Optimal Estimation of Co-Heritability in High-Dimensional Linear Models

Zijian Guo

University of Pennsylvania and Rutgers University

JSM 2017

Based on joint work with T. Tony Cai, Wanjie Wang and Hongzhe Li.

Research Problem

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \epsilon_{n\times 1} \quad w_{n\times 1} = X_{n\times p}\gamma_{p\times 1} + \delta_{n\times 1}$$

- Number of covariates $p \gg$ sample size n.
- When $p > n$, $\|\beta\|_0 \leq k$.

Research Problem

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \epsilon_{n\times 1} \quad w_{n\times 1} = X_{n\times p}\gamma_{p\times 1} + \delta_{n\times 1}$$

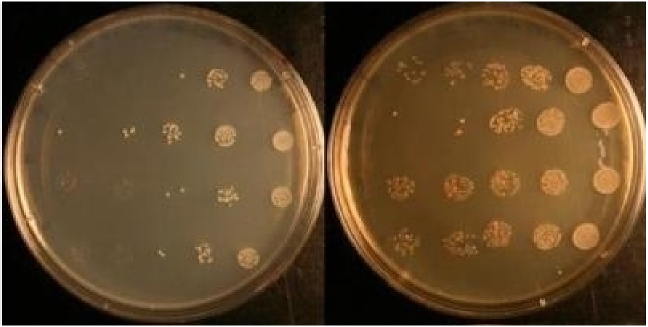
- Number of covariates $p \gg$ sample size n.
- When $p > n$, $\|\beta\|_0 \leq k$.

Research Problems:

1. $\|\beta\|_2^2$ and $\|\gamma\|_2^2$.
2. $\langle\beta, \gamma\rangle$.
3. $\langle\beta, \gamma\rangle / \|\beta\|_2 \|\gamma\|_2$.

Motivation: yeast study

Media: YNB (Yeast Nitrogen Base) v.s. YPD (Yeast extract Peptone Dextrose)



Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T. L. V., & Kruglyak, L. (2013). Finding the sources of missing heritability in a yeast cross. *Nature*, 494(7436), 234-237.

The statistical problem

$$\text{Model for YNB:} \quad y = X\beta + \epsilon$$
$$\text{Model for YPD:} \quad w = X\gamma + \delta$$

- Columns of X: SNPs.
- Number of SNPs= 4,410 $>$ sample size=1,008.
- Sparse β and γ.

The statistical problem

$$\text{Model for YNB:} \quad y = X\beta + \epsilon$$
$$\text{Model for YPD:} \quad w = X\gamma + \delta$$

- Columns of X: SNPs.
- Number of SNPs= 4,410 $>$ sample size=1,008.
- Sparse β and γ.

- Heritability: $\|\beta\|_2^2$ and $\|\gamma\|_2^2$.

Bulik-Sullivan, B., et al.(2015). An atlas of genetic correlations across human diseases and traits. *Nature genetics*, 47, 1236-1241.

The statistical problem

$$\text{Model for YNB:} \quad y = X\beta + \epsilon$$
$$\text{Model for YPD:} \quad w = X\gamma + \delta$$

- Columns of X: SNPs.
- Number of SNPs= 4,410 $>$ sample size=1,008.
- Sparse β and γ.

- Heritability: $\|\beta\|_2^2$ and $\|\gamma\|_2^2$.
- Genetic Covariance: $\langle \beta, \gamma \rangle$.
- Genetic Correlation: $\langle \beta, \gamma \rangle / \|\beta\|_2 \|\gamma\|_2$.

Bulik-Sullivan, B., et al.(2015). An atlas of genetic correlations across human diseases and traits. *Nature genetics*, 47, 1236-1241.

Plug-in Scaled Lasso?

