Atlanta 2018

The debiased Lasso

Sara van de Geer

The debiased Lasso

Sara van de Geer

August, 2018

Linear model:

$$Y = X\beta^0 + \epsilon,$$

$Y \in \mathbb{R}^n$ response,
$X \in \mathbb{R}^{n \times p}$ design matrix
$\beta^0 \in \mathbb{R}^p$ unknown coefficients
$\epsilon \in \mathbb{R}^n$ noise
$p \gg n$: high-dimensional case
Parameter of interest: $\beta_1^0 \in \mathbb{R}$

Assumptions

- ∘ The rows of $(X, Y) \in \mathbb{R}^{n\times(p+1)}$ are i.i.d.$\sim$ $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{p+1}$
- ∘ $(\mathbf{x}, \mathbf{y})$ is Gaussian with mean zero
- ∘ $\mathbb{E}(\mathbf{y}|\mathbf{x}) = \mathbf{x}\beta^0$
- ∘ $\Sigma := \mathbb{E}\mathbf{x}^T\mathbf{x}$ has inverse Θ.
- ∘ $\|\Sigma\|_\infty = \mathcal{O}(1)$, say $\mathrm{diag}(\Sigma) = I$
- ∘ $\Lambda^2_{\min} \gg 0$ where $\Lambda^2_{\min}$ is the smallest eigenvalue of Σ

The Lasso [Tibshirani, 1996] is

$$\hat{\beta} := \arg\min_{b \in \mathbb{R}^p} \left\{ \| Y - Xb \|_2^2 / n + 2\lambda \| b \|_1 \right\}$$

with $\lambda > 0$ a tuning parameter

A debiased Lasso is

$$\hat{b}_1 = \hat{\beta}_1 + \tilde{\Theta}_1^T X^T (Y - X\hat{\beta})/n$$

where $\tilde{\Theta}_1 \in \mathbb{R}^p$ is soem estimate of the first column Θ_1 of Θ.

[Zhang & Zhang, 2014], [Javanmard & Montenari, 2014], [Belloni, Chernozhukov,& Kato, 2015] ...

Lemma

Suppose $\|\hat{\beta} - \beta^0\|_1 = o_{\mathbb{P}_{\beta^0}}(1/\sqrt{\log p})$ *and that* Σ *is known. Take* $\tilde{\Theta}_1 = \Theta_1$*. Then*

$$\frac{\sqrt{n}(\hat{b}_1 - \beta_1^0)}{\sqrt{\Theta_{1,1}}} \xrightarrow{\mathcal{D}_{\beta^0}} \mathcal{N}(0,1)$$

Proof.

$$\begin{aligned}
\hat{b}_1 - \beta_1^0 &= \hat{\beta}_1 - \beta_1^0 + \Theta_1^T X^T (Y - X\hat{\beta})/n \\
&= \hat{\beta}_1 - \beta_1^0 + \Theta_1^T X^T (\underbrace{\epsilon + X\beta^0}_{=Y} - X\hat{\beta}))/n \\
&= \underbrace{\hat{\beta}_1 - \beta_1^0}_{= \mathrm{e}_1{}^T(\hat{\beta} - \beta^0)} + \Theta_1^T X^T \epsilon/n - \Theta_1^T \underbrace{X^T X/n}_{:=\hat{\Sigma}}(\hat{\beta} - \beta^0) \\
&= \Theta_1^T X^T \epsilon/n + (\underbrace{\mathrm{e}_1 - \hat{\Sigma}\Theta_1}_{=(\Sigma - \hat{\Sigma})\Theta_1})^T (\hat{\beta} - \beta^0)
\end{aligned}$$

We have

$$\Theta_1^T X^T \epsilon / \sqrt{n\Theta_{1,1}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

where we use $\Theta_1^T \Sigma \Theta_1 = \Theta_1^T \mathrm{e}_1 = \Theta_{1,1}$. Moreover

$$|((\Sigma - \hat{\Sigma})\Theta_1)^T (\hat{\beta} - \beta^0)| \le \underbrace{\|(\Sigma - \hat{\Sigma})\Theta_1\|_\infty}_{= \mathcal{O}_{\mathbb{P}}(\sqrt{\log p/n})} \underbrace{\|\hat{\beta} - \beta^0\|_1}_{= o_{\mathbb{P}}(1/\sqrt{\log p})} = o_{\mathbb{P}}(1/\sqrt{n})$$

