Testing Many and Possibly Singular Polynomial Constraints

at the 2023 German Probability and Statistics Days

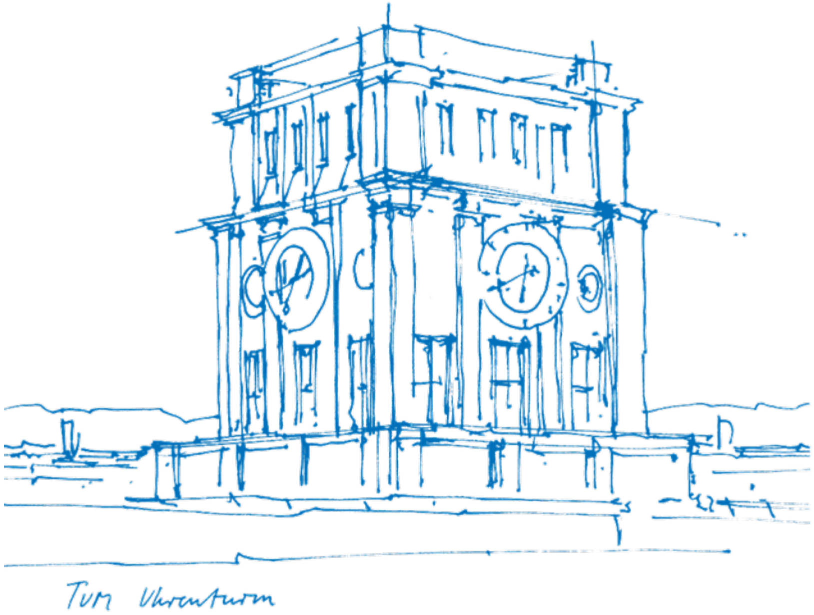
Nils Sturma

Research group Mathematical Statistics

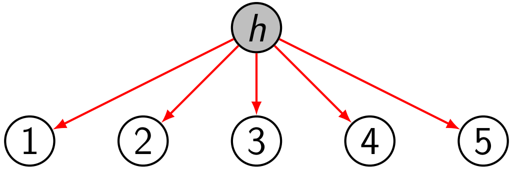
TUM School of Computation, Information and Technology

Technical University of Munich

(joint work with Mathias Drton and Dennis Leung)



Motivation: One-Factor Analysis Model



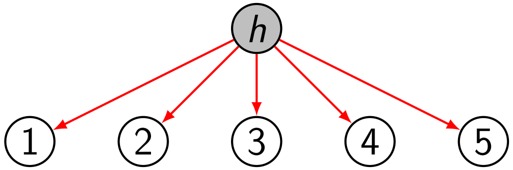
Model:

The family of multivariate normal distributions $N_k(0, \Sigma)$ whose covariance matrix lies in the set

$$\left\{\Omega + \Gamma\Gamma^\top : \Omega > 0 \text{ diagonal}, \Gamma \in \mathbb{R}^{k \times 1}\right\}.$$

Topic of the talk: Testing the goodness-of-fit based on samples $X_1, \ldots, X_n \sim N_k(0, \Sigma)$.

Algebraic Characterization



$$\Sigma = \begin{pmatrix} \omega_1 + \gamma_1^2 & \gamma_1\gamma_2 & \gamma_1\gamma_3 & \gamma_1\gamma_4 & \gamma_1\gamma_5 \\ \gamma_1\gamma_2 & \omega_2 + \gamma_2^2 & \gamma_2\gamma_3 & \gamma_2\gamma_4 & \gamma_2\gamma_5 \\ \gamma_1\gamma_3 & \gamma_2\gamma_3 & \omega_3 + \gamma_3^2 & \gamma_3\gamma_4 & \gamma_3\gamma_5 \\ \gamma_1\gamma_4 & \gamma_2\gamma_4 & \gamma_3\gamma_4 & \omega_4 + \gamma_4^2 & \gamma_4\gamma_5 \\ \gamma_1\gamma_5 & \gamma_2\gamma_5 & \gamma_3\gamma_5 & \gamma_4\gamma_5 & \omega_5 + \gamma_5^2 \end{pmatrix}$$

Observation:

Off-diagonal 2×2 minors (=tetrads) vanish:

$$\det(\Sigma_{\{12\},\{3,4\}}) = \sigma_{13}\sigma_{24} - \sigma_{23}\sigma_{14} = \gamma_1\gamma_3\gamma_2\gamma_4 - \gamma_2\gamma_3\gamma_1\gamma_4 = 0$$

If Σ is in the one-factor analysis model, then all tetrads vanish simultaneously. That is,

$$\sigma_{ij}\sigma_{kl} - \sigma_{ik}\sigma_{jl} = 0$$

for four distinct indices i, j, k, l.

