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

$$\text{if} \quad a_n(\hat{\beta}_{\mathrm{Lasso};j} - \beta_j^0) \Longrightarrow Z, Z \sim \text{non-Gaussian distribution } F$$

$$\text{then} \quad a_n(\underbrace{\hat{\beta}^*_{\mathrm{Lasso};j}}_{\text{bootstrapped}} - \hat{\beta}_{\mathrm{Lasso},j}) \not\Longrightarrow Z \text{ in probability}$$

De-biased or De-sparsified Lasso

consider (as in partialing-out):

$$Z^{(j)} = X^{(j)} - X^{(-j)}\hat{\gamma}^{(j)}$$
$$= \text{Lasso residuals from } X^{(j)} \text{ vs. } X^{(-j)} = \{X^{(k)};\ k \neq j\}$$
$$\hat{\gamma}^{(j)} = \text{argmin}_{\gamma}\|X^{(j)} - X^{(-j)}\gamma\|_2^2 + \lambda_j\|\gamma\|_1$$

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:

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

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

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

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

computation: for computing all $\hat{b}_j,\ j = 1, \ldots, p$
$\leadsto$ $p + 1$ 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 $\|Z^{(j)}\|_2^2/n \ge L > 0$
- $s_0 = o(\sqrt{n}/\log(p))$ (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:

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

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