$\square$

The asymptotic Cramér Rao lower bound

Model: $\beta^0 \in \mathcal{B}$.
Model directions: $\mathcal{H}_{\beta^0} := \{h \in \mathbb{R}^p : \ \beta^0 + h/\sqrt{n} \in \mathcal{B}\}$.
Definition An estimator T is called regular at β^0 if for all fixed $\rho > 0$ and $R > 0$ and all sequences $h \in \mathcal{H}_{\beta^0}$ with $|h_1| \ge \rho$ and $h^T \Sigma h \le R^2$, it holds that

$$\sqrt{n}\left(\frac{T - (\beta_1^0 + h_1/\sqrt{n})}{V_{\beta^0}}\right) \stackrel{\mathcal{D}_{\beta^0 + h/\sqrt{n}}}{\longrightarrow} \mathcal{N}(0,1)$$

where $V_{\beta^0}^2 = \mathcal{O}(1)$ is some constant called the asymptotic variance.

Proposition *Suppose T is asymptotically linear at β^0 with influence function* $\mathbf{i}_{\beta^0} : \mathbb{R}^{p+1} \to \mathbb{R}$*:*

$$T - \beta_1^0 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{i}_{\beta^0}(X_i, Y_i) + o_{\mathbb{P}_{\beta^0}}(1/\sqrt{n})$$

where $\mathbb{E}_{\beta^0}\mathbf{i}_{\beta^0}(\mathbf{x}, \mathbf{y}) = 0$ *and* $V_{\beta^0}^2 := \mathbb{E}_{\beta^0}\mathbf{i}_{\beta^0}^2(\mathbf{x}, \mathbf{y}) = \mathcal{O}(1)$*. Assume the Lindeberg condition*

$$\lim_{n\to\infty} \mathbb{E}_{\beta^0}\mathbf{i}_{\beta^0}^2(\mathbf{x}, \mathbf{y})\mathrm{l}\left\{\mathbf{i}_{\beta^0}^2(\mathbf{x}, \mathbf{y}) > \eta n V_{\beta^0}^2\right\} = 0 \ \forall\ \eta > 0.$$

Assume further that T is regular at β^0. Then for all fixed $\rho > 0$ and $R > 0$

$$V_{\beta^0}^2 + o(1) \geq \max_{h\in\mathcal{H}_{\beta^0}:\ |h_1|\geq\rho,\ h^T\Sigma h\leq R^2} \frac{h_1^2}{h^T\Sigma h}.$$

Corollary *Assume the conditions of the proposition and that for some fixed* $\rho > 0$ *and* $R > 0$ *and some sequence* $h \in \mathcal{H}_{\beta^0}$, *with* $|h_1| \geq \rho$ *and* $h^T \Sigma h \leq R^2$, *it is true that*

$$V_{\beta^0}^2 = \frac{h_1^2}{h^T \Sigma h} + o(1).$$

Then T *is asymptotically efficient.*

Example Suppose

$$\mathcal{B} := \{b \in \mathbb{R}^p : \underbrace{\|b\|_0^0}_{:=\# b_j \neq 0} \leq s\}.$$

Let $S_0 := \{\beta_j^0 \neq 0\}$ be the set of active coefficients in β^0 and $s_0 = |S_0|$.

Some special cases

a) If $\|\Theta_1\|_0^0 \leq s - s_0$, then $V_{\beta^0}^2 + o(1) = \Theta_{1,1}$.

b) Suppose $\{1\} \in S_0$, $s = s_0$ and that the following "betamin" condition holds: $|\beta_j^0| > m_n/\sqrt{n}$ for all $j \in S_0$, where $m_n \to \infty$. Then

$$V_{\beta^0}^2 + o(1) \geq (\Sigma_{S_0,S_0}^{-1})_{1,1}.$$

c) More generally, if $\{1\} \in S_0$ and $|\beta_j^0| > m_n/\sqrt{n}$ for all $j \in S_0$, then the lower bound corresponds to knowing the set S_0 up to $s - s_0$ additional variables.