General Setup: Testing Constraints on Statistical Models

Parametric family:

$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, where $\Theta \in \mathbb{R}^d$.

Model:

$\Theta_0 = \{\theta \in \Theta : \; f_j(\theta) \leq 0, \; 1 \leq j \leq p\}$. Main interest: Polynomial constraints f_j.

Based on samples $X_1, \ldots, X_n \sim P_\theta$ test

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta \setminus \Theta_0.$$

Challenges:

Many constraints, irregular points, inequalities, ...

Likelihood-Ratio Test

$$\lambda_n = -2 \log \left(\frac{\sup_{\theta \in \Theta_0} \mathcal{L}_n(\theta)}{\sup_{\theta \in \Theta} \mathcal{L}_n(\theta)} \right).$$

Limitations

✗ Likelihood function is not available or is difficult to maximize under Θ_0.

✗ Slow convergence if dimension of Θ is very large.
(In particular, larger than the sample size n.)

✗ Asymptotic distribution depends on the true parameter.
(Polynomials: Irregular points of Θ_0 are algebraic singularities.)

Likelihood-Ratio Test

$$\lambda_n = -2 \log \left(\frac{\sup_{\theta \in \Theta_0} \mathcal{L}_n(\theta)}{\sup_{\theta \in \Theta} \mathcal{L}_n(\theta)} \right).$$

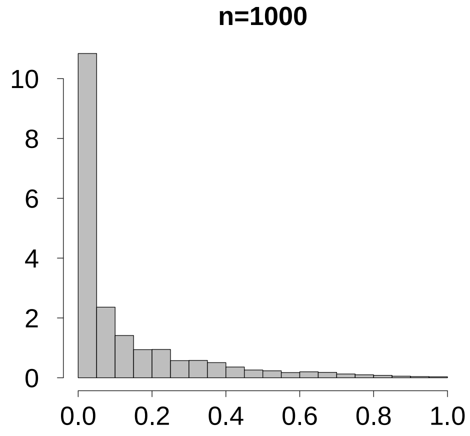
Limitations

✗ Likelihood function is not available or is difficult to maximize under Θ_0.

✗ Slow convergence if dimension of Θ is very large.
(In particular, larger than the sample size n.)

✗ Asymptotic distribution depends on the true parameter.
(Polynomials: Irregular points of Θ_0 are algebraic singularities.)

Invalidity at singularities



Simulated p-values for testing the one-factor analysis model with $k = 15$ observed variables close to a singular point.

“Plug-in” Test

$$M_n = \max_{1 \leq j \leq p} \frac{\sqrt{n}\, f_j(\hat{\theta}_n)}{\left(\widehat{\mathrm{var}}[f_j(\hat{\theta}_n)]\right)^{1/2}}, \qquad \text{where } \hat{\theta}_n \text{ is a “good” estimator of } \theta.$$

Tetrads: Gaussian approximation to derive critical values.

- ✔ High-dimensional approximation ($p \gg n$).
- ✔ Inequality constraints.
- ✔ Optimization free.
- ✗ Asymptotic distribution depends on the true parameter.

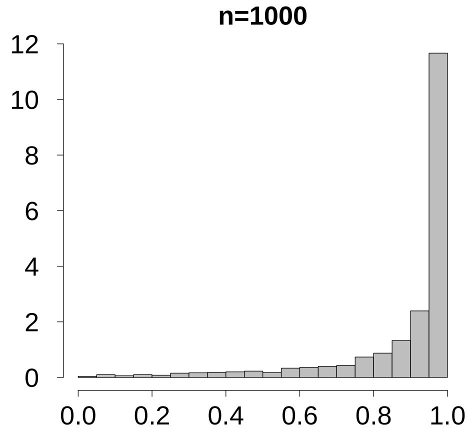
"Plug-in" Test

$$M_n = \max_{1 \le j \le p} \frac{\sqrt{n}\, f_j(\hat{\theta}_n)}{\left(\widehat{\mathrm{var}}[f_j(\hat{\theta}_n)]\right)^{1/2}}, \qquad \text{where } \hat{\theta}_n \text{ is a "good" estimator of } \theta.$$

Tetrads: Gaussian approximation to derive critical values.

