Rank tests for PCA under weak identifiability

Davy Paindaveine

Joint work with Laura Peralvo Maroto, Julien Remy, and Thomas Verdebout

Université libre de Bruxelles

November 2024, WU Wien

Flury (1988) conducted a PCA of the Swiss banknotes data, with four variables:

- the width L on the left side
- the width R on the right side
- the size B of the bottom margin
- the size T of the top margin,

all measured in mm$\times 10^{-1}$ on $n = 85$ counterfeit bills.

The resulting sample covariance matrix S provides eigenvalues $\hat{\lambda}_1 = 102.69$, $\hat{\lambda}_2 = 13.05$, $\hat{\lambda}_3 = 10.23$, and $\hat{\lambda}_4 = 2.66$, and corresponding unit eigenvectors

$$\hat{\theta}_1 = \begin{pmatrix} .032 = L \\ -.012 = R \\ .820 = B \\ -.571 = T \end{pmatrix}, \hat{\theta}_2 = \begin{pmatrix} .593 \\ .797 \\ .057 \\ .097 \end{pmatrix}, \hat{\theta}_3 = \begin{pmatrix} -.015 \\ -.129 \\ .566 \\ .814 \end{pmatrix}, \text{ and } \hat{\theta}_4 = \begin{pmatrix} .804 \\ -.590 \\ -.064 \\ -.035 \end{pmatrix}.$$

Flury concludes that the 1st PC is a contrast between B and T.

It is tempting to interpret the 2nd PC as an aggregate of L and R... But Flury writes *"beware: the 2nd and 3rd roots are quite close and so the computation of standard errors for the coefficients of* $\hat{\theta}_2$ *and* $\hat{\theta}_3$ *may be hazardous".*

The resulting sample covariance matrix S provides eigenvalues $\hat{\lambda}_1 = 102.69$, $\hat{\lambda}_2 = 13.05$, $\hat{\lambda}_3 = 10.23$, and $\hat{\lambda}_4 = 2.66$, and corresponding unit eigenvectors

$$\hat{\theta}_1 = \begin{pmatrix} .032 = L \\ -.012 = R \\ .820 = B \\ -.571 = T \end{pmatrix}, \ \hat{\theta}_2 = \begin{pmatrix} .593 \\ .797 \\ .057 \\ .097 \end{pmatrix}, \ \hat{\theta}_3 = \begin{pmatrix} -.015 \\ -.129 \\ .566 \\ .814 \end{pmatrix}, \text{ and } \hat{\theta}_4 = \begin{pmatrix} .804 \\ -.590 \\ -.064 \\ -.035 \end{pmatrix}.$$

Flury concludes that the 1st PC is a contrast between B and T.

It is tempting to interpret the 2nd PC as an aggregate of L and R... But Flury writes *"beware: the 2nd and 3rd roots are quite close and so the computation of standard errors for the coefficients of $\hat{\theta}_2$ and $\hat{\theta}_3$ may be hazardous".*

"Computation of standard errors for the coefficients of $\hat{\theta}_2$ and $\hat{\theta}_3$ may be hazardous"

The covariance matrix in the asymptotic (normal) distribution of $\sqrt{n}(\hat{\theta}_2 - \theta_2)$ is

$$\sum_{\substack{\ell=1 \\ \ell\neq 2}}^{p} \frac{\lambda_2 \lambda_\ell}{(\lambda_2 - \lambda_\ell)^2}\, \theta_\ell \theta_\ell',$$

which is huge if λ_2 is close to λ_3!

On this basis, Flury refrains from drawing any conclusion about θ_2...

Question: what can we say about θ_2 in setups where $\lambda_2 \approx \lambda_3$?

For a fixed unit vector θ_1^0 of $\mathbb{R}^p$, we consider the problem of testing

$$\mathcal{H}_0 : \theta_1 = \theta_1^0 \qquad \text{vs} \qquad \mathcal{H}_1 : \theta_1 \neq \theta_1^0,$$

based on $X_{n1}, \ldots, X_{nn}$ i.i.d. $\mathcal{N}_p(\mu_n, \Sigma_n)$, with the single-spiked structure

$$\Sigma_n = \sigma_n^2(I_p + r_n v\, \theta_1\theta_1') = \underbrace{\sigma_n^2(1 + r_n v)}_{\lambda_{1n}}\, \theta_1\theta_1' + \underbrace{\sigma_n^2}_{\lambda_{2n}=\ldots=\lambda_{pn}} (I_p - \theta_1\theta_1'),$$

where $v > 0$ and (r_n) is a positive real sequence that may be $o(1)$.

If $r_n \equiv 1$, then this is the usual asymptotic framework.

If $r_n = o(1)$, then $\lambda_{1n}/\lambda_{2n} \to 1$ (weak identifiability: challenging!)

We consider two parametric tests:

- the LR test ϕ_{LR} rejects $\mathcal{H}_0 : \theta_1 = \theta_1^0$ at asymptotic level α if

$$Q_{\mathrm{LR}} := \frac{n}{\hat{\lambda}_{1n}} \sum_{j=2}^{p} \frac{(\hat{\lambda}_{jn} - \hat{\lambda}_{1n})^2}{\hat{\lambda}_{jn}} \left(\hat{\theta}_{jn}' \theta_1^0\right)^2 > \chi^2_{p-1,1-\alpha};$$

- the score test ϕ_{S} rejects $\mathcal{H}_0$ at asymptotic level α if

$$Q_{\mathrm{S}} := \frac{n}{\hat{\lambda}_{1n}} \sum_{j=2}^{p} \frac{1}{\hat{\lambda}_{jn}} \left(\tilde{\theta}_{jn}' S_n \theta_1^0\right)^2 > \chi^2_{p-1,1-\alpha},$$

  where $\tilde{\theta}_{2n}, \ldots, \tilde{\theta}_{pn}$ result from Gram-Schmidt-ing $\theta_1^0, \hat{\theta}_{2n}, \ldots, \hat{\theta}_{pn}$.

If $r_n \equiv 1$ (so that $\lambda_{jn} = \lambda_j$ for any j), then these tests are asymptotically equivalent under $\mathcal{H}_0$, hence also under contiguous alternatives.

We may assume that $\mu_n \equiv 0$ (from translation invariance)

We may assume that $\sigma_n \equiv 1$ (from scale invariance)

$\rightsquigarrow$ Denote the resulting hypothesis as $\mathrm{P}_{\theta_1, r_n, v}$.

Theorem 1

Assume that (r_n) is $O(1)$. Then, under the sequence of null hypotheses $\mathrm{P}_{\theta_1^0, r_n, v}$,

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

as $n \to \infty$.

