INFERENCE ON DIRECTIONS UNDER WEAK IDENTIFIABILITY

Davy PAINDAVEINE, Julien REMY, Thomas VERDEBOUT

21 December 2022

Weak identifiability

Let $(\mathbf{X}_{ni})_n$ be a triangular array of idd observations with a joint probability distribution $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\boldsymbol{\xi}_n}$ indexed by a sequence of parameters $(\boldsymbol{\theta}, \boldsymbol{\xi}_n) \in \boldsymbol{\Theta} \times \boldsymbol{\Xi}$.

Weak identifiability

Let $(\mathbf{X}_{ni})_n$ be a triangular array of idd observations with a joint probability distribution $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\boldsymbol{\xi}_n}$ indexed by a sequence of parameters $(\boldsymbol{\theta},\boldsymbol{\xi}_n) \in \boldsymbol{\Theta} \times \boldsymbol{\Xi}$. Now, assume that

(i) for any n, the parameter $\boldsymbol{\theta}$ is well identified in the sense that, for any n, if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_1,\boldsymbol{\xi}_n} \neq \mathrm{P}^{(n)}_{\boldsymbol{\theta}_2,\boldsymbol{\xi}_n}$

Weak identifiability

Let $(\mathbf{X}_{ni})_n$ be a triangular array of idd observations with a joint probability distribution $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\boldsymbol{\xi}_n}$ indexed by a sequence of parameters $(\boldsymbol{\theta},\boldsymbol{\xi}_n) \in \boldsymbol{\Theta} \times \boldsymbol{\Xi}$. Now, assume that

(i) for any n, the parameter $\boldsymbol{\theta}$ is well identified in the sense that, for any n, if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_1,\boldsymbol{\xi}_n} \neq \mathrm{P}^{(n)}_{\boldsymbol{\theta}_2,\boldsymbol{\xi}_n}$ while

(ii) the parameter $\boldsymbol{\theta}$ is not properly identified in $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\boldsymbol{\xi}_0}$ where $\boldsymbol{\xi}_0 := \lim_{n\to\infty} \boldsymbol{\xi}_n$.

The parameter $\boldsymbol{\theta}$ will then be said to be *weakly identified*.

WI : Example

Consider random directions sharing a common rotationally symmetric distribution over $\mathcal{S}^{p-1}$, i.e. having a density of the form

$$\mathbf{x} \mapsto c_{p,\kappa,f} f(\kappa \mathbf{x}'\boldsymbol{\theta}), \tag{1}$$

where $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$ is a location parameter, $\kappa > 0$ is a concentration parameter, and the angular function f belongs to the collection $\mathcal{F}$ of monotone increasing functions from $\mathbb{R}$ to $\mathbb{R}^+$.

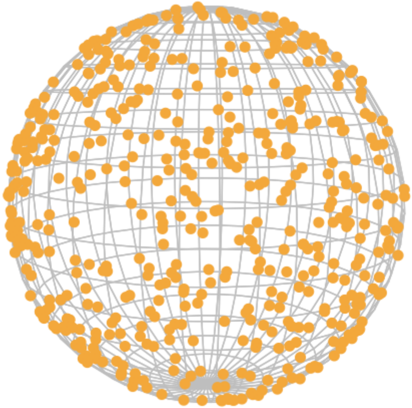
WI : Example

$\kappa = 10$

WI : Example

$\kappa = 3$

WI : Example



$\kappa \to 0$

Section 1

Testing for principal component directions under weak identifiability

Motivation example

Flury (1988) conducted a Principal Component Analysis (PCA) of the (celebrated) Swiss banknotes data. Flury (1988) focused on four measurements, namely the width L of the left side of the banknote, the width R on its right side, the width B of the bottom margin and the width T of the top margin, all measured in mm$\times 10^{-1}$ on $n = 85$ counterfeit bills made by the same forger.

The resulting sample covariance matrix is

$$\mathbf{S} = \begin{pmatrix} 6.41 & 4.89 & 2.89 & -1.30 \\ 4.89 & 9.40 & -1.09 & 0.71 \\ 2.89 & -1.09 & 72.42 & -43.30 \\ -1.30 & 0.71 & -43.30 & 40.39 \end{pmatrix},$$

with eigenvalues of $\hat{\lambda}_1 = 102.69$, $\hat{\lambda}_2 = 13.05$, $\hat{\lambda}_3 = 10.23$ and $\hat{\lambda}_4 = 2.66$, and corresponding eigenvectors :

$$\hat{\boldsymbol{\theta}}_1 = \begin{pmatrix} .032 \\ -.012 \\ .820 \\ -.571 \end{pmatrix} \quad \hat{\boldsymbol{\theta}}_2 = \begin{pmatrix} .593 \\ .797 \\ .057 \\ .097 \end{pmatrix}$$

$$\hat{\boldsymbol{\theta}}_3 = \begin{pmatrix} -.015 \\ -.129 \\ .566 \\ .814 \end{pmatrix} \quad \hat{\boldsymbol{\theta}}_4 = \begin{pmatrix} .804 \\ -.590 \\ -.064 \\ -.035 \end{pmatrix}$$

Flury concludes that the first principal component is a contrast between B and T. It is tempting to interpret the second principal component as an aggregate of L and R. Flury, however, explicitly writes *"beware : the second and third roots are quite close and so the computation of standard errors for the coefficients of* $\hat{\boldsymbol{\theta}}_2$ *and* $\hat{\boldsymbol{\theta}}_3$ *may be hazardous"*. In other words, Flury, due to the structure of the spectrum, refrains from drawing any conclusion about the second principal component.

