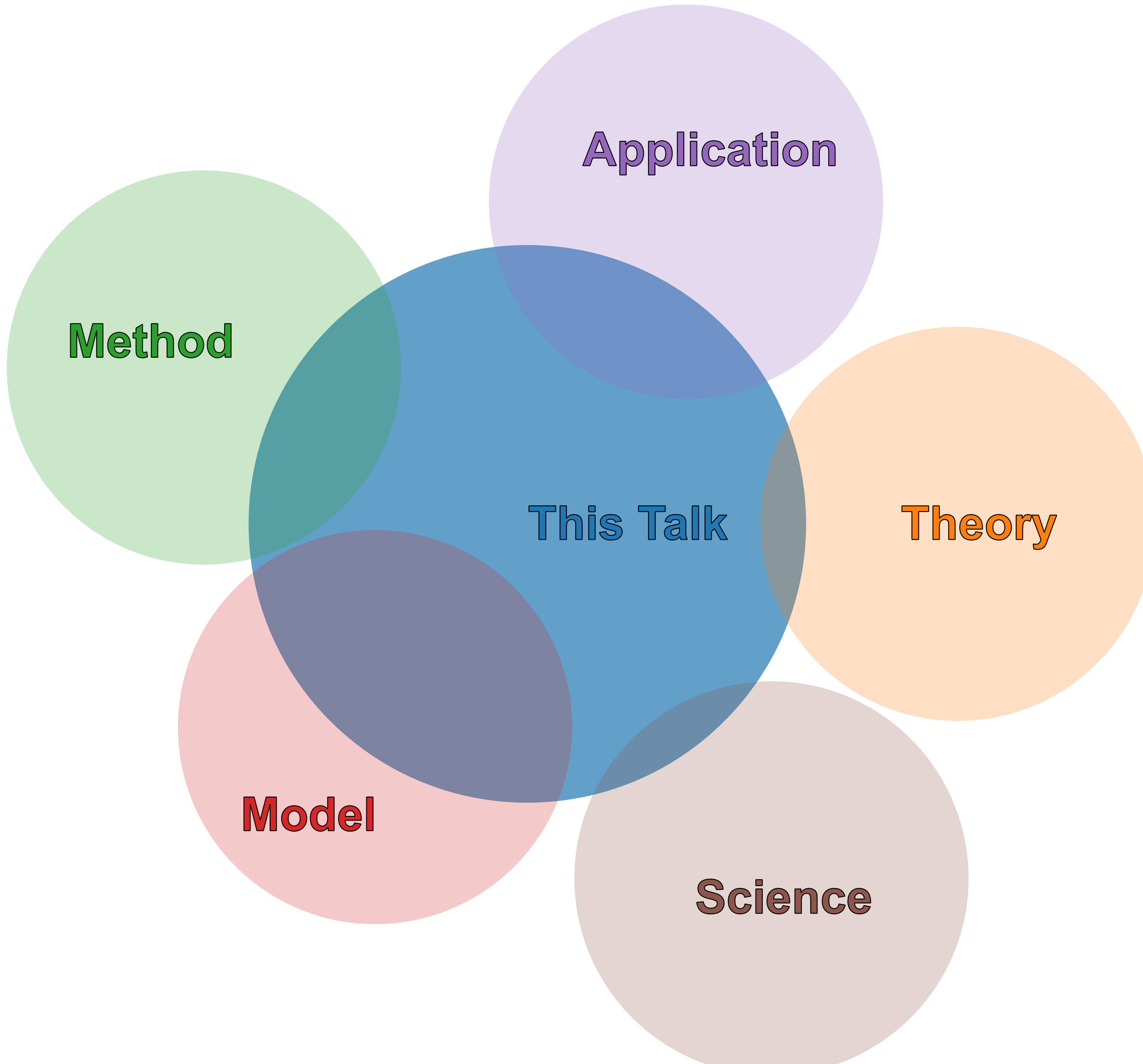
Bingkai Wang
Johns Hopkins Biostat

Stewart Mostofsky
Johns Hopkins Medicine

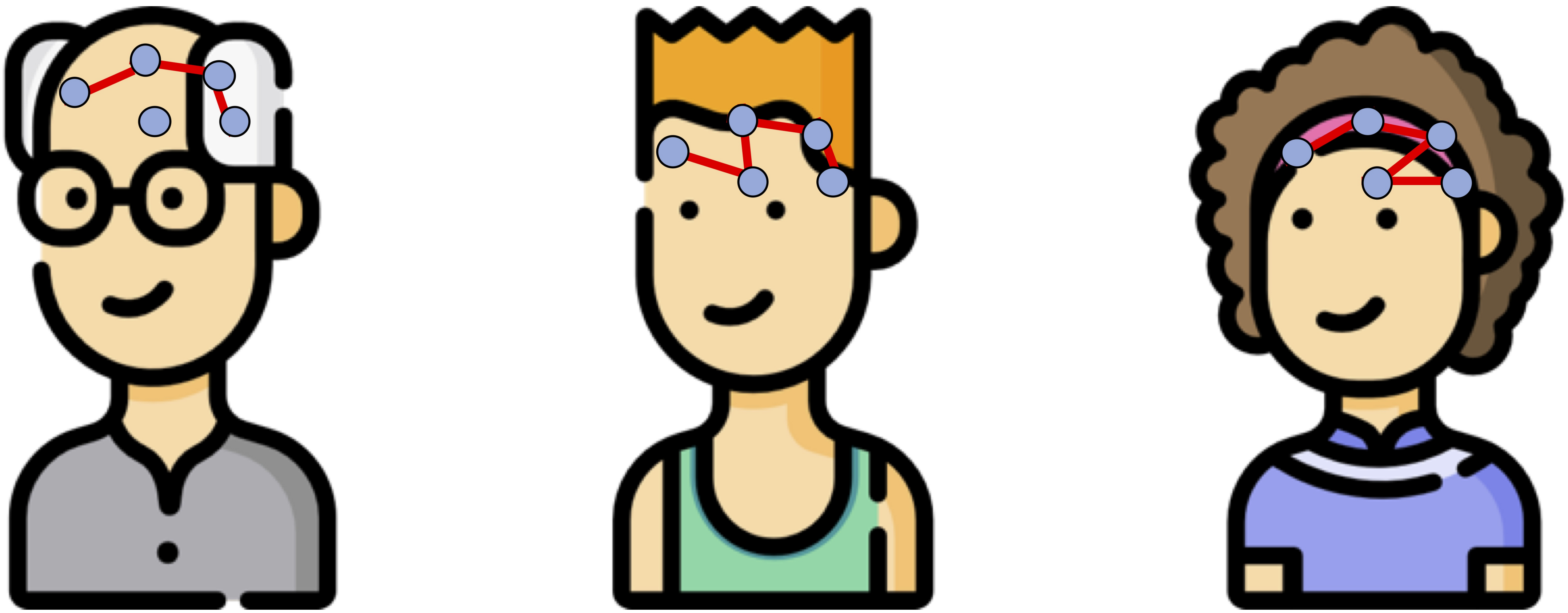
Brian Caffo
Johns Hopkins Biostat

Slides viewable on web:
bit.ly/icsahz19

Statistics/Data Science Focuses



Motivating Example



Brain network connections vary by covariates (e.g. age/sex)

Goal: model how covariates change network connections

Image source: flaticon.com

$$\text{function}(\mathbf{graph}) = \mathbf{age} \times \beta_1 + \mathbf{sex} \times \beta_2 + \cdots$$

Resting-state fMRI Networks

- fMRI measures brain activities over time
- Resting-state: "do nothing" during scanning