<u>Scaled Lasso estimators</u>

$$\{\widehat{\beta}, \hat{\sigma}_1\} = \underset{\beta \in \mathbb{R}^p, \sigma_1 \in \mathbb{R}^+}{\arg\min} \frac{\|y - X\beta\|_2^2}{2n\sigma_1} + \frac{\sigma_1}{2} + \sqrt{\frac{2.01 \log p}{n}} \sum_{j=1}^{p} \frac{\|X_{\cdot j}\|_2}{\sqrt{n}} |\beta_j|;$$

$$\{\widehat{\gamma}, \hat{\sigma}_2\} = \underset{\gamma \in \mathbb{R}^p, \sigma_2 \in \mathbb{R}^+}{\arg\min} \frac{\|w - Z\gamma\|_2^2}{2n\sigma_2} + \frac{\sigma_2}{2} + \sqrt{\frac{2.01 \log p}{n}} \sum_{j=1}^{p} \frac{\|Z_{\cdot j}\|_2}{\sqrt{n}} |\gamma_j|.$$

Q: How about $\langle \widehat{\beta}, \widehat{\gamma} \rangle$?

Plug-in Scaled Lasso?

Scaled Lasso estimators

$$\{\widehat{\beta}, \hat{\sigma}_1\} = \underset{\beta \in \mathbb{R}^p, \sigma_1 \in \mathbb{R}^+}{\arg\min} \frac{\|y - X\beta\|_2^2}{2n\sigma_1} + \frac{\sigma_1}{2} + \sqrt{\frac{2.01 \log p}{n}} \sum_{j=1}^{p} \frac{\|X_{\cdot j}\|_2}{\sqrt{n}} |\beta_j|;$$

$$\{\widehat{\gamma}, \hat{\sigma}_2\} = \underset{\gamma \in \mathbb{R}^p, \sigma_2 \in \mathbb{R}^+}{\arg\min} \frac{\|w - Z\gamma\|_2^2}{2n\sigma_2} + \frac{\sigma_2}{2} + \sqrt{\frac{2.01 \log p}{n}} \sum_{j=1}^{p} \frac{\|Z_{\cdot j}\|_2}{\sqrt{n}} |\gamma_j|.$$

Q: How about $\langle \widehat{\beta}, \widehat{\gamma} \rangle$?

A: Too much bias!

Plug-in De-biased Estimator?

De-biased Estimator:

$$\widetilde{\beta} = \widehat{\beta} + \text{Correction}; \quad \widetilde{\gamma} = \widehat{\gamma} + \text{Correction}.$$

- Zhang & Zhang (2014);
- Javanmard & Montanari (2014);
- van de Geer, Bühlmann, Ritov & Dezeure (2014).

Q: How about $\langle \widetilde{\beta}, \widetilde{\gamma} \rangle$?

Plug-in De-biased Estimator?

De-biased Estimator:

$$\widetilde{\beta} = \widehat{\beta} + \text{Correction}; \quad \widetilde{\gamma} = \widehat{\gamma} + \text{Correction}.$$

- Zhang & Zhang (2014);
- Javanmard & Montanari (2014);
- van de Geer, Bühlmann, Ritov & Dezeure (2014).

Q: How about $\langle \widetilde{\beta}, \widetilde{\gamma} \rangle$?

A: Too much variance!

Plug-in De-biased Estimator?

De-biased Estimator:

$$\widetilde{\beta} = \widehat{\beta} + \text{Correction}; \quad \widetilde{\gamma} = \widehat{\gamma} + \text{Correction}.$$

- Zhang & Zhang (2014);
- Javanmard & Montanari (2014);
- van de Geer, Bühlmann, Ritov & Dezeure (2014).

Q: How about $\langle \widetilde{\beta}, \widetilde{\gamma} \rangle$?

A: Too much variance!

Propose an estimator balancing bias and variance!

Bias Correction

Error decomposition of $\langle \widehat{\beta}, \widehat{\gamma} \rangle$:

$$\langle \widehat{\beta}, \widehat{\gamma} \rangle - \langle \beta, \gamma \rangle = -\underbrace{\left(\langle \widehat{\gamma}, \beta - \widehat{\beta} \rangle + \langle \widehat{\beta}, \gamma - \widehat{\gamma} \rangle \right)}_{\text{Main Error}} - \langle \widehat{\beta} - \beta, \widehat{\gamma} - \gamma \rangle \quad (1)$$