Question : can we say something about the true underlying eigenvector $\boldsymbol{\theta}_2$ when the true underlying eigenvalues λ_2 and λ_3 are "very close to each other" ? That is under a situation of *weak identifiability* of $\boldsymbol{\theta}_2$?

Question : can we say something about the true underlying eigenvector $\boldsymbol{\theta}_2$ when the true underlying eigenvalues λ_2 and λ_3 are "very close to each other" ? That is under a situation of *weak identifiability* of $\boldsymbol{\theta}_2$?

Testing problem : we consider the problem of testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$ against the alternative $\mathcal{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^0$, where $\boldsymbol{\theta}_1^0$ is a given unit vector of $\mathbb{R}^p$. We will consider situations where $\lambda_1 - \lambda_2$ is small.

Working context

- Triangular array of Gaussian vectors.
- Single spiked spectra :

$$\begin{aligned}\boldsymbol{\Sigma}_n := \sigma_n^2 \left(\mathbf{I}_p + r_n v \boldsymbol{\theta}_1 \boldsymbol{\theta}_1'\right) \\ = \sigma_n^2 (1 + r_n v) \boldsymbol{\theta}_1 \boldsymbol{\theta}_1' + \sigma_n^2 \left(\mathbf{I}_p - \boldsymbol{\theta}_1 \boldsymbol{\theta}_1'\right)\end{aligned}$$

- Weak identifiability occurs when $r_n \to 0$, since this implies $\lambda_1 / \lambda_2 \to 1$.

A torough asymptotic investigation of this problem requires to discuss four different regimes :

(i) $r_n \equiv 1$;

(ii) $r_n = o(1)$ with $\sqrt{n} r_n \to \infty$;

(iii) $r_n = 1/\sqrt{n}$;

(iv) $r_n = o(1/\sqrt{n})$.

Under the null : Anderson's test

The likelihood ratio test rejects the null at asymptotic level α when

$$Q_{\mathrm{A}}^{(n)} := n\big(\hat{\lambda}_1 \boldsymbol{\theta}_1^{0\prime} \mathbf{S}^{-1} \boldsymbol{\theta}_1^0 + \hat{\lambda}_1^{-1} \boldsymbol{\theta}_1^{0\prime} \mathbf{S}\, \boldsymbol{\theta}_1^0 - 2\big) > \chi^2_{p-1,1-\alpha}.$$

Under the null : Anderson's test

The likelihood ratio test rejects the null at asymptotic level α when

$$Q_{\mathrm{A}}^{(n)} := n\big(\hat{\lambda}_1 \boldsymbol{\theta}_1^{0\prime} \mathbf{S}^{-1} \boldsymbol{\theta}_1^0 + \hat{\lambda}_1^{-1} \boldsymbol{\theta}_1^{0\prime} \mathbf{S}\, \boldsymbol{\theta}_1^0 - 2\big) > \chi^2_{p-1,1-\alpha}.$$

Theorem

Fix a unit p-vector $\boldsymbol{\theta}_1^0$, $v > 0$ and a nonnegative real sequence (r_n) satisfying (i) $r_n \equiv 1$ or (ii) $r_n = o(1)$ with $\sqrt{n} r_n \to \infty$. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_1^0, r_n, v}$,

$$Q_{\mathrm{A}}^{(n)} \xrightarrow{\mathcal{D}} \chi^2_{p-1},$$

so that, in regimes (i)-(ii), the test ϕ_{A} has asymptotic size α under the null.

Under the null : Le Cam optimal test

This test rejects the null at asymptotic level α when

$$Q_{\mathrm{HPV}}^{(n)} := \frac{n}{\hat{\lambda}_1} \sum_{j=2}^{p} \hat{\lambda}_j^{-1} \big(\tilde{\boldsymbol{\theta}}_j' \mathbf{S} \boldsymbol{\theta}_1^0 \big)^2 > \chi^2_{p-1,1-\alpha}.$$

Under the null : Le Cam optimal test

This test rejects the null at asymptotic level α when

$$Q_{\rm HPV}^{(n)} := \frac{n}{\hat{\lambda}_1} \sum_{j=2}^{p} \hat{\lambda}_j^{-1} \left(\tilde{\boldsymbol{\theta}}_j' \mathbf{S} \boldsymbol{\theta}_1^0\right)^2 > \chi^2_{p-1,1-\alpha}.$$

Theorem

Fix a unit p-vector $\boldsymbol{\theta}_1^0$, $v > 0$ and a bounded nonnegative real sequence (r_n). Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_1^0, r_n, v}$,

$$Q_{\rm HPV}^{(n)} \xrightarrow{\mathcal{D}} \chi^2_{p-1},$$

so that, in all regimes (i)-(iv) from the previous section, the test $\phi_{\rm HPV}$ has asymptotic size α under the null.

Under the null : Simulations

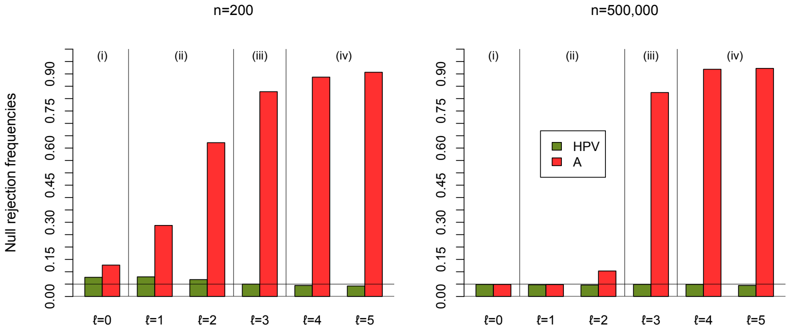


Figure – Empirical rejection frequencies of the tests $\phi^{(n)}_{\mathrm{HPV}}$ and $\phi^{(n)}_{\mathrm{A}}$ performed at nominal level 5%. Results are based on $M = 10,000$ independent ten-dimensional Gaussian random samples.

