meets

Testing for Independence of Large Dimensional Vectors

Nestor Parolya

Institute of Statistics, Leibniz University Hannover

joint work with Holger Dette and Taras Bodnar, part of the DFG-project
"Structural inference for high-dimensional covariance matrices"

Outline

The problem of testing independence I

- p-dimensional random vector $\mathbf{y}_1$
- Decomposition in two blocks

$$\mathbf{y}_1 = \begin{pmatrix} \left.\begin{matrix} y_{11} \\ \vdots \\ y_{1p_1} \end{matrix}\right\} p_1 \\ \left.\begin{matrix} y_{1p_1+1} \\ \vdots \\ y_{1p_1+p_2} \end{matrix}\right\} p_2 \end{pmatrix}$$

- **Question:** Are $y_{11}, \dots, y_{1p_1}$ independent of $y_{1p_1+1}, \dots, y_{1p_1+p_2}$?

The problem of testing independence II

- **Question:** Are $y_{11}, \ldots, y_{1p_1}$ independent of $y_{1p_1+1}, \ldots, y_{1p_1+p_2}$?
- Alternative formulation: if $\mathbf{y} \sim \mathcal{N}(\mu, \boldsymbol{\Sigma})$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

$$\underbrace{\qquad\quad}_{p_2 \times p_1} \quad \underbrace{\qquad\quad}_{p_2 \times p_2}$$

- **Is the covariance matrix block diagonal?**

$$\mathbf{H_0}: \ \boldsymbol{\Sigma}_{12} = \mathbf{0} \quad \in \mathbb{R}^{p_1 \times p_2} \quad \text{versus} \quad \mathbf{H_1}: \ \boldsymbol{\Sigma}_{12} = \mathbf{0}$$

Likelihood ratio test

- Sample covariance matrix of an *i.i.d* sample $\mathbf{y}_1, \ldots, \mathbf{y}_n \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\mathbf{S}_n = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$$

- Likelihood ratio test (Wilks, 1939) rejects the null hypothesis, if

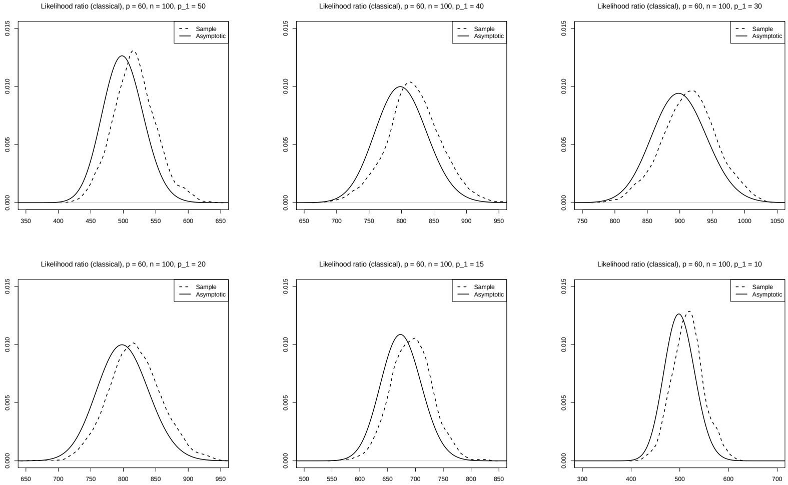
$$-2\rho_{p_1,p_2} \log V_n > \chi^2_{1-\alpha,df}$$

where

$$V_n = \frac{|\mathbf{S}_n|}{|\mathbf{S}_{11}||\mathbf{S}_{22}|}$$

$$df = \frac{1}{2}\big((p_1+p_2)(p_1+p_2+1) - p_1(p_1+1) - p_2(p_2+1)\big) = p_1 p_2$$

χ^2-approximation ($n = 100,\ p_1 + p_2 = 60$)



Simulated distribution of the LR-test statistic $-2\rho_{p_1,p_2} \log V_n$ under the null hypo... *s*

Remarks:

- There is a systematic bias in the approximation
- The χ^2- approximation is based on “classical” theory:

$$p_1, p_2, p \text{ are fixed, and} \quad n \to \infty$$

- **Can we get better approximations using a different point of view, that is:**

$$\lim_{n\to\infty} \frac{p_i}{n} = c_i \in (0, 1)$$

Alternative asymptotic distribution theory

Dimension increases with sample size n:

- $\mathbf{y}_1, \dots, \mathbf{y}_n \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_n)$
  In general, we allow normal mixtures in form $\mathbf{y}_i \sim R\mathbf{x}$ with $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_n)$ and R is a pos. def. random variable ind. of $\mathbf{x}$ (so called *generating variable)*
- $\boldsymbol{\Sigma}_n \in \mathbb{R}^{p\times p}$ is the positive definite population covariance matrix with **bounded** spectrum
  $$0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p$$
  as $p \to \infty$.
- p_i dimension of block i $(i = 1, 2)$
- $p = p_1 + p_2$ the total number of variables
- asymptotic regime:
  $$\lim_{n\to\infty} \frac{p_i}{n} = c_i \in (0, 1)$$

