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
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.

<table>
<thead>
<tr>
<th>Name</th>
<th>School</th>
<th>Year</th>
<th>Descendants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shi, Jian Qing</td>
<td>Chinese University of Hong Kong</td>
<td>1996</td>
<td></td>
</tr>
<tr>
<td>Song, Xin-Yuan</td>
<td>Chinese University of Hong Kong</td>
<td>2001</td>
<td>1</td>
</tr>
<tr>
<td>Zhang, Wenyang</td>
<td>Chinese University of Hong Kong</td>
<td>1999</td>
<td>6</td>
</tr>
<tr>
<td>Zhu, Hongtu</td>
<td>Chinese University of Hong Kong</td>
<td>2000</td>
<td>23</td>
</tr>
</tbody>
</table>

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
High-dimensional linear regression

\[ y_i = X_{i.}^T \beta + \epsilon_i, \quad \text{for } 1 \leq i \leq n. \]

where \( X_{i.}, \beta \in \mathbb{R}^p. \)

- 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

\[ H_0 : \beta_G = 0, \quad (1) \]

where \( \beta_G = \{ \beta_j; j \in G \}. \)
The null $H_0 : \beta_G = 0$ can be written as

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

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

Two special cases

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

with $\Sigma$ denoting the covariance matrix of $X_i$.

$$H_{0,1} : \beta_G^T \beta_G = 0.$$
For a group of **highly correlated** variables,
1. It is ambitious to detect significant single variable $\beta_i$
   - Inaccurate estimator of $\beta_i$
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.
Model with interaction (Tian, Alizadeh, Gentles and Tibshirani, 2014)

\[ y_i = X_i^T \beta + D_i (\gamma_0 + X_i^T \gamma) + \epsilon_i. \]

\[ H_0 : \gamma = 0 \]

- Interaction test
- Detection of Effect Heterogeneity (\(D_i\) is treatment)
Model with interaction (Tian, Alizadeh, Gentles and Tibshirani, 2014)

\[ y_i = X_i^T \beta + D_i (\gamma_0 + X_i^T \gamma) + \epsilon_i. \]

\[ H_0 : \gamma = 0 \]

- Interaction test
- Detection of Effect Heterogeneity \((D_i \text{ is treatment})\)

Equivalent model,

\[ y_i = W_i^T \eta + \epsilon_i. \]

with \( W_i = (D_i X_i^T, 1, X_i^T)^T \) and \( \eta = (\gamma^T, \gamma_0, \beta^T)^T \).

\[ H_0 : \eta_G = 0 \] with \( G = \{1, 2, \cdots, p\} \).
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

$$\beta_G^T \Sigma_{G,G} \beta_G = \mathbb{E}|X_{i,G} \beta_G|^2.$$
Inference for $Q_{\Sigma} = \beta_G^T \Sigma_{G,G} \beta_G$
Bias Correction

Initial estimators

\[ \hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{2n} \| y - X\beta \|_2^2 + \lambda \| \beta \|_1, \text{ with } \lambda \asymp \sqrt{\log p/n\sigma} \]

\[ \hat{\Sigma} = \frac{1}{n} X^T X. \]

Decompose \( \hat{\beta}_G^T \hat{\Sigma}_{G,G} \hat{\beta}_G - \beta_G^T \Sigma_{G,G} \beta_G \) as

\[-2 \hat{\beta}_G^T \hat{\Sigma}_{G,G}(\beta_G - \hat{\beta}_G) + \beta_G^T (\hat{\Sigma}_{G,G} - \Sigma_{G,G}) \beta_G - (\hat{\beta}_G - \beta_G)^T \hat{\Sigma}_{G,G}(\hat{\beta}_G - \beta_G)\]

Estimate \( \hat{\beta}_G^T \hat{\Sigma}_{G,G}(\beta_G - \hat{\beta}_G) \) and correct \( \hat{\beta}_G^T \hat{\Sigma}_{G,G} \hat{\beta}_G \).
Construction of Projection Direction

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

\[
\begin{align*}
    u^T \frac{1}{n} X^T (y - X\hat{\beta}) - \beta_G^T \hat{\Sigma}_{G,G} (\beta_G - \hat{\beta}_G) \\
    = \frac{1}{n} u^T X^T \epsilon + \left[ \hat{\Sigma} u - \left( \beta_G^T \hat{\Sigma}_{G,G} \mathbf{0} \right)^T \right]^T (\beta - \hat{\beta}).
\end{align*}
\]