Under the null : Anderson's test

Theorem

Fix $p = 2$, a unit p-vector $\boldsymbol{\theta}_1^0$, $v > 0$ and a nonnegative real sequence (r_n) such that $\sqrt{n} r_n \to 0$. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_1^0, r_n, v}$,

$$Q_{\mathrm{A}}^{(n)} \xrightarrow{\mathcal{D}} 4\chi^2_{p-1},$$

so that, irrespective of $\alpha \in (0,1)$, the test $\phi_{\mathrm{A}}^{(n)}$ has an asymptotic size under the null that is strictly larger than α.

Summary

- Unlike $\phi_{\mathrm{A}}^{(n)}$, the test $\phi_{\mathrm{HPV}}^{(n)}$ is validity-robust to weak identifiability ;

Summary

- Unlike $\phi_{\mathrm{A}}^{(n)}$, the test $\phi_{\mathrm{HPV}}^{(n)}$ is validity-robust to weak identifiability ;
- but the trivial level-α test, that randomly rejects the null with probability α, enjoys the same robustness property ;

Summary

- Unlike $\phi_{\rm A}^{(n)}$, the test $\phi_{\rm HPV}^{(n)}$ is validity-robust to weak identifiability ;
- but the trivial level-α test, that randomly rejects the null with probability α, enjoys the same robustness property ;

$\Rightarrow$ it motivates to investigate whether or not the validity-robustness of $\phi_{\rm HPV}^{(n)}$ is obtained at the expense of efficiency.

Non-null results

Theorem

Fix $\boldsymbol{\theta}_1^0 \in S^{p-1}$. Let $(\boldsymbol{\tau}_n)_n$ be a sequence converging to $\boldsymbol{\tau}$ and such that $\boldsymbol{\theta}_0 + \nu_n \boldsymbol{\tau}_n \in S^{p-1}$ for any n. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_1^0 + \nu_n \boldsymbol{\tau}_n, r_n, v}$, we have, as $n \to \infty$, that $Q^{(n)}_{\mathrm{HPV}}$ is asymptotically non-central chi-square with $p-1$ degrees of freedom and with non-centrality parameter equal to :

- if $r_n \equiv 1$
  - $\looparrowright$ $(v^2/(1+v))\|\boldsymbol{\tau}\|^2$,
- if $r_n = o(1)$ with $\sqrt{n} r_n \to \infty$
  - $\looparrowright$ $v^2\|\boldsymbol{\tau}\|^2$,
- if $r_n = \frac{1}{\sqrt{n}}$
  - $\looparrowright$ $\frac{v^2}{16}\|\boldsymbol{\tau}\|^2\left(4 - \|\boldsymbol{\tau}\|^2\right)\left(2 - \|\boldsymbol{\tau}\|^2\right)^2$,
- if $r_n\sqrt{n} \to 0$
  - $\looparrowright$ it has no non-centrality parameter.

What about optimality?

By studying the present hypothesis testing context through the Le Cam theory, one can show that the sequence of models is LAN in regimes (i), (ii) and (iv).

This leads to the conclusion that $\phi^{(n)}_{\mathrm{HPV}}$ is optimal (locally and asymptotically) in these regimes. Note that the optimality in regime (iv) is trivial, in the sense that no test can detect the most severe alternatives.

For the regime (iii), unfortunately we don't have such a LAN situation. But $\phi^{(n)}_{\mathrm{HPV}}$ is rate-consistent .

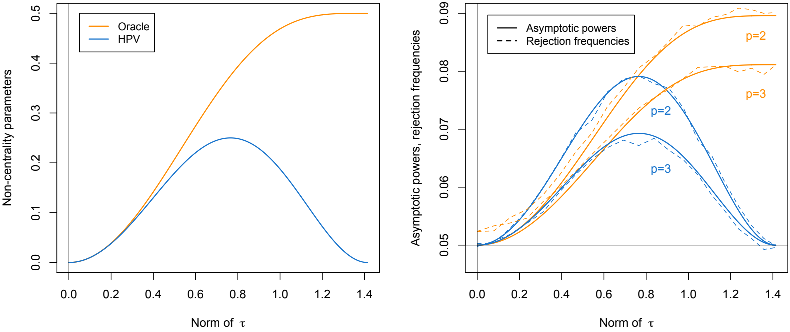


Figure – (Left :) Non-centrality parameters, as a function of $\|\boldsymbol{\tau}\| (\in [0, \sqrt{2}])$, in the asymptotic non-central chi-square distributions of the test statistics of $\phi^{(n)}_{\mathrm{HPV}}$ and $\phi^{(n)}_{\mathrm{oracle}}$, respectively, under alternatives of the form $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^0_1+\boldsymbol{\tau},1/\sqrt{n},1}$. (Right :) The corresponding asymptotic power curves in dimensions $p = 2$ and $p = 3$.

Conclusion

We saw here that the test $\phi^{(n)}_{\mathrm{HPV}}$ is

- validity-robust to weak identifiability,
- essentially locally and asymptotically optimal.

A possible research perspective is to look at this problem in the high-dimensional case.

End of the story ?

A new angle

The previous tests are based on the sample covariance matrix

A new angle

The previous tests are based on the sample covariance matrix

$$\Downarrow$$

They can't deal with heavy tails and are very sensitive to outliers

A new angle

The previous tests are based on the sample covariance matrix

$$\Downarrow$$

They can't deal with heavy tails and are very sensitive to outliers

$$\Downarrow$$

Sign tests

Section 2

Sign Tests for Weak Principal Directions

Angular Gaussian Distribution

Let $\mathbf{X}$ be a centered elliptical distributed variable with covariance matrix $\mathbf{\Sigma}_n$.

