The Lasso with general Gaussian designs and its applications

Yuting Wei

Carnegie Mellon University

JSM, Aug 4th, 2020

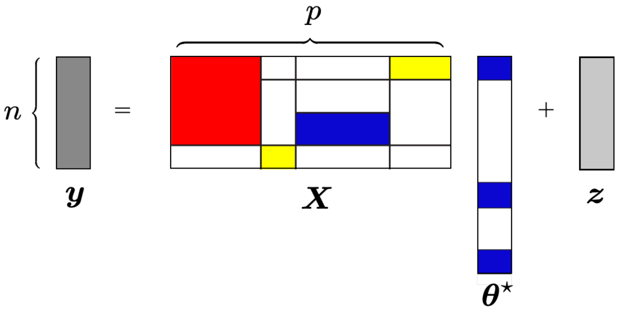
Michael Celentano
Stanford Stat

Andrea Montanari
Stanford Stat & EE

"The Lasso with general Gaussian designs with application to hypothesis testing,"

M. Celentano, A. Montanari, Y. Wei, 2020. https://arxiv.org/abs/2007.13716

Lasso estimator



$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

[Tibshirani, 1996]

Prior work: Lasso risk

Suppose $\boldsymbol{\theta}^*$ is s-sparse, $z \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$. Under restricted eigenvalue condition of design matrix $\boldsymbol{X}$,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2 \leq C\sigma\sqrt{\frac{s\log(p)}{n}}$$

[Bickel et al., 2009, Bühlmann and Van De Geer, 2011, Negahban et al., 2012, Zhao and Yu, 2006, Zhang and Zhang, 2014, Bellec et al., 2018]...

Prior work: Lasso risk

Suppose $\boldsymbol{\theta}^*$ is s-sparse, $z \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$. Under restricted eigenvalue condition of design matrix $\boldsymbol{X}$,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2 \leq C\sigma\sqrt{\frac{s\log(p)}{n}}$$

- unspecified constant

[Bickel et al., 2009, Bühlmann and Van De Geer, 2011, Negahban et al., 2012, Zhao and Yu, 2006, Zhang and Zhang, 2014, Bellec et al., 2018]...

Prior work: Lasso risk

Suppose $\boldsymbol{\theta}^*$ is s-sparse, $z \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$. Under restricted eigenvalue condition of design matrix $\boldsymbol{X}$,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2 \leq C\sigma\sqrt{\frac{s\log(p)}{n}}$$

- unspecified constant
- no distributional characterization of $\widehat{\boldsymbol{\theta}}$

[Bickel et al., 2009, Bühlmann and Van De Geer, 2011, Negahban et al., 2012, Zhao and Yu, 2006, Zhang and Zhang, 2014, Bellec et al., 2018]...

Prior work: Lasso risk

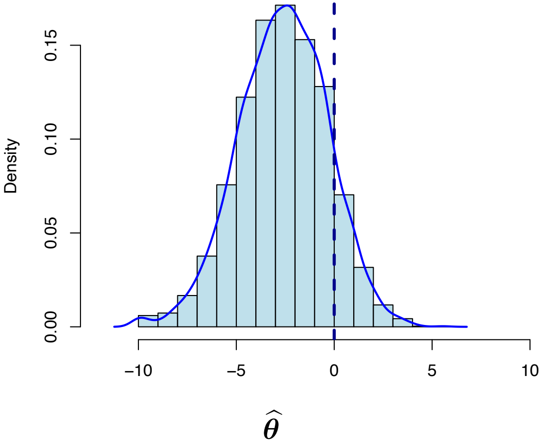
Suppose $\boldsymbol{\theta}^*$ is s-sparse, $z \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$. Under restricted eigenvalue condition of design matrix $\boldsymbol{X}$,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2 \le C\sigma\sqrt{\frac{s\log(p)}{n}}$$

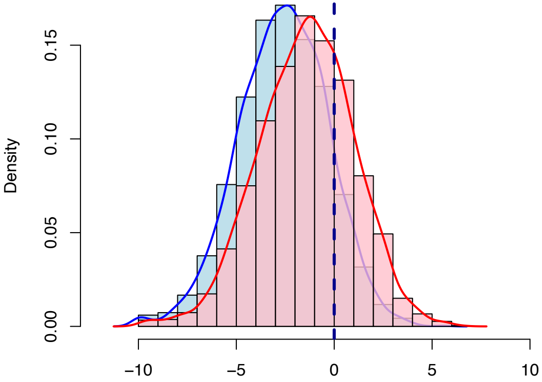
- unspecified constant
- no distributional characterization of $\widehat{\boldsymbol{\theta}}$
- inadequate for statistical inference

[Bickel et al., 2009, Bühlmann and Van De Geer, 2011, Negahban et al., 2012, Zhao and Yu, 2006, Zhang and Zhang, 2014, Bellec et al., 2018]...