\( \| \beta - \hat{\beta} \|_1 \) is small

\[
\left| \left[ \hat{\Sigma} u - \left( \beta_G^T \hat{\Sigma}_{G,G} \mathbf{0} \right)^T \right]^T (\beta - \hat{\beta}) \right| \leq \| \beta - \hat{\beta} \|_1 \left\| \hat{\Sigma} u - \left( \beta_G^T \hat{\Sigma}_{G,G} \mathbf{0} \right)^T \right\|_\infty.
\]

Minimize/Constrained \( u^T \hat{\Sigma} u \) and

\[
\left\| \hat{\Sigma} u - \left( \beta_G^T \hat{\Sigma}_{G,G} \mathbf{0} \right)^T \right\|_\infty = \max_{1 \leq j \leq p} \left| \langle e_j, \hat{\Sigma} u - \left( \beta_G^T \hat{\Sigma}_{G,G} \mathbf{0} \right)^T \rangle \right|
\]
Construction of Projection Direction

Initial proposal

\[ \hat{u} = \arg \min_{u} \; u^\top \hat{\Sigma} u \]

s.t. \[ \max_{w \in C_0} \left| \left<w, \hat{\Sigma} u - \left(\hat{\beta}_G^\top \hat{\Sigma}_{G,G} 0\right)^\top\right> \right| \leq \|\hat{\Sigma}_{G,G} \hat{\beta}_G\|_2 \lambda_n \]

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

\[ C_0 = \{e_1, \cdots, e_p\}. \]

- Constrain bias and minimize variance: Zhang & Zhang ’14; Javanmard & Montanari ’14;
- Only work for small \(|G|\).
\[ \hat{u} = \arg \min_u u^T \hat{\Sigma} u \]

subject to
\[ \max_{w \in C} \left| \left\langle w, \hat{\Sigma} u - \left( \hat{\beta}_G^T \hat{\Sigma}_{G,G} \ 0 \right)^T \right\rangle \right| \leq \| \hat{\Sigma}_{G,G} \hat{\beta}_G \|_2 \lambda_n \]

\[ C = \left\{ e_1, \cdots, e_p, \frac{1}{\| \hat{\Sigma}_{G,G} \hat{\beta}_G \|_2} \left( \hat{\beta}_G^T \hat{\Sigma}_{G,G} \ 0 \right)^T \right\}. \]

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

\[ \hat{Q}_\Sigma = \hat{\beta}_G^T \hat{\Sigma}_{G,G} \hat{\beta}_G + \frac{2}{n} \hat{u}^T X^T (y - X \hat{\beta}). \] (2)
Inference Procedure

We estimate the variance of the proposed estimator \( \hat{Q}_\Sigma \) by

\[
\hat{V}_\Sigma(\tau) = \frac{4\hat{\sigma}^2}{n} \hat{u}^T \hat{\Sigma} \hat{u} + \frac{1}{n^2} \sum_{i=1}^{n} \left( \beta_G^T X_{iG} X_{iG}^T \beta_G - \beta_G^T \hat{\Sigma}_{G,G} \beta_G \right)^2 + \frac{\tau}{n},
\]

for some positive constant \( \tau > 0 \).

\[
\phi_{\Sigma}(\tau) = 1 \left( \hat{Q}_\Sigma \geq z_{1-\alpha} \sqrt{\hat{V}_\Sigma(\tau)} \right)
\]

\[
CI_{\Sigma}(\tau) = \left( \hat{Q}_\Sigma - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_\Sigma(\tau)}, \hat{Q}_\Sigma + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_\Sigma(\tau)} \right)
\]

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 $\hat{Q}_\Sigma$ satisfies*

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

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

No condition on $G$!

Super-efficiency for $\beta_G$ close to 0, $\sqrt{V^0_\Sigma} \ll 1/\sqrt{n}$. Enlarge variance by adding $\tau/n$. 
Theoretical Justification

**Theorem 1.**

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

$$\limsup_{n,p \to \infty} P \left( \left| \hat{Q}_\Sigma - Q_\Sigma \right| \geq z_{1-\frac{\alpha}{2}} \sqrt{V_\Sigma} \right) \leq \alpha \quad \text{with} \quad V_\Sigma = V_0^\Sigma + \frac{\tau}{n}$$

$$V_0^\Sigma = \frac{4\sigma^2}{n} \hat{u}^T \hat{\Sigma} \hat{u} + \frac{1}{n^2} \sum_{i=1}^{n} \left( \beta_G^T X_{iG} X_{iG}^T \beta_G - \beta_G^T \hat{\Sigma}_{G,G} \beta_G \right)^2$$

