Regularized testing of linear hypotheses in high dimensions

Debashis Paul

Department of Statistics, University of California, Davis

joint work with

Haoran Li (Auburn U.), Alexander Aue (UC Davis), Jie Peng (UC Davis)

Consortium in Data Analytics for Risk (University of California, Berkeley)

November 8, 2022

Research supported by U.S. National Science Foundation and National Institute of Health grants

Themes

- Regularization in linear models

Themes

- Regularization in linear models
- Decision theory

Themes

- Regularization in linear models
- Decision theory
- Random matrix theory

Themes

- Regularization in linear models
- Decision theory
- Random matrix theory
- Spiked covariance model

General linear hypothesis

Multivariate linear model

$$\mathbf{Y} = BX + \Sigma_p^{1/2}\mathbf{Z}$$

$$H_0 : BC = 0 \quad \text{v.s.} \quad H_a : BC \neq 0.$$

- $\mathbf{Y}$: p-variates $\times$ N-subjects observation;

General linear hypothesis

Multivariate linear model

$$\mathbf{Y} = BX + \Sigma_p^{1/2}\mathbf{Z}$$

$$H_0 : BC = 0 \ \text{ v.s. } \ H_a : BC \neq 0.$$

- $\mathbf{Y}$: p-variates $\times$ N-subjects observation;
- B: $p \times m$ parameter matrix;

General linear hypothesis

Multivariate linear model

$$\mathbf{Y} = BX + \Sigma_p^{1/2}\mathbf{Z}$$

$$H_0 : BC = 0 \quad \text{v.s.} \quad H_a : BC \neq 0.$$

- $\mathbf{Y}$: p-variates $\times$ N-subjects observation;
- B: $p \times m$ parameter matrix;
- X: $m \times N$ design matrix of rank m;

General linear hypothesis

Multivariate linear model

$$\mathbf{Y} = BX + \Sigma_p^{1/2}\mathbf{Z}$$

$$H_0 : BC = 0 \;\; \text{v.s.} \;\; H_a : BC \neq 0.$$

- $\mathbf{Y}$: p-variates $\times$ N-subjects observation;
- B: $p \times m$ parameter matrix;
- X: $m \times N$ design matrix of rank m;
- Σ_p: population covariance;

General linear hypothesis

Multivariate linear model

$$\mathbf{Y} = BX + \Sigma_p^{1/2}\mathbf{Z}$$

$$H_0 : BC = 0 \quad \text{v.s.} \quad H_a : BC \neq 0.$$

- $\mathbf{Y}$: p-variates $\times$ N-subjects observation;
- B: $p \times m$ parameter matrix;
- X: $m \times N$ design matrix of rank m;
- Σ_p: population covariance;
- $\mathbf{Z}$: $p \times N$ error matrix; i.i.d., mean 0 and variance 1;

General linear hypothesis

Multivariate linear model

$$\mathbf{Y} = BX + \Sigma_p^{1/2}\mathbf{Z}$$

$$H_0 : BC = 0 \;\; \text{v.s.} \;\; H_a : BC \neq 0.$$

- $\mathbf{Y}$: p-variates $\times$ N-subjects observation;
- B: $p \times m$ parameter matrix;
- X: $m \times N$ design matrix of rank m;
- Σ_p: population covariance;
- $\mathbf{Z}$: $p \times N$ error matrix; i.i.d., mean 0 and variance 1;
- C: $m \times q$ constraints matrix of rank q; estimable.

Classical solution

Traditional test statistics are functionals of the spectrum of $\widehat{\mathbf{H}}_p \widehat{\mathbf{\Sigma}}_p^{-1}$.

$$\widehat{\mathbf{\Sigma}}_p = \frac{1}{n}\mathbf{Y}(I - X^T(XX^T)^{-1}X)\mathbf{Y}^T, \quad n = N - m.$$

$$\widehat{\mathbf{H}}_p = \frac{1}{n}\mathbf{Y}X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1}C^T(XX^T)^{-1}X\mathbf{Y}^T.$$

- Residual covariance of the full model

Classical solution

Traditional test statistics are functionals of the spectrum of $\widehat{\mathbf{H}}_p \widehat{\boldsymbol{\Sigma}}_p^{-1}$.

$$\widehat{\boldsymbol{\Sigma}}_p = \frac{1}{n}\mathbf{Y}(I - X^T(XX^T)^{-1}X)\mathbf{Y}^T, \quad n = N - m.$$

$$\widehat{\mathbf{H}}_p = \frac{1}{n}\mathbf{Y}X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1}C^T(XX^T)^{-1}X\mathbf{Y}^T.$$

- Residual covariance of the full model
- The hypothesis sums of squares and cross-products matrix

Classical solution

Traditional test statistics are functionals of the spectrum of $\widehat{\mathbf{H}}_p \widehat{\boldsymbol{\Sigma}}_p^{-1}$.

$$\widehat{\boldsymbol{\Sigma}}_p = \frac{1}{n}\mathbf{Y}(I - X^T(XX^T)^{-1}X)\mathbf{Y}^T, \quad n = N - m.$$

$$\widehat{\mathbf{H}}_p = \frac{1}{n}\mathbf{Y}X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1}C^T(XX^T)^{-1}X\mathbf{Y}^T.$$

- Residual covariance of the full model
- The hypothesis sums of squares and cross-products matrix
- OLS of BC

Classical solution

Traditional test statistics are functionals of the spectrum of $\widehat{\mathbf{H}}_p \widehat{\mathbf{\Sigma}}_p^{-1}$.

$$\widehat{\mathbf{\Sigma}}_p = \frac{1}{n}\mathbf{Y}(I - X^T(XX^T)^{-1}X)\mathbf{Y}^T, \quad n = N - m.$$

$$\widehat{\mathbf{H}}_p = \frac{1}{n}\mathbf{Y}X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1}C^T(XX^T)^{-1}X\mathbf{Y}^T.$$

- Residual covariance of the full model
- The hypothesis sums of squares and cross-products matrix
- OLS of BC
- The "hat matrix" of the reduced model

Classical solution

Traditional test statistics are functionals of the spectrum of $\widehat{\mathbf{H}}_p \widehat{\mathbf{\Sigma}}_p^{-1}$.

$$\widehat{\mathbf{\Sigma}}_p = \frac{1}{n}\mathbf{Y}(I - X^T(XX^T)^{-1}X)\mathbf{Y}^T, \quad n = N - m.$$

$$\widehat{\mathbf{H}}_p = \frac{1}{n}\mathbf{Y}X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1}C^T(XX^T)^{-1}X\mathbf{Y}^T.$$

- Residual covariance of the full model
- The hypothesis sums of squares and cross-products matrix
- OLS of BC
- The “hat matrix” of the reduced model
- one-way MANOVA: $\widehat{\mathbf{\Sigma}}_p = \frac{1}{n} SS_{within}$ $\quad \widehat{\mathbf{H}}_p = \frac{1}{n} SS_{between}$

Classical solution

Define $Q_n = X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1/2}$

$$\widehat{\mathbf{H}}_p \widehat{\mathbf{\Sigma}}_p^{-1} = \frac{1}{n}\mathbf{Y} Q_n Q_n^T \mathbf{Y}^T \widehat{\mathbf{\Sigma}}_p^{-1}$$

Classical solution

Define $Q_n = X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1/2}$

$$\widehat{\mathbf{H}}_p\widehat{\mathbf{\Sigma}}_p^{-1} = \frac{1}{n}\mathbf{Y}Q_nQ_n^T\mathbf{Y}^T\widehat{\mathbf{\Sigma}}_p^{-1} \xLeftrightarrow[\textit{eigenvalues}]{\textit{positive}} \frac{1}{n}Q_n^T\mathbf{Y}^T\widehat{\mathbf{\Sigma}}_p^{-1}\mathbf{Y}Q_n =: M_0$$

Classical solution

Define $Q_n = X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1/2}$

$$\widehat{\mathbf{H}}_p\widehat{\mathbf{\Sigma}}_p^{-1} = \frac{1}{n}\mathbf{Y}Q_nQ_n^T\mathbf{Y}^T\widehat{\mathbf{\Sigma}}_p^{-1} \xLeftrightarrow[\text{eigenvalues}]{\text{positive}} \frac{1}{n}Q_n^T\mathbf{Y}^T\widehat{\mathbf{\Sigma}}_p^{-1}\mathbf{Y}Q_n =: M_0$$

Invariant tests:

Classical solution

Define $Q_n = X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1/2}$

$$\widehat{\mathbf{H}}_p\widehat{\boldsymbol{\Sigma}}_p^{-1} = \frac{1}{n}\mathbf{Y}Q_nQ_n^T\mathbf{Y}^T\widehat{\boldsymbol{\Sigma}}_p^{-1} \xLeftrightarrow[\text{eigenvalues}]{\text{positive}} \frac{1}{n}Q_n^T\mathbf{Y}^T\widehat{\boldsymbol{\Sigma}}_p^{-1}\mathbf{Y}Q_n =: M_0$$

Invariant tests:

- Likelihood-ratio (LR): $T_0^{LR} = \sum_{i=1}^{q}\log\{1+\lambda_i(M_0)\}$

Classical solution

Define $Q_n = X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1/2}$

$$\widehat{\mathbf{H}}_p\widehat{\mathbf{\Sigma}}_p^{-1} = \frac{1}{n}\mathbf{Y}Q_nQ_n^T\mathbf{Y}^T\widehat{\mathbf{\Sigma}}_p^{-1} \xLeftrightarrow[\textit{eigenvalues}]{\textit{positive}} \frac{1}{n}Q_n^T\mathbf{Y}^T\widehat{\mathbf{\Sigma}}_p^{-1}\mathbf{Y}Q_n =: M_0$$