- ✔ High-dimensional approximation ($p \gg n$).
- ✔ Inequality constraints.
- ✔ Optimization free.
- ✗ Asymptotic distribution depends on the true parameter.

Invalidity at singularities



Simulated p-values for testing tetrads with $k = 15$ observed variables close to a singular point.

Connection to U-statistics

Tetrad: $f_1(\Sigma) = \sigma_{13}\sigma_{24} - \sigma_{23}\sigma_{14}$.

Observation:

$\widehat{f}_1 = \frac{n}{n-1} f_1(\hat{\Sigma}_n) = \frac{1}{\binom{n}{2}} \sum_{i<j} h_1(X_i, X_j)$ is a U-statistic with kernel

$$h_1(X_i, X_j) = \frac{1}{2}\{(X_{i1}X_{i3}X_{j2}X_{j4} - X_{i2}X_{i3}X_{j1}X_{j4}) + (X_{j1}X_{j3}X_{i2}X_{i4} - X_{j2}X_{j3}X_{i1}X_{i4})\}.$$

Connection to U-statistics

Tetrad: $f_1(\Sigma) = \sigma_{13}\sigma_{24} - \sigma_{23}\sigma_{14}$.

Observation:

$\widehat{f}_1 = \frac{n}{n-1} f_1(\hat{\Sigma}_n) = \frac{1}{\binom{n}{2}} \sum_{i<j} h_1(X_i, X_j)$ is a U-statistic with kernel

$$h_1(X_i, X_j) = \frac{1}{2}\{(X_{i1}X_{i3}X_{j2}X_{j4} - X_{i2}X_{i3}X_{j1}X_{j4}) + (X_{j1}X_{j3}X_{i2}X_{i4} - X_{j2}X_{j3}X_{i1}X_{i4})\}.$$

Asymptotics (one dimensional):

Gaussian approximation: $\sqrt{n}(\widehat{f}_1 - f_1(\Sigma)) \longrightarrow N(0, m^2\sigma_{g_1}^2)$

where m is the degree of the kernel h_1 and $\sigma_{g_1}^2$ is the variance of the Hájek projection

$$g_1(X_i) = \mathbb{E}[h_1(X_i, X_j)|X_i] = \frac{1}{2}\{(X_{i1}X_{i3}\sigma_{24} - X_{i2}X_{i3}\sigma_{14}) + (\sigma_{13}X_{i2}X_{i4} - \sigma_{23}X_{i1}X_{i4})\}.$$

Irregular points: $\sigma_{g_1}^2 = 0 \implies$ U-statistic is degenerate $\implies$ *Gaussian approximations fails.*

Proposal: Incomplete U-statistics

Assumption: $f(\theta) = (f_1(\theta), \ldots, f_p(\theta))^\top$ is *estimable*, i.e., there exists a symmetric kernel $h(x_1, \ldots, x_m)$ s.t.

$$\mathbb{E}[h(X_1, \ldots, X_m)] = f(\theta) \quad \text{for all } \theta \in \Theta,$$

whenever $X_1, \ldots, X_m$ are i.i.d. with distribution P_θ.

Randomized incomplete U-statistics:

$$U'_{n,N} = \frac{1}{\hat{N}} \sum_{\iota=(i_1,\ldots,i_m)\in I_{n,m}} Z_\iota h(X_{i_1}, \ldots, X_{i_m})$$

- $I_{n,m} = \{(i_1, \ldots, i_m) : 1 \leq i_1 < \ldots < i_m \leq n\}$.
- Computational budget parameter $N \leq \binom{n}{m}$.
- $\{Z_\iota : \iota \in I_{n,m}\}$ are i.i.d. $\mathrm{Ber}(p_n)$ with $p_n = N/\binom{n}{m}$.
- $\hat{N} = \sum_{\iota \in I_{n,m}} Z_\iota$ is the number of successes.

Proposal: Incomplete U-statistics

Assumption: $f(\theta) = (f_1(\theta), \ldots, f_p(\theta))^\top$ is *estimable*, i.e., there exists a symmetric kernel $h(x_1, \ldots, x_m)$ s.t.

$$\mathbb{E}[h(X_1, \ldots, X_m)] = f(\theta) \quad \text{for all } \theta \in \Theta,$$

whenever $X_1, \ldots, X_m$ are i.i.d. with distribution P_θ.