$\rightsquigarrow$ Asymptotic null size α irrespective of the rate at which $r_n \to 0$ (if it does).

Hence, ϕ_{S} is robust to weak identifiability.

Simulations with covariance matrix $\Sigma_n^{(\ell)} := (1 + n^{-\ell/6})\theta_1^0\theta_1^{0\prime} + 1 \times (I_p - \theta_1^0\theta_1^{0\prime})$

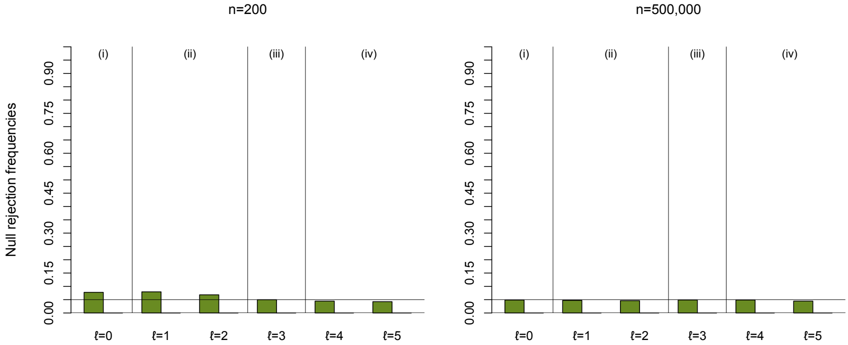


Figure: Null empirical rejection frequencies of ϕ_S and ϕ_{LR} performed at nominal level 5% (results are based on $M = 10{,}000$ independent 10-dimensional Gaussian random samples of size $n = 200$ and size $n = 500{,}000$). The larger ℓ, the weaker the identifiability.

Simulations with covariance matrix $\Sigma_n^{(\ell)} := (1 + n^{-\ell/6})\theta_1^0\theta_1^{0\prime} + 1 \times (I_p - \theta_1^0\theta_1^{0\prime})$

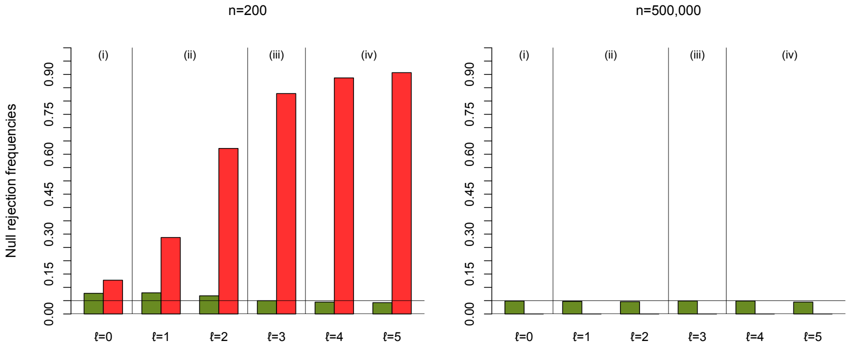


Figure: Null empirical rejection frequencies of $\phi_{\rm S}$ and $\phi_{\rm LR}$ performed at nominal level 5% (results are based on $M = 10{,}000$ independent 10-dimensional Gaussian random samples of size $n = 200$ and size $n = 500{,}000$). The larger ℓ, the weaker the identifiability.

Simulations with covariance matrix $\Sigma_n^{(\ell)} := (1 + n^{-\ell/6})\theta_1^0\theta_1^{0\prime} + 1 \times (I_p - \theta_1^0\theta_1^{0\prime})$

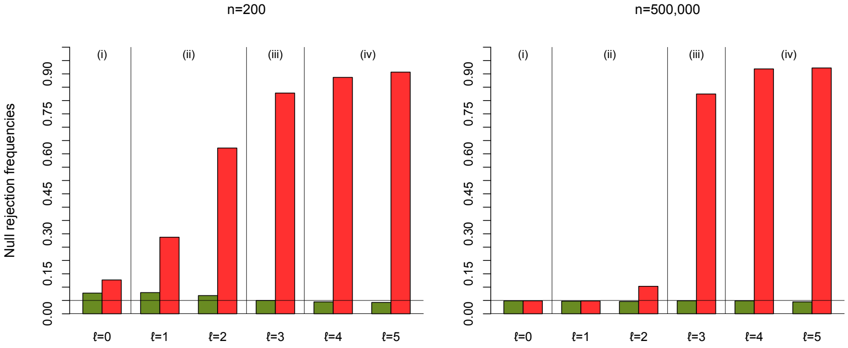


Figure: Null empirical rejection frequencies of ϕ_{S} and ϕ_{LR} performed at nominal level 5% (results are based on $M = 10{,}000$ independent 10-dimensional Gaussian random samples of size $n = 200$ and size $n = 500{,}000$). The larger ℓ, the weaker the identifiability.

Theorem 2

Let Z be a $p \times p$ random matrix such that $\mathrm{vec}(Z) \sim \mathcal{N}(0, I_{p^2} + K_p)$. Then, under the sequence of null hypotheses $\mathrm{P}_{\theta_1^0, r_n, v}$:

(i) if $r_n \equiv 1$ or (ii) if r_n is $o(1)$ with $\sqrt{n} r_n \to \infty$, then $Q_{\mathrm{LR}} \xrightarrow{\mathcal{D}} \chi^2_{p-1}$;

Theorem 2

Let Z be a $p \times p$ random matrix such that $\mathrm{vec}(Z) \sim \mathcal{N}(0, I_{p^2} + K_p)$. Then, under the sequence of null hypotheses $\mathrm{P}_{\theta_1^0, r_n, v}$:

(i) if $r_n \equiv 1$ or (ii) if r_n is $o(1)$ with $\sqrt{n} r_n \to \infty$, then $Q_{\mathrm{LR}} \xrightarrow{\mathcal{D}} \chi^2_{p-1}$;