- **No condition on $G$!**
Theoretical Justification

**Theorem 1.**

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

\[
\limsup_{n,p \to \infty} P \left( \left| \hat{Q}_\Sigma - Q_\Sigma \right| \geq z_{1-\frac{\alpha}{2}} \sqrt{V_\Sigma} \right) \leq \alpha \quad \text{with} \quad V_\Sigma = V^0_\Sigma + \frac{\tau}{n}
\]

\[
V^0_\Sigma = \frac{4\sigma^2}{n} \hat{u}^T \hat{\Sigma} \hat{u} + \frac{1}{n^2} \sum_{i=1}^{n} \left( \beta_G^T X_i G X_i^T \beta_G - \beta_G^T \hat{\Sigma} G \beta_G \right)^2
\]

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

\[ H_0 = \left\{ (\beta, \Sigma, \sigma) \in \Theta (k) : \|\beta_G\|_2 = 0 \right\} \]

Under the same assumption as Theorem 1,

\[ \sup_{\theta \in H_0} \liminf_{n,p \to \infty} P_{\theta} (\phi_\Sigma(\tau) = 1) \leq \alpha \]
Corollary 2.

For $\theta \in \mathcal{H}_1, A(\delta(t)) = \{ (\beta, \Sigma, \sigma) \in \Theta(k) : \beta_G^T A \beta_G = \delta(t) \},$

$$\liminf_{n, p \to \infty} P_{\theta} (\phi_{\Sigma}(\tau) = 1) \geq 1 - \Phi(-t)$$

- $\delta(t) = (1.01z_{1-\alpha} + t)\sqrt{V_{\Sigma}} \asymp \frac{1+t}{\sqrt{n}}(\sqrt{\tau} + \|\beta_G\|_2 + \|\beta_G\|_2^2)$

- $\phi_{\Sigma}$ is of asymptotic power 1 as long as $t \to \infty$. 
**Corollary 2.**

For $\theta \in \mathcal{H}_1$, $A(\delta(t)) = \{ (\beta, \Sigma, \sigma) \in \Theta(k) : \beta^T G A \beta_G = \delta(t) \}$,

$$\lim inf_{n,p \to \infty} P_{\theta} (\phi_{\Sigma}(\tau) = 1) \geq 1 - \Phi(-t)$$

- $\delta(t) = (1.01 z_{1-\alpha} + t) \sqrt{V_{\Sigma}} \approx \frac{1+t}{\sqrt{n}} (\sqrt{\tau} + \| \beta_G \|_2 + \| \beta_G \|_2^2)$
- $\phi_{\Sigma}$ is of asymptotic power 1 as long as $t \to \infty$.
- $\chi^2$-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

\[ y_i = X_i^T \beta + \epsilon_i, \quad \text{for } 1 \leq i \leq n. \]

- \( p = 500 \)
- \( \beta_j = \delta \) 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 \( \delta \) over \( \{0, 0.04\} \) and \( n \) over \( \{250, 350, 500, 800\} \).

\[ H_{0,G} : \beta_i = 0 \text{ for } i \in G, \text{ where } G = \{30, 31, \ldots, 200\}. \]
Implemented Methods

Maximum test based on the debiased estimator

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

\[
\phi_{\text{FD}} = 1 \left( \max_{j \in G} |\hat{\beta}_j^{\text{FD}}| \geq q_{\alpha}^{\text{FD}} \right) \quad \text{and} \quad \phi_{\text{hdi}} = 1 \left( \max_{j \in G} |\hat{\beta}_j^{\text{hdi}}| \geq q_{\alpha}^{\text{hdi}} \right),
\]

where \( q_{\alpha}^{\text{FD}} \) and \( q_{\alpha}^{\text{hdi}} \) are computed by bootstrap or sampling.
Maximum test based on the debiased estimator

1. **Fast Debiased (FD):** \( \{ \hat{\beta}^{FD}_j \}_{1 \leq j \leq p} \) (Javanmard & Montanari ’14)