Randomized incomplete U-statistics:

$$U'_{n,N} = \frac{1}{\hat{N}} \sum_{\iota=(i_1,\ldots,i_m)\in I_{n,m}} Z_\iota h(X_{i_1}, \ldots, X_{i_m})$$

- $I_{n,m} = \{(i_1, \ldots, i_m) : 1 \leq i_1 < \ldots < i_m \leq n\}$.
- Computational budget parameter $N \leq \binom{n}{m}$.
- $\{Z_\iota : \iota \in I_{n,m}\}$ are i.i.d. $\mathrm{Ber}(p_n)$ with $p_n = N/\binom{n}{m}$.
- $\hat{N} = \sum_{\iota \in I_{n,m}} Z_\iota$ is the number of successes.

Asymptotics: $\sqrt{n}(U'_{n,N,1} - f_1(\Sigma)) \longrightarrow N(0, m^2\sigma^2_{g_1} + \frac{n}{N}\sigma^2_{h_1})$.

Choose $N = \mathcal{O}(n)$ to guard against degeneracy!

Proposed Test

Test statistic

$$\mathcal{T} = \max_{1 \le j \le p} (\sqrt{n}\, \hat{\sigma}_j^{-1}) U'_{n,N,j}.$$

Critical value

1. Approximate test statistic by maximum of Gaussian random vector $Y \sim N_p(0, \Gamma)$, where $\Gamma = m^2 \Gamma_g + \frac{n}{N} \Gamma_h$.
2. Construct an estimate $\widehat{\Gamma}$ of the true asymptotic covariance matrix Γ by a Gaussian multiplier bootstrap method. Then $W \sim N_p(0, \widehat{\Gamma})$ is "close" to $Y \sim N_p(0, \Gamma)$.
3. Critical value: Quantile $c_{W_0}(1-\alpha)$ of $W_0 = \max_{1 \le j \le p} \hat{\sigma}_j^{-1} W_j$.

Our theoretical contribution

If $N = \mathcal{O}(n)$ then the proposed test based on an incomplete U-statistic is asymptotically valid (controls type I error) in high dimensions $p \gg n$ and under *mixed degeneracy*:

$$P(\mathcal{T} > c_{W_0}(1-\alpha)) \le \alpha.$$

Mixed Degeneracy

Background on high-dimensional Gaussian approximation

Chernozhukov, Chetverikov, Kato (2013). *Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors.* Ann. Statist., 41(6):2786–2819.

Mixed Degeneracy

Background on high-dimensional Gaussian approximation

Chernozhukov, Chetverikov, Kato (2013). *Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors.* Ann. Statist., 41(6):2786–2819.

Chen (2018). *Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications.* Ann. Statist., 46(2):642–678.

Assumption: Non-degenerate: There exists $c > 0$ such that $\sigma_{g_j}^2 \geq c$ for all $j = 1, \ldots, p$.

Mixed Degeneracy

Background on high-dimensional Gaussian approximation

Chernozhukov, Chetverikov, Kato (2013). *Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors.* Ann. Statist., 41(6):2786–2819.

Chen (2018). *Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications.* Ann. Statist., 46(2):642–678.

Assumption: Non-degenerate: There exists $c > 0$ such that $\sigma_{g_j}^2 \geq c$ for all $j = 1, \ldots, p$.

Chen, Kato (2019). *Randomized incomplete U-statistics in high dimensions.* Ann. Statist., 47(6):3127–3156.

Assumption: **Either** non-degenerate: There exists $c > 0$ such that $\sigma_{g_j}^2 \geq c$ for all $j = 1, \ldots, p$.

Or degenerate: $\sigma_{g_j}^2 = 0$ for all $j = 1, \ldots, p$.

Mixed Degeneracy

Background on high-dimensional Gaussian approximation

Chernozhukov, Chetverikov, Kato (2013). *Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors.* Ann. Statist., 41(6):2786–2819.

Chen (2018). *Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications.* Ann. Statist., 46(2):642–678.

Assumption: Non-degenerate: There exists $c > 0$ such that $\sigma_{g_j}^2 \geq c$ for all $j = 1, \dots, p$.

Chen, Kato (2019). *Randomized incomplete U-statistics in high dimensions.* Ann. Statist., 47(6):3127–3156.