(iii) if $r_n = 1/\sqrt{n}$, then

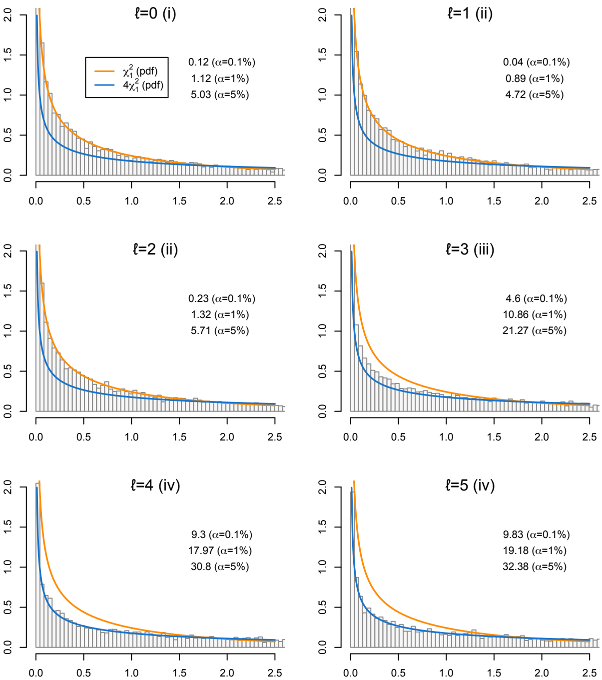
$$Q_{\mathrm{LR}} \xrightarrow{\mathcal{D}} \sum_{j=2}^{p} (\ell_1(v) - \ell_j(v))^2 (w_{j1}(v))^2,$$

where $\ell_1(v) \geq \ldots \geq \ell_p(v)$ are the eigenvalues of $Z + \mathrm{diag}(v, 0, \ldots, 0)$ and $w_j(v) = (w_{j1}(v), \ldots, w_{jp}(v))'$ is an arbitrary unit eigenv. associated with $\ell_j(v)$;

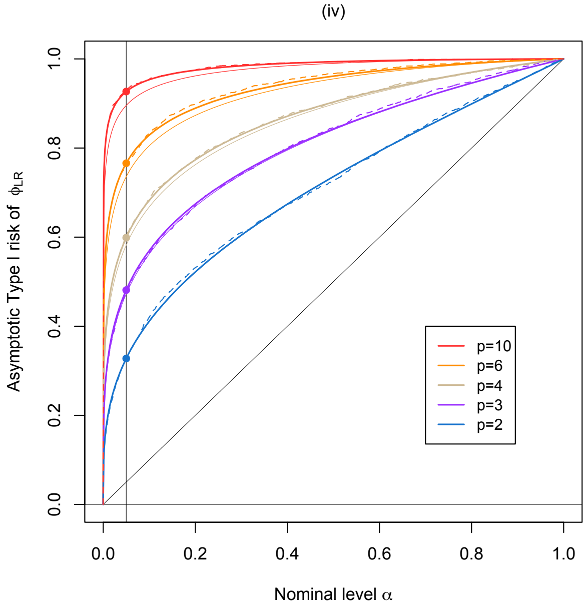
(iv) if $r_n = o(1/\sqrt{n})$, then

$$Q_{\mathrm{LR}} \xrightarrow{\mathcal{D}} \sum_{j=2}^{p} (\ell_1 - \ell_j)^2 w_{j1}^2 \ \left(\stackrel{\mathcal{D}}{=} 4\chi^2_{p-1}, \mathit{for}\ p = 2\right),$$

where $\ell_1 \geq \ldots \geq \ell_p$ are the eigenvalues of Z and $w_j = (w_{j1}, \ldots, w_{jp})'$ is an arbitrary unit eigenvector associated with ℓ_j.



ℓ=0 (i)
χ_1^2 (pdf)
$4\chi_1^2$ (pdf)
0.12 (α=0.1%)
1.12 (α=1%)
5.03 (α=5%)
ℓ=1 (ii)
0.04 (α=0.1%)
0.89 (α=1%)
4.72 (α=5%)
ℓ=2 (ii)
0.23 (α=0.1%)
1.32 (α=1%)
5.71 (α=5%)
ℓ=3 (iii)
4.6 (α=0.1%)
10.86 (α=1%)
21.27 (α=5%)
ℓ=4 (iv)
9.3 (α=0.1%)
17.97 (α=1%)
30.8 (α=5%)
ℓ=5 (iv)
9.83 (α=0.1%)
19.18 (α=1%)
32.38 (α=5%)



(iv)
Asymptotic Type I risk of ϕ_{LR}
Nominal level α
0.0
0.2
0.4
0.6
0.8
1.0
p=10
p=6
p=4
p=3
p=2

We have shown that ϕ_{S} is "validity-robust" ($><$ ϕ_{LR}).

Is this achieved at the expense of power?

We have shown that $\phi_{\rm S}$ is "validity-robust" ($>< \phi_{\rm LR}$).
Is this achieved at the expense of power?

Theorem 3

Assume that (i) $r_n \equiv 1$ or (ii) $r_n = o(1)$ with $\sqrt{n}r_n \to \infty$. Then, under $\mathrm{P}_{\theta_1^0+(\sqrt{n}r_n)^{-1}\tau_n, r_n, v}$, with $(\tau_n) \to \tau$,

$$Q_{\rm S} \xrightarrow{\mathcal{D}} \chi^2_{p-1}\left(\frac{v^2\|\tau\|^2}{1+\delta v}\right)$$

as $n \to \infty$, where $\delta = 1$ in regime (i) and $\delta = 0$ in regime (ii).

The faster r_n goes to zero, the more severe the corresponding local alternatives.

Theorem 3

(iii)

(iv) *Assume that* $r_n = o(1/\sqrt{n})$. *Then, under* $\mathrm{P}_{\theta_1^0+\tau_n, r_n, v}$, *with* $(\tau_n) \to \tau$,

$$Q_{\mathrm{S}} \xrightarrow{\mathcal{D}} \chi^2_{p-1}(0)$$

as $n \to \infty$.

Theorem 3

(iii) *Assume that* $r_n = 1/\sqrt{n}$*. Then, under* $\mathrm{P}_{\theta_1^0+\tau_n, r_n, v}$*, with* $(\tau_n) \to \tau$*,*

