Group Inference in High Dimensions with Applications to Hierarchical Testing

Zijian Guo

Rutgers University

The International Statistical Conference
In Memory of Professor Sik-Yum Lee

In Memory of Professor Sik-Yum Lee

- Structural equation model

Mathematics Genealogy Project

Sik-Yum Lee

MathSciNet

Ph.D. University of California, Los Angeles 1977

Dissertation:

Advisor: Robert Irving Jennrich

Students:
Click here to see the students listed in chronological order.

Name	School	Year	Descendants
Shi, Jian Qing	Chinese University of Hong Kong	1996	
Song, Xin-Yuan	Chinese University of Hong Kong	2001	1
Zhang, Wenyang	Chinese University of Hong Kong	1999	6
Zhu, Hongtu	Chinese University of Hong Kong	2000	23

According to our current on-line database, Sik-Yum Lee has 4 students and 34 descendants.
We welcome any additional information.

Group Inference in High Dimensions with Applications to Hierarchical Testing

Zijian Guo

Rutgers University

The International Statistical Conference
In Memory of Professor Sik-Yum Lee

Collaborators

Claude Renaux

Peter Bühlmann

Tony Cai

Reference

Guo, Z., Renaux, C., Bühlmann., P, Cai, T. T.(2019) Group Inference in High Dimensions with Applications to Hierarchical Testing. arXiv preprint arXiv:1909.01503

Group Inference

High-dimensional linear regression

$$y_i = X_{i\cdot}^{\intercal}\beta + \epsilon_i, \quad \text{for } 1 \leq i \leq n.$$

where $X_{i\cdot}, \beta \in \mathbb{R}^p$.

- high dimension: $p \gg n$
- sparse model: $\|\beta\|_0 \ll n$

For a given set $G \subset \{1, 2, \cdots, p\}$, group significance test is

$$H_0 : \beta_G = 0, \tag{1}$$

where $\beta_G = \{\beta_j;\ j \in G\}$.

Group Inference v.s. Quadratic Functional

The null $H_0 : \beta_G = 0$ can be written as

$$H_{0,A}: \; \beta_G^{\intercal} A \beta_G = 0,$$

for some positive definite matrix $A \in \mathbb{R}^{|G|\times|G|}$.

Two special cases

$$H_{0,\Sigma}: \; \beta_G^{\intercal} \Sigma_{G,G} \beta_G = 0.$$

with Σ denoting the covariance matrix of $X_{i\cdot}$.

$$H_{0,\mathrm{I}}: \; \beta_G^{\intercal} \beta_G = 0.$$

Group v.s. Individual Significance

For a group of **highly correlated** variables,

1. It is ambitious to detect significant single variable β_i
   - Inaccurate estimator of β_i
2. Significance and high correlation
   - Significant variables can treated as non-significant
3. The group significance

Hierarchical Testing (Meinshausen, 2008)
Divide variables into sub-groups + group significance

- Variables inside a group tend to be highly correlated.
- Between groups, not highly correlated.

Other Motivation I: Interaction Test

Model with interaction (Tian, Alizadeh, Gentles and Tibshirani, 2014)

$$y_i = X_{i\cdot}^{\intercal}\beta + D_i\left(\gamma_0 + X_{i\cdot}^{\intercal}\gamma\right) + \epsilon_i.$$

$$H_0 : \gamma = 0$$

- Interaction test
- Detection of Effect Heterogeneity (D_i is treatment)

Other Motivation I: Interaction Test

Model with interaction (Tian, Alizadeh, Gentles and Tibshirani, 2014)

$$y_i = X_{i\cdot}^{\intercal}\beta + D_i\left(\gamma_0 + X_{i\cdot}^{\intercal}\gamma\right) + \epsilon_i.$$

$$H_0 : \gamma = 0$$

- Interaction test
- Detection of Effect Heterogeneity (D_i is treatment)