Assumption: **Either** non-degenerate: There exists $c > 0$ such that $\sigma_{g_j}^2 \geq c$ for all $j = 1, \dots, p$.

Or degenerate: $\sigma_{g_j}^2 = 0$ for all $j = 1, \dots, p$.

Mixed degeneracy assumption

Let $p_1, p_2 \in \mathbb{N}$ such that $p_1 + p_2 = p$ and assume:

(A) There exists $c > 0$ such that $\sigma_{g_j}^2 \geq c$ for all $j = 1, \dots, p_1$.

(B) There exists $k > 0$ and $\beta > 0$ such that $\|g_j(X_1) - f_j(\theta)\|_{\psi_\beta} \leq Cn^{-k}$ for all $j = p_1 + 1, \dots, p$.

$\implies \sigma_{g_j}^2 \leq \tilde{C} n^{-2k}$

High-dimensional Bootstrap Approximation

Theorem

Under mixed degeneracy (and additional moment conditions on h), we have the ***Gaussian approximation***

$$\sup_{R \in \mathbb{R}^p_{\mathrm{re}}} |P(\sqrt{n}(U'_{n,N} - f(\theta)) \in R) - P(Y \in R)| \leq C\{\omega_{n,1} + \omega_{n,2} + \omega_{n,3}\},$$

where $Y \sim N_p(0, m^2\Gamma_g + \frac{n}{N}\Gamma_h)$ *and*

$$\omega_{n,1} = \left(\frac{m^{2/\beta}\log(pn)^{1+6/\beta}}{n \wedge N}\right)^{1/6}, \qquad \omega_{n,2} = \frac{N^{1/2}m^2\log(pn)^{1/2+2/\beta}}{n^{\min\{1/2+k,\, 5/6,\, m/3\}}}, \qquad \omega_{n,3} = \left(\frac{Nm^2\log(p)^2}{n^{1+k}}\right)^{1/3}.$$

Note: If $N = \mathcal{O}(n)$ and $m \geq 3$, $k \geq 1/3$ are fixed constants, then the bound vanishes asymptotically if $\log(pn)^{3/2+6/\beta} = o(n)$.

High-dimensional Bootstrap Approximation

Theorem

Under mixed degeneracy (and additional moment conditions on h), we have the ***Gaussian approximation***

$$\sup_{R \in \mathbb{R}^p_{\mathrm{re}}} |P(\sqrt{n}(U'_{n,N} - f(\theta)) \in R) - P(Y \in R)| \le C\{\omega_{n,1} + \omega_{n,2} + \omega_{n,3}\},$$

where $Y \sim N_p(0, m^2\Gamma_g + \frac{n}{N}\Gamma_h)$ *and*

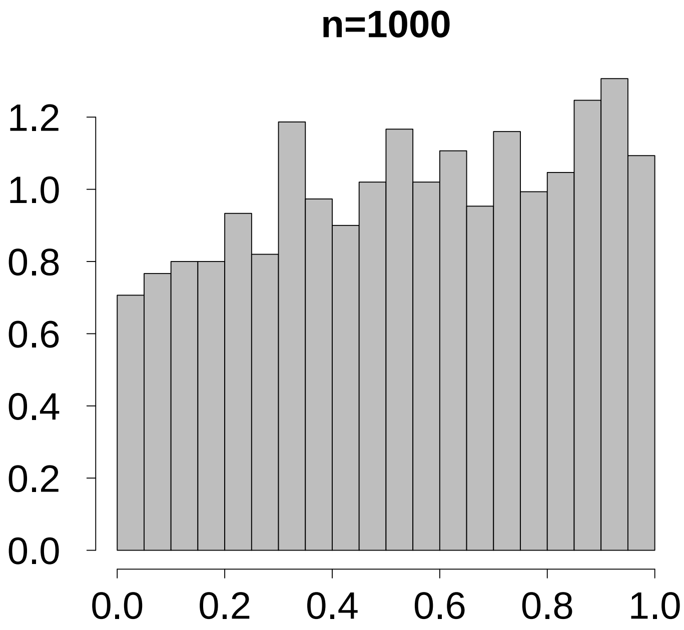
$$\omega_{n,1} = \left(\frac{m^{2/\beta} \log(pn)^{1+6/\beta}}{n \wedge N}\right)^{1/6}, \qquad \omega_{n,2} = \frac{N^{1/2} m^2 \log(pn)^{1/2+2/\beta}}{n^{\min\{1/2+k,\, 5/6,\, m/3\}}}, \qquad \omega_{n,3} = \left(\frac{Nm^2 \log(p)^2}{n^{1+k}}\right)^{1/3}.$$

Note: If $N = \mathcal{O}(n)$ and $m \ge 3$, $k \ge 1/3$ are fixed constants, then the bound vanishes asymptotically if $\log(pn)^{3/2+6/\beta} = o(n)$.