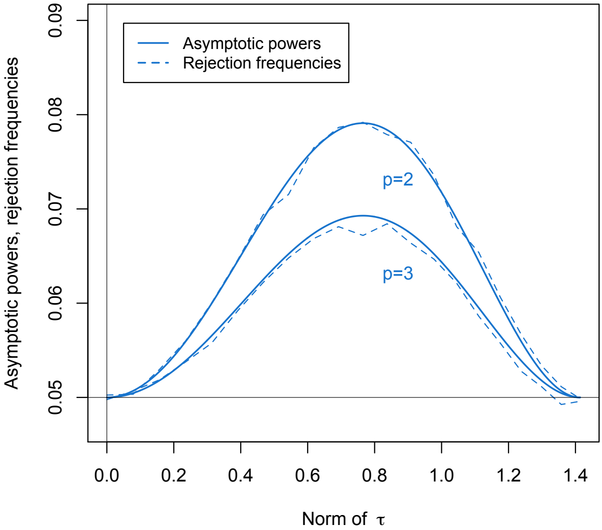
$$Q_{\mathrm{S}} \xrightarrow{\mathcal{D}} \chi^2_{p-1}\left(\frac{1}{16} v^2 \|\tau\|^2 \left(4 - \|\tau\|^2\right)\left(2 - \|\tau\|^2\right)^2\right)$$

as $n \to \infty$*.*

(iv) *Assume that* $r_n = o(1/\sqrt{n})$*. Then, under* $\mathrm{P}_{\theta_1^0+\tau_n, r_n, v}$*, with* $(\tau_n) \to \tau$*,*

$$Q_{\mathrm{S}} \xrightarrow{\mathcal{D}} \chi^2_{p-1}(0)$$

as $n \to \infty$*.*


Asymptotic powers
Rejection frequencies
p=2
p=3
0.05
0.06
0.07
0.08
0.09
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Asymptotic powers, rejection frequencies
Norm of τ

Non-trivial local powers in regimes (i)–(iii).

What about rate-optimality, what about plain optimality?

Theorem 4

Assume that (i) $r_n \equiv 1$ or that (ii) $\sqrt{n} r_n \to \infty$. Then, with $\nu_n = 1/(\sqrt{n} r_n)$,

$$\log \frac{d\mathrm{P}^{(n)}_{\theta_1^0 + \nu_n \tau_n, r_n, v}}{d\mathrm{P}^{(n)}_{\theta_1^0, r_n, v}} = \tau_n' \Delta_n - \frac{1}{2} \tau_n' \Gamma \tau_n + o_{\mathrm{P}}(1) \quad \textit{and} \quad \Delta_n^{(n)} \xrightarrow{\mathcal{D}} \mathcal{N}_p(0, \Gamma)$$

as $n \to \infty$ under $\mathrm{P}^{(n)}_{\theta_1^0, r_n, v}$, where

$$\Delta_n := \frac{\sqrt{n} v}{1 + \delta v} \big(I_p - \theta_1^0 \theta_1^{0\prime} \big) (S_n - \Sigma_n) \theta_1^0 \quad \textit{and} \quad \Gamma := \frac{v^2}{1 + \delta v} \big(I_p - \theta_1^0 \theta_1^{0\prime} \big)$$

($\delta = 1$ in case (i) and 0 in case (ii)).

In particular,

- $\nu_n = 1/(\sqrt{n} r_n)$ is the contiguity rate.
- The optimal test rejects the null for large values of $\Delta_n' \Gamma^- \Delta_n = Q_{\mathrm{S}} + o_{\mathrm{P}}(1)$.

Theorem 4

Assume that (iv) $r_n = o(1/\sqrt{n})$*. Then, even with* $\nu_n \equiv 1$,

$$\log \frac{d\mathrm{P}^{(n)}_{\theta_1^0+\nu_n\tau_n,r_n,v}}{d\mathrm{P}^{(n)}_{\theta_1^0,r_n,v}} = o_{\mathrm{P}}(1)$$

as $n \to \infty$ *under* $\mathrm{P}^{(n)}_{\theta_1^0,r_n,v}$.

$\rightsquigarrow$ No tests can show non-trivial asymptotic powers against the most severe alternatives $\theta_1 = \theta_1^0 + \tau$.

Theorem 4

Assume that (iii) $r_n = 1/\sqrt{n}$. *Then, with* $\nu_n \equiv 1$,

$$\log \frac{d\mathrm{P}^{(n)}_{\theta_1^0+\nu_n\tau_n,r_n,v}}{d\mathrm{P}^{(n)}_{\theta_1^0,r_n,v}} = \tau_n'\Big[v\sqrt{n}(S_n - \Sigma_n)\big(\theta_1^0 + \tfrac{1}{2}\tau_n\big)\Big]$$

$$-\frac{v^2}{2}\|\tau_n\|^2 + \frac{v^2}{8}\|\tau_n\|^4 + o_{\mathrm{P}}(1),$$

as $n \to \infty$ *under* $\mathrm{P}^{(n)}_{\theta_1^0,r_n,v}$.

- $\nu_n \equiv 1$ is the contiguity rate, so that ϕ_{S} is rate-consistent here as well.

Theorem 4

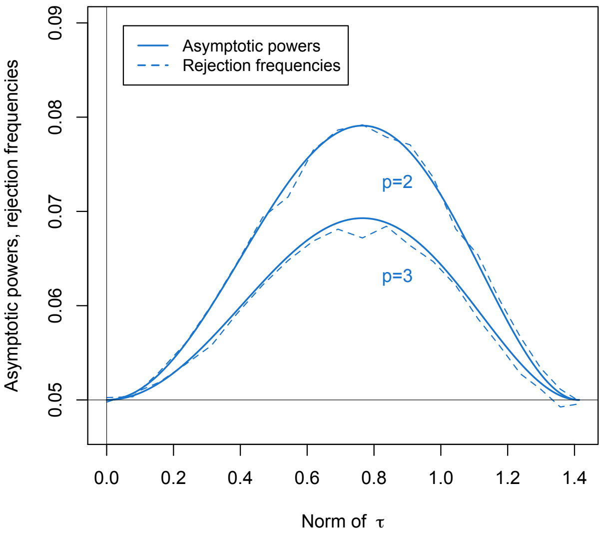
Assume that (iii) $r_n = 1/\sqrt{n}$. *Then, with* $\nu_n \equiv 1$,

$$\log \frac{d\mathrm{P}^{(n)}_{\theta_1^0+\nu_n\tau_n, r_n, v}}{d\mathrm{P}^{(n)}_{\theta_1^0, r_n, v}} = \tau_n' \Big[v\sqrt{n}(S_n - \Sigma_n)\big(\theta_1^0 + \tfrac{1}{2}\tau_n\big)\Big]$$

$$- \frac{v^2}{2}\|\tau_n\|^2 + \frac{v^2}{8}\|\tau_n\|^4 + o_{\mathrm{P}}(1),$$

as $n \to \infty$ *under* $\mathrm{P}^{(n)}_{\theta_1^0, r_n, v}$.

- $\nu_n \equiv 1$ is the contiguity rate, so that ϕ_{S} is rate-consistent here as well.
- For small τ_n, one recovers LAN, which will imply that ϕ_{S} is locally asymptotically optimal.