Invariant tests:

- Likelihood-ratio (LR): $T_0^{LR} = \sum_{i=1}^{q}\log\{1+\lambda_i(M_0)\}$
- Lawley-Hotelling trace (LH): $T_0^{LH} = \sum_{i=1}^{q}\lambda_i(M_0)$

Classical solution

Define $Q_n = X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1/2}$

$$\widehat{\mathbf{H}}_p\widehat{\boldsymbol{\Sigma}}_p^{-1} = \frac{1}{n}\mathbf{Y}Q_nQ_n^T\mathbf{Y}^T\widehat{\boldsymbol{\Sigma}}_p^{-1} \xLeftrightarrow[\text{eigenvalues}]{\text{positive}} \frac{1}{n}Q_n^T\mathbf{Y}^T\widehat{\boldsymbol{\Sigma}}_p^{-1}\mathbf{Y}Q_n =: M_0$$

Invariant tests:

- Likelihood-ratio (LR): $T_0^{LR} = \sum_{i=1}^{q} \log\{1 + \lambda_i(M_0)\}$
- Lawley-Hotelling trace (LH): $T_0^{LH} = \sum_{i=1}^{q} \lambda_i(M_0)$
- Bartlett-Nanda-Pillai trace (BNP): $T_0^{BNP} = \sum_{i=1}^{q} \frac{\lambda_i(M_0)}{1 + \lambda_i(M_0)}$

When p $\asymp$ n

- Throughout, consider q to be *fixed*.

When p $\asymp$ n

- Throughout, consider q to be *fixed*.
- Let $p, n \to \infty$ and p be comparable to n

When p $\asymp$ n

- Throughout, consider q to be *fixed*.
- Let $p, n \to \infty$ and p be comparable to n
- $\widehat{\mathbf{H}}_p$ is a $p \times p$ matrix of rank q

When p $\asymp$ n

- Throughout, consider q to be *fixed*.
- Let $p, n \to \infty$ and p be comparable to n
- $\widehat{\mathbf{H}}_p$ is a $p \times p$ matrix of rank q
- $\widehat{\boldsymbol{\Sigma}}_p$ is not a consistent estimator of Σ_p

When p $\asymp$ n

- Throughout, consider q to be *fixed*.
- Let $p, n \to \infty$ and p be comparable to n
- $\widehat{\mathbf{H}}_p$ is a $p \times p$ matrix of rank q
- $\widehat{\boldsymbol{\Sigma}}_p$ is not a consistent estimator of Σ_p
- Performance of invariant tests studied in Fujikoshi et al. (2004) and Bai et al. (2017) when $\gamma_n := p/n \to \gamma \in (0, 1)$

Tests with spectral shrinkage

- We propose a family of statistics that are functionals of the spectrum of

$$M(f) = \frac{1}{n} Q_n^T \mathbf{Y}^T f(\widehat{\mathbf{\Sigma}}_p) \mathbf{Y} Q_n$$

Tests with spectral shrinkage

- We propose a family of statistics that are functionals of the spectrum of

$$M(f) = \frac{1}{n} Q_n^T \mathbf{Y}^T f(\widehat{\boldsymbol{\Sigma}}_p) \mathbf{Y} Q_n$$

- $f(\cdot)$ is any function on $\mathbb{R}$, analytic on a certain interval

$$f(\widehat{\boldsymbol{\Sigma}}_p) := \widehat{\mathbf{P}} \mathrm{diag}[f(\lambda_1(\widehat{\boldsymbol{\Sigma}}_p)), f(\lambda_2(\widehat{\boldsymbol{\Sigma}}_p)), \cdots, f(\lambda_p(\widehat{\boldsymbol{\Sigma}}_p))] \widehat{\mathbf{P}}^T$$

with $\widehat{\mathbf{P}}$ being the matrix of eigenvectors of $\widehat{\boldsymbol{\Sigma}}_p$.

Tests with spectral shrinkage

- We propose a family of statistics that are functionals of the spectrum of

$$M(f) = \frac{1}{n} Q_n^T \mathbf{Y}^T f(\widehat{\boldsymbol{\Sigma}}_p) \mathbf{Y} Q_n$$

- $f(\cdot)$ is any function on $\mathbb{R}$, analytic on a certain interval

$$f(\widehat{\boldsymbol{\Sigma}}_p) := \widehat{\mathbf{P}} \mathrm{diag}[f(\lambda_1(\widehat{\boldsymbol{\Sigma}}_p)), f(\lambda_2(\widehat{\boldsymbol{\Sigma}}_p)), \cdots, f(\lambda_p(\widehat{\boldsymbol{\Sigma}}_p))] \widehat{\mathbf{P}}^T$$

  with $\widehat{\mathbf{P}}$ being the matrix of eigenvectors of $\widehat{\boldsymbol{\Sigma}}_p$.

- $M(f)$ is rotation-invariant.

Tests under general Σ_p when $p/n \to \gamma \in (0, \infty)$

- Under fairly general Σ_p (empirical distribution of eigenvalues of Σ_p converges to a nondegenerate limit) the regularization scheme involving $M(f)$ was studied in Li, Aue and Paul (2020).

Tests under general Σ_p when $p/n \to \gamma \in (0, \infty)$

- Under fairly general Σ_p (empirical distribution of eigenvalues of Σ_p converges to a nondegenerate limit) the regularization scheme involving $M(f)$ was studied in Li, Aue and Paul (2020).
- An asymptotic normal limit (under H_0) for the normalized versions of $T^{LR}(f)$, $T^{LH}(f)$ and $T^{BNP}(f)$ was established, where for each test statistic, we replace M_0 by $M(f)$.

Tests under general Σ_p when $p/n \to \gamma \in (0, \infty)$

- Under fairly general Σ_p (empirical distribution of eigenvalues of Σ_p converges to a nondegenerate limit) the regularization scheme involving $M(f)$ was studied in Li, Aue and Paul (2020).
- An asymptotic normal limit (under H_0) for the normalized versions of $T^{LR}(f)$, $T^{LH}(f)$ and $T^{BNP}(f)$ was established, where for each test statistic, we replace M_0 by $M(f)$.
- Power function of these tests under a class of local probabilistic alternatives was also derived.

Tests under general Σ_p when $p/n \to \gamma \in (0, \infty)$

- Under fairly general Σ_p (empirical distribution of eigenvalues of Σ_p converges to a nondegenerate limit) the regularization scheme involving $M(f)$ was studied in Li, Aue and Paul (2020).
- An asymptotic normal limit (under H_0) for the normalized versions of $T^{LR}(f)$, $T^{LH}(f)$ and $T^{BNP}(f)$ was established, where for each test statistic, we replace M_0 by $M(f)$.
- Power function of these tests under a class of local probabilistic alternatives was also derived.
- The functional form of the power function was used to determine an appropriate regularization scheme f, depending on the class of local alternatives.

Two sample problem: Hotelling's T^2 test

- Let $X_{j1}, X_{j2}, \ldots, X_{jn_j}$ be i.i.d. p-dimensional $\mathcal{N}(\mu_j, \Sigma)$, for $j = 1, 2$.

Two sample problem: Hotelling's T^2 test

- Let $X_{j1}, X_{j2}, \ldots, X_{jn_j}$ be i.i.d. p-dimensional $\mathcal{N}(\mu_j, \Sigma)$, for $j = 1, 2$.
- $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$.

Two sample problem: Hotelling's T^2 test

- Let $X_{j1}, X_{j2}, \ldots, X_{jn_j}$ be i.i.d. p-dimensional $\mathcal{N}(\mu_j, \Sigma)$, for $j = 1, 2$.
- $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$.
- Hotelling's T^2 (HT) statistic:

$$\mathrm{HT} = \frac{n_1 n_2}{n_1 + n_2}(\bar{X}_1 - \bar{X}_2)^T \widehat{\mathbf{\Sigma}}_p^{-1}(\bar{X}_1 - \bar{X}_2)$$

  where $\overline{X}_j$ is the sample mean for the j-th sample, and $\widehat{\mathbf{\Sigma}}_p$ is the pooled sample covariance.
  - Equivalent to likelihood ratio statistics;
  - Invariant under rotation of the coordinates.
  - Null distribution does not depend on Σ_p.

Two sample problem: Hotelling's T^2 test

- Let $X_{j1}, X_{j2}, \ldots, X_{jn_j}$ be i.i.d. p-dimensional $\mathcal{N}(\mu_j, \Sigma)$, for $j = 1, 2$.
- $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$.
- Hotelling's T^2 (HT) statistic:

$$\mathrm{HT} = \frac{n_1 n_2}{n_1 + n_2}(\bar{X}_1 - \bar{X}_2)^T \widehat{\boldsymbol{\Sigma}}_p^{-1}(\bar{X}_1 - \bar{X}_2)$$

  where $\overline{X}_j$ is the sample mean for the j-th sample, and $\widehat{\boldsymbol{\Sigma}}_p$ is the pooled sample covariance.
  - Equivalent to likelihood ratio statistics;
  - Invariant under rotation of the coordinates.
  - Null distribution does not depend on Σ_p.
- Issues in high dimension:
  - When $p \geq n$ (with $n = n_1 + n_2$) HT is not defined.
  - Poor power even when $p < n$ but $p/n \approx 1$.