This is the basis for the **bootstrap approximation:**

1. Further approximate Y by a Gaussian multiplier bootstrap $W \longrightarrow$ Similar bound under $N = \mathcal{O}(n)$.
2. Control studentization.
3. Establish asymptotic validity (control of type I error).

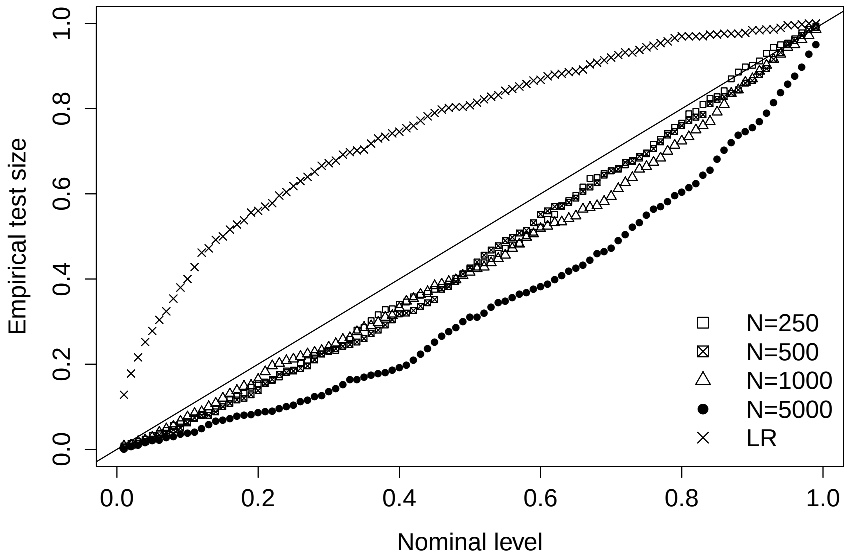
Our Test at Irregular Points



Simulated p-values for testing tetrads with $k = 15$ observed variables close to a singular point.
Computational budget parameter $N = 2n$.