- Brain networks constructed using **cov/cor** matrices of time series

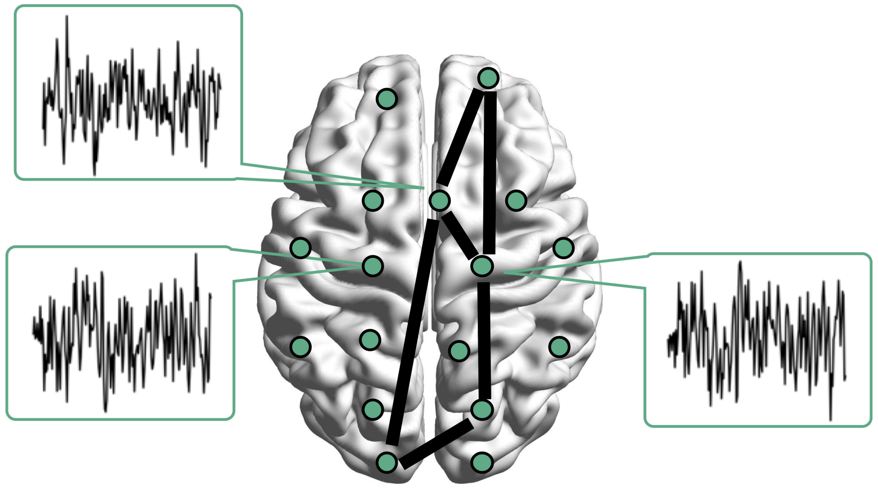


Image source: NSF

Mathematical Problem

- Given n (semi-)positive matrix outcomes, $\Sigma_i \in \mathbb{R}^{p \times p}$
- Given n corresponding vector covariates, $x_i \in \mathbb{R}^q$
- Find function $g(\Sigma_i) = x_i\beta, i = 1, \ldots, n$
- In essense, **regress positive matrices on vectors**

Some Related Problems

- Heterogeneous regression or weighted LS:
  - Usually for scalar variance σ_i, find $g(\sigma_i) = f(x_i)$
  - Goal: to improve efficiency, not to interpret $x_i\beta$
- Covariance models [Anderson, 73; Pourahmadi, 99; Hoff, Niu, 12; Fox, Dunson, 15; Zou, 17]
  - Model $\Sigma_i = g(x_i)$, sometimes $n = i = 1$
  - Goal: better models for Σ_i
- Multi-group PCA [Flury, 84, 88; Boik 02; Hoff 09; Franks, Hoff, 16]
  - No regression model, cannot handle vector x_i
  - Goal: find common/uncommon parts of multiple Σ_i
- Tensor-on-scalar regression [Li, Zhang, 17; Sun, Li, 17]
  - No guarantees for positive matrix outcomes

Massive Edgewise Regressions

- Intuitive method by mostly neuroscientists
- Try $g_{j,k}(\Sigma_i) = \Sigma_i[j,k] = x_i\beta$
- Repeat for all $(j,k) \in \{1,\ldots,p\}^2$ pairs
- Essentially $O(p^2)$ regressions for each connection
- Limitations: multiple testing $O(p^2)$, failure to accout for dependencies between regressions

Our CAP in a Nutshell

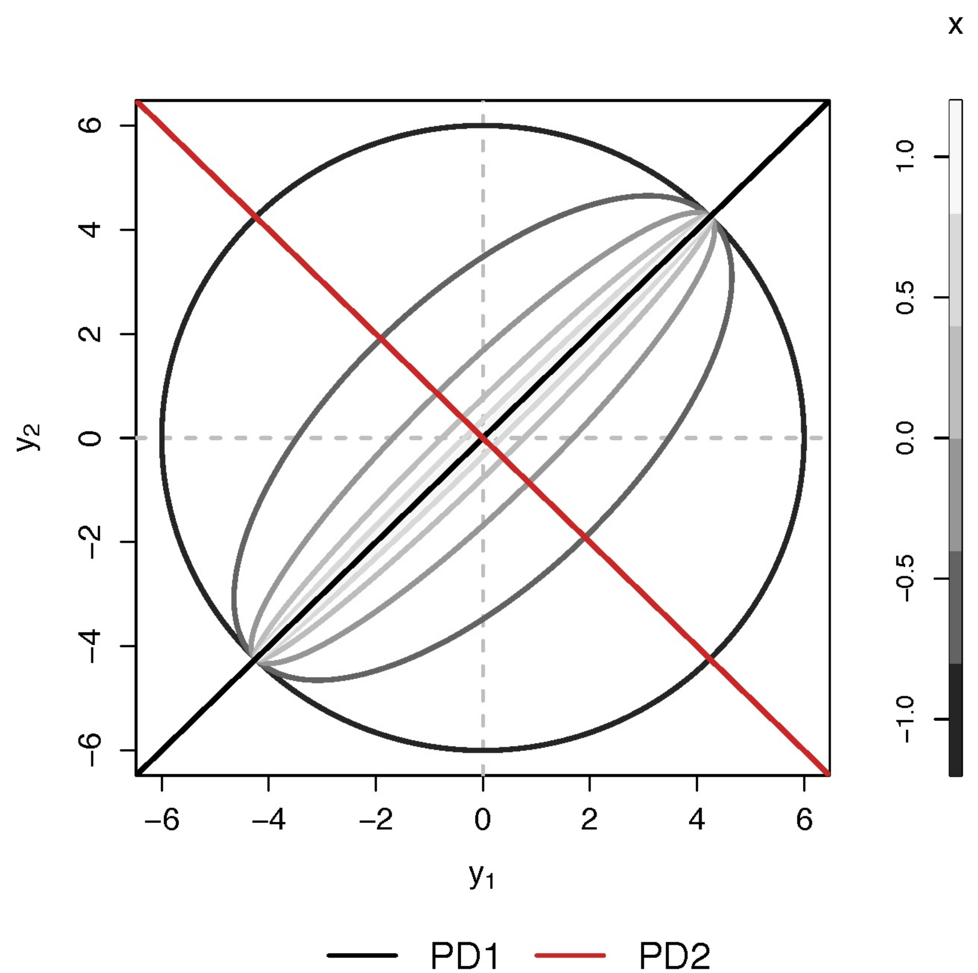
$$\mathbf{PCA}(\boldsymbol{\Sigma}_{\boldsymbol{i}}) = \boldsymbol{x}_{\boldsymbol{i}}\boldsymbol{\beta}$$