Equivalent model,

$$y_i = W_{i\cdot}^{\intercal}\eta + \epsilon_i.$$

with $W_i = (D_i X_{i\cdot}^{\intercal}, 1, X_{i\cdot}^{\intercal})^{\intercal}$ and $\eta = (\gamma^{\intercal}, \gamma_0, \beta^{\intercal})^{\intercal}$.

$$H_0 : \eta_G = 0 \quad \text{with} \quad G = \{1, 2, \cdots, p\}.$$

Other Motivation II: Local Heritability in Genetics

Local heritability: the proportion of variance explained by a subset of genotypes indexed by the group G. (Shi et. al., 2016)

1. G: the set of SNPs located in on the same chromosome.
2. Then the local heritability is

$$\beta_G^{\intercal}\Sigma_{G,G}\beta_G = \mathbb{E}|X_{i,G}^{\intercal}\beta_G|^2.$$

Inference for $Q_\Sigma = \beta_G^\intercal \Sigma_{G,G} \beta_G$

Bias Correction

Initial estimators

- $\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{2n}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$, with $\lambda \asymp \sqrt{\log p/n}\sigma$
- $\widehat{\Sigma} = \frac{1}{n}X^{\intercal}X$.

Decompose $\widehat{\beta}_G^{\intercal}\widehat{\Sigma}_{G,G}\widehat{\beta}_G - \beta_G^{\intercal}\Sigma_{G,G}\beta_G$ as

$$-2\widehat{\beta}_G^{\intercal}\widehat{\Sigma}_{G,G}(\beta_G - \widehat{\beta}_G) + \beta_G^{\intercal}(\widehat{\Sigma}_{G,G} - \Sigma_{G,G})\beta_G - (\widehat{\beta}_G - \beta_G)^{\intercal}\widehat{\Sigma}_{G,G}(\widehat{\beta}_G - \beta_G)$$

Estimate $\widehat{\beta}_G^{\intercal}\widehat{\Sigma}_{G,G}(\beta_G - \widehat{\beta}_G)$ and correct $\widehat{\beta}_G^{\intercal}\widehat{\Sigma}_{G,G}\widehat{\beta}_G$.

Construction of Projection Direction

For any $u \in \mathbb{R}^p$,

$$u^\intercal \frac{1}{n} X^\intercal (y - X\widehat{\beta}) - \widehat{\beta}_G^\intercal \widehat{\Sigma}_{G,G}(\beta_G - \widehat{\beta}_G)$$
$$= \frac{1}{n} u^\intercal X^\intercal \epsilon + \left[\widehat{\Sigma} u - \begin{pmatrix} \widehat{\beta}_G^\intercal \widehat{\Sigma}_{G,G} & \mathbf{0} \end{pmatrix}^\intercal\right]^\intercal (\beta - \widehat{\beta}).$$

- $\|\beta - \widehat{\beta}\|_1$ is small

$$\left| \left[\widehat{\Sigma} u - \begin{pmatrix} \widehat{\beta}_G^\intercal \widehat{\Sigma}_{G,G} & \mathbf{0} \end{pmatrix}^\intercal\right]^\intercal (\beta - \widehat{\beta}) \right| \leq \|\beta - \widehat{\beta}\|_1 \left\| \widehat{\Sigma} u - \begin{pmatrix} \widehat{\beta}_G^\intercal \widehat{\Sigma}_{G,G} & \mathbf{0} \end{pmatrix}^\intercal \right\|_\infty .$$

- Minimize/Constrained $u^\intercal \widehat{\Sigma} u$ and

$$\left\| \widehat{\Sigma} u - \begin{pmatrix} \widehat{\beta}_G^\intercal \widehat{\Sigma}_{G,G} & \mathbf{0} \end{pmatrix}^\intercal \right\|_\infty = \max_{1 \leq j \leq p} \left| \left\langle e_j, \widehat{\Sigma} u - \begin{pmatrix} \widehat{\beta}_G^\intercal \widehat{\Sigma}_{G,G} & \mathbf{0} \end{pmatrix}^\intercal \right\rangle \right|$$