Prior work: debiased Lasso



Prior work: debiased Lasso

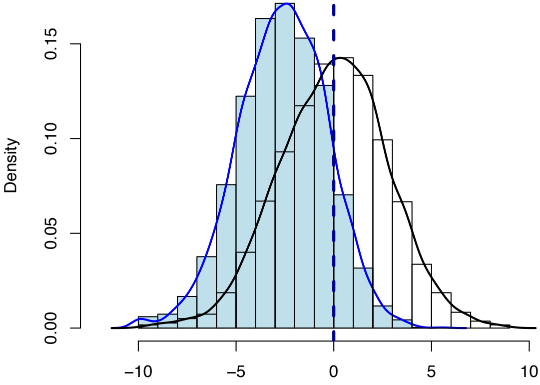


$$\widehat{\boldsymbol{\theta}}^d = \widehat{\boldsymbol{\theta}} + \boldsymbol{M}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})$$

$\boldsymbol{M}$ surrogate for $\boldsymbol{\Sigma}^{-1} = \mathbb{E}[\boldsymbol{x}_i\boldsymbol{x}_i^\top]^{-1}$

[Zhang and Zhang, 2014, Van de Geer et al., 2014, Javanmard and Montanari, 2014a, Javanmard and Montanari, 2014b]

Prior work: debiased Lasso



$$\widehat{\boldsymbol{\theta}}^d = \widehat{\boldsymbol{\theta}} + \boldsymbol{M}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})$$

$\boldsymbol{M}$ scaled version of $\boldsymbol{\Sigma}^{-1} = \mathbb{E}[\boldsymbol{x}_i \boldsymbol{x}_i^\top]^{-1}$

[Javanmard et al., 2018, Miolane and Montanari, 2018, Bellec and Zhang, 2019a, Bellec and Zhang, 2019b]

Inadequacy of current theory

- not applicable for $s/p = \text{const}$ regime
- precise characterization developed for uncorrelated designs
  [Javanmard and Montanari, 2014b, Miolane and Montanari, 2018]
- for correlated designs with $n > p$
  [Bellec and Zhang, 2019a, Bellec and Zhang, 2019b]

Towards a general design

Prior work: i.i.d. Gaussian design: $\boldsymbol{x}_i \sim \mathcal{N}(0, \mathbf{I}_p)$

[Bayati and Montanari, 2011, Thrampoulidis et al., 2015, Miolane and Montanari, 2018]

Towards a general design

Prior work: i.i.d. Gaussian design: $\boldsymbol{x}_i \sim \mathcal{N}(0, \mathbf{I}_p)$

[Bayati and Montanari, 2011, Thrampoulidis et al., 2015, Miolane and Montanari, 2018]

What happens with general Gaussian design $\boldsymbol{x}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma})$?

Towards a general design

Prior work: i.i.d. Gaussian design: $\boldsymbol{x}_i \sim \mathcal{N}(0, \mathbf{I}_p)$

[Bayati and Montanari, 2011, Thrampoulidis et al., 2015, Miolane and Montanari, 2018]

What happens with general Gaussian design $\boldsymbol{x}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma})$?

— **difficulty:** non-isometry of $\| \cdot \|_1$ penalty.