- Essentially, we aim to turn unsupervised PCA to a supervised PCA
- Ours differs from existing PCA methods:
  - Supervised PCA [Bair et al, 06] models **scalar-on-vector**

Model and Method

Model

- Find principal direction (PD) $\gamma \in \mathbb{R}^p$, such that:
$$\log(\gamma^\top \Sigma_i \gamma) = \beta_0 + x_i^\top \beta_1, \quad i = 1, \ldots, n$$



Example (p=2): PD1 largest variation but not related to x

PCA selects PD1, Ours selects **PD2**

Advantages

- Scalability: potentially for $p \sim 10^6$ or larger
- Interpretation: covariate assisted PCA
  - Turn **unsupervised** PCA into **supervised**
- Sensitivity: target those covariate-related variations
  - **Covariate assisted** SVD?
- Applicability: other big data problems besides fMRI

Method

- MLE with constraints:

$$\underset{\boldsymbol{\beta},\boldsymbol{\gamma}}{\text{minimize}}\ \ell(\boldsymbol{\beta},\boldsymbol{\gamma}) := \frac{1}{2}\sum_{i=1}^{n}(x_i^\top \boldsymbol{\beta})\cdot T_i + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\gamma}^\top \Sigma_i \boldsymbol{\gamma}\cdot \exp(-x_i^\top \boldsymbol{\beta}),$$

$$\text{such that } \boldsymbol{\gamma}^\top H \boldsymbol{\gamma} = 1$$

- Two obvious constriants:
  - C1: $H = I$
  - C2: $H = n^{-1}(\Sigma_1 + \cdots + \Sigma_n)$

Choice of H

Proposition: *When (C1) $H = \boldsymbol{I}$ in the optimization problem, for any fixed $\boldsymbol{\beta}$, the solution of $\boldsymbol{\gamma}$ is the eigenvector corresponding to the minimum eigenvalue of matrix*

$$\sum_{i=1}^{n} \frac{\Sigma_i}{exp(x_i^\top \boldsymbol{\beta})}$$

Will focus on the constraint (C2)

Algoirthm

- Iteratively update β and then γ
- Prove explicit updates
- Extension to multiple γ:
  - After finding $\gamma^{(1)}$, we will update Σ_i by removing its effect
  - Search for the next PD $\gamma^{(k)}$, $k = 2, \ldots$
  - Impose the orthogonal constraints such that γ^k is orthogonal to all $\gamma^{(t)}$ for $t < k$

Theory for β

Theorem: *Assume* $\sum_{i=1}^{n} x_i x_i^T / n \to Q$ *as* $n \to \infty$ *. Let* $T = min_i T_i$, $M_n = \sum_{i=1}^{n} T_i$, *under the true* $\boldsymbol{\gamma}$, *we have*

$$\sqrt{M_n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, 2Q^{-1}\right), \quad as\, n, T \to \infty,$$

where $\hat{\boldsymbol{\beta}}$ *is the maximum likelihood estimator when the true* $\boldsymbol{\gamma}$ *is known.*