Construction of Projection Direction

Initial proposal

$$\widehat{u} = \arg\min \; u^{\intercal}\widehat{\Sigma}u$$
$$\text{s.t.} \quad \max_{w\in\mathcal{C}_0}\left|\left\langle w, \widehat{\Sigma}u - \begin{pmatrix}\widehat{\beta}_G^{\intercal}\widehat{\Sigma}_{G,G} & \mathbf{0}\end{pmatrix}^{\intercal}\right\rangle\right| \leq \|\widehat{\Sigma}_{G,G}\widehat{\beta}_G\|_2\lambda_n$$

where $\lambda_n = C\sqrt{\log p/n}$ and

$$\mathcal{C}_0 = \{e_1, \cdots, e_p\}.$$

- Constrain bias and minimize variance: Zhang & Zhang '14; Javanmard & Montanari '14;
- Only work for small $|G|$.

Construction of Projection Direction

$$\widehat{u} = \arg\min \; u^{\intercal}\widehat{\Sigma}u$$

$$\text{s.t.} \; \max_{w \in \mathcal{C}} \left| \left\langle w, \widehat{\Sigma}u - \begin{pmatrix} \widehat{\beta}_G^{\intercal}\widehat{\Sigma}_{G,G} & \mathbf{0} \end{pmatrix}^{\intercal} \right\rangle \right| \leq \|\widehat{\Sigma}_{G,G}\widehat{\beta}_G\|_2 \lambda_n$$

$$\mathcal{C} = \left\{ e_1, \cdots, e_p, \frac{1}{\|\widehat{\Sigma}_{G,G}\widehat{\beta}_G\|_2} \begin{pmatrix} \widehat{\beta}_G^{\intercal}\widehat{\Sigma}_{G,G} & \mathbf{0} \end{pmatrix}^{\intercal} \right\}.$$

- **Work for any $|G|$.**
- Constrain bias, minimize variance and Constrain Variance

$$\widehat{Q}_{\Sigma} = \widehat{\beta}_G^{\intercal}\widehat{\Sigma}_{G,G}\widehat{\beta}_G + \frac{2}{n}\widehat{u}^{\intercal}X^{\intercal}(y - X\widehat{\beta}). \quad (2)$$

Inference Procedure

We estimate the variance of the proposed estimator $\widehat{\mathrm{Q}}_{\Sigma}$ by

$$\widehat{\mathrm{V}}_{\Sigma}(\tau)=\frac{4\widehat{\sigma}^2}{n}\widehat{u}^{\intercal}\widehat{\Sigma}\widehat{u}+\frac{1}{n^2}\sum_{i=1}^{n}\left(\widehat{\beta}_G^{\intercal}X_{iG}X_{iG}^{\intercal}\widehat{\beta}_G-\widehat{\beta}_G^{\intercal}\widehat{\Sigma}_{G,G}\widehat{\beta}_G\right)^2+\frac{\tau}{n},$$

for some positive constant $\tau > 0$.

$$\phi_{\Sigma}(\tau)=\mathbf{1}\left(\widehat{\mathrm{Q}}_{\Sigma}\geq z_{1-\alpha}\sqrt{\widehat{\mathrm{V}}_{\Sigma}(\tau)}\right)$$

$$\mathrm{CI}_{\Sigma}(\tau)=\left(\widehat{\mathrm{Q}}_{\Sigma}-z_{1-\frac{\alpha}{2}}\sqrt{\widehat{\mathrm{V}}_{\Sigma}(\tau)},\ \widehat{\mathrm{Q}}_{\Sigma}+z_{1-\frac{\alpha}{2}}\sqrt{\widehat{\mathrm{V}}_{\Sigma}(\tau)}\right)$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal

Theoretical Justification

Theorem 1.