Bias Correction

Error decomposition of $\langle\widehat{\beta},\widehat{\gamma}\rangle$:

$$\langle\widehat{\beta},\widehat{\gamma}\rangle-\langle\beta,\gamma\rangle=-\underbrace{\left(\langle\widehat{\gamma},\beta-\widehat{\beta}\rangle+\langle\widehat{\beta},\gamma-\widehat{\gamma}\rangle\right)}_{\text{Main Error}}-\langle\widehat{\beta}-\beta,\widehat{\gamma}-\gamma\rangle \quad (1)$$

Bias Correction Idea:

$$\langle\widehat{\beta},\widehat{\gamma}\rangle+\underbrace{\langle\widehat{\gamma},\beta-\widehat{\beta}\rangle+\langle\widehat{\beta},\gamma-\widehat{\gamma}\rangle}_{\text{Main Error}}-\langle\beta,\gamma\rangle=-\langle\widehat{\beta}-\beta,\widehat{\gamma}-\gamma\rangle. \quad (2)$$

Estimation of $\langle \widehat{\gamma}, \beta - \widehat{\beta} \rangle + \langle \widehat{\beta}, \gamma - \widehat{\gamma} \rangle$

$$\widehat{u}_1^\top \frac{1}{n} X^\top \left(y - X\widehat{\beta} \right) \Rightarrow \langle \widehat{\gamma}, \beta - \widehat{\beta} \rangle; \;\; \widehat{u}_2^\top \frac{1}{n} Z^\top \left(w - Z\widehat{\gamma} \right) \Rightarrow \langle \widehat{\beta}, \gamma - \widehat{\gamma} \rangle.$$

Estimation of $\langle\widehat{\gamma}, \beta - \widehat{\beta}\rangle + \langle\widehat{\beta}, \gamma - \widehat{\gamma}\rangle$

$$\widehat{u}_1^\top \frac{1}{n} X^\top \left(y - X\widehat{\beta}\right) \Rightarrow \langle\widehat{\gamma}, \beta - \widehat{\beta}\rangle; \;\; \widehat{u}_2^\top \frac{1}{n} Z^\top (w - Z\widehat{\gamma}) \Rightarrow \langle\widehat{\beta}, \gamma - \widehat{\gamma}\rangle.$$

$$\sqrt{n}\left(u^\top \frac{1}{n} X^\top (y - X\widehat{\beta}) - \langle\widehat{\gamma}, \beta - \widehat{\beta}\rangle\right) = \underbrace{u^\top X^\top \epsilon/\sqrt{n}}_{\text{Variance}} + \underbrace{\sqrt{n}(u^\top \widehat{\Sigma} - \widehat{\gamma})(\beta - \widehat{\beta})}_{\text{Remaining Bias}}$$

1. $\frac{1}{\sqrt{n}} u^\top X^\top \epsilon \mid X \sim N(0, u^\top \widehat{\Sigma} u)$ where $\widehat{\Sigma} = \frac{1}{n} X^\top X$.
2. $\left|\sqrt{n}\left(u^\top \widehat{\Sigma} - \widehat{\gamma}\right)\left(\beta - \widehat{\beta}\right)\right| \leq \sqrt{n}\|u^\top \widehat{\Sigma} - \widehat{\gamma}\|_\infty \|\beta - \widehat{\beta}\|_1$

Estimation of $\langle \widehat{\gamma}, \beta - \widehat{\beta} \rangle + \langle \widehat{\beta}, \gamma - \widehat{\gamma} \rangle$

$$\widehat{u}_1^{\top} \frac{1}{n} X^{\top} \left(y - X\widehat{\beta} \right) \Rightarrow \langle \widehat{\gamma}, \beta - \widehat{\beta} \rangle; \;\; \widehat{u}_2^{\top} \frac{1}{n} Z^{\top} (w - Z\widehat{\gamma}) \Rightarrow \langle \widehat{\beta}, \gamma - \widehat{\gamma} \rangle.$$

$$\sqrt{n} \left(u^{\top} \frac{1}{n} X^{\top} (y - X\widehat{\beta}) - \langle \widehat{\gamma}, \beta - \widehat{\beta} \rangle \right) = \underbrace{u^{\top} X^{\top} \epsilon / \sqrt{n}}_{\text{Variance}} + \underbrace{\sqrt{n} (u^{\top} \widehat{\Sigma} - \widehat{\gamma})(\beta - \widehat{\beta})}_{\text{Remaining Bias}}$$