Asymptotic normality (Yao, Bai and Zheng, 2015)

Theorem

Under the null hypothesis

$$\frac{\log V_n - p_2 s_{LR,n} - \mu_{LR,n}}{\sigma_{LR,n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)\,,$$

where $s_{LR,n}$, $\mu_{LR,n}$ *and* $\sigma_{LR,n}$ *depend only on* p_1, p_2 *and* n*.*

Details on the constants

$$\mu_{LR} = 1/2 \log\left[\frac{(w_n^{*\,2} - \gamma_{2,n}^2) w_n^{*\,2}}{(w_n^{*\,2} - \gamma_{2,n}^{3/2})^2}\right], \qquad \sigma_{LR}^2 = 2\log\left[\frac{w_n^{*\,2}}{w_n^{*\,2} - \gamma_{2,n}}\right],$$

$$s_{LR} = \log\left(\frac{\gamma_{1,n}}{\gamma_{2,n}}(1-\gamma_{2,n})^2\right) + \frac{1-\gamma_{2,n}}{\gamma_{2,n}}\log(w_n^*) - \frac{\gamma_{1,n}+\gamma_{2,n}}{\gamma_{1,n}\gamma_{2,n}}\log(w_n^* - \gamma_{2,n}^2/w_n^*)$$

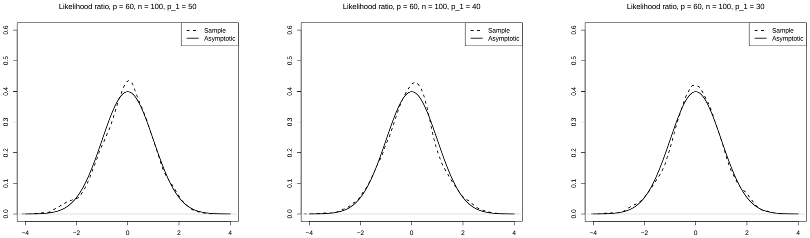
$$+ \begin{cases} \frac{1-\gamma_{1,n}}{\gamma_{1,n}}\log(w_n^* - w_n^*\gamma_{2,n}), & \gamma_{1,n} \in (0,1) \\ 0, & \gamma_{1,n} = 1 \\ -\frac{1-\gamma_{1,n}}{\gamma_{1,n}}\log(w_n^* - \gamma_{2,n}/w_n^*), & \gamma_{1,n} > 1\,. \end{cases},$$

where

$$\gamma_{1,n} = \frac{p_2}{p_1} \in (0,+\infty)\ , \quad \gamma_{2,n} = \frac{p_2}{n-p_1} \in (0,1),$$

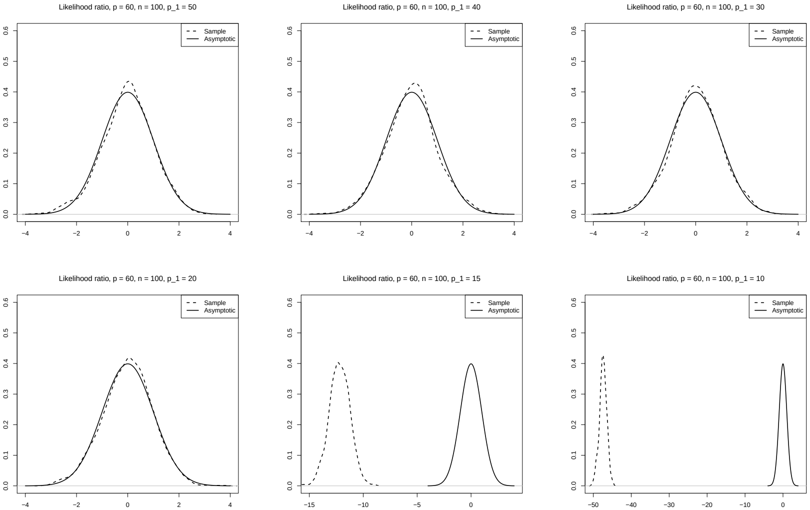
$$w_n^* = \sqrt{\frac{\gamma_{1,n}+\gamma_{2,n}-\gamma_{1,n}\gamma_{2,n}}{\gamma_{2,n}}}\,.$$

Normal approximation ($n = 100$, $p_1 + p_2 = 60$)



Normal approximation for the distribution of $\left(\log V_n - p_2 s_{LR,n} - \mu_{LR,n} \right) / \sigma_{LR,n}$ *under the null hypothesis*

Finite sample properties of the normal approximation



Normal approximation for the distribution of $\left(\log V_n - p_2 s_{LR,n} - \mu_{LR,n}\right)/\sigma_{LR,n}$ *under the null hypothesis*