2. **hdi:** \( \{ \hat{\beta}^{\text{hdi}}_j \}_{1 \leq j \leq p} \) (van de Geer, Bühlmann, Ritov & Dezeure ’14)

\[
\phi_{\text{FD}} = 1 \left( \max_{j \in G} |\hat{\beta}^{\text{FD}}_j| \geq q^{\text{FD}}_{\alpha} \right)
\text{ and } \phi_{\text{hdi}} = 1 \left( \max_{j \in G} |\hat{\beta}^{\text{hdi}}_j| \geq q^{\text{hdi}}_{\alpha} \right).
\]

where \( q^{\text{FD}}_{\alpha} \) and \( q^{\text{hdi}}_{\alpha} \) are computed by bootstrap or sampling.

\[
\phi_{\Sigma}(\tau) = 1 \left( \hat{Q}_{\Sigma} \geq z_{1-\alpha} \sqrt{\hat{V}_{\Sigma}(\tau)} \right)
\]

Compare \( \phi_{I}(0), \phi_{I}(1), \phi_{\Sigma}(0), \phi_{\Sigma}(1) \) and \( \phi_{\text{FD}}, \phi_{\text{hdi}} \).
One computational unit: a $p$-dimensional LASSO regression.

1. $\phi_{FD}, \phi_{hdi}: |G| + 1$ 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.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$n$</th>
<th>$\phi_I(0)$</th>
<th>$\phi_I(1)$</th>
<th>$\phi_\Sigma(0)$</th>
<th>$\phi_\Sigma(1)$</th>
<th>$\phi_{FD}$</th>
<th>$\phi_{hdi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>250</td>
<td>0.962</td>
<td>0.002</td>
<td>0.994</td>
<td>0.008</td>
<td>0.112</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>350</td>
<td>0.992</td>
<td>0.000</td>
<td>0.998</td>
<td>0.004</td>
<td>0.086</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.996</td>
<td>0.002</td>
<td>1.000</td>
<td>0.002</td>
<td>0.078</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.980</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.064</td>
<td>0.038</td>
</tr>
<tr>
<td>0.04</td>
<td>250</td>
<td>0.986</td>
<td>0.230</td>
<td>1.000</td>
<td>0.618</td>
<td>0.226</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>350</td>
<td>1.000</td>
<td>0.188</td>
<td>1.000</td>
<td>0.854</td>
<td>0.184</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1.000</td>
<td>0.292</td>
<td>1.000</td>
<td>0.946</td>
<td>0.128</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>1.000</td>
<td>0.374</td>
<td>1.000</td>
<td>1.000</td>
<td>0.128</td>
<td>0.180</td>
</tr>
</tbody>
</table>

$\phi_\Sigma(1)$ controls Type I error and more powerful than $\phi_{FD}$, $\phi_{hdi}$.
Simulation II: High Correlation
- $p = 500$
- $\beta_1 = \beta_3 = \delta$ and $\beta_j = 0$ for $j \neq 1, 3$
- High correlation among the first 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 $\delta$ 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

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$n$</th>
<th>$\phi_I(0)$</th>
<th>$\phi_I(1)$</th>
<th>$\phi_\Sigma(0)$</th>
<th>$\phi_\Sigma(1)$</th>
<th>$\phi_{FD}$</th>
<th>$\phi_{hdi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>250</td>
<td>0.034</td>
<td>0.000</td>
<td>0.072</td>
<td>0.000</td>
<td>0.070</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>350</td>
<td>0.058</td>
<td>0.000</td>
<td>0.104</td>
<td>0.000</td>
<td>0.082</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.052</td>
<td>0.000</td>
<td>0.092</td>
<td>0.000</td>
<td>0.082</td>
<td>0.056</td>
</tr>
<tr>
<td>0.2</td>
<td>250</td>
<td>0.682</td>
<td>0.134</td>
<td>0.982</td>
<td>0.590</td>
<td>0.998</td>
<td>0.936</td>
</tr>
<tr>
<td></td>
<td>350</td>
<td>0.704</td>
<td>0.138</td>
<td>1.000</td>
<td>0.822</td>
<td>1.000</td>
<td>0.972</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.634</td>
<td>0.234</td>
<td>1.000</td>
<td>0.960</td>
<td>1.000</td>
<td>0.994</td>
</tr>
</tbody>
</table>

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

- $\phi_\Sigma(1)$ has valid confidence property;
- **No confidence property for $\phi_{FD}, \phi_{hdi}$.**
### High Correlation: Coverage Property