Regularized Hotelling's T^2 test (RHT)

- The RHT statistic for testing $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ is given by

$$RHT(\lambda) = \frac{n_1 n_2}{n_1 + n_2}(\bar{X}_1 - \bar{X}_2)^T(\widehat{\mathbf{\Sigma}}_p + \lambda I_p)^{-1}(\bar{X}_1 - \bar{X}_2), \qquad (1)$$

with $n = n_1 + n_2$, and

$$\widehat{\mathbf{\Sigma}}_p = \frac{1}{n-2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}(X_{ji} - \bar{X}_j)(X_{ji} - \bar{X}_j)^T.$$

Regularized Hotelling's T^2 test (RHT)

- The RHT statistic for testing $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ is given by

$$RHT(\lambda) = \frac{n_1 n_2}{n_1 + n_2}(\bar{X}_1 - \bar{X}_2)^T(\widehat{\mathbf{\Sigma}}_p + \lambda I_p)^{-1}(\bar{X}_1 - \bar{X}_2), \qquad (1)$$

with $n = n_1 + n_2$, and

$$\widehat{\mathbf{\Sigma}}_p = \frac{1}{n-2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}(X_{ji} - \bar{X}_j)(X_{ji} - \bar{X}_j)^T.$$

- This corresponds to the general linear hypothesis testing problem with $k = 2$ and $q = 1$.

Regularized Hotelling's T^2 test (RHT)

- The RHT statistic for testing $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ is given by

$$RHT(\lambda) = \frac{n_1 n_2}{n_1 + n_2}(\bar{X}_1 - \bar{X}_2)^T(\widehat{\boldsymbol{\Sigma}}_p + \lambda I_p)^{-1}(\bar{X}_1 - \bar{X}_2), \qquad (1)$$

with $n = n_1 + n_2$, and

$$\widehat{\boldsymbol{\Sigma}}_p = \frac{1}{n-2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}(X_{ji} - \bar{X}_j)(X_{ji} - \bar{X}_j)^T.$$

- This corresponds to the general linear hypothesis testing problem with $k = 2$ and $q = 1$.
- A special case of the general regularization scheme with $f(x) = 1/(x + \lambda)$.

Asymptotic null distribution

When $\mu = \mu_1 - \mu_2 = 0$, as $p, n \to \infty$ so that $\gamma_n = p/n \to \gamma \in (0, \infty)$, (under additional technical assumptions)

$$T_{n,p}(\lambda) := \frac{\sqrt{p}\left(\frac{1}{p}\mathtt{RHT}(\lambda) - \widehat{\Theta}_{1,n}(\lambda, \gamma_n)\right)}{\left(2\widehat{\Theta}_{2,n}(\lambda, \gamma_n)\right)^{1/2}} \Longrightarrow N(0, 1), \qquad (2)$$

Asymptotic null distribution

When $\mu = \mu_1 - \mu_2 = 0$, as $p, n \to \infty$ so that $\gamma_n = p/n \to \gamma \in (0, \infty)$, (under additional technical assumptions)

$$T_{n,p}(\lambda) := \frac{\sqrt{p}\left(\frac{1}{p}\mathtt{RHT}(\lambda) - \widehat{\Theta}_{1,n}(\lambda, \gamma_n)\right)}{\left(2\widehat{\Theta}_{2,n}(\lambda, \gamma_n)\right)^{1/2}} \Longrightarrow N(0, 1), \qquad (2)$$

where

$$\begin{aligned}
\widehat{\Theta}_{1,n}(\lambda, \gamma_n) &= \frac{1 - \lambda m_n(-\lambda)}{1 - \gamma_n(1 - \lambda m_n(-\lambda))} \\
\widehat{\Theta}_{2,n}(\lambda, \gamma_n) &= \frac{1 - \lambda m_n(-\lambda)}{(1 - \gamma_n + \gamma_n \lambda m_n(-\lambda))^3} - \lambda \frac{m_n(-\lambda) - \lambda m_n'(-\lambda)}{(1 - \gamma_n + \gamma_n \lambda m_n(-\lambda))^4} \\
m_n(\mathbb{z}) &= \frac{1}{p}\mathrm{tr}((\widehat{\mathbf{\Sigma}}_p - \mathbb{z} I_p)^{-1}).
\end{aligned}$$

Notice that $m_n(\mathbb{z})$, $\mathbb{z} \in \mathbb{C}$, is the *Stieltjes transform* of the empirical spectral distribution of $\widehat{\mathbf{\Sigma}}_p$.

A Bayesian framework: Probabilistic local alternatives

$\mu := \mu_1 - \mu_2 = n^{-1/4} p^{-1/2} B\nu$ where B is a $p \times p$ matrix, and ν is random vector with independent entries,

- $\mathbb{E}[\nu_i] = 0$,
- $\mathbb{E}[|\nu_i|^2] = 1$,
- $\max_i \mathbb{E}[|\nu_i|^4] \leq p^{c_\nu}$ for some $c_\nu \in (0, 1)$.

Power under local alternatives

- Let $\boldsymbol{B} = BB^T$ such that $\|\boldsymbol{B}\| \le C < \infty$ and, as $n, p \to \infty$,

$$p^{-1}\mathrm{tr}\{D_p(-\lambda)\boldsymbol{B}\} \to q(\lambda, \gamma)$$

for some finite, positive constant $q(\lambda, \gamma)$, where

$$D_p(-\lambda) = \left(\frac{1}{1 + \gamma\Theta_1(\lambda, \gamma)}\Sigma_p + \lambda I_p\right)^{-1}$$

is the *deterministic equivalent* of $(\widehat{\boldsymbol{\Sigma}}_p + \lambda I_p)^{-1}$.

Power under local alternatives

- Let $\mathbf{B} = BB^T$ such that $\|\mathbf{B}\| \leq C < \infty$ and, as $n, p \to \infty$,

$$p^{-1}\mathrm{tr}\{D_p(-\lambda)\mathbf{B}\} \to q(\lambda, \gamma)$$

  for some finite, positive constant $q(\lambda, \gamma)$, where

$$D_p(-\lambda) = \left(\frac{1}{1 + \gamma\Theta_1(\lambda, \gamma)}\Sigma_p + \lambda I_p\right)^{-1}$$

  is the *deterministic equivalent* of $(\widehat{\mathbf{\Sigma}}_p + \lambda I_p)^{-1}$.

- As $n, p \to \infty$ the asymptotic level α RHT test has power

$$\beta_n(\mu, \lambda) \xrightarrow{P} \Phi\left(-\xi_\alpha + \kappa(1-\kappa)\frac{q(\lambda, \gamma)}{\{2\gamma\Theta_2(\lambda, \gamma)\}^{1/2}}\right).$$

Bayes selection of λ

- Given a sequence of local probabilistic alternatives (prior for μ), the strategy is to choose λ by maximizing the **local power** function $\beta_n(\mu, \lambda)$.

Bayes selection of λ

- Given a sequence of local probabilistic alternatives (prior for μ), the strategy is to choose λ by maximizing the **local power** function $\beta_n(\mu, \lambda)$.

- Need to specify $q(\lambda, \gamma)$ in the expression for local power.

$$\beta_n(\mu, \lambda) \xrightarrow{L_1} \Phi\Big(-\xi_\alpha + \kappa(1-\kappa)\frac{q(\lambda,\gamma)}{\{2\gamma\Theta_2(\lambda,\gamma)\}^{1/2}}\Big).$$

Estimate $q(\lambda, \gamma)$ under a "polynomial prior"

- It is convenient to have $q(\lambda, \gamma)$ in a closed form.

Estimate $q(\lambda, \gamma)$ under a "polynomial prior"

- It is convenient to have $q(\lambda, \gamma)$ in a closed form.
- Feasible if $\boldsymbol{B}$ is a polynomial in Σ_p, i.e., $\boldsymbol{B} = \sum_{k=0}^{r} \pi_k \Sigma_p^k$.

Estimate $q(\lambda, \gamma)$ under a "polynomial prior"

- It is convenient to have $q(\lambda, \gamma)$ in a closed form.
- Feasible if $\boldsymbol{B}$ is a polynomial in Σ_p, i.e., $\boldsymbol{B} = \sum_{k=0}^{r} \pi_k \Sigma_p^k$.
- Under this structural assumption,

$$q(\lambda, \gamma) = \sum_{k=0}^{r} \pi_k \rho_k(-\lambda, \gamma),$$

with $\rho_k(-\lambda, \gamma)$ satisfying the recursive formula

$$\rho_{k+1}(-\lambda, \gamma) = \{1 + \gamma \Theta_1(\lambda, \gamma)\}\{\int \tau^k dF^{\Sigma}(\tau) - \lambda \rho_k(-\lambda, \gamma)\}.$$

Algorithm (for $r = 2$)

- Specify prior weights $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2)$;

Algorithm (for $r = 2$)

- Specify prior weights $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2)$;
- For each λ, compute the estimates

$$\hat{\rho}_0(-\lambda, \gamma_n) = m_n(-\lambda),$$
$$\hat{\rho}_1(-\lambda, \gamma_n) = \hat{\Theta}_1(\lambda, \gamma_n),$$
$$\hat{\rho}_2(-\lambda, \gamma_n) = \{1 + \gamma_n \hat{\Theta}_1(\lambda, \gamma_n)\}\{p^{-1}\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_p) - \lambda \hat{\rho}_1(-\lambda, \gamma_n)\};$$

Algorithm (for $r = 2$)

- Specify prior weights $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2)$;
- For each λ, compute the estimates

$$\hat{\rho}_0(-\lambda, \gamma_n) = m_n(-\lambda),$$
$$\hat{\rho}_1(-\lambda, \gamma_n) = \hat{\Theta}_1(\lambda, \gamma_n),$$
$$\hat{\rho}_2(-\lambda, \gamma_n) = \{1 + \gamma_n \hat{\Theta}_1(\lambda, \gamma_n)\}\{p^{-1}\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_p) - \lambda \hat{\rho}_1(-\lambda, \gamma_n)\};$$