This talk

$$\boldsymbol{y} = \boldsymbol{X}\theta^* + \boldsymbol{z} \ \in \mathbb{R}^n$$

This talk

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}^* + \boldsymbol{z} \;\; \in \mathbb{R}^n$$

- $\boldsymbol{\theta}^* \in \mathbb{R}^p$: s-sparse

This talk

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}^* + \boldsymbol{z} \ \in \mathbb{R}^n$$

- $\boldsymbol{\theta}^* \in \mathbb{R}^p$: s-sparse
- proportional regime: $p/n = \text{const}$, $s/p = \text{const}$

This talk

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}^* + \boldsymbol{z} \ \in \mathbb{R}^n$$

- $\boldsymbol{\theta}^* \in \mathbb{R}^p$: s-sparse
- proportional regime: $p/n = \text{const}$, $s/p = \text{const}$
- Gaussian noise: $\boldsymbol{z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$; Gaussian design: $\boldsymbol{x}_i \sim \mathcal{N}(0, \underbrace{\boldsymbol{\Sigma}}_{\text{known}})$

This talk

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}^* + \boldsymbol{z} \ \in \mathbb{R}^n$$

- $\boldsymbol{\theta}^* \in \mathbb{R}^p$: s-sparse
- proportional regime: $p/n = \text{const}$, $s/p = \text{const}$
- Gaussian noise: $\boldsymbol{z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$; Gaussian design: $\boldsymbol{x}_i \sim \mathcal{N}(0, \underbrace{\boldsymbol{\Sigma}}_{\text{known}})$

Goal: a distributional theory for general Gaussian design

Key observation

original model
$\widehat{\boldsymbol{\theta}}$

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\theta + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

Key observation

original model
$\widehat{\boldsymbol{\theta}}$

fixed design model
$\widehat{\boldsymbol{\theta}}^f$

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\theta + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \left\{ \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^* + \tau^* \boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, \mathbf{I}_p)$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \left\{ \frac{\zeta^*}{2} \|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1 \right\}$$

Key observation

original model
$\widehat{\boldsymbol{\theta}}$

fixed design model
$\widehat{\boldsymbol{\theta}}^f$

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\theta + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^* + \tau^*\boldsymbol{g},\ \boldsymbol{g} \sim \mathcal{N}(0, \mathbf{I}_p)$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^*}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

τ^*: effective risk level; ζ^*: effective non-sparsity

Key observation

original model $\widehat{\boldsymbol{\theta}}$	$\overset{\text{distribution}}{\approx}$	fixed design model $\widehat{\boldsymbol{\theta}}^f$

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\theta + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^* + \tau^*\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, \mathbf{I}_p)$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^*}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

τ^*: effective risk level; ζ^*: effective non-sparsity

Fixed point equations

$$(\tau^*, \zeta^*) \xrightarrow{\text{solution}} \begin{aligned} \tau^2 &= \sigma^2 + \mathsf{R}(\tau^2, \zeta) \\ \zeta &= 1 - \mathsf{df}(\tau^2, \zeta) \end{aligned}$$

Fixed point equations

$$(\tau^*, \zeta^*) \quad \xrightarrow{\text{solution}} \quad \begin{aligned} \tau^2 &= \sigma^2 + \mathsf{R}(\tau^2, \zeta) \\ \zeta &= 1 - \mathsf{df}(\tau^2, \zeta) \end{aligned}$$

$$\mathsf{R}(\tau^2, \zeta) := \underbrace{\frac{1}{n}\,\mathbb{E}\left[\left\|\boldsymbol{\Sigma}^{1/2}(\widehat{\boldsymbol{\theta}}^f(\tau, \zeta) - \boldsymbol{\theta}^*)\right\|_2^2\right]}_{\text{in-sample prediction risk}}$$

$$\mathsf{df}(\tau^2, \zeta) := \underbrace{\frac{1}{n}\,\mathbb{E}\left[\left\|\widehat{\boldsymbol{\theta}}^f(\tau, \zeta)\right\|_0\right]}_{\text{degrees of freedom}}$$

Fixed point equations

$$(\tau^*, \zeta^*) \quad \xrightarrow{\text{solution}} \quad \begin{aligned} \tau^2 &= \sigma^2 + \mathsf{R}(\tau^2, \zeta) \\ \zeta &= 1 - \mathsf{df}(\tau^2, \zeta) \end{aligned}$$

$$\mathsf{R}(\tau^2, \zeta) := \underbrace{\frac{1}{n}\mathbb{E}\left[\left\|\boldsymbol{\Sigma}^{1/2}(\widehat{\boldsymbol{\theta}}^f(\tau,\zeta) - \boldsymbol{\theta}^*)\right\|_2^2\right]}_{\text{in-sample prediction risk}}$$

$$\mathsf{df}(\tau^2, \zeta) := \underbrace{\frac{1}{n}\mathbb{E}\left[\left\|\widehat{\boldsymbol{\theta}}^f(\tau,\zeta)\right\|_0\right]}_{\text{degrees of freedom}}$$

Property: solution is unique and bounded for reasonably sparse $\boldsymbol{\theta}^*$.