Then the projected observation $\mathbf{U} = \mathbf{X}/\|\mathbf{X}\|$ follows the p-variate angular Gaussian distribution with shape matrix $\mathbf{V}_n = p\mathbf{\Sigma}_n / \operatorname{tr}(\mathbf{\Sigma}_n)$.

Working context

- Triangular array of Angular Gaussian vectors.
- Single spiked spectra :

$$\mathbf{V}_n = \left(1 - \frac{\delta_n \epsilon}{p}\right) \mathbf{I}_p + \delta_n \epsilon \boldsymbol{\theta}_1 \boldsymbol{\theta}_1'$$

- Weak identifiability occurs when $\delta_n \to 0$, since this implies $\lambda_1/\lambda_2 \to 1$.

We could discern four different regimes :

(i) $\delta_n \equiv 1$;

(ii) $\delta_n = o(1)$ with $\sqrt{n}\delta_n \to \infty$;

(iii) $\delta_n = 1/\sqrt{n}$;

(iv) $\delta_n = o(1/\sqrt{n})$.

Our proposed sign test

This test rejects the null hypothesis at constraint level α if

$$Q_{\mathrm{Sgn}}^{(n)}(\tilde{\mathbf{V}}_{0n}) = np(p+2)\|(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}_0')\mathbf{S}_n(\tilde{\mathbf{V}}_{0n})\boldsymbol{\theta}_0\|^2 > \chi^2_{p-1,1-\alpha},$$

where

$$\mathbf{S}_n(\mathbf{V}) = \frac{1}{n}\sum_1^n \frac{\mathbf{V}^{-\frac{1}{2}}\mathbf{U}_{ni}\mathbf{U}_{ni}'\mathbf{V}^{-\frac{1}{2}}}{\|\mathbf{V}^{-\frac{1}{2}}\mathbf{U}_{ni}\|^2},$$

and $\tilde{\mathbf{V}}$ is the Tyler's M-estimator of $\mathbf{V}$.

Our proposed sign test

This test rejects the null hypothesis at constraint level α if

$$Q_{\mathrm{Sgn}}^{(n)}(\tilde{\mathbf{V}}_{0n}) = np(p+2)\|(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}_0')\mathbf{S}_n(\tilde{\mathbf{V}}_{0n})\boldsymbol{\theta}_0\|^2 > \chi^2_{p-1,1-\alpha},$$

where

$$\mathbf{S}_n(\mathbf{V}) = \frac{1}{n}\sum_{1}^{n} \frac{\mathbf{V}^{-\frac{1}{2}}\mathbf{U}_{ni}\mathbf{U}_{ni}'\mathbf{V}^{-\frac{1}{2}}}{\|\mathbf{V}^{-\frac{1}{2}}\mathbf{U}_{ni}\|^2},$$

and $\tilde{\mathbf{V}}$ is the Tyler's M-estimator of $\mathbf{V}$.

Theorem

Let $\mathbf{V}_{0n}$ be any sequence of null shape matrix of the form $\lambda_1\boldsymbol{\theta}_0\boldsymbol{\theta}_0' + \sum_2^j \lambda_j\boldsymbol{\theta}_j\boldsymbol{\theta}_j'$. Then under $\mathrm{P}^{(n)}_{\mathbf{V}_{0n}}$, $Q_{\mathrm{Sgn}}^{(n)}(\tilde{\mathbf{V}}_{0n})$ is asymptotically chi-square with $p-1$ degrees of freedom.

Our proposed sign test

This test rejects the null hypothesis at constraint level α if

$$Q^{(n)}_{\mathrm{Sgn}}(\tilde{\mathbf{V}}_{0n}) = np(p+2)\|(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}_0')\mathbf{S}_n(\tilde{\mathbf{V}}_{0n})\boldsymbol{\theta}_0\|^2 > \chi^2_{p-1,1-\alpha},$$

where

$$\mathbf{S}_n(\mathbf{V}) = \frac{1}{n}\sum_{1}^{n} \frac{\mathbf{V}^{-\frac{1}{2}}\mathbf{U}_{ni}\mathbf{U}_{ni}'\mathbf{V}^{-\frac{1}{2}}}{\|\mathbf{V}^{-\frac{1}{2}}\mathbf{U}_{ni}\|^2},$$

and $\tilde{\mathbf{V}}$ is the Tyler's M-estimator of $\mathbf{V}$.

Theorem

Let $\mathbf{V}_{0n}$ be any sequence of null shape matrix of the form $\lambda_1\boldsymbol{\theta}_0\boldsymbol{\theta}_0' + \sum_2^j \lambda_j\boldsymbol{\theta}_j\boldsymbol{\theta}_j'$. Then under $\mathrm{P}^{(n)}_{\mathbf{V}_{0n}}$, $Q^{(n)}_{\mathrm{Sgn}}(\tilde{\mathbf{V}}_{0n})$ is asymptotically chi-square with $p-1$ degrees of freedom.

Note that this remains true for multispike setting !