New tests for independence

- **Recall:**

$$\boldsymbol{\Sigma}_n = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

- **Note:** the hypothesis

$$H_0: \ \boldsymbol{\Sigma}_{12} = \mathbf{O} \quad \text{versus} \quad H_1: \ \boldsymbol{\Sigma}_{12} \neq \mathbf{O}\,,$$

is equivalent to

$$H_0: \ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} = \mathbf{O} \quad \text{versus} \quad H_1: \ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \neq \mathbf{O}$$

Fisher matrix I

- **Decompose the sample covariance matrix**

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$$
$$\underbrace{\qquad}_{p_2 \times p_1} \quad \underbrace{\qquad}_{p_2 \times p_2}$$

- Estimate the matrix $\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$ by

$$\mathbf{W} = \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$$

Fisher matrix II

- **Central Wishart** distribution under the null hypothesis, i.e.

$$\mathbf{W} = \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} \sim W_{p_2}(p_1, \mathbf{\Sigma}_{22\cdot 1}),$$

where $\mathbf{\Sigma}_{22\cdot 1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$ is the corresponding Schur complement (Muirhead, 1982).

- **Non-central Wishart** distribution under the alternative **conditionally on $\mathbf{S}_{11}$**, that is

$$\mathbf{W}|\mathbf{S}_{11} \sim W_{p_2}(p_1, \mathbf{\Sigma}_{22\cdot 1}, \mathbf{\Omega}_1),$$

where

$$\mathbf{\Omega}_1 = \mathbf{\Sigma}_{22\cdot 1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}.$$

Fisher matrix III

- Estimate the Schur complement $\mathbf{\Sigma}_{22\cdot1}$ by

$$\mathbf{T} = \mathbf{S}_{22\cdot1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} \sim W_{p_2}(n - p_1, \mathbf{\Sigma}_{22\cdot1})$$

- **Note:** under the null hypothesis and alternative
  - $\mathbf{T} \sim W_{p_2}(n - p_1, \mathbf{\Sigma}_{22\cdot1})$
  - The matrices $\mathbf{W}$ and $\mathbf{T}$ are independent.

Fisher matrix V

- **Note:** Under the null hypothesis of independence

$$\begin{aligned} \mathbf{T} &\sim W_{p_2}(n - p_1, \mathbf{\Sigma}_{22}) \\ \mathbf{W} &\sim W_{p_2}(p_1, \mathbf{\Sigma}_{22}) \end{aligned}$$

- **In particular:** Under the null hypothesis the distribution of $\mathbf{F} = \mathbf{W}\mathbf{T}^{-1}$ does not depend on $\mathbf{\Sigma}$ (distribution free).
- The matrix $\mathbf{F} = \mathbf{W}\mathbf{T}^{-1}$ is called **Fisher matrix** (central under the null hypothesis and non-central under the alternative)
- We will use **linear combinations** of the eigenvalues to test the hypothesis of independence!

Example: eigenvalues of Fisher matrix IV

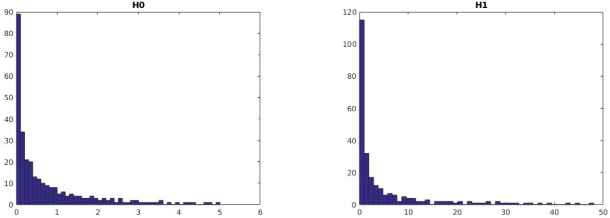
- $A \in \mathbb{R}^{p \times p}$ is a matrix with i.i.d. standard normal distributed variables
- Covariance matrix under H_1

$$\mathbf{\Sigma}_{H_1} = AA^T = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$$

- Covariance matrix under H_0

$$\mathbf{\Sigma}_{H_0} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22} \end{pmatrix}$$

- Empirical eigenvalue distribution of $\mathbf{F}$ based on a sample of $n = 1000$ i.i.d. $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{H_0})$ and $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{H_1})$ random variables ($p_1 = 300$, $p_2 = 300$, $p = 600$)



Alternative test statistics (MANOVA)

(1) Wilks' Λ statistics:

$$T_W = -\log(|\mathbf{T}|/|\mathbf{T}+\mathbf{W}|) = \log(|\mathbf{I}+\mathbf{W}\mathbf{T}^{-1}|) = \log(|\mathbf{I}+\mathbf{F}|)$$

(2) Lawley-Hotelling's trace criterion:

$$T_{LH} = tr(\mathbf{W}\mathbf{T}^{-1}) = tr(\mathbf{F})$$

(3) Bartlett-Nanda-Pillai's trace criterion:

$$T_{BNP} = tr(\mathbf{W}\mathbf{T}^{-1}(\mathbf{I}+\mathbf{W}\mathbf{T}^{-1})^{-1}) = tr(\mathbf{F}(\mathbf{I}+\mathbf{F})^{-1})$$

Note: all statistics depend on the eigenvalues $v_1 \geq v_2 \geq \ldots \geq v_{p_2}$ of the matrix $\mathbf{F}$

Representation as linear spectral statistics

Note: all statistics depend on the eigenvalues $v_1 \geq v_2 \geq ... \geq v_{p_2}$ of the matrix **F**