Asymptotic powers
Rejection frequencies
p=2
p=3
0.05
0.06
0.07
0.08
0.09
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Asymptotic powers, rejection frequencies
Norm of τ

To summarize optimality results:

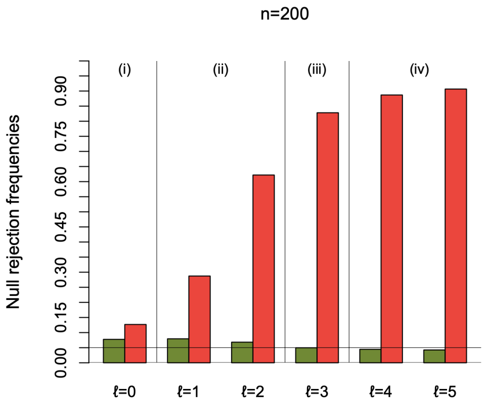
In regimes (i)–(ii): LAN and ϕ_S is Le Cam optimal.
In regime (iii): not LAN, but ϕ_S is still rate-consistent (and locally optimal).
In regime (iv): LAN and ϕ_S is Le Cam optimal, but optimality is degenerate

The sample covariance matrix S provides eigenvalues $\hat{\lambda}_1 = 102.69$, $\hat{\lambda}_2 = 13.05$, $\hat{\lambda}_3 = 10.23$, and $\hat{\lambda}_4 = 2.66$, and corresponding eigenvectors

$$\hat{\theta}_1 = \begin{pmatrix} .032 \\ -.012 \\ .820 \\ -.571 \end{pmatrix}, \ \hat{\theta}_2 = \begin{pmatrix} .593 = L \\ .797 = R \\ .057 = B \\ .097 = T \end{pmatrix}, \ \hat{\theta}_3 = \begin{pmatrix} -.015 \\ -.129 \\ .566 \\ .814 \end{pmatrix}, \ \text{and } \hat{\theta}_4 = \begin{pmatrix} .804 \\ -.590 \\ -.064 \\ -.035 \end{pmatrix}.$$

It is tempting to interpret the 2nd PC as an aggregate of L and R, but Flury was anxious about it.

For the null hypothesis $\mathcal{H}_0 : \theta_2 := (1, 1, 0, 0)'/\sqrt{2}$, the p-value of ϕ_{S} is .177 and the p-value of ϕ_{LR} is .099. Since ϕ_{LR} overrejects under weak identifiability, we are confident that one should not reject the null hypothesis here.


n=200
Null rejection frequencies
0.00
0.15
0.30
0.45
0.60
0.75
0.90
(i)
(ii)
(iii)
(iv)
ℓ=0
ℓ=1
ℓ=2
ℓ=3
ℓ=4
ℓ=5

A natural question: are the above phenomenons Gaussian accidents?

A natural question: are the above phenomenons Gaussian accidents? No.

A natural question: are the above phenomenons Gaussian accidents? No.

(a) They hold, without any modification, at any elliptical distribution for which the kurtosis coefficient

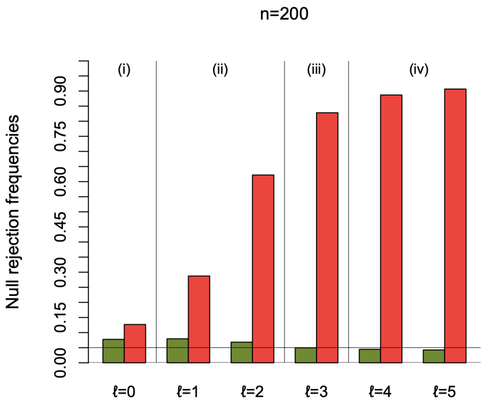
$$\kappa_p(f) := \frac{p\mathrm{E}[d^4]}{(p+2)(\mathrm{E}[d^2])^2} - 1, \quad \text{with } d := \sqrt{(X-\mu)'\Sigma^{-1}(X-\mu)},$$

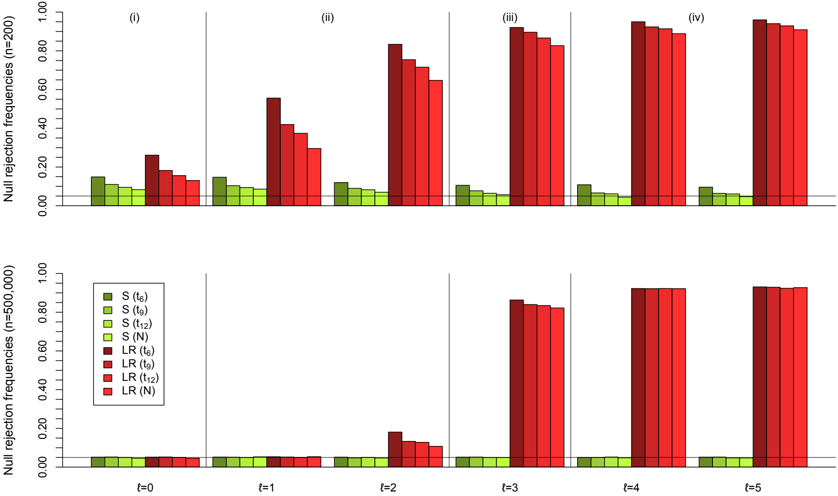
takes the same value ($\kappa_p(\phi) = 0$) as in the Gaussian case.

(b) At any elliptical distribution with $\kappa_p(f) < \infty$, they apply to the modified tests

$$\frac{Q_{\mathrm{LR}}}{1+\hat{\kappa}_p} > \chi^2_{p-1,1-\alpha} \quad \text{and} \quad \frac{Q_{\mathrm{S}}}{1+\hat{\kappa}_p} > \chi^2_{p-1,1-\alpha},$$

where $\hat{\kappa}_p$ is an arbitrary consistent estimator of $\kappa_p(f)$.


n=200
Null rejection frequencies
0.00
0.15
0.30
0.45
0.60
0.75
0.90
(i)
(ii)
(iii)
(iv)
ℓ=0
ℓ=1
ℓ=2
ℓ=3
ℓ=4
ℓ=5



Null rejection frequencies (n=200)
Null rejection frequencies (n=500,000)
0.00
0.20
0.40
0.60
0.80
1.00
(i)
(ii)
(iii)
(iv)
S (t_6)
S (t_9)
S (t_{12})
S (N)
LR (t_6)
LR (t_9)
LR (t_{12})
LR (N)
ℓ=0
ℓ=1
ℓ=2
ℓ=3
ℓ=4
ℓ=5

A few severe limitations in the previous results. . .

(i) Non-Gaussian results are under the null hypothesis only (what about the power of pseudo-Gaussian tests away from the Gaussian case?)