Under regularity conditions and $\|\beta\|_0 \ll \sqrt{n}/\log p$, *then the proposed estimator* $\widehat{\mathrm{Q}}_\Sigma$ *satisfies*

$$\limsup_{n,p\to\infty} \mathbf{P}\left(\left|\widehat{\mathrm{Q}}_\Sigma - \mathrm{Q}_\Sigma\right| \geq z_{1-\frac{\alpha}{2}}\sqrt{\mathrm{V}_\Sigma}\right) \leq \alpha \quad \textit{with } \mathrm{V}_\Sigma = \mathrm{V}_\Sigma^0 + \frac{\tau}{n}$$

$$\mathrm{V}_\Sigma^0 = \frac{4\sigma^2}{n}\widehat{u}^\intercal \widehat{\Sigma}\widehat{u} + \frac{1}{n^2}\sum_{i=1}^{n}\left(\widehat{\beta}_G^\intercal X_{iG}X_{iG}^\intercal \widehat{\beta}_G - \widehat{\beta}_G^\intercal \widehat{\Sigma}_{G,G}\widehat{\beta}_G\right)^2$$

Theoretical Justification

Theorem 1.

Under regularity conditions and $\|\beta\|_0 \ll \sqrt{n}/\log p$, *then the proposed estimator* $\widehat{\mathrm{Q}}_\Sigma$ *satisfies*

$$\limsup_{n,p\to\infty} \mathbf{P}\left(\left|\widehat{\mathrm{Q}}_\Sigma - \mathrm{Q}_\Sigma\right| \geq z_{1-\frac{\alpha}{2}}\sqrt{\mathrm{V}_\Sigma}\right) \leq \alpha \quad \textit{with } \mathrm{V}_\Sigma = \mathrm{V}^0_\Sigma + \frac{\tau}{n}$$

$$\mathrm{V}^0_\Sigma = \frac{4\sigma^2}{n}\widehat{u}^\intercal\widehat{\Sigma}\widehat{u} + \frac{1}{n^2}\sum_{i=1}^{n}\left(\widehat{\beta}_G^\intercal X_{iG}X_{iG}^\intercal\widehat{\beta}_G - \widehat{\beta}_G^\intercal\widehat{\Sigma}_{G,G}\widehat{\beta}_G\right)^2$$

- **No condition on G!**

Theoretical Justification

Theorem 1.

Under regularity conditions and $\|\beta\|_0 \ll \sqrt{n}/\log p$, *then the proposed estimator* $\widehat{\mathrm{Q}}_\Sigma$ *satisfies*

$$\limsup_{n,p\to\infty} \mathbf{P}\left(\left|\widehat{\mathrm{Q}}_\Sigma - \mathrm{Q}_\Sigma\right| \geq z_{1-\frac{\alpha}{2}}\sqrt{\mathrm{V}_\Sigma}\right) \leq \alpha \quad \textit{with } \mathrm{V}_\Sigma = \mathrm{V}^0_\Sigma + \frac{\tau}{n}$$

$$\mathrm{V}^0_\Sigma = \frac{4\sigma^2}{n}\widehat{u}^\intercal \widehat{\Sigma}\widehat{u} + \frac{1}{n^2}\sum_{i=1}^{n}\left(\widehat{\beta}_G^\intercal X_{iG} X_{iG}^\intercal \widehat{\beta}_G - \widehat{\beta}_G^\intercal \widehat{\Sigma}_{G,G}\widehat{\beta}_G\right)^2$$

- **No condition on G!**
- Super-efficiency
  - for β_G close to 0, $\sqrt{\mathrm{V}^0_\Sigma} \ll 1/\sqrt{n}$.
  - Enlarge variance by adding τ/n.