(1) Wilks' Λ statistics:

$$T_W = \log(|\mathbf{I} + \mathbf{F}|) = \sum_{i=1}^{p_2} \log(1 + v_i)$$

(2) Lawley-Hotelling's trace criterion:

$$T_{LH} = tr(\mathbf{W}\mathbf{T}^{-1}) = tr(\mathbf{F}) = \sum_{i=1}^{p_2} v_i$$

(3) Bartlett-Nanda-Pillai's trace criterion:

$$T_{BNP} = tr(\mathbf{F}(\mathbf{I} + \mathbf{F})^{-1})) = \sum_{i=1}^{p_2} \frac{v_i}{1 + v_i}$$

Linear spectral statistics

- Eigenvalues of the matrix $\mathbf{F} = \mathbf{W}\mathbf{T}^{-1}$;

$$v_1 \geq v_2 \geq \ldots \geq v_{p_2}$$

- Empirical spectral distribution function:

$$F_n(x) = \frac{1}{p_2} \sum_{i=1}^{p_2} \mathbb{1}_{(-\infty, v_i]}(x)$$

- Linear spectral statistic: let $f : \mathbb{R} \to \mathbb{R}$ denote a "suitable" function

$$LSS_n = p_2 \int_0^\infty f(x) dF_n(x) = \sum_{i=1}^{p_2} f(v_i)$$

Linear spectral statistics

- **Question:** Can we find the asymptotic distribution of the linear spectral statistic

$$LSS_n = p_2 \int_0^\infty f(x) dF_n(x) = \sum_{i=1}^{p_2} f(v_i)$$

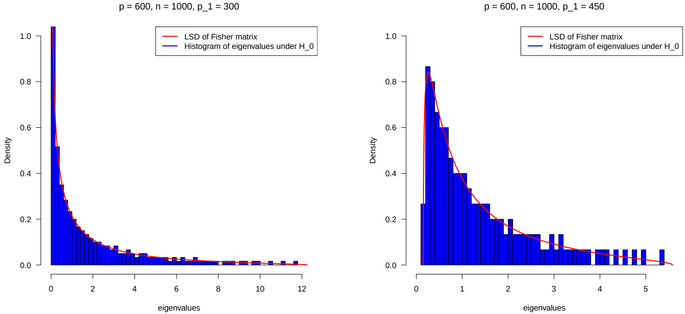
for many functions of f?

- This is a very difficult problem in random matrix theory
- For this purpose we need to have knowledge about the asymptotic properties of the eigenvalues $v_1, \dots v_{p_2}$ as $n, p_1, p_2 \to \infty$.
- In this talk:

$$\lim_{n, p_i \to \infty} \frac{p_i}{n} = c_i \in (0,1) \ , \ \ i = 1, 2.$$

Asymptotic properties of the spectrum

Example: Empirical spectral distribution of the Fisher matrix and the limiting density



$$q(x) = \frac{1-\gamma_2}{2\pi x(\gamma_1+\gamma_2 x)}\sqrt{(b-x)(x-a)}, \quad a = \frac{(1-h)^2}{(1-\gamma_2)^2}, \quad b = \frac{(1+h)^2}{(1-\gamma_2)^2}$$

$$\gamma_1 = \lim_{n\to\infty}\gamma_{1,n} = \lim_{n\to\infty}\frac{p_2}{p_1}, \quad \gamma_2 = \lim_{n\to\infty}\gamma_{2,n} = \lim_{n\to\infty}\frac{p_2}{n-p_1}$$

$$h = \lim_{n\to\infty} h_n = \sqrt{\gamma_{1,n}+\gamma_{2,n}-\gamma_{1,n}\gamma_{2,n}}$$

Asymptotic properties of the spectrum

Take home message I:

- The empirical spectrum of a Fisher matrix converges almost surely to a well defined density.
- This distribution appears in the standardisation of the linear spectral statistic.

Asymptotic distribution of linear spectral statistics I

- **Take home message II:**
  - Under the null hypothesis standardised versions of linear spectral statistics are asymptotically normal distributed.
  - The constants in this standardisation are very complicated (and depend on the limiting distribution of the the spectrum).
  - **For a more precise statement** we need the definition of the **Stieltjes transform**

$$m_G(z) = \int \frac{G(dt)}{t - z}$$

    of a distribution function G.
  - The Stieltjes transform has similar properties as the characteristic function, for example:
    - G is determined by m_G
    - Convergence in distribution can be characterised in terms of convergence of the Stieltjes transforms

Asymptotic distribution of linear spectral statistics I

- **A more precise statement of asymptotic normality of linear spectral** as $n, p_1, p_2 \to \infty$

$$\sum_{i=1}^{p_2} f(v_i) - p_2 \int_0^\infty f(x) q_n(x) dx$$
$$= p_2 \Big(\int_0^\infty f(x) dF_n(x) - \int_0^\infty f(x) q_n(x) dx \Big) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{\mu}, \sigma^2)$$

where

$$q_n(x) = \frac{1-\gamma_{2,n}}{2\pi x(\gamma_{1,n} + \gamma_{2,n} x)} \sqrt{(b_n - x)(x - a_n)}, \;\; a_n = \frac{(1-h_n)^2}{(1-\gamma_{2,n})^2}, \;\; b_n = \frac{(1+h_n)^2}{(1-\gamma_{2,n})^2}$$