- For each λ, compute

$$R_n(\lambda, \gamma_n; \boldsymbol{\pi}) = \frac{\sum_{k=0}^{2} \pi_k \hat{\rho}_k(-\lambda, \gamma_n)}{\{\gamma_n \hat{\Theta}_2(\lambda, \gamma_n)\}^{1/2}}$$

Algorithm (for $r = 2$)

- Specify prior weights $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2)$;
- For each λ, compute the estimates

$$\begin{aligned}
\hat{\rho}_0(-\lambda, \gamma_n) &= m_n(-\lambda), \\
\hat{\rho}_1(-\lambda, \gamma_n) &= \hat{\Theta}_1(\lambda, \gamma_n), \\
\hat{\rho}_2(-\lambda, \gamma_n) &= \{1 + \gamma_n \hat{\Theta}_1(\lambda, \gamma_n)\}\{p^{-1}\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_p) - \lambda\hat{\rho}_1(-\lambda, \gamma_n)\};
\end{aligned}$$

- For each λ, compute

$$R_n(\lambda, \gamma_n; \boldsymbol{\pi}) = \frac{\sum_{k=0}^{2} \pi_k \hat{\rho}_k(-\lambda, \gamma_n)}{\{\gamma_n \hat{\Theta}_2(\lambda, \gamma_n)\}^{1/2}}$$

- Select $\lambda_{\boldsymbol{\pi}} = \arg\max_{\lambda} R_n(\lambda, \gamma_n; \boldsymbol{\pi})$.

Linear hypothesis testing under spiked covariance model

- Assume that, there is a fixed $K \geq 1$ such that

$$\Sigma_p = \sum_{j=1}^{K} \ell_j \mathbf{p}_j \mathbf{p}_j^T + \sigma^2 (I_p - \sum_{j=1}^{K} \mathbf{p}_j \mathbf{p}_j^T) \qquad (3)$$

where $\mathbf{p}_j$'s are the orthonormal eigenvectors, and $\ell_1 > \cdots > \ell_K > \sigma^2$ are the distinct eigenvalues of Σ_p.

Linear hypothesis testing under spiked covariance model

- Assume that, there is a fixed $K \geq 1$ such that

$$\Sigma_p = \sum_{j=1}^{K} \ell_j \mathbf{p}_j \mathbf{p}_j^T + \sigma^2 (I_p - \sum_{j=1}^{K} \mathbf{p}_j \mathbf{p}_j^T) \tag{3}$$

where $\mathbf{p}_j$'s are the orthonormal eigenvectors, and $\ell_1 > \cdots > \ell_K > \sigma^2$ are the distinct eigenvalues of Σ_p.

- When $p/n \to \gamma \in (0, \infty)$ there is a phase transition in the behavior of sample eigenvalues and eigenvectors.

Linear hypothesis testing under spiked covariance model

- Assume that, there is a fixed $K \geq 1$ such that

$$\Sigma_p = \sum_{j=1}^{K} \ell_j \mathbf{p}_j \mathbf{p}_j^T + \sigma^2 (I_p - \sum_{j=1}^{K} \mathbf{p}_j \mathbf{p}_j^T) \quad (3)$$

  where $\mathbf{p}_j$'s are the orthonormal eigenvectors, and $\ell_1 > \cdots > \ell_K > \sigma^2$ are the distinct eigenvalues of Σ_p.
- When $p/n \to \gamma \in (0, \infty)$ there is a phase transition in the behavior of sample eigenvalues and eigenvectors.
- Phase transition limit depends on p/n and ℓ_j/σ^2. If $\ell_j/\sigma^2 > 1 + \sqrt{\gamma}$, the eigenvalue ℓ_j can be consistently estimated from data.

Linear hypothesis testing under spiked covariance model

- Assume that, there is a fixed $K \geq 1$ such that

$$\Sigma_p = \sum_{j=1}^{K} \ell_j \mathbf{p}_j \mathbf{p}_j^T + \sigma^2 (I_p - \sum_{j=1}^{K} \mathbf{p}_j \mathbf{p}_j^T) \qquad (3)$$

  where $\mathbf{p}_j$'s are the orthonormal eigenvectors, and $\ell_1 > \cdots > \ell_K > \sigma^2$ are the distinct eigenvalues of Σ_p.
- When $p/n \to \gamma \in (0, \infty)$ there is a phase transition in the behavior of sample eigenvalues and eigenvectors.
- Phase transition limit depends on p/n and ℓ_j/σ^2. If $\ell_j/\sigma^2 > 1 + \sqrt{\gamma}$, the eigenvalue ℓ_j can be consistently estimated from data.
- Assume that K is known (can be estimated consistently), and $\ell_K/\sigma^2 > 1 + \sqrt{\gamma}$ (all the spiked eigenvalues can be estimated consistently).

Asymptotic limits of $\widehat{\ell}_j$ and $\widehat{\mathbf{p}}_j$

For $x > 1$, define

$$\psi(x,\gamma) = x\left(1 + \frac{\gamma}{x-1}\right) \quad \text{and} \quad \zeta(x,\gamma) = \left[\frac{1-\gamma/(x-1)^2}{1+\gamma/(x-1)}\right]^{1/2}.$$

Then,

$$\widehat{\ell}_j \xrightarrow{a.s.} \sigma^2\psi\left(\frac{\ell_j}{\sigma^2},\gamma\right), \qquad j = 1,\ldots,K. \tag{4}$$

and

$$|\langle \widehat{\mathbf{p}}_j, \mathbf{p}_k\rangle| \xrightarrow{a.s.} \delta_{jk}\zeta\left(\frac{\ell_j}{\sigma^2},\gamma\right), \qquad 1 \le j,k \le K. \tag{5}$$

Regularized tests under spiked covariance

- For $\Lambda = (\lambda_0, \lambda_1, \ldots, \lambda_K)$, define statistic

$$M(\Lambda) = \frac{1}{n} Q_n^T \mathbf{Y}^T \Omega(\Lambda) \mathbf{Y} Q_n$$

where

$$\Omega(\Lambda) = \sum_{j=1}^{K} \lambda_j \widehat{\mathbf{p}}_j \widehat{\mathbf{p}}_j^T + \lambda_0 I_p$$

Regularized tests under spiked covariance

- For $\Lambda = (\lambda_0, \lambda_1, \ldots, \lambda_K)$, define statistic

$$M(\Lambda) = \frac{1}{n} Q_n^T \mathbf{Y}^T \Omega(\Lambda) \mathbf{Y} Q_n$$

where

$$\Omega(\Lambda) = \sum_{j=1}^{K} \lambda_j \widehat{\mathbf{p}}_j \widehat{\mathbf{p}}_j^T + \lambda_0 I_p$$

- $\Omega(\Lambda)$ can be viewed as a "regularized estimator" of Σ_p^{-1}.

Regularized tests under spiked covariance

- For $\Lambda = (\lambda_0, \lambda_1, \ldots, \lambda_K)$, define statistic

$$M(\Lambda) = \frac{1}{n} Q_n^T \mathbf{Y}^T \Omega(\Lambda) \mathbf{Y} Q_n$$

where

$$\Omega(\Lambda) = \sum_{j=1}^{K} \lambda_j \widehat{\mathbf{p}}_j \widehat{\mathbf{p}}_j^T + \lambda_0 I_p$$

- $\Omega(\Lambda)$ can be viewed as a "regularized estimator" of Σ_p^{-1}.
- Define regularized test statistics $T_0^{LR}(\Lambda)$, $T_0^{LH}(\Lambda)$ and $T_0^{BNP}(\Lambda)$ by replacing M_0 with $M(\Lambda)$.

Other regularization schemes under spiked model

- **Bai & Saranadasa (1996):** (For two-sample tests) Replace $\widehat{\Sigma}_p^{-1}$ by I_p, which corresponds to taking $\lambda_0 = 1$ and $\lambda_j = 0$ for $j = 1, \ldots, K$. Nontrivial modifications due to Chen & Qin (2010).

Other regularization schemes under spiked model

- **Bai & Saranadasa (1996):** (For two-sample tests) Replace $\widehat{\Sigma}_p^{-1}$ by I_p, which corresponds to taking $\lambda_0 = 1$ and $\lambda_j = 0$ for $j = 1, \ldots, K$. Nontrivial modifications due to Chen & Qin (2010).
- **Aoshima & Yata (2018):** If $\ell_j \geq C\sqrt{p}$ for $1 \leq j \leq K$, for some $C > 0$, replace $\widehat{\Sigma}_p^{-1}$ by an estimate of $I_p - \sum_{j=1}^{K} \mathbf{p}_j \mathbf{p}_j^T$, of the form $I_p - \sum_{j=1}^{K} \alpha_j^2 \widehat{\mathbf{p}}_j \widehat{\mathbf{p}}_j^T$ where $(\alpha_j \widehat{\mathbf{p}}_j^T \mathbf{p}_j)^2 \to 1$.