$\rightsquigarrow$ We considered (essentially arbitrary) elliptical densities of the form

$$x \mapsto \frac{c_{p,f}}{\sqrt{\det \Sigma}} f\left(\sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}\right)$$

(only suitable derivability assumptions for f + finite Fisher information) and studied the asymptotic behavior of the corresponding local log-likelihood ratios under double-asymptotic scenarios involving weak identifiability.

This requires new results on quadratic mean differentiable families in triangular array frameworks.

The Le Cam third lemma then provides the local asymptotic powers of pseudo-Gaussian tests under virtually any f (these are poor under heavy tails!)

(ii) Pseudo-Gaussian tests are based on the empirical covariance matrix

$$S_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})',$$

hence require finite fourth moments and are sensitive to possible outliers.

We consider rank-based covariance matrices of the form

$$S_{n,K} = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{R_i(\hat{\mu}, \hat{V})}{n+1}\right) U_i(\hat{\mu}, \hat{V}) U_i'(\hat{\mu}, \hat{V}),$$

where $R_i(\mu, V)$ is the rank of $d_i(\mu, V)$ among $d_1(\mu, V), \ldots, d_n(\mu, V)$,

$$d_i(\mu, V) := \|V^{-1/2}(X_i - \mu)\| \quad \text{and} \quad U_i(\mu, V) := \frac{V^{-1/2}(X_i - \mu)}{\|V^{-1/2}(X_i - \mu)\|};$$

here, $\hat{\mu}$ and $\hat{V}$ are suitable (robust) estimators of μ and $V = \Sigma/(\det \Sigma)^{1/p}$.

Away from weak identifiability, the rank tests

$$Q_K := \frac{np(p+2)}{(\int_0^1 K^2(u)\,du)} \sum_{j=2}^{p} (\tilde{\theta}_j' S_{n,K} \theta_1^0)^2 > \chi^2_{p-1,1-\alpha},$$

that mimick the score test

$$Q_{\mathcal{S}} := \frac{n}{\hat{\lambda}_{1n}} \sum_{j=2}^{p} \frac{1}{\hat{\lambda}_{jn}} (\tilde{\theta}_{jn}' S_n \theta_1^0)^2 > \chi^2_{p-1,1-\alpha},$$

are attractive:

- They do not require moment conditions
- With Gaussian scores ($K(u) = \Psi_p^{-1}(u)$, with Ψ_p the cdf of the χ^2_p), their AREs with respect to pseudo-Gaussian tests are uniformly larger than one.

Do such properties survive weak identifiability?

Still with $R_i(\mu, V)$ the rank of $d_i(\mu, V)$ among $d_1(\mu, V), \ldots, d_n(\mu, V)$,

$$S_{n,K} = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{R_i(\hat{\mu}, \hat{V})}{n+1}\right) U_i(\hat{\mu}, \hat{V}) U_i'(\hat{\mu}, \hat{V})$$

and

$$S_{K,f} = \frac{1}{n} \sum_{i=1}^{n} K\Big(\tilde{F}(d_i(\mu, V))\Big) U_i(\mu, V) U_i'(\mu, V)$$

can be shown to be asymptotically equivalent under the null hypothesis; here,

$$\tilde{F}(t) := \left(\int_0^\infty r^{p-1} f(r)\, dr\right)^{-1} \int_0^t r^{p-1} f(r)\, dr$$

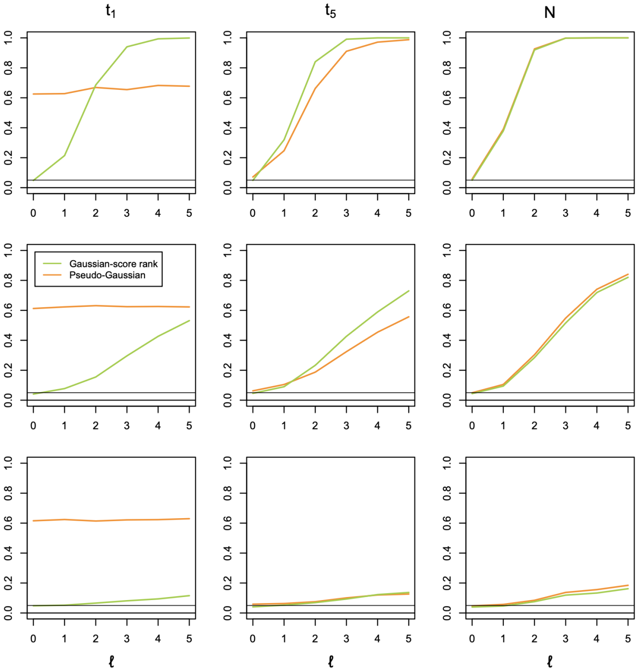
is the cdf of $d_1(\mu, V)$ under f.

Again, this result holds under arbitrarily weak identifiability.

This allows us to study the null asymptotic behavior of rank tests under weak identifiability: they are as robust to weak identifiability as pseudo-Gaussian tests, but extend validity to infinite fourth moments.

Better: from contiguity, combining this representation result with our study of the asymptotic behavior of elliptical log-likelihood ratios under weak identifiability, we can study the local powers of rank tests: for Gaussian scores, uniform dominance over pseudo-Gaussian tests in terms of AREs resists weak identifiability.

		Underlying density						
K	p	t_5	t_8	t_{12}	$\mathcal{N}$	e_2	e_3	e_5
vdW	2	2.204	1.215	1.078	1.000	1.129	1.308	1.637
	3	2.270	1.233	1.086	1.000	1.108	1.259	1.536
	4	2.326	1.249	1.093	1.000	1.093	1.223	1.462
	6	2.413	1.275	1.106	1.000	1.072	1.174	1.363
	10	2.531	1.312	1.126	1.000	1.050	1.121	1.254
	∞	3.000	1.500	1.250	1.000	1.000	1.000	1.000
L	2	1.500	0.750	0.625	0.500	0.392	0.365	0.347
	3	1.800	0.900	0.750	0.600	0.493	0.464	0.444
	4	2.000	1.000	0.833	0.667	0.565	0.537	0.517
	6	2.250	1.125	0.938	0.750	0.662	0.636	0.617
	10	2.500	1.250	1.041	0.833	0.766	0.746	0.730
	∞	3.000	1.500	1.250	1.000	1.000	1.000	1.000



t_1
t_5
N
Gaussian-score rank
Pseudo-Gaussian
0.0
0.2
0.4
0.6
0.8
1.0
0 1 2 3 4 5
ℓ

(iii) So far, we considered weak identifiability with a single-spiked spectrum

$$\lambda_1 = 1 + r_n v, \quad \lambda_2 = \ldots = \lambda_p = 1,$$

but what happens for more general spectra of the form

$$\lambda_1 = 1 + r_n v, \quad \lambda_2 = \ldots = \lambda_q = 1 < \lambda_{q+1} < \ldots < \lambda_p?$$

Both pseudo-Gaussian tests and rank tests show the same asymptotic null behavior under such more general spectra.