Example Suppose

$$\mathcal{B} := \{b \in \mathbb{R}^p : \ \|b\|_1 \le \sqrt{s}\}.$$

Suppose β^0 stays away from the boundary of the parameter space, i.e., for a fixed $0 < \eta < 1$ it holds that $\|\beta^0\|_1 \le (1-\eta)\sqrt{s}$. Then for all $M > 0$ fixed

$$V_{\beta^0}^2 + o(1) \ge \left(\min_{c \in \mathbb{R}^{p-1}: \ \|c\|_1 \le M\sqrt{ns}} \mathbb{E}(\mathbf{x}_1 - \mathbf{x}_{-1} c)^2 \right)^{-1}$$

where $\mathbf{x}_{-1} := (\mathbf{x}_2, \ldots, \mathbf{x}_p)$. To improve over $\Theta_{1,1}$ we must have that Θ_1 is rather non-sparse: $\|\Theta_1\|_1$ should be of larger order than $\sqrt{ns}$.

Example Suppose

$$\mathcal{B} := \{b \in \mathbb{R}^p : \ \|b\|_r^r \le \sqrt{s^{2-r}}\}.$$

If β^0 stays away from the boundary, then for all $M > 0$ fixed

$$V_{\beta^0}^2 + o(1) \ge \left(\min_{c \in \mathbb{R}^{p-1}: \ \|c\|_1 \le M\sqrt{n^r s^{2-r}}} \mathbb{E}(\mathbf{x}_1 - \mathbf{x}_{-1} c)^2 \right)^{-1}.$$

To improve over $\Theta_{1,1}$ we must have that $\|\Theta_1\|_1$ is of larger order than $\sqrt{n^r s^{2-r}}$.

Definition *Let* $\mathcal{B}$ *be the model for* β^0. *Let* $\{\mathbf{Z}_n\}$ *be a sequence of real-valued random variables depending on* (X, Y) *and* $\{r_n\}$ *a sequence of positive numbers. We say that* $\{\mathbf{Z}_n\}$ *is* $\mathcal{O}_{\mathbb{P}_{\beta^0}}(r_n)$ *uniformly in* $\beta^0 \in \mathcal{B}$ *if*

$$\lim_{M\to\infty} \limsup_{n\to\infty} \sup_{\beta^0\in\mathcal{B}} \mathbb{P}_{\beta^0}(|\mathbf{Z}_n| > Mr_n) = 0.$$

We say that $\mathbf{Z}_n = o_{\mathbb{P}_{\beta^0}}(r_n)$ *uniformly in* $\beta^0 \in \mathcal{B}$ *if*

$$\lim_{n\to\infty} \sup_{\beta_0\in\mathcal{B}} \mathbb{P}_{\beta^0}\left(|\mathbf{Z}_n| > \eta r_n\right) = 0,\ \forall\, \eta > 0.$$

Notation

$$\gamma^0 := \arg \min_{c \in \mathbb{R}^{p-1}} \mathbb{E}(\mathbf{x}_1 - \mathbf{x}_{-1} c)^2.$$

Remark Thus

$$\gamma^0 = \Sigma_{-1,-1}^{-1} \mathbb{E}\mathbf{x}_{-1}^T \mathbf{x}_1$$

where $\Sigma_{-1,-1} := \mathbb{E}\mathbf{x}_{-1}^T \mathbf{x}_{-1}$, and $\mathbb{E}(\mathbf{x}_1 | \mathbf{x}_{-1}) = \mathbf{x}_{-1}\gamma^0$.
We say that $\mathbf{x}_{-1}\gamma^0$ is the projection of $\mathbf{x}_1$ on $\mathbf{x}_{-1}$.