Other regularization schemes under spiked model

- **Bai & Saranadasa (1996):** (For two-sample tests) Replace $\widehat{\Sigma}_p^{-1}$ by I_p, which corresponds to taking $\lambda_0 = 1$ and $\lambda_j = 0$ for $j = 1, \ldots, K$. Nontrivial modifications due to Chen & Qin (2010).
- **Aoshima & Yata (2018):** If $\ell_j \geq C\sqrt{p}$ for $1 \leq j \leq K$, for some $C > 0$, replace $\widehat{\Sigma}_p^{-1}$ by an estimate of $I_p - \sum_{j=1}^{K} \mathbf{p}_j \mathbf{p}_j^T$, of the form $I_p - \sum_{j=1}^{K} \alpha_j^2 \widehat{\mathbf{p}}_j \widehat{\mathbf{p}}_j^T$ where $(\alpha_j \widehat{\mathbf{p}}_j^T \mathbf{p}_j)^2 \to 1$.
- **Wang & Xu (2018):** If $\ell_j \geq C\sqrt{p}$ for $1 \leq j \leq K$, for some $C > 0$, replace $\widehat{\Sigma}_p^{-1}$ by $I_p - \sum_{j=1}^{K} \widehat{\mathbf{p}}_j \widehat{\mathbf{p}}_j^T$.

Other regularization schemes under spiked model

- **Bai & Saranadasa (1996):** (For two-sample tests) Replace $\widehat{\Sigma}_p^{-1}$ by I_p, which corresponds to taking $\lambda_0 = 1$ and $\lambda_j = 0$ for $j = 1, \ldots, K$. Nontrivial modifications due to Chen & Qin (2010).
- **Aoshima & Yata (2018):** If $\ell_j \geq C\sqrt{p}$ for $1 \leq j \leq K$, for some $C > 0$, replace $\widehat{\Sigma}_p^{-1}$ by an estimate of $I_p - \sum_{j=1}^{K} \mathbf{p}_j \mathbf{p}_j^T$, of the form $I_p - \sum_{j=1}^{K} \alpha_j^2 \widehat{\mathbf{p}}_j \widehat{\mathbf{p}}_j^T$ where $(\alpha_j \widehat{\mathbf{p}}_j^T \mathbf{p}_j)^2 \to 1$.
- **Wang & Xu (2018):** If $\ell_j \geq C\sqrt{p}$ for $1 \leq j \leq K$, for some $C > 0$, replace $\widehat{\Sigma}_p^{-1}$ by $I_p - \sum_{j=1}^{K} \widehat{\mathbf{p}}_j \widehat{\mathbf{p}}_j^T$.
- **Ma, Lan and Wang (2015):** Derived χ^2 limit of a slightly modified version of Bai & Saranadasa's test under strong spikes ($\ell_j \asymp p$ for $1 \leq j \leq K$.)

Asymptotic null distribution of $M(\Lambda)$

- Under $H_0 : BC = 0$, if $\lambda_0 > 0$, and assuming Gaussianity of noise,

$$\frac{\sqrt{p}(M(\Lambda) - \Theta_{1p}(\Lambda, \gamma) I_q)}{\sqrt{2\Theta_{2p}(\Lambda, \gamma)}} \Longrightarrow \mathbf{W}, \tag{6}$$

where $\mathbf{W}$ is $q \times q$ following standard GOE (Gaussian Orthogonal Ensemble).

Asymptotic null distribution of $M(\Lambda)$

- Under $H_0 : BC = 0$, if $\lambda_0 > 0$, and assuming Gaussianity of noise,

$$\frac{\sqrt{p}(M(\Lambda) - \Theta_{1p}(\Lambda,\gamma)I_q)}{\sqrt{2\Theta_{2p}(\Lambda,\gamma)}} \Longrightarrow \mathbf{W}, \tag{6}$$

where $\mathbf{W}$ is $q \times q$ following standard GOE (Gaussian Orthogonal Ensemble).

- With $\zeta_j = \zeta(\ell_j/\sigma^2, \gamma)$,

$$\Theta_{1p}(\Lambda,\gamma) = \frac{1}{p}\sum_{j=1}^{K} \lambda_j(\zeta_j^2\ell_j + (1-\zeta_j^2)\sigma^2) + \lambda_0\left(\frac{1}{p}\mathrm{tr}(\Sigma)\right)$$

$$\Theta_{2p}(\Lambda,\gamma) = \frac{1}{p}\sum_{j=1}^{K}\left[\lambda_j^2(\zeta_j^2\ell_j + (1-\zeta_j^2)\sigma^2)^2 + 2\lambda_0\lambda_j(\zeta_j^2\ell_j^2 + (1-\zeta_j^2)\sigma^4)\right] + \lambda_0^2\left(\frac{1}{p}\mathrm{tr}(\Sigma^2)\right)$$

Asymptotic null distribution of test statistics

Under $H_0 : BC = 0$, normalized versions of $T_0^{\mathrm{LR}}(\Lambda)$, $T_0^{\mathrm{LH}}(\Lambda)$ and $T_0^{\mathrm{BNP}}(\Lambda)$ are asymptotically normal for any Λ with $\lambda_0 > 0$:

$$T^{\mathrm{LR}}(\Lambda) := \frac{\sqrt{p}\{1 + \Theta_{1p}(\Lambda, \gamma)\}}{\{2q\Theta_{2p}(\Lambda, \gamma)\}^{1/2}} \left[T_0^{\mathrm{LR}}(\Lambda) - q \log\{1 + \Theta_{1p}(\Lambda, \gamma)\} \right] \Longrightarrow \mathcal{N}(0, 1),$$

$$T^{\mathrm{LH}}(\Lambda, \gamma) := \frac{\sqrt{p}}{\{2q\Theta_{2p}(\Lambda, \gamma)\}^{1/2}} \left[T_0^{\mathrm{LH}}(\Lambda) - q\Theta_{1p}(\Lambda, \gamma) \right] \Longrightarrow \mathcal{N}(0, 1),$$

$$T^{\mathrm{BNP}}(\Lambda, \gamma) := \frac{\sqrt{p}\{1 + \Theta_{1p}(\Lambda, \gamma)\}^2}{\{2q\Theta_{2p}(\Lambda, \gamma)\}^{1/2}} \left[T_0^{\mathrm{BNP}}(\Lambda, \gamma) - \frac{q\Theta_{1p}(\Lambda, \gamma)}{1 + \Theta_{1p}(\Lambda, \gamma)} \right] \Longrightarrow \mathcal{N}(0, 1)$$

Probabilistic local alternatives

- Consider alternatives of the form

$$BC = n^{-1/4}\mathbf{P}_K\mathbf{P}_K^T\mathcal{S} + n^{-1/4}p^{-1/2}\mathbf{P}_\perp\mathbf{P}_\perp^T\mathcal{Z}$$

(i) $\mathcal{S}$ is any $p \times q$ deterministic matrix;
(ii) $\mathcal{Z}$ is a $p \times q$ matrix $\sim \mathcal{MN}(0, \mathbf{U}, \mathbf{V})$, where $\mathbf{U}$ is the $p \times p$ row-wise covariance matrix and $\mathbf{V}$ is the $q \times q$ column-wise covariance matrix.

Probabilistic local alternatives

- Consider alternatives of the form

$$BC = n^{-1/4}\mathbf{P}_K\mathbf{P}_K^T\mathcal{S} + n^{-1/4}p^{-1/2}\mathbf{P}_\perp\mathbf{P}_\perp^T\mathcal{Z}$$

(i) $\mathcal{S}$ is any $p \times q$ deterministic matrix;
(ii) $\mathcal{Z}$ is a $p \times q$ matrix $\sim \mathcal{MN}(0, \mathbf{U}, \mathbf{V})$, where $\mathbf{U}$ is the $p \times p$ row-wise covariance matrix and $\mathbf{V}$ is the $q \times q$ column-wise covariance matrix.

- Define $\mathbf{R} = C^T(n^{-1}\mathbf{X}\mathbf{X}^T)^{-1}C$, and $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_K)$ with

$$\pi_0 = p^{-1}\mathrm{tr}[\mathbf{P}_\perp^T\mathbf{U}\mathbf{P}_\perp]\mathrm{tr}[\mathbf{R}^{-1}\mathbf{V}]$$

and

$$\pi_j = \mathbf{p}_j^T\mathcal{S}\mathbf{R}^{-1}\mathcal{S}^T\mathbf{p}_j, \qquad \text{for } \; j = 1, \ldots, K.$$

Probabilistic local alternatives

- Consider alternatives of the form

$$BC = n^{-1/4}\mathbf{P}_K\mathbf{P}_K^T\mathcal{S} + n^{-1/4}p^{-1/2}\mathbf{P}_\perp\mathbf{P}_\perp^T\mathcal{Z}$$

  (i) $\mathcal{S}$ is any $p \times q$ deterministic matrix;
  (ii) $\mathcal{Z}$ is a $p \times q$ matrix $\sim \mathcal{MN}(0, \mathbf{U}, \mathbf{V})$, where $\mathbf{U}$ is the $p \times p$ row-wise covariance matrix and $\mathbf{V}$ is the $q \times q$ column-wise covariance matrix.
- Define $\mathbf{R} = C^T(n^{-1}\mathbf{XX}^T)^{-1}C$, and $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_K)$ with

$$\pi_0 = p^{-1}\mathrm{tr}[\mathbf{P}_\perp^T\mathbf{U}\mathbf{P}_\perp]\mathrm{tr}[\mathbf{R}^{-1}\mathbf{V}]$$

  and

$$\pi_j = \mathbf{p}_j^T\mathcal{S}\mathbf{R}^{-1}\mathcal{S}^T\mathbf{p}_j, \qquad \text{for } \; j = 1, \ldots, K.$$

- $\boldsymbol{\pi}$ captures the projection of the alternative in the idiosyncratic noise and spiked subspaces.