- Asymptotic mean μ and variance σ^2 depend on the Stieltjes transform

$$m_q(z) = \int \frac{q(t)dt}{t-z}$$

of the limiting density q of the spectrum in a complicated manner

Asymptotic distribution of linear spectral statistics II

$$
\begin{aligned}
\mu &= \frac{1}{2\pi i}\oint f(z) d\log\left(\frac{\frac{1-c}{1-c_1} m_0^2(z) + 2m_0(z) + 2 - c/c_1}{\frac{1-c}{1-c_1} m_0^2(z) + 2m_0(z) + 1}\right) \\
&+ \frac{1}{2\pi i}\oint f(z) d\log\left(\frac{1 - \frac{c-c_1}{1-c_1} m_0^2(z)}{(1+m_0^2(z))^2}\right) \\
\sigma^2 &= -\frac{1}{2\pi^2}\oint\oint \frac{f(z_1)f(z_2)}{(m_0(z_1)-m_0(z_2))^2} dm_0(z_1) dm_0(z_2)
\end{aligned}
$$

- The integrals are taken over arbitrary positively oriented contour which contains the interval $[a, b]$.
- For a given f (e.g. $f(z) = \log z$) μ and σ^2 can be calculated

Asymptotic distribution under H_0

Theorem

Let $\alpha \in \{W, LH, BNP\}$, then under the null hypothesis H_0

$$\frac{T_\alpha - p_2 s_{\alpha,n} - \mu_{\alpha,n}}{\sigma_{\alpha,n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

where $s_{\alpha,n}$, $\mu_{\alpha,n}$ and $\sigma^2_{\alpha,n}$ depend on p_1, p_2 and n.

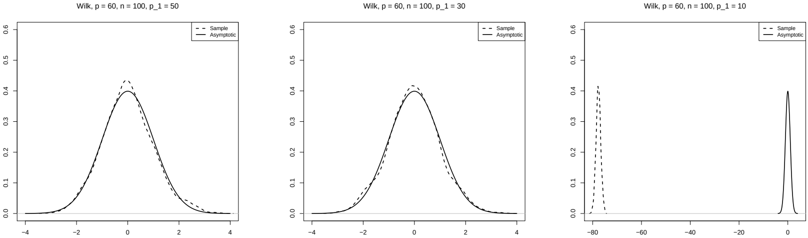
Example: Lawley-Hotelling's trace criterion:

$$\mu_{LH} = \frac{\gamma_{2,n}}{(1-\gamma_{2,n})^2}, \qquad \sigma^2_{LH} = \frac{2(\gamma_{1,n} + \gamma_{2,n} - \gamma_{1,n}\gamma_{2,n})}{(1-\gamma_{2,n})^4}, \qquad s_{LH} = \frac{1}{1-\gamma_{2,n}}$$

where

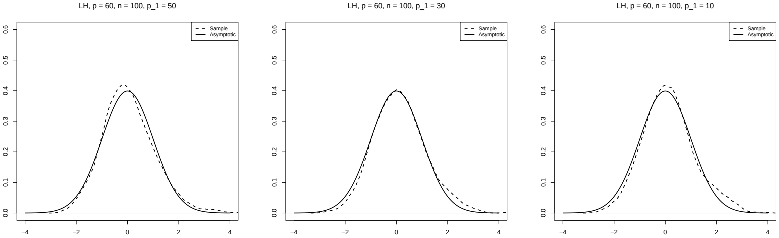
$$\gamma_{1,n} = \frac{p_2}{p_1} \in (0, +\infty), \quad \gamma_{2,n} = \frac{p_2}{n - p_1} \in (0,1)$$

Simulation under H_0: Wilks' Λ



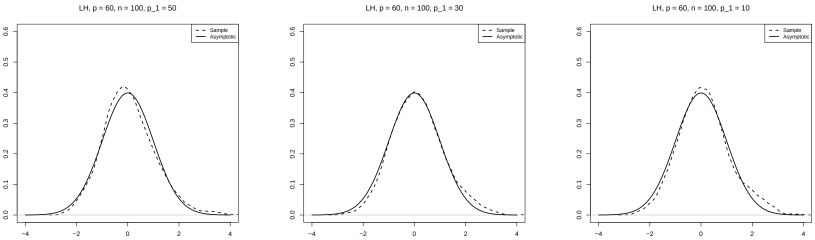
No reliable approximation if p_1 is small compared to p_2!

Simulation under H_0: Lawley-Hotelling's trace criterion



Reliable approximation in all cases!