Their local powers are affected, but not the AREs!

Let $X_{n1}, \ldots, X_{nn}$ be a random sample from an elliptical distribution with location zero and shape matrix (throughout, shapes have unit determinants)

$$V_n = \frac{I_p + r_n v \theta_1 \theta_1'}{(\det(I_p + r_n v \theta_1 \theta_1'))^{1/p}},$$

with $r_n = O(1)$ and $v > 0$. Here, radial densities may freely depend on n.

We want to estimate the leading eigenvector θ_1.

We consider $\hat{\theta}_{n1}$, the leading eigenvector of the estimator of shape $\hat{V}_n$ solving

$$\frac{p}{n} \sum_{i=1}^{n} \frac{X_{ni} X_{ni}'}{X_{ni}' \hat{V}_n^{-1} X_{ni}} = \hat{V}_n;$$

see Tyler (AoS 1987). It can be shown that, irrespective of r_n, $\sqrt{n}(\hat{V}_n - V_n)$ is asymptotically normal with mean zero and covariance matrix $g_p(V)$, $V := \lim V_n$.

Theorem 5

(i) *if* $r_n \equiv 1$, *then* $\sqrt{n}(\hat{\theta}_{n1} - \theta_1)$ *is asymptotically normal with mean zero and covariance matrix*

$$\frac{1+v}{v^2}\left(1+\frac{2}{p}\right)(I_p - \theta_1\theta_1');$$

(ii) *if* r_n *is* $o(1)$ *with* $\sqrt{n}r_n \to \infty$, *then* $\sqrt{n}r_n(\hat{\theta}_{n1} - \theta_1)$ *is asymptotically normal with mean zero and covariance matrix*

$$\frac{1}{v^2}\left(1+\frac{2}{p}\right)(I_p - \theta_1\theta_1');$$

(iii) *if* $r_n = \frac{1}{\sqrt{n}}$, *then* $\hat{\theta}_{n1}$ *converges weakly to the unit eigenvector associated with the largest eigenvalue of* $Z + v\theta_1\theta_1'$, $Z \sim \mathcal{N}_{p,p}\big(0, (1+\frac{2}{p})\{(I_{p^2} + K_p) - \frac{2}{p}J_p\}\big)$;

(iv) *if* $r_n = o(\frac{1}{\sqrt{n}})$, *then* $\hat{\theta}_{n1}$ *converges weakly to a random vector that is uniformly distributed over the unit sphere* $\mathcal{S}^{p-1}$.

Theorem 5

(i) *if* $r_n \equiv 1$, *then* $\sqrt{n}(\hat{\theta}_{n1} - \theta_1)$ *is asymptotically normal with mean zero and covariance matrix*

$$\frac{1+v}{v^2}\left(1+\frac{2}{p}\right)(I_p - \theta_1\theta_1');$$

(ii) *if* r_n *is* $o(1)$ *with* $\sqrt{n}r_n \to \infty$, *then* $\sqrt{n}r_n(\hat{\theta}_{n1} - \theta_1)$ *is asymptotically normal with mean zero and covariance matrix*

$$\frac{1}{v^2}\left(1+\frac{2}{p}\right)(I_p - \theta_1\theta_1');$$

(iii) *if* $r_n = \frac{1}{\sqrt{n}}$, *then* $\hat{\theta}_{n1}$ *converges weakly to the unit eigenvector associated with the largest eigenvalue of* $Z + v\theta_1\theta_1'$, $Z \sim \mathcal{N}_{p,p}\big(0, (1+\frac{2}{p})\{(I_{p^2} + K_p) - \frac{2}{p}J_p\}\big)$;

(iv) *if* $r_n = o(\frac{1}{\sqrt{n}})$, *then* $\hat{\theta}_{n1}$ *converges weakly to a random vector that is uniformly distributed over the unit sphere* $\mathcal{S}^{p-1}$.

Theorem 5

(i) *if* $r_n \equiv 1$, *then* $\sqrt{n}(\hat{\theta}_{n1} - \theta_1)$ *is asymptotically normal with mean zero and covariance matrix*

$$\frac{1+v}{v^2}\left(1+\frac{2}{p}\right)(I_p - \theta_1\theta_1');$$

(ii) *if* r_n *is* $o(1)$ *with* $\sqrt{n}r_n \to \infty$, *then* $\sqrt{n}r_n(\hat{\theta}_{n1} - \theta_1)$ *is asymptotically normal with mean zero and covariance matrix*

$$\frac{1}{v^2}\left(1+\frac{2}{p}\right)(I_p - \theta_1\theta_1');$$

(iii) *if* $r_n = \frac{1}{\sqrt{n}}$, *then* $\hat{\theta}_{n1}$ *converges weakly to the unit eigenvector associated with the largest eigenvalue of* $Z + v\theta_1\theta_1'$, $Z \sim \mathcal{N}_{p,p}\big(0, (1+\frac{2}{p})\big\{(I_{p^2} + K_p) - \frac{2}{p}J_p\big\}\big)$;

(iv) *if* $r_n = o(\frac{1}{\sqrt{n}})$, *then* $\hat{\theta}_{n1}$ *converges weakly to a random vector that is uniformly distributed over the unit sphere* $\mathcal{S}^{p-1}$.