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

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( n )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( \beta_5 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( \beta_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>250</td>
<td>0.972</td>
<td>0.968</td>
<td>0.970</td>
<td>0.976</td>
<td>0.976</td>
<td>0.952</td>
<td>0.950</td>
<td>0.944</td>
<td>0.950</td>
<td>0.946</td>
</tr>
<tr>
<td></td>
<td>350</td>
<td>0.968</td>
<td>0.972</td>
<td>0.962</td>
<td>0.970</td>
<td>0.968</td>
<td>0.942</td>
<td>0.942</td>
<td>0.932</td>
<td>0.966</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.974</td>
<td>0.972</td>
<td>0.964</td>
<td>0.970</td>
<td>0.982</td>
<td>0.950</td>
<td>0.936</td>
<td>0.956</td>
<td>0.950</td>
<td>0.956</td>
</tr>
<tr>
<td>0.2</td>
<td>250</td>
<td>0.400</td>
<td>0.714</td>
<td>0.418</td>
<td>0.720</td>
<td>0.758</td>
<td>0.864</td>
<td>0.798</td>
<td>0.910</td>
<td>0.828</td>
<td>0.268</td>
</tr>
<tr>
<td></td>
<td>350</td>
<td>0.464</td>
<td>0.696</td>
<td>0.414</td>
<td>0.722</td>
<td>0.680</td>
<td>0.910</td>
<td>0.822</td>
<td>0.922</td>
<td>0.844</td>
<td>0.234</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.424</td>
<td>0.702</td>
<td>0.408</td>
<td>0.686</td>
<td>0.674</td>
<td>0.876</td>
<td>0.860</td>
<td>0.916</td>
<td>0.842</td>
<td>0.298</td>
</tr>
</tbody>
</table>

#### Table: Empirical Coverage for \( \|\beta_G\|_2^2, \beta_G^T \Sigma G, G \beta_G \) in the Highly Correlated scenario

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( n )</th>
<th>( CI_{\Pi}(\tau = 0) )</th>
<th>( CI_{\Pi}(\tau = 1) )</th>
<th>( CI_{\Sigma}(\tau = 0) )</th>
<th>( CI_{\Sigma}(\tau = 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>250</td>
<td>0.104</td>
<td>1.000</td>
<td>0.090</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>350</td>
<td>0.110</td>
<td>1.000</td>
<td>0.088</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.094</td>
<td>1.000</td>
<td>0.070</td>
<td>1.000</td>
</tr>
<tr>
<td>0.2</td>
<td>250</td>
<td>0.912</td>
<td>0.992</td>
<td>0.822</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>350</td>
<td>0.916</td>
<td>0.998</td>
<td>0.822</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.924</td>
<td>0.994</td>
<td>0.842</td>
<td>0.996</td>
</tr>
</tbody>
</table>
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
**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.
**INPUT:** Hierarchical tree $T$ 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

1: repeat
2: Go top-down the tree $T$. The raw $p$-value is corrected for multiplicity using

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

where $G'$ is any group in the tree $T$.
3: If $P_{G; \text{adjusted}} \leq \alpha$, continue to consider the children of $G$ for group testing.
4: until All the children of each group $G$ when going top-down in $T$ are non-significant at level $\alpha$. 

Zijian Guo

Group Inference 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\textsubscript{2}) Production with Bacillus Subtilis

- Bacillus Subtilis: bacterium
- Response: log-transformed riboflavin production rate
- Covariates: expression levels of 4088 genes
- Sample size $n = 71$
The log-expression level of $p = 4088$ genes is tested for association with the response.

<table>
<thead>
<tr>
<th>$p$-value</th>
<th>significant cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.631e-11</td>
<td>YEBC_at</td>
</tr>
<tr>
<td>&lt; 2.2e-16</td>
<td>LYSC_at</td>
</tr>
<tr>
<td>&lt; 2.2e-16</td>
<td>XTRA_at</td>
</tr>
<tr>
<td>&lt; 2.2e-16</td>
<td>XKDS_at</td>
</tr>
<tr>
<td>0.01420</td>
<td>YXLC_at, YXLD_at, YXLG_at</td>
</tr>
<tr>
<td>0.01420</td>
<td>YOAB_at</td>
</tr>
<tr>
<td>0.04544</td>
<td>BMR_at</td>
</tr>
<tr>
<td>0.01420</td>
<td>YCKE_at</td>
</tr>
</tbody>
</table>
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 $\beta_G^T A \beta_G$

- Hierarchical testing: high correlation

- Feasible computation for millions of variables

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

Thank You!