TESTING FOR PRINCIPAL COMPONENT DIRECTIONS UNDER WEAK IDENTIFIABILITY

Davy PAINDAVEINE, Julien REMY, Thomas VERDEBOUT

19 October 2018

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 : throughout the presentation, 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

Gaussian framework :

Working context

Gaussian framework :

- triangular array of observations $\mathbf{X}_{ni}$, $i = 1, \ldots, n$, $n = 1, 2, \ldots$; where $\mathbf{X}_{n1}, \ldots, \mathbf{X}_{nn}$ form a random sample from the p-variate normal distribution with mean $\boldsymbol{\mu}_n$ and covariance matrix

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

  where v is a positive real number, (σ_n^2) is a positive real sequence, (r_n) is a bounded nonnegative real sequence, and $\mathbf{I}_\ell$ denotes the ℓ-dimensional identity matrix.

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

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

(i) $r_n \equiv 1$;

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

(i) $r_n \equiv 1$;

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

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

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

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

Under the null : Anderson's test

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

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

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 $\mathrm{P}_{\boldsymbol{\theta}_1^0, r_n, v}$,

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

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

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

where $\tilde{\boldsymbol{\theta}}_j$, $j = 2, \ldots, p$, are defined recursively through a Gram-Schmidt orthogonalization of $\boldsymbol{\theta}_1^0, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_p$, where $\hat{\boldsymbol{\theta}}_j$ is a unit eigenvector of $\mathbf{S}$ associated with the eigenvalue $\hat{\lambda}_j$, $j = 2, \ldots, p$.

Under the null : Le Cam optimal test

This test rejects the null at asymptotic level α when

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

where $\tilde{\boldsymbol{\theta}}_j$, $j = 2, \ldots, p$, are defined recursively through a Gram-Schmidt orthogonalization of $\boldsymbol{\theta}_1^0, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_p$, where $\hat{\boldsymbol{\theta}}_j$ is a unit eigenvector of $\mathbf{S}$ associated with the eigenvalue $\hat{\lambda}_j$, $j = 2, \ldots, p$.

Theorem

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

$$Q_{\mathrm{HPV}} \xrightarrow{\mathcal{D}} \chi^2_{p-1},$$

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

Under the null : Simulations

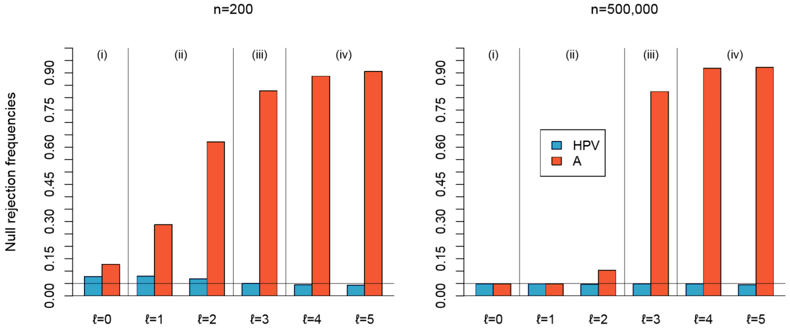

FIGURE – Empirical rejection frequencies of the tests ϕ_{HPV} and ϕ_{A} performed at nominal level 5%. Results are based on $M = 10,000$ independent ten-dimensional Gaussian random samples.

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 $\mathrm{P}_{\boldsymbol{\theta}_1^0, r_n, v}$,

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

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

Summary

- Unlike ϕ_{A}, the test ϕ_{HPV} is validity-robust to weak identifiability ;

Summary

- Unlike ϕ_{A}, the test ϕ_{HPV} 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 ϕ_{A}, the test ϕ_{HPV} 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 ϕ_{HPV} is obtained at the expense of efficiency.