Simulation under H_0: Bartlett-Nanda-Pillai's trace criterion



Reliable approximation in all cases!

Analysis under the alternative

- **Recall:** Note that under the alternative the matrix $\mathbf{W}\mathbf{T}^{-1}|\mathbf{S}_{11}$ has a non-central Fisher matrix with non-centrality parameter

$$\mathbf{\Omega}_1 = \mathbf{\Sigma}_{22\cdot 1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$$

- Proceed in two steps:
  (1) Determine the asymptotic distribution of the empirical spectral distribution (this is needed for centering - at least)
  (2) Determine the asymptotic distribution of the linear spectral statistics (extremely difficult)

For the illustration of the type of result we recall the definition of the **Stieltjes transform**

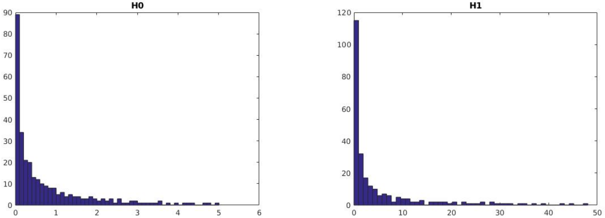
$$m_G(z) = \int \frac{G(dt)}{t - z}$$

of a distribution function G

Analysis under alternative hypothesis (take home)

Take home message III(a): The empirical spectral distribution of the matrix $\mathbf{F} = \mathbf{W}\mathbf{T}^{-1}$ converges almost surely to a deterministic distribution function F^*, which depends on the eigenvalues of the matrix

$$\mathbf{R} = \mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1/2} = \mathbf{\Sigma}_{22\cdot1}^{1/2}\mathbf{\Omega}_1\mathbf{\Sigma}_{22\cdot1}^{-1/2}$$



Analysis under the alternative

Theorem

If the empirical spectral distribution of the matrix

$$\mathbf{R} = \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot 1}^{-1/2} = \mathbf{\Sigma}_{22\cdot 1}^{1/2}\mathbf{\Omega}_1\mathbf{\Sigma}_{22\cdot 1}^{-1/2}$$

converges weakly to some function G then the empirical spectral distribution of $\mathbf{F} = \mathbf{W}\mathbf{T}^{-1}$ converges almost surely to a deterministic distribution function F^*. The Stieltjes transform

$$s(z) = m_{F^*}(z) = \int \frac{F^*(dt)}{t - z}$$

of F^* is the unique solution of the system of equations

$$\begin{aligned}
\frac{s(z)}{1+\gamma_2 zs(z)} &= m_H(z(1+\gamma_2 zs(z))),\\
\frac{m_H(z)}{1+\gamma_1 m_H(z)} &= m_{\tilde{H}}((1+\gamma_1 m_H(z))[(1+\gamma_1 m_H(z))z-(1-\gamma_1)]),\\
m_{\tilde{H}}(z)(1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z))c_1^{-1} &= m_G\left(\frac{c_1 z}{1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z)}\right)
\end{aligned}$$

Linear spectral statistics under the alternative

- **The distribution F^* is required for the centering of the linear spectral statistic** and its Stieltjes transform

$$s(z) = m_{F^*}(z) = \int \frac{F^*(dt)}{t-z}$$

is the unique solution of the system of equations

$$\begin{aligned} \frac{s(z)}{1+\gamma_2 zs(z)} &= m_H(z(1+\gamma_2 zs(z))), \\ \frac{m_H(z)}{1+\gamma_1 m_H(z)} &= m_{\tilde{H}}((1+\gamma_1 m_H(z))[(1+\gamma_1 m_H(z))z-(1-\gamma_1)]), \\ m_{\tilde{H}}(z)(1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z))c_1^{-1} &= m_G\left(\frac{c_1 z}{1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z)}\right) \end{aligned}$$

This has to be solved recursively ($\tilde{H} \to H \to F^* \to F_n^*$)

- **Empirical analogue F_n^*:** Replace
γ_1, γ_2, c_1 and c_2 by $\frac{p_2}{p_1}$, $\frac{p_2}{n-p_1}$, $\frac{p_1}{n}$ and $\frac{p_2}{n}$

CLT for linear spectral of under the alternative

Theorem

If $n, p_1, p_2 \to \infty$, then

$$p_2\Big(\int_0^\infty f(x)F_n(dx) - \int_0^\infty f(x)F_n^*(x)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(\mu, \sigma^2)$$