Parameter Spaces

$$\Theta(k) = \left\{ (\beta, \Sigma, \sigma) : \|\beta\|_0 \leq k, \ \frac{1}{M_1} \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq M_1, \ \sigma_1 \leq M_2 \right\},$$

For a fixed group G, define

$$\mathcal{H}_0 = \{(\beta, \Sigma, \sigma) \in \Theta(k) : \|\beta_G\|_2 = 0\}$$

Under the same assumption as Theorem 1,

$$\sup_{\theta \in \mathcal{H}_0} \liminf_{n,p \to \infty} \mathbf{P}_\theta\left(\phi_\Sigma(\tau) = 1\right) \leq \alpha$$

Size and Power

Corollary 2.

For $\theta \in \mathcal{H}_{1,A}(\delta(t)) = \left\{(\beta, \Sigma, \sigma) \in \Theta(k) : \beta_G^{\intercal} A \beta_G = \delta(t)\right\}$,

$$\liminf_{n,p\to\infty} \mathbf{P}_{\theta}\left(\phi_{\Sigma}(\tau) = 1\right) \geq 1 - \Phi(-t)$$

- $\delta(t) = (1.01 z_{1-\alpha} + t)\sqrt{\mathrm{V}_{\Sigma}} \asymp \frac{1+t}{\sqrt{n}}(\sqrt{\tau} + \|\beta_G\|_2 + \|\beta_G\|_2^2)$
- ϕ_{Σ} is of asymptotic power 1 as long as $t \to \infty$.

Size and Power

Corollary 2.

For $\theta \in \mathcal{H}_{1,A}(\delta(t)) = \left\{(\beta, \Sigma, \sigma) \in \Theta(k) : \beta_G^{\intercal} A \beta_G = \delta(t)\right\}$,

$$\liminf_{n,p\to\infty} \mathbf{P}_\theta\left(\phi_\Sigma(\tau) = 1\right) \geq 1 - \Phi(-t)$$

- $\delta(t) = (1.01 z_{1-\alpha} + t)\sqrt{\mathrm{V}_\Sigma} \asymp \frac{1+t}{\sqrt{n}}(\sqrt{\tau} + \|\beta_G\|_2 + \|\beta_G\|_2^2)$
- ϕ_Σ is of asymptotic power 1 as long as $t \to \infty$.
- χ^2-test will be of size $\sqrt{|G|/n}$ (Mitra, Zhang, 2016; van de Geer, Stucky, 2016)
- Large $|G|$: $\frac{1}{\sqrt{n}}(\sqrt{\tau} + \|\beta_G\|_2 + \|\beta_G\|_2^2) \ll \sqrt{|G|/n}$

Simulation I: Dense Alternatives

Simulation Setting

$$y_i = X_{i\cdot}^{\intercal}\beta + \epsilon_i, \quad \text{for } 1 \leq i \leq n.$$

- $p = 500$
- $\beta_j = \delta$ for $25 \leq j \leq 50$ and $\beta_j = 0$ otherwise;
- $\Sigma_{ij} = 0.6^{|i-j|}$ for $1 \leq i, j \leq 500$.
- Vary δ over $\{0, 0.04\}$ and n over $\{250, 350, 500, 800\}$.

$$H_{0,G} : \beta_i = 0 \text{ for } i \in G, \text{ where } G = \{30, 31, \cdots, 200\}.$$

Implemented Methods

Maximum test based on the debiased estimator

1. **Fast Debiased (FD)**: $\{\widehat{\beta}_j^{\rm FD}\}_{1\leq j\leq p}$ (Javanmard & Montanari '14)
2. **hdi**: $\{\widehat{\beta}_j^{\rm hdi}\}_{1\leq j\leq p}$ (van de Geer, Bühlmann, Ritov & Dezeure '14)

$$\phi_{\rm FD} = \mathbf{1}\left(\max_{j\in G} |\widehat{\beta}_j^{\rm FD}| \geq q_{\alpha}^{\rm FD}\right) \quad \text{and} \quad \phi_{\rm hdi} = \mathbf{1}\left(\max_{j\in G} |\widehat{\beta}_j^{\rm hdi}| \geq q_{\alpha}^{\rm hdi}\right).$$

where $q_{\alpha}^{\rm FD}$ and $q_{\alpha}^{\rm hdi}$ are computed by bootstrap or sampling.