1. $\frac{1}{\sqrt{n}} u^{\top} X^{\top} \epsilon \mid X \sim N(0, u^{\top} \widehat{\Sigma} u)$ where $\widehat{\Sigma} = \frac{1}{n} X^{\top} X$.
2. $\left| \sqrt{n} \left(u^{\top} \widehat{\Sigma} - \widehat{\gamma} \right) \left(\beta - \widehat{\beta} \right) \right| \leq \sqrt{n} \| u^{\top} \widehat{\Sigma} - \widehat{\gamma} \|_{\infty} \| \beta - \widehat{\beta} \|_1$

$$\widehat{u}_1 = \arg\min_{u \in \mathbb{R}^p} \left\{ u^{\top} \widehat{\Sigma} u : \| \widehat{\Sigma} u - \widehat{\gamma} \|_{\infty} \leq \| \widehat{\gamma} \|_2 \lambda_n \right\}, \tag{3}$$

where $\lambda_n \asymp \sqrt{\log p / n}$.

FDE for $\langle\beta,\gamma\rangle$

Functional De-biased Estimator (FDE)

$$\widehat{\langle\beta,\gamma\rangle} = \langle\widehat{\beta},\widehat{\gamma}\rangle + \underbrace{\widehat{u}_1^{\top}\frac{1}{n}X^{\top}\left(y - X\widehat{\beta}\right) + \widehat{u}_2^{\top}\frac{1}{n}Z^{\top}\left(w - Z\widehat{\gamma}\right)}_{\text{Estimation of Main Error}}. \quad (4)$$

FDE for $\langle \beta, \gamma \rangle$

Functional De-biased Estimator (FDE)

$$\widehat{\langle \beta, \gamma \rangle} = \langle \widehat{\beta}, \widehat{\gamma} \rangle + \underbrace{\widehat{u}_1^{\top} \frac{1}{n} X^{\top} \left(y - X\widehat{\beta} \right) + \widehat{u}_2^{\top} \frac{1}{n} Z^{\top} \left(w - Z\widehat{\gamma} \right)}_{\text{Estimation of Main Error}}. \quad (4)$$

- Not the plug-in of de-biased estimators.
- Center of confidence intervals for $\langle \beta, \gamma \rangle$.

Theoretical Property

Theorem 1 (G. et.al., 2016)

Suppose that $k \leq c \min\{\frac{n}{\log p}, p^{\nu}\}$ with $0 < \nu < 1/2$ and $k = \max\{\|\beta\|_0, \|\gamma\|_0\}$, then with high probability,

$$\left|\widehat{\langle\beta,\gamma\rangle} - \langle\beta,\gamma\rangle\right| \lesssim (\|\beta\|_2 + \|\gamma\|_2)\left(\frac{1}{\sqrt{n}} + \frac{k\log p}{n}\right) + \frac{k\log p}{n}.$$

- Variance + Bias.
- When $k \ll \frac{\sqrt{n}}{\log p}$, $\min\{\|\beta\|_2, \|\gamma\|_2\} \gg \frac{k\log p}{\sqrt{n}}$, rate is

$$(\|\beta\|_2 + \|\gamma\|_2)\frac{1}{\sqrt{n}}.$$

Theoretical Property

Theorem 1(G. et.al., 2016)

Suppose that $k \leq c\min\{\frac{n}{\log p}, p^{\nu}\}$ with $0 < \nu < 1/2$ and $k = \max\{\|\beta\|_0, \|\gamma\|_0\}$, then with high probability,

$$\left|\widehat{\langle\beta,\gamma\rangle} - \langle\beta,\gamma\rangle\right| \lesssim (\|\beta\|_2 + \|\gamma\|_2)\left(\frac{1}{\sqrt{n}} + \frac{k\log p}{n}\right) + \frac{k\log p}{n}.$$