- Asymptotic mean and variance are very complicated
  - a system of three equations for the Stieltjes transform has to be solved recursively ($\tilde{H} \to H \to F^* \to F_n^*$)
  - this system reduces to a quadratic equation under the null hypothesis

$$
\begin{aligned}
\mathbb{E}[X_f] &= \frac{1}{4\pi i}\oint f(z)d\log(q(z)) + \frac{1}{2\pi i}\oint f(z)B(zb(z))d(zb(z)) \\
&+ \frac{1}{2\pi i}\oint f(z)\theta_{b,H}(z)\left(\theta_{\tilde{b},\tilde{H}}(zb(z))\frac{c_1^2\int \underline{m}_{\tilde{H}}^3(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-3}dG(t)}{(1-c_1\int \underline{m}_{\tilde{H}}^2(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-2}dG(t))^2}\right)dz
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}[X_f] &= -\frac{1}{2\pi^2}\oint\oint f(z_1)f(z_2)\frac{\partial^2\log(z_1b(z_1)-z_2b(z_2))}{\partial z_1\partial z_2}dz_1dz_2 \\
&- \frac{1}{2\pi^2}\oint\oint f(z_1)f(z_2)\frac{\partial^2\log(z_1b(z_1)\eta(z_1b(z_1))-z_2b(z_2)\eta(z_2b(z_2)))}{\partial z_1\partial z_1}dz_1dz_2 \\
&- \frac{1}{2\pi^2}\oint\oint f(z_1)f(z_2)\left[\theta_{\tilde{b},\tilde{H}}(z_1b(z_1))\theta_{\tilde{b},\tilde{H}}(z_2b(z_2))\left(\frac{\partial^2\log\left[\frac{\underline{m}_{\tilde{H}}(z_2b(z_2))-\underline{m}_{\tilde{H}}(z_1b(z_1))}{(z_2b(z_2)-z_1b(z_1))}\right]}{\partial z_1\partial z_2}\right)\right]dz_1dz_2
\end{aligned}
$$

Take home message III(b):

- Linear spectral statistics (appropriately normalised) of the Fisher matrix **F** are asymptotically normal distributed
- The standardisation and limiting distribution depend on the eigenvalues the matrix

$$\mathbf{R} = \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot 1}^{-1/2}$$

(more precisely on its asymptotic properties) in a complicated way.

- **But** the asymptotic properties do not depend on the eigenvectors of the matrix **R**
- Under the null hypothesis: $\mathbf{R} = \mathbf{0}$

Why all these efforts?

- Interesting mathematics!
- A better understanding of the properties of the tests!
- **Example:** Finite rank alternatives:
  - Finite rank alternatives $\mathbf{R}$ have no influence on the asymptotic power of the tests.
  - The asymptotic means and variances coincide under the null hypothesis and alternative.
  - Heuristically: tests based on a linear spectral statistics of the Fisher matrix cannot detect the alternative if the matrix $\mathbf{R}$ has no large eigenvalues.

Finite sample properties I

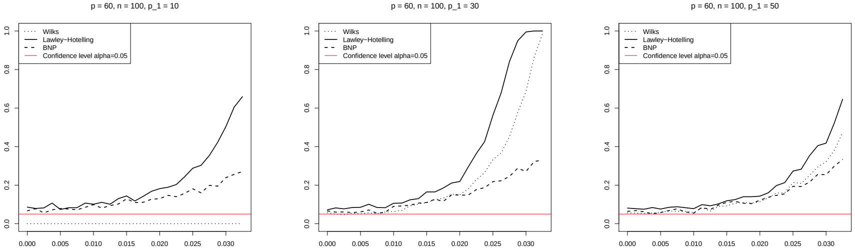
$$
\boldsymbol{\Sigma} = \begin{pmatrix}
1 & 0 & \dots & 0 & 0 & \rho & \rho & \dots & \rho & \rho \\
0 & 1 & \dots & 0 & 0 & \rho & \rho & \dots & \rho & \rho \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\
0 & 0 & \dots & 1 & 0 & \rho & \rho & \dots & \rho & \rho \\
0 & 0 & \dots & 0 & 1 & \rho & \rho & \dots & \rho & \rho \\
\rho & \rho & \dots & \rho & \rho & 1 & 0 & \dots & 0 & 0 \\
\rho & \rho & \dots & \rho & \rho & 0 & 1 & \dots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\rho & \rho & \dots & \rho & \rho & 0 & 0 & \dots & 0 & 1
\end{pmatrix}
$$

$$
\underbrace{\qquad\qquad\qquad}_{p_2 \times p_1} \qquad \underbrace{\qquad\qquad\qquad}_{p_2 \times p_2}
$$

Note:

- The correlation coefficient ρ will change in the interval $[0, 0.0325]$
- We set some elements of $\boldsymbol{\Sigma}_{12}$ (randomly) equal to zero (sparse $\boldsymbol{\Sigma}_{12}$).