Main result: Lasso distribution

Theorem (Celetano, Montanari, Wei '20)

Under mild conditions, for any 1*-Lipschitz function* ϕ *and* $\epsilon > 0$

$$\forall \lambda \in [\lambda_{\mathsf{min}}, \lambda_{\mathsf{max}}], \qquad \left| \phi\Big(\frac{\widehat{\boldsymbol{\theta}}_\lambda}{\sqrt{p}}, \frac{\boldsymbol{\theta}^*}{\sqrt{p}}\Big) - \mathbb{E}\Big[\phi\Big(\frac{\widehat{\boldsymbol{\theta}}^f_\lambda}{\sqrt{p}}, \frac{\boldsymbol{\theta}^*}{\sqrt{p}}\Big)\Big]\right| \le \epsilon,$$

with probability at least $1 - \frac{C}{\epsilon^4} e^{-cn\epsilon^4}$.

Main result: Lasso distribution

Theorem (Celetano, Montanari, Wei '20)

Under mild conditions, for any 1*-Lipschitz function* ϕ *and* $\epsilon > 0$

$$\forall \lambda \in [\lambda_{\min}, \lambda_{\max}], \qquad \left| \phi\Big(\frac{\widehat{\boldsymbol{\theta}}_\lambda}{\sqrt{p}}, \frac{\boldsymbol{\theta}^*}{\sqrt{p}} \Big) - \mathbb{E}\Big[\phi\Big(\frac{\widehat{\boldsymbol{\theta}}^f_\lambda}{\sqrt{p}}, \frac{\boldsymbol{\theta}^*}{\sqrt{p}} \Big) \Big] \right| \le \epsilon,$$

with probability at least $1 - \frac{C}{\epsilon^4} e^{-cn\epsilon^4}$.

A direct consequence:

$$\forall \lambda \in [\lambda_{\min}, \lambda_{\max}], \qquad \|\widehat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*\|_2 \approx \mathbb{E}\Big[\|\widehat{\boldsymbol{\theta}}^f_\lambda - \boldsymbol{\theta}^*\|_2 \Big]$$

Main result: properties for Lasso

- Lasso residual

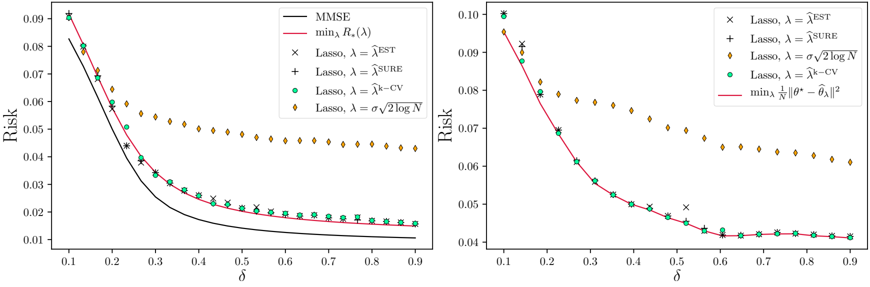
$$\mathbb{P}\left(\left|\frac{\|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\theta}}\|_2}{\sqrt{n}}-\tau^*\zeta^*\right|>\epsilon\right)\leq\frac{C}{\epsilon^2}e^{-cn\epsilon^4}\,.$$

- Lasso sparsity

$$\mathbb{P}\left(\left|\frac{\|\widehat{\boldsymbol{\theta}}\|_0}{n}-(1-\zeta^*)\right|>\epsilon\right)\leq\frac{C}{\epsilon^3}e^{-cn\epsilon^6}\,.$$

Application: model selection

$$\lambda \textbf{ selection:} \qquad \widehat{\lambda}^{\mathrm{EST}} := \min_{\lambda} \widehat{\tau}(\lambda) := \frac{\sqrt{n}\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}\|_2}{\underbrace{n - \|\widehat{\boldsymbol{\theta}}\|_0}_{\text{degrees of freedom}}}$$