- Variance + Bias.
- When $k \ll \frac{\sqrt{n}}{\log p}$, $\min\{\|\beta\|_2, \|\gamma\|_2\} \gg \frac{k\log p}{\sqrt{n}}$, rate is

$$(\|\beta\|_2 + \|\gamma\|_2)\frac{1}{\sqrt{n}}.$$

- Optimal convergence rate

Minimax Convergence Rate

Define

$$\Theta(k, M_0) = \{(\beta, \Sigma_1, \sigma_1, \gamma, \Sigma_2, \sigma_2) : (\beta, \Sigma_1, \sigma_1) \in \mathcal{G}(k, M_0), (\gamma, \Sigma_2, \sigma_2) \in \mathcal{G}(k, M_0)\},$$

$$\mathcal{G}(k, M_0) = \left\{(\beta, \Sigma, \sigma) : \|\beta\|_0 \leq k,\ \|\beta\|_2 \leq M_0,\ \frac{1}{M_1} \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq M_1,\ \sigma \leq M_2\right\}$$

with $M_1 \geq 1$ and $M_2 > 0$.

Minimax Convergence Rate

Define

$$\Theta(k, M_0) = \{(\beta, \Sigma_1, \sigma_1, \gamma, \Sigma_2, \sigma_2) : (\beta, \Sigma_1, \sigma_1) \in \mathcal{G}(k, M_0), (\gamma, \Sigma_2, \sigma_2) \in \mathcal{G}(k, M_0)\},$$

$$\mathcal{G}(k, M_0) = \left\{(\beta, \Sigma, \sigma) : \|\beta\|_0 \le k,\ \|\beta\|_2 \le M_0,\ \frac{1}{M_1} \le \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) \le M_1,\ \sigma \le M_2\right\}$$

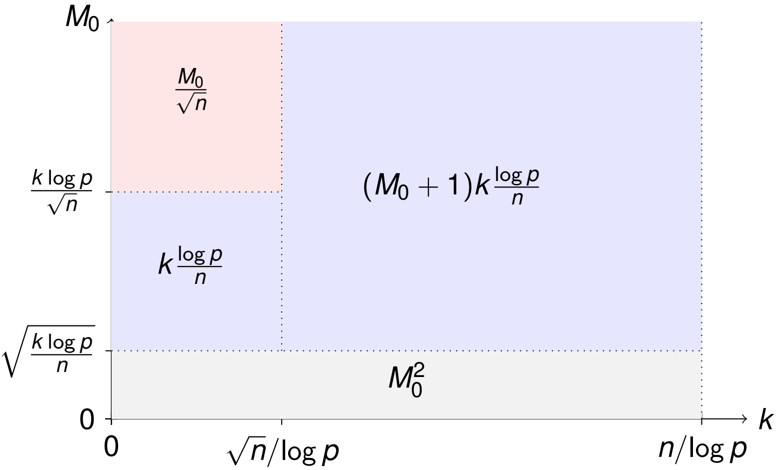
with $M_1 \ge 1$ and $M_2 > 0$.

Theorem 2(G. et.al., 2016)

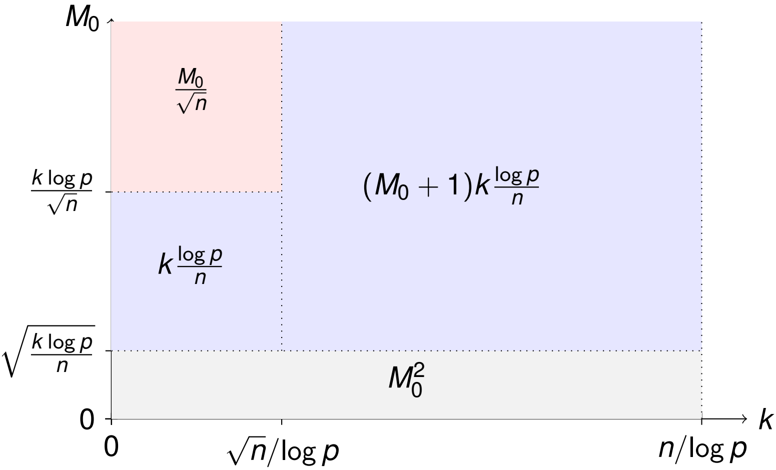
Suppose $k \le c \min\left\{\frac{n}{\log p}, p^{\nu}\right\}$ for some constants $c > 0$ and $0 \le \nu < \frac{1}{2}$. Then