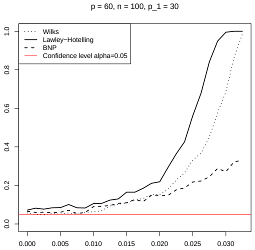
Comparison of new tests (power) I



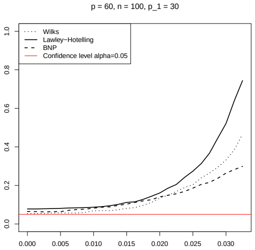
Note:

- All tests have problems to detect the alternative for small values of ρ (as predicted by our theory)
- The best power is obtained for $p_1 = p_2 = 30$
- The Lawley-Hotelling's trace criterion shows the best performance

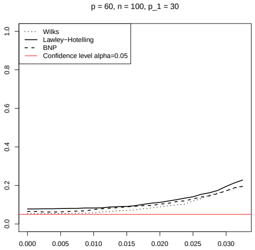
Comparison of new tests (power) II - increasing sparsity



Sparsity 0%



Sparsity 20%



Sparsity 50%

Note: $n = 100, \ p = 60, \ p_1 = 30$

- The power decreases with increasing sparsity
- The Lawley-Hotelling's trace criterion shows the best performance

Other benchmarks

(1) Trace criterion introduced by Jiang-Bai-Zheng(2013):

$$T_{JBZ} = \text{tr}\left[\mathbf{W}\mathbf{T}^{-1}\left(\mathbf{W}\mathbf{T}^{-1} + \frac{\gamma_{1,n}}{\gamma_{2,n}}\mathbf{I}_{p-p_1}\right)^{-1}\right]$$

(2) Minimum distance test of Yamada-Hyodo-Nishiyama (2017):

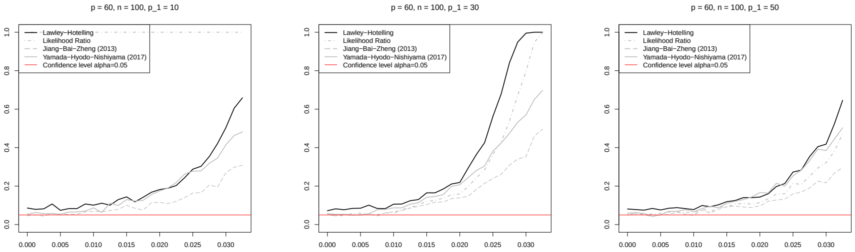
$$T_{YHN} = (n-2)(n-1)\text{tr}(\mathbf{S}^2) + (\text{tr}(\mathbf{S}))^2$$

(3) Likelihood ratio test

$$T_{LR} = \log\left(\frac{|\mathbf{S}|}{|\mathbf{S}_{11}||\mathbf{S}_{22}|}\right)$$

Note: Standardised versions of the test statistics are asymptotically normal distributed (linear spectral statistics)

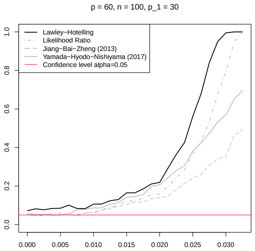
Comparison with alternative tests (power) I



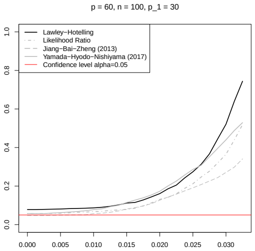
Note: $n = 100,\ p = 60,\ p_1 = 30$

- The best power is obtained for $p_1 = p_2 = 30$
- The Lawley-Hotelling's trace criterion shows the best performance

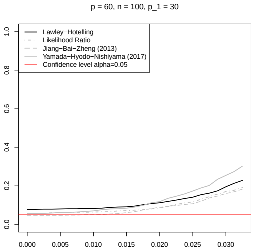
Comparison with alternative tests (power) II - increasing sparsity



Sparsity 0%



Sparsity 20%



Sparsity 50%

Note: $n = 100,\ p = 60,\ p_1 = 30$

- The power decreases with increasing sparsity
- The Lawley-Hotelling's trace criterion shows the best performance (except for 50% sparsity)

Conclusions

- We have studied the problem of testing independence in a large dimensional vector.
- The "classical" likelihood ratio test for independence does not keep its nominal level if p_1 is small compared to p_2.
- We have introduced alternative tests which yield a more reliable approximation.
- We determined asymptotic properties under the null hypothesis and alternative.
- For this purpose we investigated asymptotic properties of linear spectral statistics of central and non-central Fisher matrices. $\mathbf{W}\mathbf{T}^{-1}$, where $\mathbf{W}$ and $\mathbf{T}$ are independent Wishart matrices ($\mathbf{W}$ is conditionally Wishart).
- The theoretical results can be used for a better understanding of the finite sample properties of tests based on linear spectral statistics of the Fisher matrix.

References:

- Bai Z.D., and J. W. Silverstein (2010) Spectral Analysis of Large Dimensional Random Matrices, Springer: New York; Dordrecht; Heidelberg; London.
- Bai, Z., Jiang, D., Yao, J.-F., and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Annals of Statistics*, 37:3822-3840.
- Bodnar, T., Dette, H., and Parolya N. (2018). Testing for independence of large dimensional vectors. Appears in *Annals of Statistics*.
- Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics* 1, 457-483.
- Yao, J., Zheng, Z. and Z. Bai, (2015), Large Sample Covariance Matrices and High-Dimensional Data Analysis, Cambridge University press: New York.
- Zheng, Z. (2012), Central limit theorems for linear spectral statistics of large dimensional F-matrices. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 48(2):444-476.