Maximum test based on the debiased estimator

1. **Fast Debiased (FD)**: $\{\widehat{\beta}_j^{\rm FD}\}_{1\le j\le p}$ (Javanmard & Montanari '14)
2. **hdi**: $\{\widehat{\beta}_j^{\rm hdi}\}_{1\le j\le p}$ (van de Geer, Bühlmann, Ritov & Dezeure '14)

$$\phi_{\rm FD} = \mathbf{1}\left(\max_{j\in G}|\widehat{\beta}_j^{\rm FD}| \ge q_{\alpha}^{\rm FD}\right) \quad \text{and} \quad \phi_{\rm hdi} = \mathbf{1}\left(\max_{j\in G}|\widehat{\beta}_j^{\rm hdi}| \ge q_{\alpha}^{\rm hdi}\right).$$

where $q_{\alpha}^{\rm FD}$ and $q_{\alpha}^{\rm hdi}$ are computed by bootstrap or sampling.

$$\phi_{\Sigma}(\tau) = \mathbf{1}\left(\widehat{\rm Q}_{\Sigma} \ge z_{1-\alpha}\sqrt{\widehat{\rm V}_{\Sigma}(\tau)}\right)$$

Compare $\phi_{\rm I}(0), \phi_{\rm I}(1), \phi_{\Sigma}(0), \phi_{\Sigma}(1)$ and $\phi_{\rm FD}, \phi_{\rm hdi}$

Computational Comparison

One computational unit: a *p*-dimensional LASSO regression.

1. $\phi_{\rm FD}, \phi_{\rm hdi}$: $|G| + 1$ computational units
2. $\phi_{\Sigma}(\tau)$: 2 computational units
3. **Computational efficiency!**

Dense alternatives

Empirical Rejection Rate (ERR) out of 1000

- $\delta = 0$ (null): control ERR below 0.05.
- $\delta \neq 0$: obtain ERR close to 1.

δ	n	$\phi_{\mathrm{I}}(0)$	$\phi_{\mathrm{I}}(1)$	$\phi_{\Sigma}(0)$	$\phi_{\Sigma}(1)$	ϕ_{FD}	ϕ_{hdi}
0	250	0.962	0.002	0.994	0.008	0.112	0.044
	350	0.992	0.000	0.998	0.004	0.086	0.042
	500	0.996	0.002	1.000	0.002	0.078	0.048
	800	0.980	0.000	1.000	0.000	0.064	0.038
0.04	250	0.986	0.230	1.000	0.618	0.226	0.084
	350	1.000	0.188	1.000	0.854	0.184	0.106
	500	1.000	0.292	1.000	0.946	0.128	0.112
	800	1.000	0.374	1.000	1.000	0.128	0.180

$\phi_{\Sigma}(1)$ controls Type I error and **more powerful than** $\phi_{\mathrm{FD}}, \phi_{\mathrm{hdi}}$.

Simulation II: High Correlation

- $p = 500$
- $\beta_1 = \beta_3 = \delta$ and $\beta_j = 0$ for $j \neq 1, 3$
- High correlation among the fist five variables

$$\Sigma_{ij} = \begin{cases} 0.8 & \text{if } 1 \leq i \neq j \leq 5 \\ 1 & \text{if } 1 \leq i = j \leq 5 \\ 0.6^{|i-j|} & \text{otherwise.} \end{cases}$$

- Vary δ over $\{0, 0.2\}$ and n over $\{250, 350, 500\}$.