The case Σ known

Definition *Let* $\gamma^\sharp \in \mathbb{R}^{p-1}$ *and* $\lambda^\sharp > 0$. *We say that the pair* $(\gamma^\sharp, \lambda^\sharp)$ *is eligible if*

$$\|\Sigma_{-1,-1}(\gamma^\sharp - \gamma^0)\|_\infty \leq \lambda^\sharp \tag{1}$$

and

$$\lambda^\sharp \|\gamma^\sharp\|_1 \to 0. \tag{2}$$

We then define

$$\Theta_1^\sharp := \begin{pmatrix} 1 \\ -\gamma^\sharp \end{pmatrix} / (1 - \gamma^{0T} \Sigma_{-1,-1} \gamma^\sharp). \tag{3}$$

Sample splitting

Assume the sample size n is even. Define the matrices

$$
\begin{aligned}
(X_I, Y_I) &:= \{X_{i,1}, \ldots, X_{i,p}, Y_i\}_{1 \le i \le n/2} \in \mathbb{R}^{n/2 \times (p+1)}, \\
(X_{II}, Y_{II}) &:= \{X_{i,1}, \ldots, X_{i,p}, Y_i\}_{n/2 < i \le n} \in \mathbb{R}^{n/2 \times (p+1)}.
\end{aligned}
$$

Let $\hat{\beta}_I$ be an estimator of β^0 based on the first half (X_I, Y_I) of the sample.
Let $\hat{\beta}_{II}$ be an estimator of β^0 based on the second half (X_{II}, Y_{II}) of the sample.
Define the two debiased estimators

$$
\begin{aligned}
\hat{b}^{\sharp}_{I,1} &:= \hat{\beta}_{II,1} + 2\Theta_1^{\sharp T} X_I^T \left(Y_I - X_I \hat{\beta}_{II} \right) / n \\
\hat{b}^{\sharp}_{II,1} &:= \hat{\beta}_{I,1} + 2\Theta_1^{\sharp T} X_{II}^T \left(Y_{II} - X_{II} \hat{\beta}_{I} \right) / n
\end{aligned}
$$

and let

$$
\hat{b}^{\sharp}_1 := \frac{\hat{b}^{\sharp}_{I,1} + \hat{b}^{\sharp}_{II,1}}{2}. \tag{4}
$$

Theorem Let $(\gamma^\sharp, \lambda^\sharp)$ be an eligible pair, Suppose that uniformly in $\beta^0 \in \mathcal{B}$

$$\|\Sigma^{1/2}(\hat{\beta}_I - \beta^0)\|_2 = o_{\mathbb{P}_{\beta^0}}(1),\ \|\Sigma^{1/2}(\hat{\beta}_{II} - \beta^0)\|_2 = o_{\mathbb{P}_{\beta^0}}(1) \quad (5)$$

and

$$\sqrt{n}\lambda^\sharp\|\hat{\beta}_I - \beta^0\|_1 = o_{\mathbb{P}_{\beta^0}}(1),\ \sqrt{n}\lambda^\sharp\|\hat{\beta}_{II} - \beta^0\|_1 = o_{\mathbb{P}_{\beta^0}}(1). \quad (6)$$

Then, uniformly in $\beta^0 \in \mathcal{B}$,

$$\hat{b}_1^\sharp - \beta_1^0 = \Theta_1^{\sharp T} X^T \epsilon/n + o_{\mathbb{P}_{\beta^0}}(1/\sqrt{n}),$$

and

$$\lim_{n\to\infty} \sup_{\beta^0\in\mathcal{B}} \mathbb{P}_{\beta^0}\left(\frac{\sqrt{n}(\hat{b}_1^\sharp - \beta_1^0)}{\Theta_{1,1}^\sharp} \leq z\right) = \Phi(z),\ \forall\ z \in \mathbb{R}.$$

Corollary The theorem shows that under its conditions the estimator $\hat{b}_1^\sharp$ is uniformly asymptotically linear and regular. It means that for this estimator the Cramér Rao lower bound is relevant.