An alternative test

To test this problem, one can also consider the likelihood ratio test $\phi_{\mathrm{Tyl}}^{(n)}$, that rejects the nul hypothesis at constraint level α if

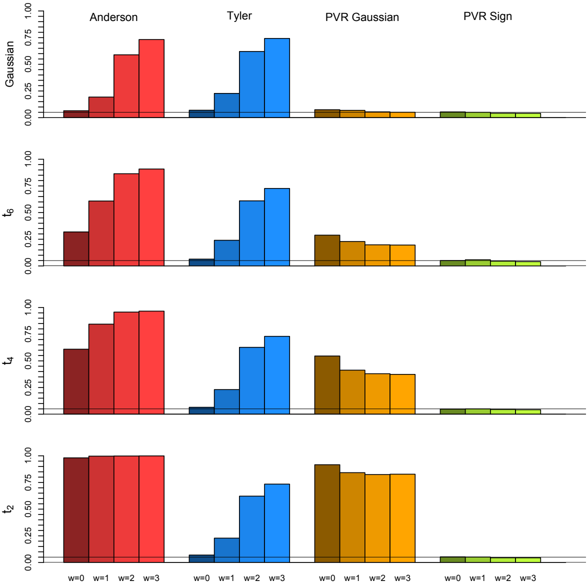
$$Q_{\mathrm{Tyl}}^{(n)} = \frac{np}{p+2}(\hat{\lambda}_{n1}\boldsymbol{\theta}_0'\hat{\mathbf{V}}_n^{-1}\boldsymbol{\theta}_0 + \hat{\lambda}_{n1}^{-1}\boldsymbol{\theta}_0'\hat{\mathbf{V}}_n^{-1}\boldsymbol{\theta}_0 - 2) > \chi^2_{p-1,1-\alpha},$$

where $\hat{\mathbf{V}}_n$ still stands for Tyler's M-estimator.

Theorem

Away from weak identifiability, both test are asymptotically equivalent.

Simulations



Asymptotics of the proposed test

Theorem

Fix $\boldsymbol{\theta}_0 \in S^{p-1}$. Let $(\boldsymbol{\tau}_n)_n$ be a sequence converging to $\boldsymbol{\tau}$ and such that $\boldsymbol{\theta}_0 + \nu_n \boldsymbol{\tau}_n \in S^{p-1}$ for any n; and take $\nu_n = 1/(\sqrt{n}\gamma_n)$. Then, under $\mathrm{P}^{(n)}_{\mathbf{V}_{1n}}$, we have, as $n \to \infty$, that $Q^{(n)}_{\mathrm{Sgn}}(\tilde{\mathbf{V}}_{0n})$ is asymptotically chi-square with $p-1$ degrees of freedom and non-centrality parameter equal to :

- if $\delta_n \equiv 1$
  - ↬ $\frac{p(p+(p-1)\xi)}{(p+2)(p-\xi)} \|\boldsymbol{\tau}\|^2$,
- if $\delta_n = o(1)$ with $\sqrt{n}\delta_n \to \infty$
  - ↬ $\frac{p}{p+2} \|\boldsymbol{\tau}\|^2$,
- if $\delta_n = \frac{1}{\sqrt{n}}$
  - ↬ $\frac{p}{p+2} \|\boldsymbol{\tau}\|^2 \left(1 - \frac{1}{2\xi^2}\|\boldsymbol{\tau}\|^2\right)^2 \left(1 - \frac{1}{4\xi^2}\|\boldsymbol{\tau}\|^2\right)$,
- if $\delta_n \sqrt{n} \to 0$
  - ↬ it has no non-centrality parameter.

Notations

- $\xi > 0$ is a locality parameter,
- δ_n is a bounded sequence,
- $\gamma_n = \frac{p\delta_n\xi}{p+(p-1)\delta_n\xi} = O(\delta_n)$.

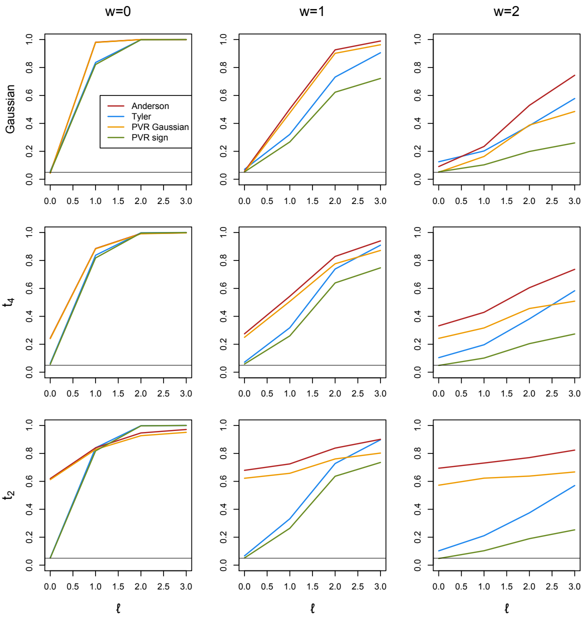


Figure – Rejection frequencies for the four tests considered, in different heavy tail/weak identifiability settings. Simulation were done on 2500 samples of 200 independent bivariate variables.

What about optimality ?

By studying the present hypothesis testing context through the Le Cam theory, one can show that the sequence of models is LAN in regimes (i), (ii) and (iv).

This leads to the conclusion that our sign test is optimal (locally and asymptotically) in these regimes. Note that the optimality in regime (iv) is trivial, in the sense that no test can detect the most severe alternatives.

For the regime (iii), unfortunately we don't have such a LAN situation. But our test is rate-consistent (it shows non-trivial asymptotics power against contiguous alternatives).

Back to the starting example

Flury (1988) conducted a Principal Component Analysis (PCA) of the (celebrated) Swiss banknotes data. Flury (1988) focused on four measurements, namely the width L of the left side of the banknote, the width R on its right side, the width B of the bottom margin and the width T of the top margin, all measured in mm$\times 10^{-1}$ on $n = 85$ counterfeit bills made by the same forger.