  $H_{0,G} : \beta_i = 0$ for $i \in G$, where $G = \{1, 2, \cdots, 5\}$.

High Correlation: ERR

δ	n	$\phi_{\rm I}(0)$	$\phi_{\rm I}(1)$	$\phi_{\Sigma}(0)$	$\phi_{\Sigma}(1)$	$\phi_{\rm FD}$	$\phi_{\rm hdi}$
	250	0.034	0.000	0.072	0.000	0.070	0.036
0	350	0.058	0.000	0.104	0.000	0.082	0.062
	500	0.052	0.000	0.092	0.000	0.082	0.056
	250	0.682	0.134	0.982	0.590	0.998	0.936
0.2	350	0.704	0.138	1.000	0.822	1.000	0.972
	500	0.634	0.234	1.000	0.960	1.000	0.994

$\phi_{\Sigma}(1)$ controls Type I error and less powerful than $\phi_{\rm FD}, \phi_{\rm hdi}$.

- $\phi_{\Sigma}(1)$ has valid confidence property;
- **No confidence property for** $\phi_{\rm FD}, \phi_{\rm hdi}$.

High Correlation: Coverage Property

		$\mathrm{CI}_{\mathrm{FD}}$					$\mathrm{CI}_{\mathrm{hdi}}$				
δ	n	β_1	β_2	β_3	β_4	β_5	β_1	β_2	β_3	β_4	β_5
	250	0.972	0.968	0.970	0.976	0.976	0.952	0.950	0.944	0.950	0.946
0	350	0.968	0.972	0.962	0.970	0.968	0.942	0.942	0.932	0.966	0.948
	500	0.974	0.972	0.964	0.970	0.982	0.950	0.936	0.956	0.950	0.956
	250	0.400	0.714	0.418	0.720	0.758	0.864	0.798	0.910	0.828	0.268
0.2	350	0.464	0.696	0.414	0.722	0.680	0.910	0.822	0.922	0.844	0.234
	500	0.424	0.702	0.408	0.686	0.674	0.876	0.860	0.916	0.842	0.298

Table: Empirical Coverage for $\{\beta_j\}_{1\leq j\leq 5}$ in the Highly Correlated scenario

δ	n	$\mathrm{CI}_{\mathrm{I}}(\tau=0)$	$\mathrm{CI}_{\mathrm{I}}(\tau=1)$	$\mathrm{CI}_{\Sigma}(\tau=0)$	$\mathrm{CI}_{\Sigma}(\tau=1)$
	250	0.104	1.000	0.090	1.000
0	350	0.110	1.000	0.088	1.000
	500	0.094	1.000	0.070	1.000
	250	0.912	0.992	0.822	0.998
0.2	350	0.916	0.998	0.822	0.996
	500	0.924	0.994	0.842	0.996

Table: Empirical Coverage for $\|\beta_G\|_2^2$, $\beta_G^{\intercal}\Sigma_{G,G}\beta_G$ in the Highly Correlated scenario

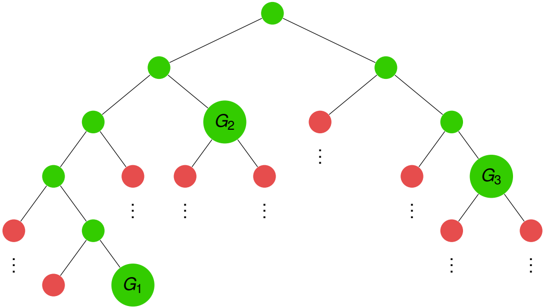
Four Key Messages

1. No requirement on G.
2. Computationally efficient.
3. Powerful test under dense alternatives.
4. Have coverage property for the highly correlated setting.