Theory for γ

Theorem: *Assume* $\Sigma_i = \Gamma \Lambda_i \Gamma^\top$, *where* $\Gamma = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)$ *is an orthogonal matrix and* $\Lambda_i = diag\{\lambda_{i1}, \dots, \lambda_{ip}\}$*with* $\lambda_{ik} \neq \lambda_{il}$ $(k \neq l)$*, for at least one* $i \in \{1, \dots, n\}$*There exists* $k \in \{1, \dots, p\}$*such that for* $\forall i \in \{1, \dots, n,\}$ $\boldsymbol{\gamma}_k^\top \Sigma_i \boldsymbol{\gamma}_k = exp(x_i^\top \boldsymbol{\beta})$*. Let* $\hat{\boldsymbol{\gamma}}$ *be the maximum likelihood estimator of* $\boldsymbol{\gamma}_k$ *in Flury, 84. Then assuming that the assumptions are satisfied,* $\hat{\boldsymbol{\beta}}$ *from our algorithm is* $\sqrt{M_n}$*-consistent estimator of* $\boldsymbol{\beta}$*.*

Simulations

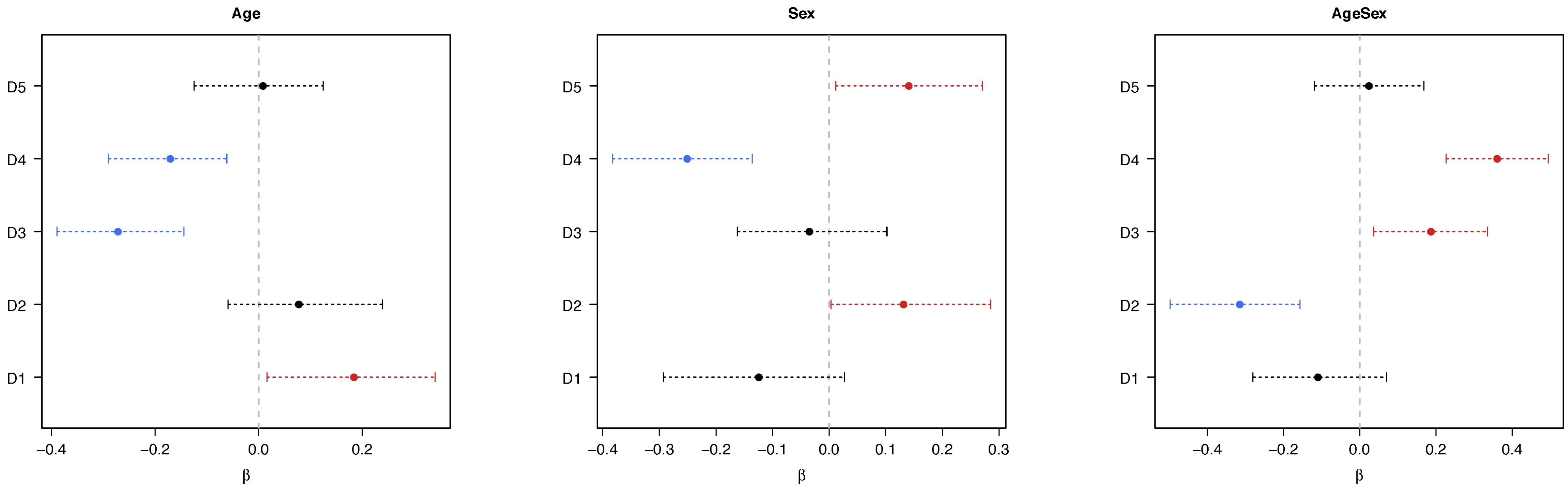
Table 1: Estimate (Est.) of β_1, as well as standard error (SE), coverage probability with asymptotic variance in Theorem 1 (CP-A) and coverage probability from 500 bootstrap samples (CP-B) from different methods under the alternative hypothesis. All values are computed with $n = 100$ and $T_i = 100$ over 200 simulations.