Probabilistic local alternatives

- Consider alternatives of the form

$$BC = n^{-1/4}\mathbf{P}_K\mathbf{P}_K^T\mathcal{S} + n^{-1/4}p^{-1/2}\mathbf{P}_\perp\mathbf{P}_\perp^T\mathcal{Z}$$

  (i) $\mathcal{S}$ is any $p \times q$ deterministic matrix;
  (ii) $\mathcal{Z}$ is a $p \times q$ matrix $\sim \mathcal{MN}(0, \mathbf{U}, \mathbf{V})$, where $\mathbf{U}$ is the $p \times p$ row-wise covariance matrix and $\mathbf{V}$ is the $q \times q$ column-wise covariance matrix.
- Define $\mathbf{R} = C^T(n^{-1}\mathbf{X}\mathbf{X}^T)^{-1}C$, and $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_K)$ with

$$\pi_0 = p^{-1}\mathrm{tr}[\mathbf{P}_\perp^T\mathbf{U}\mathbf{P}_\perp]\mathrm{tr}[\mathbf{R}^{-1}\mathbf{V}]$$

  and

$$\pi_j = \mathbf{p}_j^T\mathcal{S}\mathbf{R}^{-1}\mathcal{S}^T\mathbf{p}_j, \qquad \text{for} \;\; j = 1, \ldots, K.$$

- $\boldsymbol{\pi}$ captures the projection of the alternative in the idiosyncratic noise and spiked subspaces.
- Set $\sum_{j=0}^{K}\pi_j = 1$ so that $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_K)$ is a prob. distribution.

Power under local alternatives

- At asymptotic level α, the regularized test that rejects for large values of $M(\Lambda)$ has asymptotic power (under the local alternative) $\beta_n(B,\Lambda) := \mathbb{P}_B(M(\Lambda) > \xi_\alpha)$, with

$$\beta_n(B,\Lambda) - \Phi\left(-\xi_\alpha + Q(\Lambda,\gamma;\boldsymbol{\pi})\right) \xrightarrow{P} 0, \tag{7}$$

Power under local alternatives

- At asymptotic level α, the regularized test that rejects for large values of $M(\Lambda)$ has asymptotic power (under the local alternative) $\beta_n(B,\Lambda) := \mathbb{P}_B(M(\Lambda) > \xi_\alpha)$, with

$$\beta_n(B,\Lambda) - \Phi\left(-\xi_\alpha + Q(\Lambda,\gamma;\boldsymbol{\pi})\right) \xrightarrow{P} 0, \tag{7}$$

- where

$$Q(\Lambda,\gamma;\boldsymbol{\pi}) = \frac{\mathcal{H}(\Lambda,\gamma;\boldsymbol{\pi})}{\sqrt{2\gamma q \Theta_2(\Lambda,\gamma)}}$$

with $\mathcal{H}(\Lambda,\gamma;\boldsymbol{\pi}) = \sum_{j=1}^{K} \lambda_j \zeta_j^2 \pi_j + \lambda_0 = \sum_{j=1}^{K} (\lambda_j \zeta_j^2 + \lambda_0)\pi_j + \lambda_0 \pi_0.$

Power under local alternatives

- At asymptotic level α, the regularized test that rejects for large values of $M(\Lambda)$ has asymptotic power (under the local alternative) $\beta_n(B,\Lambda) := \mathbb{P}_B(M(\Lambda) > \xi_\alpha)$, with

$$\beta_n(B,\Lambda) - \Phi\left(-\xi_\alpha + Q(\Lambda,\gamma;\boldsymbol{\pi})\right) \xrightarrow{P} 0, \tag{7}$$

- where

$$Q(\Lambda,\gamma;\boldsymbol{\pi}) = \frac{\mathcal{H}(\Lambda,\gamma;\boldsymbol{\pi})}{\sqrt{2\gamma q\Theta_2(\Lambda,\gamma)}}$$

with $\mathcal{H}(\Lambda,\gamma;\boldsymbol{\pi}) = \sum_{j=1}^{K} \lambda_j \zeta_j^2 \pi_j + \lambda_0 = \sum_{j=1}^{K} (\lambda_j \zeta_j^2 + \lambda_0)\pi_j + \lambda_0 \pi_0.$

- Power remains invariant under scalar multiplication of Λ and so, set $\|\Lambda\| = 1$.

Bayes test

- The Bayes test (within the class of tests determined by $M(\Lambda)$) w.r.t. prior π is obtained by maximizing (estimated) $Q(\Lambda, \gamma; \pi)$ with respect to Λ.

Bayes test

- The Bayes test (within the class of tests determined by $M(\Lambda)$) w.r.t. prior $\boldsymbol{\pi}$ is obtained by maximizing (estimated) $Q(\Lambda, \gamma; \boldsymbol{\pi})$ with respect to Λ.
- When Λ is unrestricted, this is an eigenvalue problem, with solution

$$\Lambda(\boldsymbol{\pi}) = \frac{\mathbf{E}^{-1}\mathbf{b}(\boldsymbol{\pi})}{\|\mathbf{E}^{-1}\mathbf{b}(\boldsymbol{\pi})\|}$$

where $\mathbf{b}(\boldsymbol{\pi}) = (1, \zeta_1^2\pi_1, \ldots, \zeta_K^2\pi_K)$ and

$$\mathbf{E} = \begin{bmatrix} c & \mathbf{a}^T \\ \mathbf{a} & \mathbf{D} \end{bmatrix},$$

Bayes test

- The Bayes test (within the class of tests determined by $M(\Lambda)$) w.r.t. prior $\boldsymbol{\pi}$ is obtained by maximizing (estimated) $Q(\Lambda, \gamma; \boldsymbol{\pi})$ with respect to Λ.
- When Λ is unrestricted, this is an eigenvalue problem, with solution

$$\Lambda(\boldsymbol{\pi}) = \frac{\mathbf{E}^{-1}\mathbf{b}(\boldsymbol{\pi})}{\|\mathbf{E}^{-1}\mathbf{b}(\boldsymbol{\pi})\|}$$

where $\mathbf{b}(\boldsymbol{\pi}) = (1, \zeta_1^2\pi_1, \ldots, \zeta_K^2\pi_K)$ and

$$\mathbf{E} = \begin{bmatrix} c & \mathbf{a}^T \\ \mathbf{a} & \mathbf{D} \end{bmatrix},$$

- Elements of $\mathbf{E}$ have known functional forms, and can be consistently estimated from data.

Minimax test

- Let Π be a class of priors (a subset of the unit simplex in $\mathbb{R}^{K+1}$). The minimax test (within the class determined by $M(\Lambda)$, and with respect to priors Π) can be found by minimizing $Q(\Lambda(\boldsymbol{\pi}), \gamma; \boldsymbol{\pi})$ over $\boldsymbol{\pi} \in \Pi$.

Minimax test

- Let Π be a class of priors (a subset of the unit simplex in $\mathbb{R}^{K+1}$). The minimax test (within the class determined by $M(\Lambda)$, and with respect to priors Π) can be found by minimizing $Q(\Lambda(\boldsymbol{\pi}), \gamma; \boldsymbol{\pi})$ over $\boldsymbol{\pi} \in \Pi$.
- If Π is determined by hyperplane constraints, finding the minimax test is a quadratic programming problem.

Minimax test

- Let Π be a class of priors (a subset of the unit simplex in $\mathbb{R}^{K+1}$). The minimax test (within the class determined by $M(\Lambda)$, and with respect to priors Π) can be found by minimizing $Q(\Lambda(\boldsymbol{\pi}), \gamma; \boldsymbol{\pi})$ over $\boldsymbol{\pi} \in \Pi$.
- If Π is determined by hyperplane constraints, finding the minimax test is a quadratic programming problem.
- Specifically, for $\varrho \in [0, 1]$, we focus on the class of priors

$$P(\varrho) = \{\boldsymbol{\pi} : \pi_0 \in [0, \varrho], \sum_{j=0}^{K} \pi_j = 1, \pi_j \geq 0 \text{ for } j = 1, \ldots, K\}$$

Minimax test

- Let Π be a class of priors (a subset of the unit simplex in $\mathbb{R}^{K+1}$). The minimax test (within the class determined by $M(\Lambda)$, and with respect to priors Π) can be found by minimizing $Q(\Lambda(\boldsymbol{\pi}), \gamma; \boldsymbol{\pi})$ over $\boldsymbol{\pi} \in \Pi$.
- If Π is determined by hyperplane constraints, finding the minimax test is a quadratic programming problem.
- Specifically, for $\varrho \in [0, 1]$, we focus on the class of priors

$$P(\varrho) = \{\boldsymbol{\pi} : \pi_0 \in [0, \varrho], \sum_{j=0}^{K} \pi_j = 1, \pi_j \geq 0 \text{ for } j = 1, \ldots, K\}$$

- If we restrict Λ such that all the $\lambda_j \geq 0$ for $j = 0, \ldots, K$, and Π is the unit simplex in $\mathbb{R}^{K+1}$ (i.e.,$\varrho = 1$), then the minimax test corresponds to $\lambda_0 = 1$ and $\lambda_1 = \cdots = \lambda_K = 0$. This is the Bai and Saranadasa's (1996) test!

Simulation to illustrate power characteristics

- Focus on the two sample test for equality of means:
  - $B = [\boldsymbol{\mu}_1 : \boldsymbol{\mu}_2]$, $C = (1, -1)^T$ so that $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$. Here $m = 2$ and $q = 1$.

Simulation to illustrate power characteristics

- Focus on the two sample test for equality of means:
  - $B = [\boldsymbol{\mu}_1 : \boldsymbol{\mu}_2]$, $C = (1, -1)^T$ so that $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$. Here $m = 2$ and $q = 1$.
  - $N_1 = 40$, $N_2 = 60$ (sample sizes for groups); so $N = N_1 + N_2 = 100$, and $n = N - m = 98$.