Hierarchical Test

Hierarchical Testing

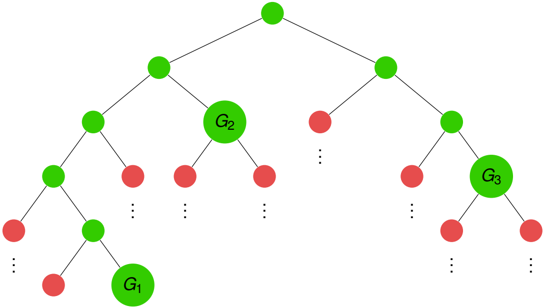
INPUT: Hierarchical tree $\mathcal{T}$ with nodes corresponding to groups of variables; Group testing returning p-values P_G for each group G.



- The nodes at each level build a partition of $\{1, \cdots, p\}$
- The upper part of the tree $\rightarrow$ Large groups.

Hierarchical Testing

INPUT: Hierarchical tree $\mathcal{T}$ with nodes corresponding to groups of variables; Group testing returning p-values P_G for each group G.



- The nodes at each level build a partition of $\{1, \cdots, p\}$
- The upper part of the tree $\rightarrow$ Large groups.
- Testing group significance in a top-down manner.

Hierarchical Testing

repeat

Go top-down the tree $\mathcal{T}$. The raw p-value is corrected for multiplicity using

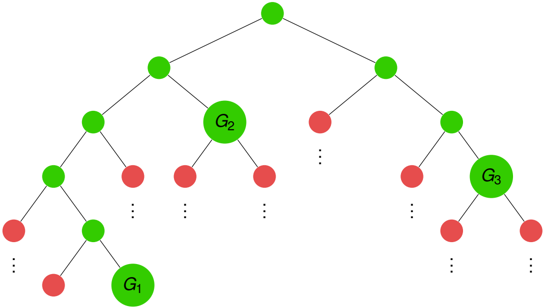
$$P_{G;\text{adjusted}} = \max_{G' \supseteq G} \tilde{P}_{G'} \quad \text{with} \quad \tilde{P}_G = P_G \cdot p/|G|,$$

where G' is any group in the tree $\mathcal{T}$.

If $P_{G;\text{adjusted}} \leq \alpha$, continue to consider the children of G for group testing.

until All the children of each group G when going top-down in $\mathcal{T}$ are non-significant at level α.

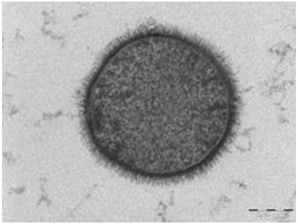
Hierarchical testing



- Significant groups and non-significant groups,
- The hierarchical procedure returns G_1, G_2, and G_3.
- The group G_1 is a leaf consisting of one variable.

Riboflavin (vitamin B_2) Production with Bacillus Subtilis

- Bacillus Subtilis: bacterium



- Response: log-transformed riboflavin production rate
- Covariates: expression levels of 4088 genes
- Sample size $n = 71$

Results

The log-expression level of $p = 4088$ genes is tested for association with the response.

p-value	significant cluster
1.631e-11	YEBC_at
$<$ 2.2e-16	LYSC_at
$<$ 2.2e-16	XTRA_at
$<$ 2.2e-16	XKDS_at
0.01420	YXLC_at, YXLD_at, YXLG_at
0.01420	YOAB_at
0.04544	BMR_at
0.01420	YCKE_at

Conclusion and Discussion

- Group Inference
  1. No requirement on G.
  2. Computationally efficient.
  3. Powerful test under dense alternatives.
  4. Have coverage property for the highly correlated setting.
- Inference for $\beta_G^{\intercal} A \beta_G$
- Hierarchical testing: high correlation
- Feasible computation for millions of variables

Reference and Acknowledgement

Guo, Z., Renaux, C., Bühlmann, P., Cai, T.T. (2019) Group Inference in High Dimensions with Applications to Hierarchical Testing

Acknowledgement to NSF and NIH, the Institute of Mathematical Research (FIM) at ETH Zurich for support.

Thank You!