$$\inf_{\widetilde{\mathrm{I}}} \sup_{\theta \in \Theta(k, M_0)} \mathbf{P}_{\theta}\left(\left|\widetilde{\mathrm{I}} - \langle \beta, \gamma \rangle\right| \gtrsim \min\left\{M_0\left(\frac{1}{\sqrt{n}} + \frac{k \log p}{n}\right) + \frac{k \log p}{n}, M_0^2\right\}\right) \ge \frac{1}{4}.$$

Optimal Convergence Rate of Estimating $\langle \beta, \gamma \rangle$



Optimal Convergence Rate of Estimating $\langle \beta, \gamma \rangle$



For $M_0 \gtrsim \sqrt{\frac{k \log p}{n}}$, the optimal rate is achieved by FDE.

FDE estimator of $\langle \beta, \gamma \rangle / \|\beta\|_2 \|\gamma\|_2$?

1. How to estimate $\|\beta\|_2^2$?
   - Correct the main error by $\|\widehat{\beta}\|_2^2$.
   - Functional De-biased Estimator (FDE) of $\|\beta\|_2^2$.

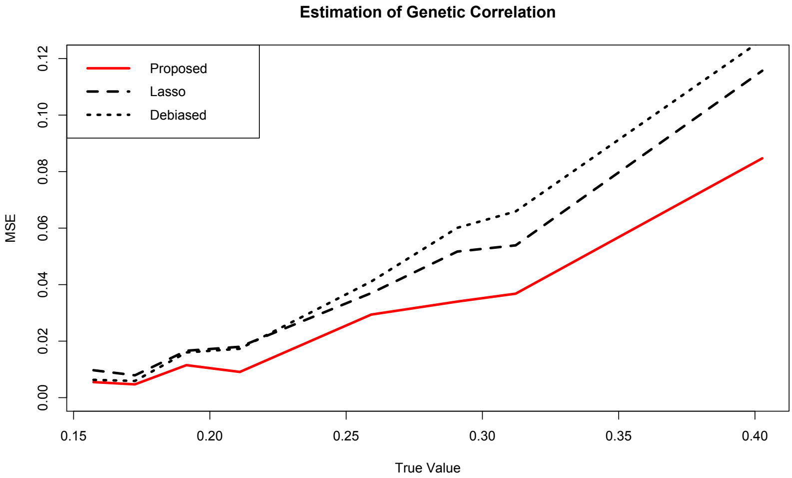
FDE estimator of $\langle \beta, \gamma \rangle / \|\beta\|_2 \|\gamma\|_2$?

1. How to estimate $\|\beta\|_2^2$?
   - Correct the main error by $\|\widehat{\beta}\|_2^2$.
   - Functional De-biased Estimator (FDE) of $\|\beta\|_2^2$.

2. How to estimate $\langle \beta, \gamma \rangle / \|\beta\|_2 \|\gamma\|_2$?

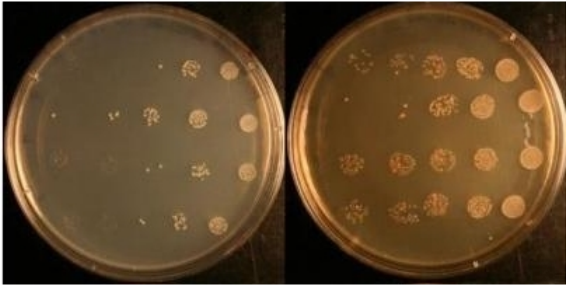
$$\text{FDE of } \frac{\langle \beta, \gamma \rangle}{\|\beta\|_2 \|\gamma\|_2} = \frac{\text{FDE of } \langle \beta, \gamma \rangle}{\sqrt{\text{FDE of } \|\beta\|_2^2} \times \sqrt{\text{FDE of } \|\gamma\|_2^2}}$$

Simulation study: $\langle \beta, \gamma \rangle / \|\beta\|_2 \|\gamma\|_2$



Real Data Analysis

Media: YNB (Yeast Nitrogen Base) v.s. YPD (Yeast extract Peptone Dextrose)