Non-null results

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* $\mathrm{P}_{\boldsymbol{\theta}_1^0 + \boldsymbol{\tau}_n/(\sqrt{n} r_n), r_n, v}$, *with* $(\boldsymbol{\tau}_n) \to \boldsymbol{\tau}$, *the statistic* Q_{HPV} *is asymptotically non-central chi-square with* $p-1$ *degrees of freedom and with non-centrality parameter* $(v^2/(1+\delta v))\|\boldsymbol{\tau}\|^2$ $(\delta = 1$ *in regime (i) and* $\delta = 0$ *in regime (ii)).*

Non-null results

Theorem

(iii) *When* $r_n = 1/\sqrt{n}$, $Q_{\rm HPV}$, *under* $\mathrm{P}_{\boldsymbol{\theta}_1^0+\boldsymbol{\tau}_n, r_n, \mathcal{V}}$, *with* $(\boldsymbol{\tau}_n) \to \boldsymbol{\tau}$, *is asymptotically non-central chi-square with* $p-1$ *degrees of freedom and with non-centrality parameter*

$$\frac{\mathcal{V}^2}{16}\|\boldsymbol{\tau}\|^2\big(4-\|\boldsymbol{\tau}\|^2\big)\big(2-\|\boldsymbol{\tau}\|^2\big)^2.$$

(iv) *When* $r_n = o(1/\sqrt{n})$, $Q_{\rm HPV}$, *under* $\mathrm{P}_{\boldsymbol{\theta}_1^0+\boldsymbol{\tau}_n, r_n, \mathcal{V}}$, *with* $(\boldsymbol{\tau}_n) \to \boldsymbol{\tau}$, *is asymptotically chi-square with* $p-1$ *degrees of freedom*

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 ϕ_{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 ϕ_{HPV} is rate-consistent .

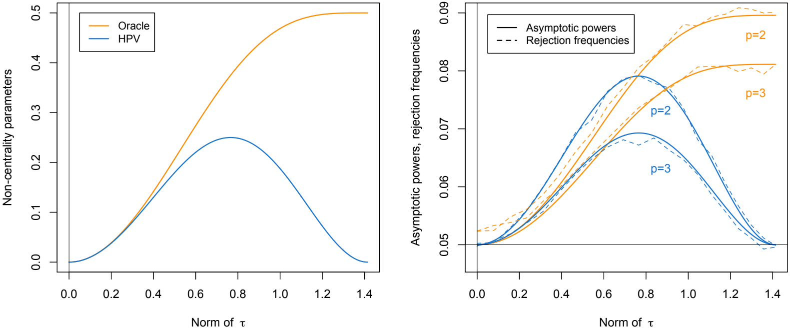

FIGURE – (Left :) Non-centrality parameters, as a function of $\|\boldsymbol{\tau}\|(\in[0,\sqrt{2}])$, in the asymptotic non-central chi-square distributions of the test statistics of ϕ_{HPV} and ϕ_{oracle}, respectively, under alternatives of the form $\mathrm{P}_{\boldsymbol{\theta}_1^0+\boldsymbol{\tau},1/\sqrt{n},1}$.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

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 HPV test provides a p-value equal to .177 $\Rightarrow$ does not lead to rejection of the null hypothesis at any usual nominal level.
- The Anderson test provides a p-value equal to 0.099 $\Rightarrow$ rejects the null at the level 10%.

But following the results that we presented before, practitioners should here be confident that the HPV test provides the right decision, since the Anderson test tends to strongly overreject the null when eigenvalues are close.

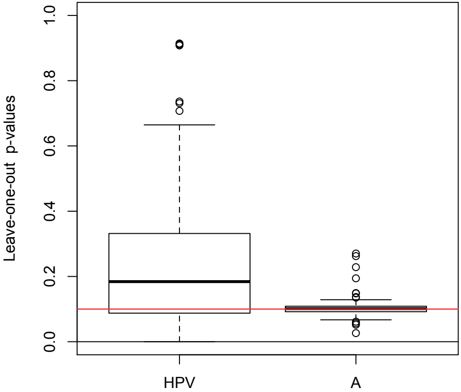

FIGURE – Boxplots of the 85 “leave-one-out” p-values of the HPV test (left) and Anderson test (right) when testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_2 := (1, 1, 0, 0)'/\sqrt{2}$.

Conclusion

We saw here that the the HPV test 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.

References

- Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, 34, 122–148.
- Hallin, M., Paindaveine, D. and Verdebout, Th. (2010). Optimal rank-based testing for principal components. *Annals of Statistics*, 38, 3245–3299.
- Paindaveine, D., Remy, J. and Verdebout, Th. (2018). Testing for principal component directions under weak identifiability. *submitted*.