The resulting sample covariance matrix is

$$\mathbf{S} = \begin{pmatrix} 6.41 & 4.89 & 2.89 & -1.30 \\ 4.89 & 9.40 & -1.09 & 0.71 \\ 2.89 & -1.09 & 72.42 & -43.30 \\ -1.30 & 0.71 & -43.30 & 40.39 \end{pmatrix},$$

with eigenvalues of $\hat{\lambda}_1 = 102.69$, $\hat{\lambda}_2 = 13.05$, $\hat{\lambda}_3 = 10.23$ and $\hat{\lambda}_4 = 2.66$, and corresponding eigenvectors :

$$\hat{\boldsymbol{\theta}}_1 = \begin{pmatrix} .032 \\ -.012 \\ .820 \\ -.571 \end{pmatrix} \quad \hat{\boldsymbol{\theta}}_2 = \begin{pmatrix} .593 \\ .797 \\ .057 \\ .097 \end{pmatrix}$$

$$\hat{\boldsymbol{\theta}}_3 = \begin{pmatrix} -.015 \\ -.129 \\ .566 \\ .814 \end{pmatrix} \quad \hat{\boldsymbol{\theta}}_4 = \begin{pmatrix} .804 \\ -.590 \\ -.064 \\ -.035 \end{pmatrix}$$

Flury concludes that the first principal component is a contrast between B and T. It is tempting to interpret the second principal component as an aggregate of L and R. Flury, however, explicitly writes *“beware : the second and third roots are quite close and so the computation of standard errors for the coefficients of* $\hat{\boldsymbol{\theta}}_2$ *and* $\hat{\boldsymbol{\theta}}_3$ *may be hazardous”*. In other words, Flury, due to the structure of the spectrum, refrains from drawing any conclusion about the second component.

The considerations above make it natural to test that L and R contribute equally to the second principal component and that they are the only variables to contribute to it. In other words, it is natural to test the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0$, with $\boldsymbol{\theta}_2^0 := (1, 1, 0, 0)'/\sqrt{2}$.

The considerations above make it natural to test that L and R contribute equally to the second principal component and that they are the only variables to contribute to it. In other words, it is natural to test the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0$, with $\boldsymbol{\theta}_2^0 := (1, 1, 0, 0)'/\sqrt{2}$.

- The Anderson test provides a p-value aqual to .099,
- The HPV test provides a p-value equal to .177,
- Our Sign test provides a p-value equal to .992,
- The Tyler test provides a p-value equal to 0.609.

Both sign tests as well as the HPV test do not reject the null hypothesis at any usual nominal level. But it not the case of the Anderson test who rejects the null at the level 10%. But in view of all our results, we can be confident in taking the good decision by not rejecting the null here.

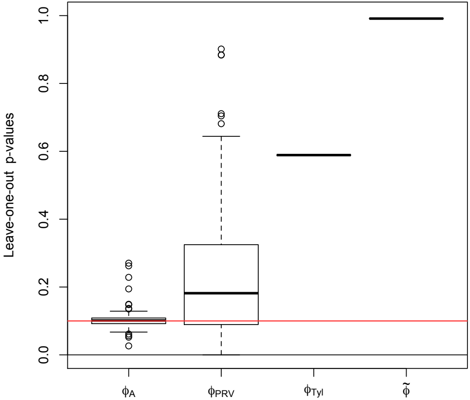


Figure – Boxplots of the 85 “leave-one-out” p-values of the four tests when testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_2 := (1, 1, 0, 0)'/\sqrt{2}$.

Conclusion

We saw here that our proposed sign test is

- validity-robust to heavy tails (as other sign tests),
- validity-robust to weak identifiability (as the HPV test),

but unlike the competitors, it achieves both, while being asymptotically locally optimal (but in the case $\delta_n = 1/\sqrt{n}$, for which it is still rate-optimal.

Beside that, our test is also robust to some departures from ellipticiy. Since it only assume that the spatial signs $\mathbf{U}_{ni} = \mathbf{X}_{ni}/\|\mathbf{X}_{ni}\|$ follows an Angular Gaussian distribution, they do not need to be independent of $\|\mathbf{X}_{ni}\|$, and thus our test can deal with some skewed distributions.

Section 3

ANOVA on weak directions

$\kappa = 10$ $\kappa = 3$ $\kappa \to 0$

Question : can we compare the location parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ of two samples when their common true underlying distribution is "very close" to the uniform one ? That is under a situation of *weak identifiability* of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$?

Question : can we compare the location parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ of two samples when their common true underlying distribution is "very close" to the uniform one ? That is under a situation of *weak identifiability* of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$?

Testing problem : we consider the problem of testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ against the alternative $\mathcal{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, where $\boldsymbol{\theta}_i$ is a unit vector of $\mathbb{R}^p$. We will consider situations where κ_n tends to 0.

Working context

- Two triangular arrays of rotationally symmetric distributed vectors.
- Both samples have the same angular function f, also the same concentration parameter $\kappa_n = \xi\sqrt{p}\eta_n + o_P(1)$ but their location parameter may differ.
- Weak identifiability occurs when $\eta_n \to 0$.

We could once again discern four different regimes :

(i) $\eta_n \equiv 1$;

(ii) $\eta_n = o(1)$ with $\sqrt{n}\eta_n \to \infty$;

(iii) $\eta_n = 1/\sqrt{n}$;

(iv) $\eta_n = o(1/\sqrt{n})$.