Example

Suppose

$$\mathcal{B} = \{\|b\|_0^0 \le s\},\ s = o(n/\log p).$$

For the Lasso estimator $\hat\beta$ uniformly in $\beta^0 \in \mathcal{B}$

$$\|\Sigma^{1/2}(\hat\beta - \beta^0)\|_2^2 = \mathcal{O}_{\mathbb{P}_{\beta^0}}(s\log p/n),\ \|\hat\beta - \beta^0\|_1 = O_{\mathbb{P}_{\beta^0}}(s\sqrt{\log p/n}).$$

So a suitable requirement on $\lambda^\sharp$ is then

$$\lambda^\sharp s\sqrt{\log p} = o(1).$$

Example Suppose

$$\mathcal{B} = \{\|b\|_1 \le \sqrt{s}\},\ s = o(n/\log p).$$

Then uniformly in $\beta^0 \in \mathcal{B}$ it is true that

$$\|\Sigma^{1/2}(\hat\beta - \beta^0)\|_2 = o_{\mathbb{P}_{\beta^0}}(1),\ \|\hat\beta - \beta^0\|_1 = \mathcal{O}_{\mathbb{P}_{\beta^0}}(\sqrt{s}).$$

So we need

$$\lambda^\sharp\sqrt{ns} = o(1).$$

Then for $\|\gamma^\sharp\|_1 = \mathcal{O}(\sqrt{ns})$ we get an eligible pair $(\gamma^\sharp, \lambda^\sharp)$ and the Cramér Rao lower bound is achieved. In order to be able to improve over $\Theta_{1,1}$ we must have $\|\gamma^0\|_1$ of order larger than $\sqrt{ns}$.

Example Suppose

$$\mathcal{B} = \{\|b\|_r^r \le \sqrt{s^{2-r}}\},\ s = o(n/\log p).$$

For the Lasso $\hat\beta$, uniformly in $\beta^0 \in \mathcal{B}$

$$\|\Sigma^{1/2}(\hat\beta - \beta^0)\|_2^2 = \mathcal{O}_{\mathbb{P}_{\beta^0}}(s\log p/n)^{\frac{2-r}{2}},$$

$$\|\hat\beta - \beta^0\|_1 = \mathcal{O}_{\mathbb{P}_{\beta^0}}((\log p/n)^{\frac{1-r}{2}} s^{\frac{2-r}{2}}).$$

Thus we need

$$\lambda^\sharp(\log p)^{\frac{1-r}{2}}\sqrt{n^r s^{2-r}} = o(1).$$

If $\|\gamma^\sharp\|_r^r = \mathcal{O}(\sqrt{n^r s^{2-r}})$ then $(\gamma^\sharp, \lambda^\sharp)$ is an eligible pair and the Cramér Rao lower bound is achieved. In order to be able to improve over $\Theta_{1,1}$ we now need $\|\gamma^0\|_r^r$ of larger order $\sqrt{n^r s^{2-r}}$.

	Σ known $\mathcal{B} = \{\|b\|_0^0 \le s\}$	Σ known $\mathcal{B} = \{\|b\|_1 \le \sqrt{s}\}$
asymptotic normality	$s = o(\frac{n}{\log p})$ $\lambda^\sharp s \log^{\frac{1}{2}} p = o(1)$ $\lambda^\sharp \|\gamma^\sharp\|_1 = o(1)$	$s = o(\frac{n}{\log p})$ $\lambda^\sharp \sqrt{ns} = o(1)$ $\lambda^\sharp \|\gamma^\sharp\|_1 = o(1)$
asymptotic linearity	yes	yes
asymptotic efficiency	$\|\gamma^\sharp\|_0^0 = \mathcal{O}(s)$	$\|\gamma^\sharp\|_1 = \mathcal{O}(\sqrt{ns})$

Table: Throughout, $(\gamma^\sharp, \lambda^\sharp)$ is required to be an eligible pair, i.e. $\|\Sigma_{-1,-1}(\gamma^\sharp - \gamma^0)\|_\infty \le \lambda^\sharp$ (and $\lambda^\sharp \|\gamma^\sharp\|_1 \to 0$). Asymptotic efficiency is established when β^0 stays away from the boundary of $\mathcal{B}$. In the case $\mathcal{B} = \{\|b\|_0^0 \le s\}$ the conditions on $\gamma^\sharp$ for asymptotic efficiency depend on β^0.