Size vs. Power

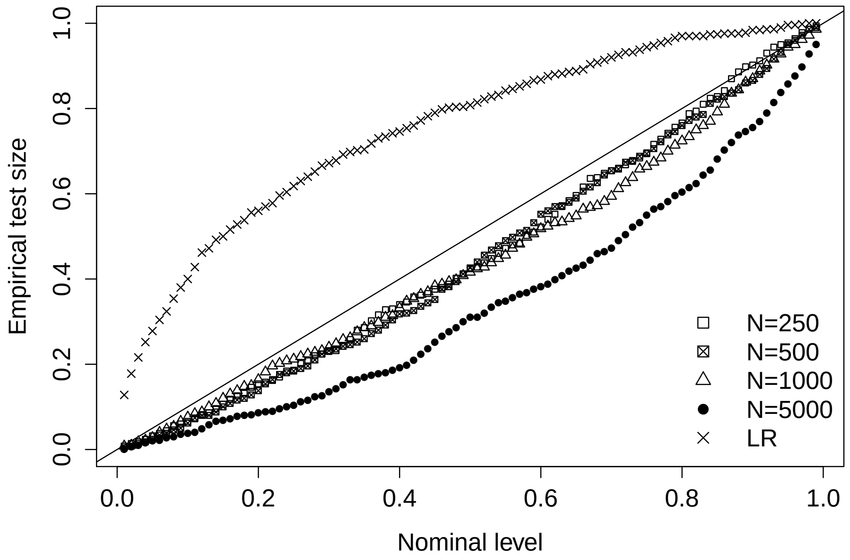
$n = 500$



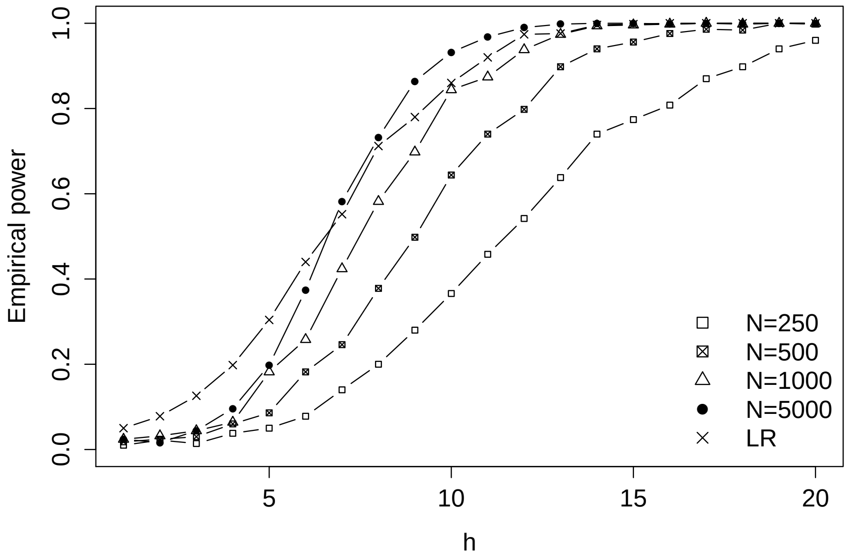
Empirical sizes vs. nominal levels for testing tetrads with $k = 15$ observed variables. True parameter is close to a **singular point**.

Size vs. Power

$n = 500$



Empirical sizes vs. nominal levels for testing tetrads with $k = 15$ observed variables. True parameter is close to a **singular point**.



Empirical power for different local alternatives for testing tetrads with $k = 15$ observed variables ($\alpha = 0.05$). True parameter is a **regular point**.

Trade-off between efficiency and guarding against singularities.

Conclusion

- ✔ General strategy for simultaneous testing of many constraints ($p \gg n$).
- ✔ Equality and inequality constraints.
- ✔ Optimization free.

  Although computationally demanding for large p and large computational budget N.
- ✔ Accommodate irregular settings where the incomplete U-statistics is mixed degenerate by choosing $N = \mathcal{O}(n)$.

Our paper and background reading:

Sturma, Drton, Leung (2022).
Testing Many and Possibly Singular Polynomial Constraints. arXiv:2208.11756.

Leung, Drton (2018).
Algebraic tests of general Gaussian latent tree models. NeurIPS 2018.

Drton (2009).
Likelihood ratio tests and singularities. Ann. Statist., 37(2):979-1012

European Research Council
Established by the European Commission

Appendix: Kernels for Polynomial Hypotheses

Polynomial of total degree s:

$$f_j(\theta) = a_0 + \sum_{r=1}^{s} \sum_{\substack{(i_1,\ldots,i_r) \\ i_l \in \{1,\ldots,d\}}} a_{(i_1,\ldots,i_r)} \theta_{i_1} \cdots \theta_{i_r}$$

Appendix: Kernels for Polynomial Hypotheses

Polynomial of total degree s:

$$f_j(\theta) = a_0 + \sum_{r=1}^{s} \sum_{\substack{(i_1,\dots,i_r) \\ i_l \in \{1,\dots,d\}}} a_{(i_1,\dots,i_r)} \theta_{i_1} \cdots \theta_{i_r}$$

Construct kernel h_j:

1) For a fixed integer $\eta \geq 1$, find unbiased estimators $\widehat{\theta}_i(X_1^\eta)$ of θ_i for all $i = 1, \dots, d$.
2) For the degree $m = \eta s$, define the unbiased estimator

$$\check{h}_j(X_1^m) = a_0 + \sum_{r=1}^{s} \sum_{\substack{(i_1,\dots,i_r) \\ i_l \in \{1,\dots,d\}}} a_{(i_1,\dots,i_r)} \widehat{\theta}_{i_1}(X_1^\eta) \widehat{\theta}_{i_2}(X_{\eta+1}^{2\eta}) \cdots \widehat{\theta}_{i_r}(X_{(r-1)\eta+1}^{r\eta}).$$

3) Symmetrizing: Average over all permutations of $\{1, \dots, m\}$: $h_j(X_1^m) = \frac{1}{m!} \sum_{\pi \in S_m} \check{h}_j(X_{\pi(1)}, \dots, X_{\pi(m)})$.

Polynomials are estimable.