Colony sizes under YNB and YPD

- Heritability of YNB: 0.4594
- Heritability of YPD: 0.6680
- Genetic Correlation/Covariance: 0.5195/0.4246

A Related Problem – Accuracy assessment

Given data (X, y) and an estimator $\widehat{\beta}$,

$$\ell_q \text{ Accuracy Functional: } \quad \|\widehat{\beta} - \beta\|_q^q.$$

A Related Problem – Accuracy assessment

Given data (X, y) and an estimator $\widehat{\beta}$,

$$\ell_q \text{ Accuracy Functional: } \quad \|\widehat{\beta} - \beta\|_q^q.$$

- The difficulty depends on the estimator $\widehat{\beta}$.
  1. Lasso estimator.
  2. Zero estimator.

A Related Problem – Accuracy assessment

Given data (X, y) and an estimator $\widehat{\beta}$,

$$\ell_q \text{ Accuracy Functional: } \quad \|\widehat{\beta} - \beta\|_q^q.$$

- The difficulty depends on the estimator $\widehat{\beta}$.
  1. Lasso estimator.
  2. Zero estimator.
- Inference for $\|\widehat{\beta} - \beta\|_q^q$: general lower bound tool.

References

Guo, Z., Wang, W., Cai, T.T., & Li, H.(2016). Optimal estimation of co-heritability in high-dimensional linear models. *Submitted*.

Cai, T.T., & Guo, Z.(2016). Accuracy assessment for high-dimensional linear regression. *Annals of Statistics*, to appear.

Thank You!

Estimation of $\langle \beta, \gamma \rangle / \|\beta\|_2 \|\gamma\|_2$?

The optimal convergence rate of estimating $\frac{\langle \beta, \gamma \rangle}{\|\beta\|_2 \|\gamma\|_2}$ over

$$\{\|\beta\|_2 \geq \eta_0, \|\gamma\|_2 \geq \eta_0\}$$

is

$$\min\left\{\frac{1}{\eta_0}\left(\frac{1}{\sqrt{n}} + \frac{k \log p}{n}\right) + \frac{1}{\eta_0^2}\frac{k \log p}{n}, 1\right\}.$$

Estimation of $\langle\beta,\gamma\rangle/\|\beta\|_2\|\gamma\|_2$?

The optimal convergence rate of estimating $\frac{\langle\beta,\gamma\rangle}{\|\beta\|_2\|\gamma\|_2}$ over

$$\{\|\beta\|_2 \geq \eta_0, \|\gamma\|_2 \geq \eta_0\}$$

is

$$\min\left\{\frac{1}{\eta_0}\left(\frac{1}{\sqrt{n}}+\frac{k\log p}{n}\right)+\frac{1}{\eta_0^2}\frac{k\log p}{n}, 1\right\}.$$

1. If $\eta_0 \gg \sqrt{\frac{k\log p}{n}}$, the optimal procedure is to plug in FDE of $\langle\beta,\gamma\rangle$, $\|\beta\|_2^2$ and $\|\gamma\|_2^2$.

Estimation of $\langle \beta, \gamma \rangle / \|\beta\|_2 \|\gamma\|_2$?

The optimal convergence rate of estimating $\frac{\langle \beta, \gamma \rangle}{\|\beta\|_2 \|\gamma\|_2}$ over

$$\{\|\beta\|_2 \geq \eta_0, \|\gamma\|_2 \geq \eta_0\}$$

is

$$\min \left\{ \frac{1}{\eta_0} \left(\frac{1}{\sqrt{n}} + \frac{k \log p}{n} \right) + \frac{1}{\eta_0^2} \frac{k \log p}{n}, 1 \right\}.$$

1. If $\eta_0 \gg \sqrt{\frac{k \log p}{n}}$, the optimal procedure is to plug in FDE of $\langle \beta, \gamma \rangle$, $\|\beta\|_2^2$ and $\|\gamma\|_2^2$.
2. If $\eta_0 \lesssim \sqrt{\frac{k \log p}{n}}$, just estimate it by 0.