

Recap

Inference in high-dimensional linear models: p-values and confidence intervals

one cannot use the bootstrap or subsampling for approximating the distribution of the Lasso

it is inconsistent due to non-Gaussian limiting distribution of the Lasso

if $a_n(\hat{\beta}_{\text{Lasso};j} - \beta_j^0) \implies Z, Z \sim \text{non-Gaussian distribution } F$
then $a_n(\underbrace{\hat{\beta}_{\text{Lasso};j}^*}_{\text{bootstrapped}} - \hat{\beta}_{\text{Lasso};j}) \not\Rightarrow Z \text{ in probability}$

De-biased or De-sparsified Lasso

consider (as in partialing-out):

$$Z^{(j)} = X^{(j)} - X^{(-j)}\hat{\gamma}^{(j)}$$

= Lasso residuals from $X^{(j)}$ vs. $X^{(-j)} = \{X^{(k)}; k \neq j\}$

$$\hat{\gamma}^{(j)} = \operatorname{argmin}_{\gamma} \|X^{(j)} - X^{(-j)}\gamma\|_2^2 + \lambda_j \|\gamma\|_1$$

build projection of Y onto $Z^{(j)}$:

$$\frac{Y^T Z^{(j)}}{(X^{(j)})^T Z^{(j)}} \underbrace{=}_{Y=X\beta^0+\varepsilon} \beta_j^0 + \underbrace{\sum_{k \neq j} \frac{(X^{(k)})^T Z^{(j)}}{(X^{(j)})^T Z^{(j)}} \beta_k^0}_{\text{bias}} + \frac{\varepsilon^T Z^{(j)}}{(X^{(j)})^T Z^{(j)}}$$

estimate bias and subtract it:

$$\widehat{\text{bias}} = \sum_{k \neq j} \frac{(X^{(k)})^T Z^{(j)}}{(X^{(j)})^T Z^{(j)}} \underbrace{\hat{\beta}_k}_{\text{standard Lasso}}$$

$\leadsto$ de-biased/de-sparsified Lasso estimator

$$\hat{b}_j = \frac{Y^T Z^{(j)}}{(X^{(j)})^T Z^{(j)}} - \sum_{k \neq j} \frac{(X^{(k)})^T Z^{(j)}}{(X^{(j)})^T Z^{(j)}} \hat{\beta}_k \quad (j = 1, \dots, p)$$

not sparse! Never equal to zero for all $j = 1, \dots, p$

computation: for computing all $\hat{b}_j$, $j = 1, \dots, p$

$\leadsto p + 1$ Lasso fits

i.e. $O(p \underbrace{np \min(n, p)}_{\text{comp. Compl. of Lasso}}) = O(p^2 n^2)$ comp. complexity

if $p \gg n$

Theorem 10.1 in the notes

assume:

- ▶ $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$
- ▶ $\lambda_j = C_j \sqrt{\log(p)/n}$ and $\|Z^{(j)}\|_2^2/n \geq L > 0$
- ▶ $s_0 = o(\sqrt{n}/\log(p))$ (a bit more sparse than “usual”)
- ▶ $\|\hat{\beta} - \beta^0\|_1 = O_P(s_0 \sqrt{\log(p)/n})$
(i.e., compatibility constant ϕ_o^2 bounded away from zero)

Then:

$$\sigma^{-1} \sqrt{n} \frac{(X^{(j)})^T Z^{(j)} / n}{\|Z^{(j)}\|_2 / \sqrt{n}} (\hat{b}_j - \beta_j^0) \implies \mathcal{N}(0, 1) \quad (j = 1, \dots, p)$$

plugging-in $\hat{\sigma}$: $\rightsquigarrow$ confidence intervals/hypothesis testing for β_j^0

more precisely (since ε is assumed to be Gaussian):

$$\sigma^{-1} \sqrt{n} \frac{(X^{(j)})^T Z^{(j)} / n}{\|Z^{(j)}\|_2 / \sqrt{n}} (\hat{b}_j - \beta_j^0) = W_j + \Delta_j$$

$$(W_1, \dots, W_p)^T \sim \mathcal{N}_p(0, \Omega), \quad \Omega_{jj} = 1$$

$$\max_{j=1, \dots, p} |\Delta_j| = o_P(1)$$

can test

$$H_{0,j} : \beta_j^0 = 0 \text{ versus } H_{A,j} : \beta_j^0 \neq 0$$

can also test group hypothesis: for $G \subseteq \{1, \dots, p\}$

$$H_{0,G} : \beta_j^0 \equiv 0 \forall j \in G$$

$$H_{A,G} : \exists j \in G \text{ such that } \beta_j^0 \neq 0$$

under $H_{0,G}$:

$$\max_{j \in G} \sigma^{-1} \sqrt{n} \frac{|(X^{(j)})^T Z^{(j)} / n|}{\|Z^{(j)}\|_2 / \sqrt{n}} |\hat{b}_j| = \max_{j \in G} |W_j + \Delta_j| \asymp \underbrace{\max_{j \in G} |W_j|}_{\text{distr. simulated}}$$

and plug-in $\hat{\sigma}$ for σ