(a) Independent columns

(b) Correlated columns

[Miolane and Montanari, 2018]

Debiased Lasso

- classical debiased Lasso

$$\widehat{\boldsymbol{\theta}}_0^{\mathrm{d}} = \widehat{\boldsymbol{\theta}} + \boldsymbol{M}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}), \qquad \boldsymbol{M} = \boldsymbol{\Sigma}^{-1}$$

Debiased Lasso

- classical debiased Lasso

$$\widehat{\boldsymbol{\theta}}_0^{\mathrm{d}} = \widehat{\boldsymbol{\theta}} + \boldsymbol{M}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}), \qquad \boldsymbol{M} = \boldsymbol{\Sigma}^{-1}$$

- debiased Lasso with degrees-of-freedom (DOF) adjustment

$$\widehat{\boldsymbol{\theta}}^{\mathrm{d}} := \widehat{\boldsymbol{\theta}} + \boldsymbol{M}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}), \qquad \boldsymbol{M} = \frac{\boldsymbol{\Sigma}^{-1}}{1 - \|\widehat{\boldsymbol{\theta}}\|_0/n}$$

[Javanmard and Montanari, 2014b, Miolane and Montanari, 2018, Bellec and Zhang, 2019a, Bellec and Zhang, 2019b]

Main result: $\widehat{\boldsymbol{\theta}}^{\mathrm{d}}$ behaves like $\boldsymbol{\theta}^* + \tau^*\boldsymbol{\Sigma}^{-1/2}\boldsymbol{g}$

Intuition for DOF adjustment

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\theta + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^* + \tau^*\boldsymbol{g}$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^*}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

Intuition for DOF adjustment

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^* + \tau^*\boldsymbol{g}$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^*}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

$$\widehat{\boldsymbol{\theta}}^{\mathrm{d}} := \widehat{\boldsymbol{\theta}} + \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})}{1 - \|\widehat{\boldsymbol{\theta}}\|_0/n}$$

$\widehat{\boldsymbol{\theta}}^f$

Intuition for DOF adjustment

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^* + \tau^*\boldsymbol{g}$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^*}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

$$\widehat{\boldsymbol{\theta}}^{\mathrm{d}} := \underbrace{\widehat{\boldsymbol{\theta}}}_{\widehat{\boldsymbol{\theta}}^f} + \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})}{\underbrace{1 - \|\widehat{\boldsymbol{\theta}}\|_0/n}_{\zeta^*}}$$

Intuition for DOF adjustment

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^* + \tau^*\boldsymbol{g}$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^*}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

$$\boldsymbol{\Sigma}^{-1}\cdot\zeta^*\boldsymbol{\Sigma}^{1/2}(\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\widehat{\boldsymbol{\theta}}^f) = \zeta^*(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}^f - \widehat{\boldsymbol{\theta}}^f)$$

$$\widehat{\boldsymbol{\theta}}^{\mathrm{d}} := \widehat{\boldsymbol{\theta}} + \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})}{1 - \|\widehat{\boldsymbol{\theta}}\|_0/n}$$

$\widehat{\boldsymbol{\theta}}^f$

ζ^*

Intuition for DOF adjustment

- **original model:** $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{z}$

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

- **fixed design model:** $\boldsymbol{y}^f = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}^* + \tau^*\boldsymbol{g}$

$$\widehat{\boldsymbol{\theta}}^f := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left\{ \frac{\zeta^*}{2}\|\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \right\}$$

