Recap

De-biased or De-sparsified Lasso

consider

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

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:

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

$\rightsquigarrow$ 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 $\hat{b}_j,\ j = 1, \ldots, p$
$\rightsquigarrow$ $p + 1$ Lasso fits
i.e. $O(pnp \min(n, p)) = O(p^2 n^2)$ 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:

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

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

more precisely:

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

$$(W_1, \ldots, W_p)^T \sim \mathcal{N}_p(0, \Omega),\ \Omega_{jj} \equiv 1\ \forall j,\ \max_{j=1,\ldots,p} |\Delta_j| = o_P(1)$$

test for group hypothesis: for $G \subseteq \{1, \ldots, p\}$

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

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

Why the $1/\sqrt{n}$ convergence rate?

de-biased/de-sparsified Lasso is considering

- low-dimensional components $\{\beta^0_j; j \in A\}$ with $|A|$ small

  $$\sum_{j \in A} c_j \sqrt{n}(\hat{b}_j - \beta^0_j) \Rightarrow \mathcal{N}(0, \sum_{j,j' \in A} c_j c_{j'} V_{j,j'}), V = \lim_n n\mathrm{Cov}(\hat{b})$$

  for large $|A|$: the sum would blow up the variance and the scaling with $\sqrt{n}$ is not correct
- high-dimensional β^0 and ℓ_∞-norm:

  $$\sqrt{n}\|\hat{b} - \beta^0\|_\infty \sim \underbrace{\text{maximum of } p \text{ dependent Gaussian r.v.'s}}_{\sim \underbrace{C\sqrt{\log(p)}}_{\text{under independence/weak dependence}}}$$

  $\leadsto \sqrt{\log(p)/n}$ convergence rate