Simulation to illustrate power characteristics

- Focus on the two sample test for equality of means:
  - $B = [\boldsymbol{\mu}_1 : \boldsymbol{\mu}_2]$, $C = (1, -1)^T$ so that $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$. Here $m = 2$ and $q = 1$.
  - $N_1 = 40$, $N_2 = 60$ (sample sizes for groups); so $N = N_1 + N_2 = 100$, and $n = N - m = 98$.
  - $p = 50, 200, 1000$;

Simulation to illustrate power characteristics

- Focus on the two sample test for equality of means:
  - $B = [\boldsymbol{\mu}_1 : \boldsymbol{\mu}_2]$, $C = (1, -1)^T$ so that $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$. Here $m = 2$ and $q = 1$.
  - $N_1 = 40$, $N_2 = 60$ (sample sizes for groups); so $N = N_1 + N_2 = 100$, and $n = N - m = 98$.
  - $p = 50, 200, 1000$;
  - $K = 3$ spikes evenly spaced from $100\sigma^2(1 + \sqrt{\gamma})$ to $10\sigma^2(1 + \sqrt{\gamma})$.

Simulation to illustrate power characteristics

- Focus on the two sample test for equality of means:
  - $B = [\boldsymbol{\mu}_1 : \boldsymbol{\mu}_2]$, $C = (1, -1)^T$ so that $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$. Here $m = 2$ and $q = 1$.
  - $N_1 = 40$, $N_2 = 60$ (sample sizes for groups); so $N = N_1 + N_2 = 100$, and $n = N - m = 98$.
  - $p = 50, 200, 1000$;
  - $K = 3$ spikes evenly spaced from $100\sigma^2(1 + \sqrt{\gamma})$ to $10\sigma^2(1 + \sqrt{\gamma})$.
  - True distribution of signal $\mu = \mu_1 - \mu_2$ is given by $\boldsymbol{\pi} = h \times (\frac{1-\pi_0}{K}, \ldots, \frac{1-\pi_0}{K}, \pi_0)$ with $\pi_0 = 0, 0.5, 0.8, 1$, and $h > 0$.

Simulation to illustrate power characteristics

- Focus on the two sample test for equality of means:
  - $B = [\boldsymbol{\mu}_1 : \boldsymbol{\mu}_2]$, $C = (1, -1)^T$ so that $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$. Here $m = 2$ and $q = 1$.
  - $N_1 = 40$, $N_2 = 60$ (sample sizes for groups); so $N = N_1 + N_2 = 100$, and $n = N - m = 98$.
  - $p = 50, 200, 1000$;
  - $K = 3$ spikes evenly spaced from $100\sigma^2(1 + \sqrt{\gamma})$ to $10\sigma^2(1 + \sqrt{\gamma})$.
  - True distribution of signal $\mu = \mu_1 - \mu_2$ is given by $\boldsymbol{\pi} = h \times (\frac{1 - \pi_0}{K}, \ldots, \frac{1 - \pi_0}{K}, \pi_0)$ with $\pi_0 = 0, 0.5, 0.8, 1$, and $h > 0$.
- Consider the proposed minimax test with $\varrho = 0.2, 0.5, 0.8$.

Simulation to illustrate power characteristics

- Focus on the two sample test for equality of means:
  - $B = [\boldsymbol{\mu}_1 : \boldsymbol{\mu}_2]$, $C = (1, -1)^T$ so that $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$. Here $m = 2$ and $q = 1$.
  - $N_1 = 40$, $N_2 = 60$ (sample sizes for groups); so $N = N_1 + N_2 = 100$, and $n = N - m = 98$.
  - $p = 50, 200, 1000$;
  - $K = 3$ spikes evenly spaced from $100\sigma^2(1 + \sqrt{\gamma})$ to $10\sigma^2(1 + \sqrt{\gamma})$.
  - True distribution of signal $\mu = \mu_1 - \mu_2$ is given by $\boldsymbol{\pi} = h \times (\frac{1 - \pi_0}{K}, \ldots, \frac{1 - \pi_0}{K}, \pi_0)$ with $\pi_0 = 0, 0.5, 0.8, 1$, and $h > 0$.
- Consider the proposed minimax test with $\varrho = 0.2, 0.5, 0.8$.
- Compare with *LRT under factor model* (Anderson & Rubin, 1956): Test statistic: $\log\det(\widehat{\Sigma}_{red}\widehat{\Sigma}_{full}^{-1})$ [Bartlett approximation for null distribution if $p < n$].

Simulation to illustrate power characteristics

- Focus on the two sample test for equality of means:
  - $B = [\boldsymbol{\mu}_1 : \boldsymbol{\mu}_2]$, $C = (1, -1)^T$ so that $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$. Here $m = 2$ and $q = 1$.
  - $N_1 = 40$, $N_2 = 60$ (sample sizes for groups); so $N = N_1 + N_2 = 100$, and $n = N - m = 98$.
  - $p = 50, 200, 1000$;
  - $K = 3$ spikes evenly spaced from $100\sigma^2(1 + \sqrt{\gamma})$ to $10\sigma^2(1 + \sqrt{\gamma})$.
  - True distribution of signal $\mu = \mu_1 - \mu_2$ is given by $\boldsymbol{\pi} = h \times (\frac{1-\pi_0}{K}, \ldots, \frac{1-\pi_0}{K}, \pi_0)$ with $\pi_0 = 0, 0.5, 0.8, 1$, and $h > 0$.
- Consider the proposed minimax test with $\varrho = 0.2, 0.5, 0.8$.
- Compare with *LRT under factor model* (Anderson & Rubin, 1956): Test statistic: $\log\det(\widehat{\Sigma}_{red}\widehat{\Sigma}_{full}^{-1})$ [Bartlett approximation for null distribution if $p < n$].
- Number of spikes K is estimated using the eigenvalue thresholding scheme of Kritchman & Nadler (2008).

Empirical size of the tests

p	$\varrho = 0.2$			$\varrho = 0.5$			$\varrho = 0.8$			LRT*
	BNP	LH	LR	BNP	LH	LR	BNP	LH	LR	
50	3.55	6.41	4.90	2.93	5.70	4.30	2.93	5.70	4.30	0.05
200	5.82	6.79	6.28	4.54	6.22	5.38	4.23	6.02	5.08	NA
1000	5.78	5.97	5.88	5.09	5.47	5.32	4.60	5.23	4.96	NA

Table: Empirical sizes (as percentage) at 5% nominal significance level with normal critical values.

p	$\varrho = 0.2$			$\varrho = 0.5$			$\varrho = 0.8$			LRT*
	BNP	LH	LR	BNP	LH	LR	BNP	LH	LR	
50	4.80	4.80	4.80	4.42	4.42	4.42	4.40	4.40	4.40	0.05
200	4.68	4.68	4.68	4.84	4.84	4.84	4.80	4.80	4.80	NA
1000	3.70	3.70	3.70	3.68	3.68	3.68	3.84	3.84	3.84	NA

Table: Empirical sizes (as percentage) at 5% nominal significance level with (parametric) bootstrapped critical values.

Size-adjusted power: $p = 50$

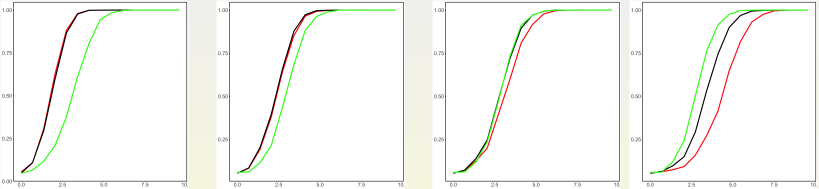

Figure: Size-adjusted empirical power when $p = 50$ and $(N_1, N_2) = (40, 60)$. From left to right: $\pi_0^{\mathrm{true}} = 0, 0.5, 0.8, 1$. LRT (Green), T^{LR} with $\varrho = 0.2$ (Red), $\varrho = 0.5$ (Blue), $\varrho = 0.8$ (Black).

Size-adjusted power: $p = 200$

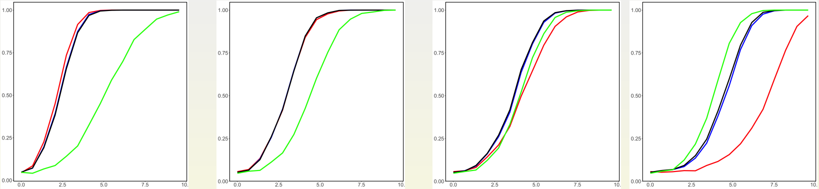

Figure: Size-adjusted empirical power when $p = 200$ and $(N_1, N_2) = (40, 60)$. From left to right: $\pi_0^{\mathrm{true}} = 0, 0.5, 0.8, 1$. LRT (Green), T^{LR} with $\varrho = 0.2$ (Red), $\varrho = 0.5$ (Blue), $\varrho = 0.8$ (Black).

Size-adjusted power: $p = 1000$

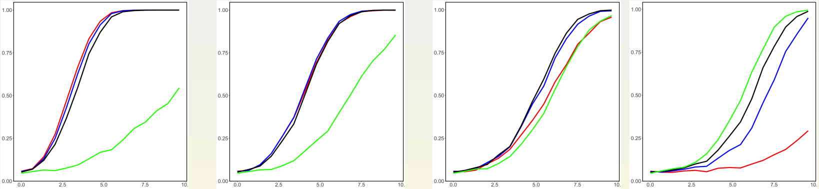

Figure: Size-adjusted empirical power when $p = 1000$ and $(N_1, N_2) = (40, 60)$. From left to right: $\pi_0^{\mathrm{true}} = 0, 0.5, 0.8, 1$. LRT (Green), T^{LR} with $\varrho = 0.2$ (Red), $\varrho = 0.5$ (Blue), $\varrho = 0.8$ (Black).