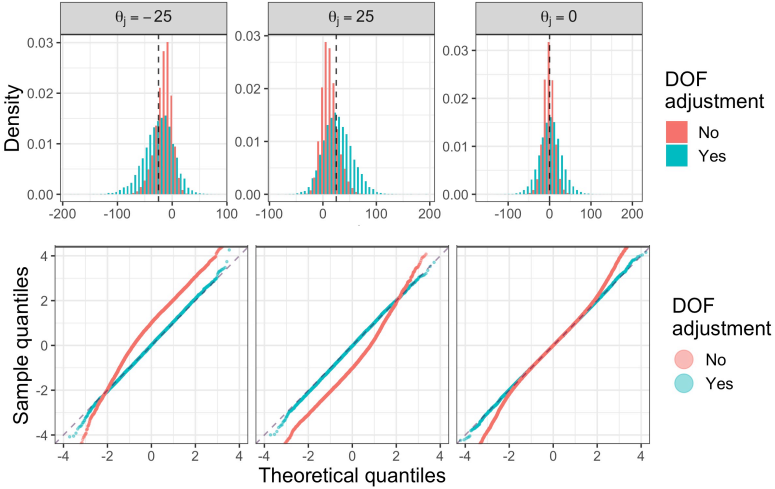
$$\boldsymbol{\Sigma}^{-1}\cdot\zeta^*\boldsymbol{\Sigma}^{1/2}(\boldsymbol{y}^f - \boldsymbol{\Sigma}^{1/2}\widehat{\boldsymbol{\theta}}^f) = \zeta^*(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}^f - \widehat{\boldsymbol{\theta}}^f)$$

$$\widehat{\boldsymbol{\theta}}^{\mathrm{d}} := \widehat{\boldsymbol{\theta}} + \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})}{1 - \|\widehat{\boldsymbol{\theta}}\|_0/n} \approx \boldsymbol{\theta}^* + \tau^*\boldsymbol{\Sigma}^{-1/2}\boldsymbol{g}$$

$\widehat{\boldsymbol{\theta}}^f$

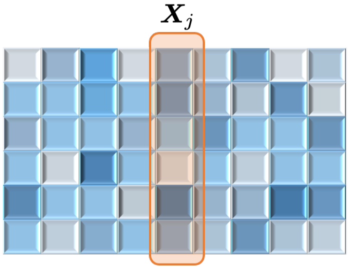
ζ^*

Debiased Lasso with DOF adjustment



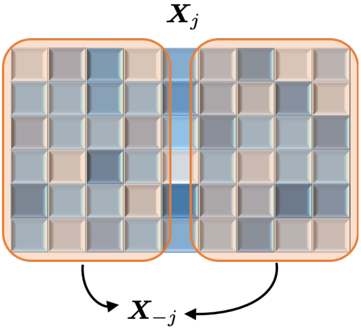
Here $p = 100,\ n = 25,\ s = 20,\ \Sigma_{ij} = 0.5^{|i-j|},\ \sigma = 1$

Confidence interval for a single coordinate



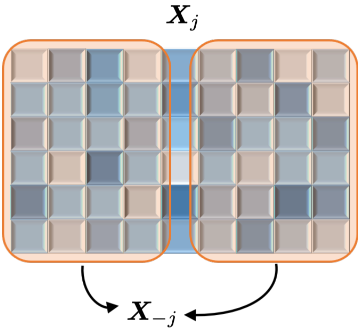
- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$

Confidence interval for a single coordinate



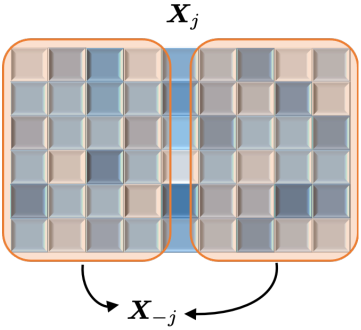
- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$

Confidence interval for a single coordinate



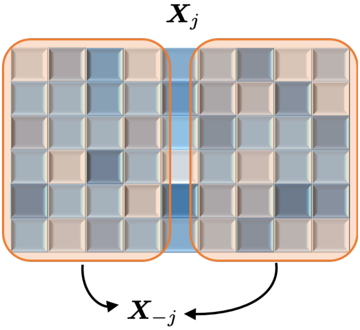
- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$ $\longrightarrow$ residual $\boldsymbol{X}_j^{\perp}$

Confidence interval for a single coordinate



- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$ $\longrightarrow$ residual $\boldsymbol{X}_j^{\perp}$
- obtain leave-j^{th}-coordinate-out Lasso $\widehat{\boldsymbol{\theta}}_{\mathrm{loo}}$