Method	First Direction			Second Direction		
	Est. (SE)	CP-A	CP-B	Est. (SE)	CP-A	CP-B
Truth	-1.00	-	-	1.00	-	-
CAP	-1.00 (0.03)	0.950	0.950	0.81 (0.58)	0.885	0.870
CAP-OC	-1.00 (0.03)	0.950	0.950	0.52 (0.84)	0.730	0.715
CAP-C	-1.00 (0.03)	0.950	0.955	1.00 (0.03)	0.975	0.960
PCA	-0.02 (0.10)	-	0	-0.98 (0.03)	-	0
CPCA	-0.01 (0.11)	-	0	-1.00 (0.03)	-	0

PCA and common PCA do not find the first principal direction, because they don't model covariates

Resting-state fMRI

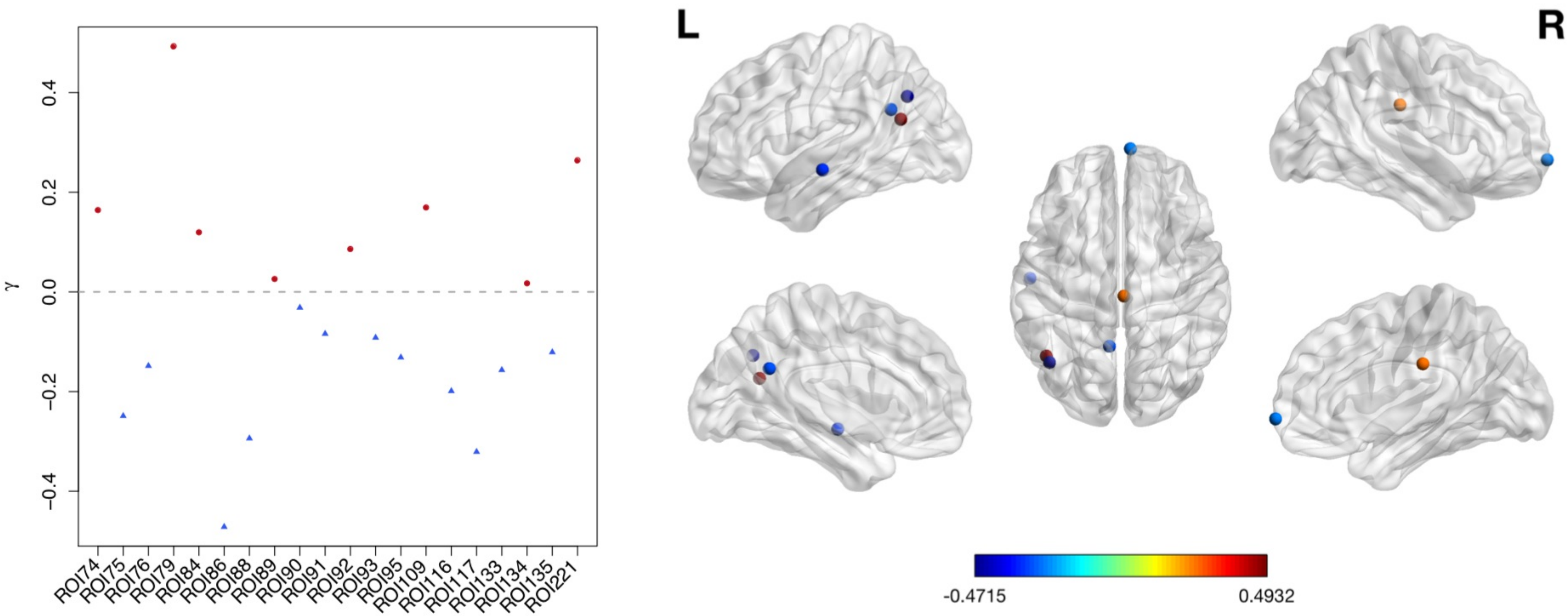
Regression Coefficients



Age Sex Age*Sex

No statistical significant changes were found by massive edgewise regression

Brain Map of γ



(a) The loadings.

(b) Regions with $|\gamma_j|$ above 0.2 in brain map.

Figure 4: The loading profile and brain regions with absolute loading greater than 0.2 in projection direction D1 identified by CAP.

Discussion

- Regress **PD** matrices on vectors
- Method to identify covariate-related (supervised) directions vs (unsupervised) PCA
- Theorectical justifications
- Paper: Biostatistics (10.1093/biostatistics/kxz057)
- R pkg: **cap** 

Image source: flaticon.com

Thank you!

Comments? Questions?

BigComplexData.com

or **BrainDataScience.com**