HCP data: regression model

- **Response:** $p = 127$ variables (denoted by $\mathbf{y}$) representing the test scores on different behavioral traits.

HCP data: regression model

- **Response:** $p = 127$ variables (denoted by $\mathbf{y}$) representing the test scores on different behavioral traits.

- **Predictors:** (i) Age and Gender groups (categorical, denoted by D); (ii) Volumetric measurements (SA = surface area, AT = average thickness, GV = gray-matter volume) of 14 cerebral lobes; (iii) Volumes (SC) of 44 subcortical structures. Total $m = 1 + 3 + 1 + 3 \times 14 + 38 = 85$ predictors.

HCP data: regression model

- **Response:** $p = 127$ variables (denoted by $\mathbf{y}$) representing the test scores on different behavioral traits.
- **Predictors:** (i) Age and Gender groups (categorical, denoted by D); (ii) Volumetric measurements (SA = surface area, AT = average thickness, GV = gray-matter volume) of 14 cerebral lobes; (iii) Volumes (SC) of 44 subcortical structures. Total $m = 1 + 3 + 1 + 3 \times 14 + 38 = 85$ predictors.
- **Model**:

$$\mathbf{y}_i = \beta_0 + \beta_1 \mathbf{D}_i + \beta_2 \mathbf{SA}_i + \beta_3 \mathbf{AT}_i + \beta_4 \mathbf{GV}_i + \beta_5 \mathbf{SC}_i + \varepsilon_i. \quad (8)$$

Spiked covariance structure of noise

We detect 12 spikes by using Kritchman & Nadler (2008) algorithm and estimate ℓ_j's and σ^2 using a method by Passemier, Li & Yao (2017).

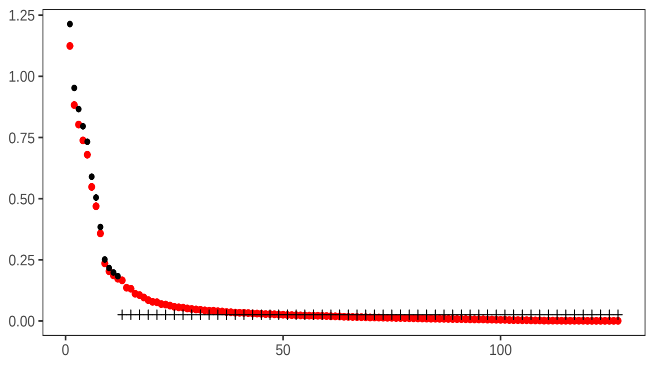

Figure: Empirical eigenvalues (Red), estimated spiked eigenvalues (Black dot), estimated noise variance (Black cross).

Detection of significant (marginal) association

- Individually, test for possible significance of each of the volumetric measurements. This corresponds to tests of the form $H_0^{kj} : \beta_{k,j} = 0$ where $1 \leq j \leq 14$ for $k = 2, 3, 4$ and $1 \leq j \leq 38$ for $k = 5$.

Detection of significant (marginal) association

- Individually, test for possible significance of each of the volumetric measurements. This corresponds to tests of the form $H_0^{kj} : \beta_{k,j} = 0$ where $1 \leq j \leq 14$ for $k = 2, 3, 4$ and $1 \leq j \leq 38$ for $k = 5$.

- We use the test $\widehat{T}^{\mathrm{LR}}(\Lambda)$ with minimax selection of Λ for two (extremal) choices of ϱ, viz., $\varrho = 0.001$ and $\varrho = 1$.

Detection of significant (marginal) association

- Individually, test for possible significance of each of the volumetric measurements. This corresponds to tests of the form $H_0^{kj} : \beta_{k,j} = 0$ where $1 \leq j \leq 14$ for $k = 2, 3, 4$ and $1 \leq j \leq 38$ for $k = 5$.
- We use the test $\widehat{T}^{\mathrm{LR}}(\Lambda)$ with minimax selection of Λ for two (extremal) choices of ϱ, viz., $\varrho = 0.001$ and $\varrho = 1$.
- Also, compare with the likelihood ratio test using the factor model structure by Anderson and Rubin (1956).

Significant marginal associations: *p*-values

Name	Type	test	*p*-value
Left Medial Temporal Lobe Area	Cortical	$\varrho = 1$	1.54×10^{-2}
		$\varrho = 0.001$	2.38×10^{-2}
		LRT	0.12
Left Occipital lobe Thickness	Cortical	$\varrho = 1$	4.8×10^{-2}
		$\varrho = 0.001$	4.18×10^{-2}
		LRT	0.79
Left Amygdala Volume	Subcortical	$\varrho = 1$	5.4×10^{-2}
		$\varrho = 0.001$	2.82×10^{-2}
		LRT	0.98
Left Caudate Volume	Subcortical	$\varrho = 1$	9×10^{-4}
		$\varrho = 0.001$	7.4×10^{-4}
		LRT	0.57
Right Caudate Volume	Subcortical	$\varrho = 1$	1.3×10^{-4}
		$\varrho = 0.001$	1.4×10^{-4}
		LRT	0.25
Right Cerebellum White-matter Volume	Subcortical	$\varrho = 1$	8.8×10^{-3}
		$\varrho = 0.001$	1.2×10^{-2}
		LRT	0.16

Significant marginal associations: *p*-values

Name	Type	test	*p*-value
Right Hippocampus Volume	Subcortical	$\varrho = 1$	3.2×10^{-3}
		$\varrho = 0.001$	5.3×10^{-3}
		LRT	0.07
Right Choroid Plexus Volume	Subcortical	$\varrho = 1$	0.52
		$\varrho = 0.001$	2.8×10^{-2}
		LRT	0.95
Brain Stem Volume	Subcortical	$\varrho = 1$	9×10^{-7}
		$\varrho = 0.001$	1.94×10^{-7}
		LR	0.53
Corpus Callosum Anterior Volume	Subcortical	$\varrho = 1$	2.38×10^{-2}
		$\varrho = 0.001$	3.84×10^{-2}
		LR	0.11
White-matter Hypointensity Volume	Subcortical	$\varrho = 1$	2.84×10^{-2}
		$\varrho = 0.001$	2.59×10^{-2}
		LR	0.77

Interpreting the findings in light of neuroscience

- *Amygdala* performs primary roles in the formation and storage of memories associated with emotional events (Maren, 1999).

Interpreting the findings in light of neuroscience

- *Amygdala* performs primary roles in the formation and storage of memories associated with emotional events (Maren, 1999).
- *Hippocampus* plays an important role in the formation of new memories about experienced events (Eichenbaum et al., 1993).

Interpreting the findings in light of neuroscience

- *Amygdala* performs primary roles in the formation and storage of memories associated with emotional events (Maren, 1999).
- *Hippocampus* plays an important role in the formation of new memories about experienced events (Eichenbaum et al., 1993).
- In the set of behavioral variables (i.e., the response), there are emotion processing tasks that may lead to activation of *amygdala* and *hippocampus* (Barch et al., 2013).

Interpreting the findings in light of neuroscience

- *Amygdala* performs primary roles in the formation and storage of memories associated with emotional events (Maren, 1999).
- *Hippocampus* plays an important role in the formation of new memories about experienced events (Eichenbaum et al., 1993).
- In the set of behavioral variables (i.e., the response), there are emotion processing tasks that may lead to activation of *amygdala* and *hippocampus* (Barch et al., 2013).
- *Medial temporal lobe* contains *parahippocampal cortex* and *entorhinal cortex* that are among the primary regions deemed responsible for the formation of memories and spatial cognition. These cortices are anatomically adjacent to and functionally communicate with *amygdala* and *hippocampus* (Koob et al., 2010).

Summary

- Regularized tests for linear hypotheses based on shrinkage of the eigenvalues of the sample residual covariance matrix.

Summary

- Regularized tests for linear hypotheses based on shrinkage of the eigenvalues of the sample residual covariance matrix.
- Asymptotic theory for the proposed tests under the $p \asymp n$ regime.

Summary

- Regularized tests for linear hypotheses based on shrinkage of the eigenvalues of the sample residual covariance matrix.
- Asymptotic theory for the proposed tests under the $p \asymp n$ regime.
- A Bayesian paradigm for selecting the regularization parameter using the notion of locally most powerful tests.

Summary

- Regularized tests for linear hypotheses based on shrinkage of the eigenvalues of the sample residual covariance matrix.
- Asymptotic theory for the proposed tests under the $p \asymp n$ regime.
- A Bayesian paradigm for selecting the regularization parameter using the notion of locally most powerful tests.
- Derivation of a minimax test within the proposed family of regularized tests.

Summary

- Regularized tests for linear hypotheses based on shrinkage of the eigenvalues of the sample residual covariance matrix.
- Asymptotic theory for the proposed tests under the $p \asymp n$ regime.
- A Bayesian paradigm for selecting the regularization parameter using the notion of locally most powerful tests.
- Derivation of a minimax test within the proposed family of regularized tests.
- Surprising near-optimality of Bai & Sarandasa (1996)'s test.

Summary

- Regularized tests for linear hypotheses based on shrinkage of the eigenvalues of the sample residual covariance matrix.
- Asymptotic theory for the proposed tests under the $p \asymp n$ regime.
- A Bayesian paradigm for selecting the regularization parameter using the notion of locally most powerful tests.
- Derivation of a minimax test within the proposed family of regularized tests.
- Surprising near-optimality of Bai & Sarandasa (1996)'s test.
- R software package `ARHT` available on CRAN