Confidence interval for a single coordinate

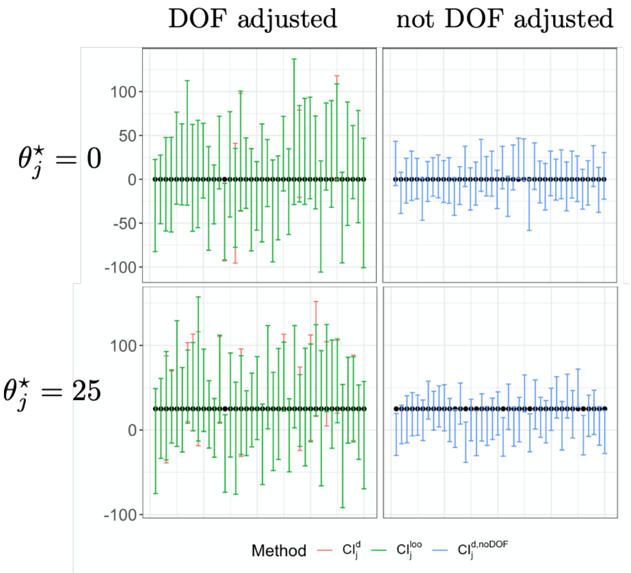


- regress $\boldsymbol{X}_j$ on $\boldsymbol{X}_{-j}$ $\longrightarrow$ residual $\boldsymbol{X}_j^{\perp}$
- obtain leave-j^{th}-coordinate-out Lasso $\widehat{\boldsymbol{\theta}}_{\mathrm{loo}}$
- construct confidence interval

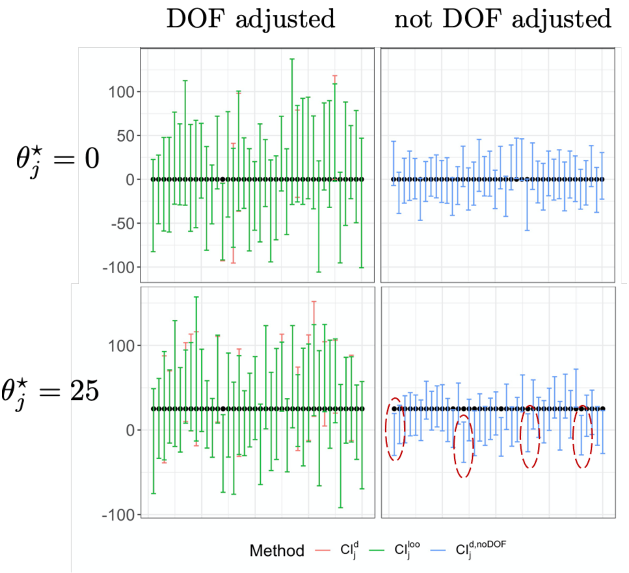
$$\mathsf{CI}_j^{\mathrm{loo}} := \left[\xi_j \ \pm \ \widehat{\mathsf{sd}} \cdot z_{1-\alpha/2}\right]$$

$$\xi_j = \text{correlation between } \boldsymbol{X}_j^{\perp} \text{ and } \boldsymbol{y} - \boldsymbol{X}_{-j}\widehat{\boldsymbol{\theta}}_{\mathrm{loo}}$$

Confidence interval for a single coordinate



Confidence interval for a single coordinate



Concluding remarks

Summary

- distributional theory of Lasso under general Gaussian design
- applications
  - theoretical support for model selection
  - study debiased Lasso and propose single confidence interval

Concluding remarks

Summary

- distributional theory of Lasso under general Gaussian design
- applications
  - theoretical support for model selection
  - study debiased Lasso and propose single confidence interval

Future directions

- distributional theory beyond Gaussian design
  [Bayati et al., 2015, Oymak and Tropp, 2018, Montanari and Nguyen, 2017]
- theoretical limit if $\boldsymbol{\Sigma}$ is unknown

Concluding remarks

Summary

- distributional theory of Lasso under general Gaussian design
- applications
  - theoretical support for model selection
  - study debiased Lasso and propose single confidence interval

Future directions

- distributional theory beyond Gaussian design
  [Bayati et al., 2015, Oymak and Tropp, 2018, Montanari and Nguyen, 2017]
- theoretical limit if $\boldsymbol{\Sigma}$ is unknown

"The Lasso with general Gaussian designs with application to hypothesis testing,"
M. Celentano, A. Montanari, Y. Wei, 2020. https://arxiv.org/abs/2007.13716