	Σ known $\mathcal{B} = \{\|b\|_0^0 \le s\}$	Σ known $\mathcal{B} = \{\|b\|_r^r \le \sqrt{s^{2-r}}\}$
asymptotic normality	$s = o(\frac{n}{\log p})$ $\lambda^\sharp s \log^{\frac{1}{2}} p = o(1)$ $\lambda^\sharp \|\gamma^\sharp\|_1 = o(1)$	$s = o(\frac{n}{\log p})$ $\lambda^\sharp n^{\frac{r}{2}} s^{\frac{2-r}{2}} \log^{\frac{1-r}{2}} p = o(1)$ $\lambda^\sharp \|\gamma^\sharp\|_1 = o(1)$
asymptotic linearity	yes	yes
asymptotic efficiency	$\|\gamma^\sharp\|_0^0 = \mathcal{O}(s)$	$\|\gamma^\sharp\|_r^r = \mathcal{O}(n^{\frac{r}{2}} s^{\frac{2-r}{2}})$

Table: Throughout, $(\gamma^\sharp, \lambda^\sharp)$ is required to be an eligible pair i.e. $\|\Sigma_{-1,-1}(\gamma^\sharp - \gamma^0)\|_\infty \le \lambda^\sharp$ (and $\lambda^\sharp \|\gamma^\sharp\|_1 \to 0$). Asymptotic efficiency is established when β^0 stays away from the boundary of $\mathcal{B}$. In the case $\mathcal{B} = \{\|b\|_0^0 \le s\}$ the conditions on $\gamma^\sharp$ for asymptotic efficiency depend on β^0.

Finding eligible pairs

Remark The pair $(\gamma^\sharp, \lambda^\sharp)$ is eligible iff

$$\mathbf{x}_{-1}\gamma^0 = \mathbf{x}_{-1}\gamma^\sharp + \varepsilon^0,$$

where $|\mathrm{cov}(\mathbf{x}_j, \varepsilon^0)| \leq \lambda^\sharp$ and $\lambda^\sharp \|\gamma^\sharp\|_1 \to 0$.

Moreover, $\Theta_{1,1} \gg \Theta^\sharp_{1,1}$ iff $\mathbb{E}(\varepsilon^0)^2 \gg 0$.

Lemma *Suppose there exists a vector* $z \in \mathbb{R}^{p-1}$ *with*

$$\|z\|_\infty \leq 1,$$

and

$$1 - \lambda^{\sharp 2} \|\Sigma_{-1,-1}^{-1/2} z\|_2^2 \gg 0, \ \lambda^{\sharp 2} \|\Sigma_{-1,-1}^{-1/2} z\|_2^2 \gg 0.$$

Let $\gamma^\sharp$ *be a vector in* $\mathbb{R}^{p-1}$ *with* $\gamma^\sharp_{-S} = 0$ *(i.e.* $\gamma^\sharp = \gamma^\sharp_S$*) and*

$$1 - \lambda^{\sharp 2} \|\Sigma_{-1,-1}^{-1/2} z\|_2^2 - \|\Sigma_{-1,-1}^{1/2} \gamma^\sharp_S\|_2^2 \gg 0.$$

Define

$$\gamma^0 := \gamma^\sharp + \lambda^\sharp \Sigma_{-1,-1}^{-1} z.$$

Then, if $\lambda^\sharp \sqrt{|S|} \to 0$*, the pair* $(\gamma^\sharp, \lambda^\sharp)$ *is an eligible pair. Moreover,* γ^0 *is eventually allowed,* $\lambda^\sharp \|\gamma^0\|_1 \not\to 0$ *and in fact*

$$\Theta_{1,1} - \Theta^\sharp_{1,1} \gg 0.$$

The case Σ unknown

We use the noisy Lasso

$$\hat{\gamma} \in \arg \min_{c \in \mathbb{R}^{p-1}} \left\{ \|X_{-1} - X_{-1} c\|_2^2 / n + 2\lambda^{\text{Lasso}} \|c\|_1 \right\}. \tag{7}$$