The two tests

- $Q^{(n)}_{\mathrm{LSV}} := \frac{(p-1)n_1 n_2}{\hat{\mathcal{J}}_p n}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)'\mathbf{P}^{\perp}_{\hat{\boldsymbol{\theta}}}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) > \chi^2_{p-1,1-\alpha}$
  where $\hat{\mathcal{J}}_p$ is a natural estimator of
  $$\mathcal{J}_p := \mathrm{E}[(1 - ((\mathbf{X}^{(j)}_{ni})'\boldsymbol{\theta})^2)].$$
- $Q^{(n)}_{\mathrm{H}} := \frac{n_1 n_2}{n}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)'\mathbf{S}_n^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) > \chi^2_{p,1-\alpha}$
  where $\mathbf{S}_n$ is a pooled empirical covariance matrix.

Asymptotics under the null

Theorem

Fix $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \boldsymbol{\theta}')' \in (\mathcal{S}^{p-1})^2$. *Then, for any bounded sequence* κ_n *and any angular density* f, *we have that, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\kappa_n,f}$,

(i) $Q^{(n)}_{\mathrm{LSV}}$ *converges weakly to a chi-square random variable with* $p-1$ *degrees of freedom, and*

(ii) $Q^{(n)}_{\mathrm{H}}$ *converges weakly to a chi-square random variable with* p *degrees of freedom,*

as $n \to \infty$.

Asymptotics under the null

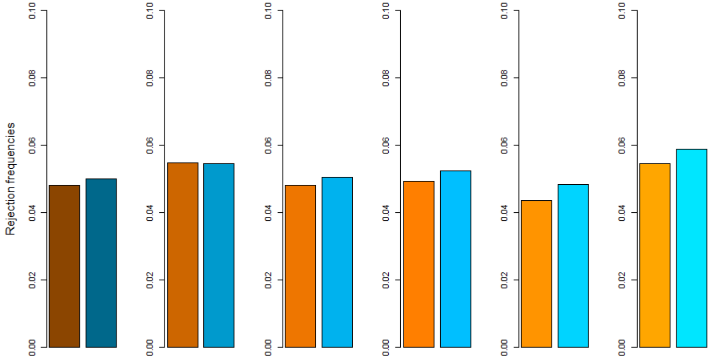


Figure – Rejection frequencies of the Hotelling test (orange) and of the pseudo FvML test (blue). Simulations were made with two 3-dimensional FvML samples of size $20,000$ with $2,500$ replications. The lighter the color (the larger w), the faster κ_n converges to 0.

Asymptotics under the alt

Theorem

Fix $\boldsymbol{\vartheta} \in \mathcal{H}_0$ *and let* $\kappa_n = \sqrt{p}\eta_n\xi$ *be a bounded (potentially* $o(1)$*) sequence. Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}+\ell_n\boldsymbol{\nu}_n\boldsymbol{\tau}_n,\kappa_n,f}$*, we have the following as* $n \to \infty$ *:*

Asymptotics under the alt

Theorem

Fix $\boldsymbol{\vartheta} \in \mathcal{H}_0$ and let $\kappa_n = \sqrt{p}\eta_n\xi$ be a bounded (potentially $o(1)$) sequence. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}+\ell_n\boldsymbol{\nu}_n\boldsymbol{\tau}_n,\kappa_n,f}$, we have the following as $n \to \infty$:

(i) *if $\eta_n \equiv 1$, with $\ell_n = 1/\sqrt{n}$,*

(a) *$Q^{(n)}_{\mathrm{LSV}}$ converges weakly to a chi-square random variable with $p-1$ degrees of freedom and non-centrality parameter*

$$\frac{\mathcal{K}^2_{p,f}\xi^2 p}{\mathcal{J}_p(p-1)}\|\sqrt{r_2}\boldsymbol{\tau}_1 - \sqrt{r_1}\boldsymbol{\tau}_2\|^2,$$

with $\mathcal{K}_{p,f} := \mathrm{E}[\varphi_f(\kappa_n(\mathbf{X}^{(j)}_{ni})'\boldsymbol{\theta})(1-((\mathbf{X}^{(j)}_{ni})'\boldsymbol{\theta})^2)]$ (the expectation is taken under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\kappa_n,f}$);

(b) *$Q^{(n)}_{\mathrm{H}}$ converges weakly to a chi-square random variable with p degrees of freedom and non-centrality parameter*

$$\frac{\mathcal{K}^2_{p,f}\xi^2 p}{\mathcal{J}_p(p-1)}\|\sqrt{r_2}\boldsymbol{\tau}_1 - \sqrt{r_1}\boldsymbol{\tau}_2\|^2;$$

Asymptotics under the alt

Theorem

Fix $\boldsymbol{\vartheta} \in \mathcal{H}_0$ *and let* $\kappa_n = \sqrt{p}\eta_n\xi$ *be a bounded (potentially* $o(1)$*) sequence. Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}+\ell_n\boldsymbol{\nu}_n\boldsymbol{\tau}_n,\kappa_n,f}$*, we have the following as* $n \to \infty$ *:*

(ii) *if* $\eta_n = o(1)$ *with* $\sqrt{n}\eta_n \to \infty$ *as* $n \to \infty$*, with* $\ell_n = \frac{1}{\sqrt{n}\eta_n}$*,*

(a) $Q^{(n)}_{\mathrm{LSV}}$ *converges weakly to a chi-square random variable with* $p-1$ *degrees of freedom and non-centrality parameter*

$$\xi^2\|\sqrt{r_2}\boldsymbol{\tau}_1 - \sqrt{r_1}\boldsymbol{\tau}_2\|^2;$$

(b) $Q^{(n)}_{\mathrm{H}}$ *converges weakly to a chi-square random variable with* p *degrees of freedom and non-centrality parameter*

$$\xi^2\|\sqrt{r_2}\boldsymbol{\tau}_1 - \sqrt{r_1}\boldsymbol{\tau}_2\|^2;$$

Asymptotics under the alt

Theorem

Fix $\boldsymbol{\vartheta} \in \mathcal{H}_0$ and let $\kappa_n = \sqrt{p}\eta_n\xi$ be a bounded (potentially $o(1)$) sequence. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}+\ell_n\boldsymbol{\nu}_n\boldsymbol{\tau}_n,\kappa_n,f}$, we have the following as $n \to \infty$:

(iii) *if $\eta_n \equiv 1/\sqrt{n}$, with $\ell_n = 1$,*

(a) *$Q^{(n)}_{\mathrm{LSV}}$ converges weakly to*

$$\mathbf{Y}'(\mathbf{I}_p - \frac{\mathbf{Z}\mathbf{Z}'}{\|\mathbf{Z}\|\|\mathbf{Z}\|})\mathbf{Y},$$

where $\mathbf{Y}$ and $\mathbf{Z}$ are two mutually independent Gaussian random p-vectors, respectively with mean $\xi(r_2^{1/2}\boldsymbol{\tau}_1 - r_1^{1/2}\boldsymbol{\tau}_2)$ and mean $\xi(\boldsymbol{\theta} + {r_1}^{1/2}\boldsymbol{\tau}_1 + {r_2}^{1/2}\boldsymbol{\tau}_2)$ and both with covariance matrix $\mathbf{I}_p$;

(b) *$Q^{(n)}_{\mathrm{H}}$ converges weakly to a chi-square random variable with p degrees of freedom and non-centrality parameter $\xi^2\|\sqrt{r_2}\boldsymbol{\tau}_1 - \sqrt{r_1}\boldsymbol{\tau}_2\|^2$;*

Asymptotics under the alt

Theorem

Fix $\boldsymbol{\vartheta} \in \mathcal{H}_0$ *and let* $\kappa_n = \sqrt{p}\eta_n\xi$ *be a bounded (potentially* $o(1)$*) sequence. Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}+\ell_n\boldsymbol{\nu}_n\boldsymbol{\tau}_n,\kappa_n,f}$*, we have the following as* $n \to \infty$ *:*

(iv) *if* $\eta_n = o(1)$ *with* $\sqrt{n}\eta_n \to 0$ *as* $n \to \infty$*, with* $\ell_n = 1$*,*

- (a) $Q^{(n)}_{\mathrm{LSV}}$ *converges weakly to a chi-square random variable with* $p-1$ *degrees of freedom ;*
- (b) $Q^{(n)}_{\mathrm{H}}$ *converges weakly to a chi-square random variable with* p *degrees of freedom.*

Asymptotics under the alt

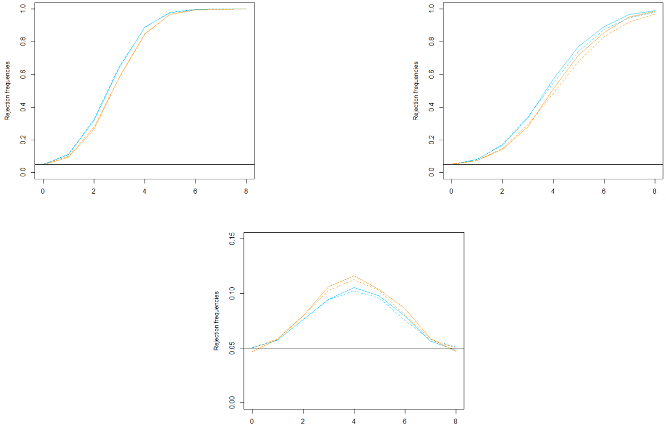


Figure – Theoretical powers (dotted lines) vs empirical powers (plain lines) for both tests (as before Hotelling is in orange and pseudo-FvML in blue). Simulations were made with two 3-dimensional FvML samples of size respectively 10, 000 and 15, 000 with 2, 500 replications.

Conclusion

The two test we compared here are both valid under weak identifiability, and shows non-trivial powers. But there isn't one that is outperforming the other.

Conclusion

The two test we compared here are both valid under weak identifiability, and shows non-trivial powers. But there isn't one that is outperforming the other.

Nice but not as striking results as the two first papers where

Conclusion

The two test we compared here are both valid under weak identifiability, and shows non-trivial powers. But there isn't one that is outperforming the other.

Nice but not as striking results as the two first papers where

(i) we saw that the test used in practice is not valid at all in case of weak identifiability, while there is an asymptotically equivalent test that is robust to weak identifiability ;

Conclusion

The two test we compared here are both valid under weak identifiability, and shows non-trivial powers. But there isn't one that is outperforming the other.

Nice but not as striking results as the two first papers where

(i) we saw that the test used in practice is not valid at all in case of weak identifiability, while there is an asymptotically equivalent test that is robust to weak identifiability ;

(ii) we proposed a sign test that is robust to weak identifiability AND heavy tails/outliers.

Wrap up

At the end of the day, we saw :

Wrap up

At the end of the day, we saw :

- what is weak identifiability (and why it makes inference on the parameter harder) ;

Wrap up

At the end of the day, we saw :

- what is weak identifiability (and why it makes inference on the parameter harder) ;
- that it is not purely theoretically interesting but it has some uses in practice too !

Wrap up

At the end of the day, we saw :

- what is weak identifiability (and why it makes inference on the parameter harder) ;
- that it is not purely theoretically interesting but it has some uses in practice too !
- Take-home message :

Wrap up

At the end of the day, we saw :

- what is weak identifiability (and why it makes inference on the parameter harder) ;
- that it is not purely theoretically interesting but it has some uses in practice too !
- Take-home message : Statistic is cool and awesome O:)