Theorem 5

(i) *if* $r_n \equiv 1$, *then* $\sqrt{n}(\hat{\theta}_{n1} - \theta_1)$ *is asymptotically normal with mean zero and covariance matrix*

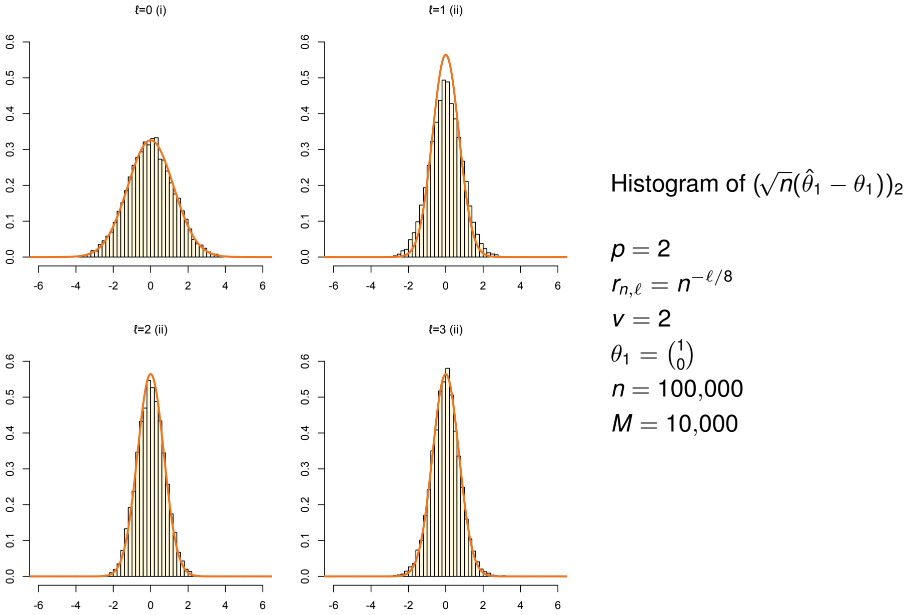
$$\frac{1+v}{v^2}\left(1+\frac{2}{p}\right)(I_p - \theta_1\theta_1');$$

(ii) *if* r_n *is* $o(1)$ *with* $\sqrt{n}r_n \to \infty$, *then* $\sqrt{n}r_n(\hat{\theta}_{n1} - \theta_1)$ *is asymptotically normal with mean zero and covariance matrix*

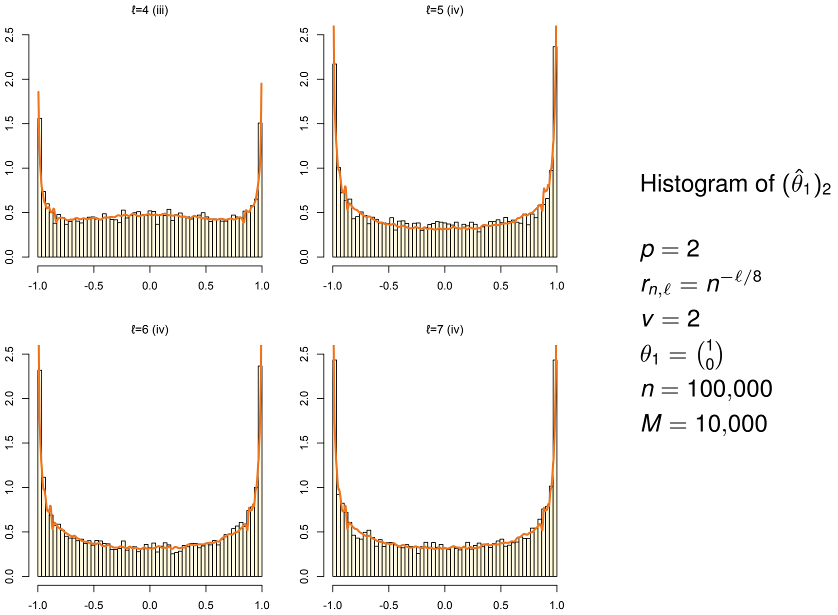
$$\frac{1}{v^2}\left(1+\frac{2}{p}\right)(I_p - \theta_1\theta_1');$$

(iii) *if* $r_n = \frac{1}{\sqrt{n}}$, *then* $\hat{\theta}_{n1}$ *converges weakly to the unit eigenvector associated with the largest eigenvalue of* $Z + v\theta_1\theta_1'$, $Z \sim \mathcal{N}_{p,p}\left(0, \left(1+\frac{2}{p}\right)\left\{(I_{p^2} + K_p) - \frac{2}{p}J_p\right\}\right)$;

(iv) *if* $r_n = o(\frac{1}{\sqrt{n}})$, *then* $\hat{\theta}_{n1}$ *converges weakly to a random vector that is uniformly distributed over the unit sphere* $\mathcal{S}^{p-1}$.



ℓ=0 (i)
ℓ=1 (ii)
ℓ=2 (ii)
ℓ=3 (ii)
0.0 0.1 0.2 0.3 0.4 0.5 0.6
-6 -4 -2 0 2 4 6
Histogram of $(\sqrt{n}(\hat{\theta}_1 - \theta_1))_2$
$p = 2$
$r_{n,\ell} = n^{-\ell/8}$
$v = 2$
$\theta_1 = \binom{1}{0}$
$n = 100{,}000$
$M = 10{,}000$



ℓ=4 (iii)
ℓ=5 (iv)
ℓ=6 (iv)
ℓ=7 (iv)
0.0
0.5
1.0
1.5
2.0
2.5
-1.0
-0.5
0.0
0.5
1.0
Histogram of $(\hat{\theta}_1)_2$
$p = 2$
$r_{n,\ell} = n^{-\ell/8}$
$v = 2$
$\theta_1 = \binom{1}{0}$
$n = 100{,}000$
$M = 10{,}000$

Weak identifiability (WI) may hurt!

But some procedures are validity-robust to weak identifiability.
They may even show adaptively optimal Type 2 risks.

In some problems, robustness to WI might therefore be a further point to consider when selecting a statistical procedure(?)

To do 1: high-dimensional case ($p = p_n \to \infty$)
To do 2: LR vs score tests in a generic WI problem?

Weak identifiability (WI) may hurt!

But some procedures are validity-robust to weak identifiability.
They may even show adaptively optimal Type 2 risks.

In some problems, robustness to WI might therefore be a further point to consider when selecting a statistical procedure(?)

To do 1: high-dimensional case ($p = p_n \to \infty$)
To do 2: LR vs score tests in a generic WI problem?

Thank you!

Paindaveine, D., Remy, J., and Verdebout, Th. (2020). Testing for principal component directions under weak identifiability. *Annals of Statistics* 48, 324–345.

Paindaveine, D., Remy, J., and Verdebout, Th. (2020). Sign tests for weak principal directions. *Bernoulli* 26, 2987–3016.

Paindaveine, D., and Verdebout, Th. (2023). On the asymptotic behavior of the leading eigenvector of Tyler's shape estimator under weak identifiability. In: Nordhausen, K., Yi, M. (eds), Robust and Multivariate Statistical Methods. Festschrift in honor of David Tyler. Springer, Cham., 45–64.

Paindaveine, D., Peralvo Maroto, L. and Verdebout, Th. (2024+). Rank tests for PCA under weak identifiability. *Submitted.*