Then in the debiased Lasso we apply

$$\tilde{\Theta}_1 := \hat{\Theta}_1$$

where

$$\hat{\Theta}_1 := \begin{pmatrix} 1 \\ -\hat{\gamma} \end{pmatrix} / (\|X_1 - X_{-1}\hat{\gamma}\|_2^2 / n + \lambda^{\text{Lasso}} \|\hat{\gamma}\|_1). \tag{8}$$

Theorem *Let $\hat{b}_1$ be the debiased Lasso*

$$\hat{b}_1 := \hat{\beta}_1 + \hat{\Theta}_1 X^T (Y - X\hat{\beta})/n.$$

Assume that uniformly in $\beta^0 \in \mathcal{B}$

$$\|\hat{\beta} - \beta^0\|_1 = o_{\mathbb{P}_{\beta^0}}(1/\sqrt{\log p}). \tag{9}$$

Let $(\gamma^\sharp, \lambda^\sharp)$ be an eligible pair, with $\lambda^\sharp = \mathcal{O}(\sqrt{\log p/n})$ and $\sqrt{\log p/n}\|\gamma^\sharp\|_1 = o(1)$. Then for suitable $\lambda^{\text{Lasso}} \asymp \sqrt{\log p/n}$ uniformly in $\beta^0 \in \mathcal{B}$

$$\hat{b}_1 - \beta_1^0 = \hat{\Theta}_1^T X^T \epsilon/n + o_{\mathbb{P}_{\beta^0}}(1/\sqrt{n}).$$

Moreover,

$$\lim_{n\to\infty} \sup_{\beta^0 \in \mathcal{B}} \mathbb{P}\left(\sqrt{n}\frac{(\hat{b}_1 - \beta_1^0)}{\sqrt{\hat{\Theta}_1^T \hat{\Sigma} \hat{\Theta}_1}} \le z\right) = \Phi(z)\ \forall\ z \in \mathbb{R},$$

and

$$\hat{\Theta}_1^T \hat{\Sigma} \hat{\Theta}_1 = \Theta_{1,1}^\sharp + o_{\mathbb{P}}(1).$$

Remark By Slutsky's Theorem we conclude that the asymptotic variance of $\hat{b}_1$ is (up to smaller order terms) equal to $\Theta^\sharp_{1,1}$.

Remark Assume that in fact

$$\|\hat{\gamma} - \gamma^\sharp\|_1 = o_{\mathbb{P}}(1/\sqrt{\log p}). \tag{10}$$

Then

$$\hat{\Theta}_1 X^T \epsilon / n = \Theta^\sharp_1 X^T \epsilon / n + o_{\mathbb{P}}(1/\sqrt{n}).$$

Thus, then the estimator $\hat{b}_1$ is asymptotically linear, uniformly in $\beta^0 \in \mathcal{B}$. The asymptotic linearity of $\hat{b}_1$ implies in turn that the Cramér Rao lower bound applies.

Example Suppose

$$\mathcal{B} = \{b \in \mathbb{R}^p : \ \|b\|_0^0 \leq s\}, \ s = o(\sqrt{n}/\log p).$$

For the Lasso estimator $\hat{\beta}$ with appropriate choice of the tuning parameter $\lambda \asymp \sqrt{\log p/n}$, one has uniformly in $\beta^0 \in \mathcal{B}$

$$\|\hat{\beta} - \beta^0\|_1 = O_{\mathbb{P}_{\beta^0}}\left(s\sqrt{\log p/n}\right) = o_{\mathbb{P}_{\beta^0}}(1/\sqrt{\log p}).$$

Suppose $\|\gamma^\sharp\|_0^0 = \mathcal{O}(s)$. Then

$$\|\hat{\gamma} - \gamma^\sharp\|_1 = O_{\mathbb{P}}\left(s\sqrt{\log p/n}\right) = o_{\mathbb{P}}(1/\sqrt{\log p}).$$

Thus then we have asymptotic linearity. It means that the Cramér Rao lower bound applies and is achieved.

Example Suppose

$$\mathcal{B} = \{b \in \mathbb{R}^p : \ \|b\|_r^r \leq \sqrt{s^{2-r}}\}, \ s = o(n^{\frac{1-r}{2-r}} / \log p).$$

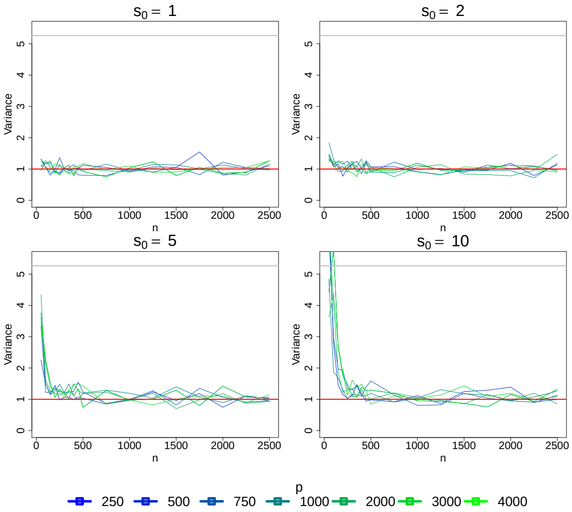
Then for the Lasso $\hat{\beta}$

$$\|\hat{\beta} - \beta^0\|_1 = o_{\mathbb{P}}((\log p/n)^{\frac{1-r}{2}} s^{\frac{2-r}{2}}) = o_{\mathbb{P}_{\beta^0}}(1/\sqrt{\log p}).$$

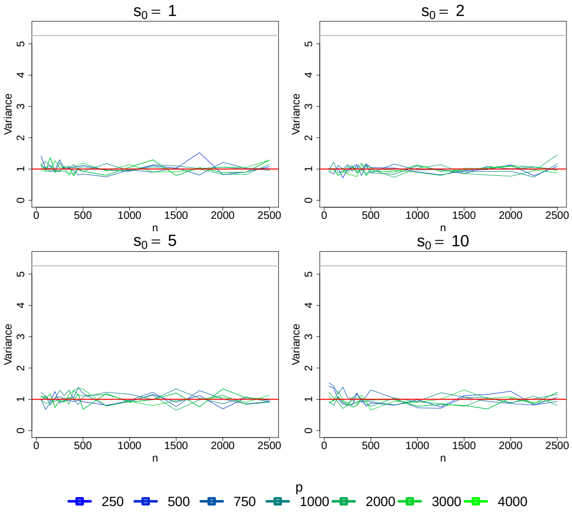
	Σ unknown $\mathcal{B} = \{\|b\|_0 \leq s\}$	Σ unknown $\mathcal{B} = \{\|b\|_r^r \leq \sqrt{s^{2-r}}\}$ $0 \leq r < 1$
asymptotic normality	$s = o(\frac{\sqrt{n}}{\log p})$ $\lambda^\sharp = \mathcal{O}(\sqrt{\frac{\log p}{n}})$ $\sqrt{\frac{\log p}{n}}\|\gamma^\sharp\|_1 = o(1)$	$s = o(n^{\frac{1-r}{2-r}} / \log p)$ $\lambda^\sharp = \mathcal{O}(\sqrt{\frac{\log p}{n}})$ $\sqrt{\frac{\log p}{n}}\|\gamma^\sharp\|_1 = o(1)$
asymptotic linearity	$\|\gamma^\sharp\|_r^r = o(n^{\frac{1-r}{2}} / \log^{\frac{2-r}{2}} p)$	$\|\gamma^\sharp\|_r^r = o(n^{\frac{1-r}{2}} / \log^{\frac{2-r}{2}} p)$
asymptotic efficiency	$\|\gamma^\sharp\|_0^0 = \mathcal{O}(s)$	$\|\gamma^\sharp\|_r^r = \mathcal{O}(n^{\frac{r}{2}} s^{\frac{2-r}{2}})$

Some simulations ...

Acknowledgement: Francesco Ortelli



$$\Theta_{1,1} = 5.26,\ \Theta^{\sharp}_{1,1} = 1,\ \ \beta^0_1 = 1$$



$$\Theta_{1,1} = 5.26,\ \Theta^{\sharp}_{1,1} = 1,\ \beta^0_1 = 0$$


THANK YOU!