High-dimensional inference

Consider the normal linear model $Y = X\beta^0 + \varepsilon$ (no intercept for simplicity), $\varepsilon \sim N_n(0, \sigma^2 I)$.

Low-dim setting: inference based on $\sqrt{n}(\hat\beta^{OLS} - \beta^0)$

$$\sqrt{n}(\hat\beta^{OLS} - \beta^0) \sim N_p\left(0, \sigma^2 \left(\tfrac{1}{n}X^TX\right)^{-1}\right)$$

High-dim setting: natural analog would be $\sqrt{n}(\hat\beta^L_\lambda - \beta^0)$ has an intractable distribution (in particular, the k-th component will place positive mass on $\sqrt{n}\beta^0_k$)

The recently introduced debiased Lasso (Zhang & Zhang 2014; van de Geer et al. 2014) overcomes this problem.

Let $\hat\beta$ be the Lasso soln at fixed $\lambda > 0$. Recall the KKT conditions

$$\tfrac{1}{n}X^T(Y - X\hat\beta) = \lambda\hat\nu \quad \text{where } \|\hat\nu\|_\infty \le 1 \text{ and writing } \hat S = \{k : \hat\beta_k \ne 0\},$$
$$\hat\nu_{\hat S} = \mathrm{sgn}(\hat\beta_{\hat S}).$$

Set $\hat\Sigma = \frac{1}{n}X^TX$. Have

$$\hat\Sigma(\hat\beta - \beta^0) + \lambda\hat\nu = \tfrac{1}{n}X^T\varepsilon$$

Key idea: use an approximate inverse $\hat\Theta$ of $\hat\Sigma$ (to be specified). Have

$$\sqrt{n}(\underbrace{\hat\beta + \lambda\hat\Theta\hat\nu}_{\hat b} - \beta^0) = \frac{1}{\sqrt{n}}\hat\Theta X^T\varepsilon + \Delta \quad \text{where } \Delta = (\hat\Theta\hat\Sigma - I)(\beta^0 - \hat\beta)\sqrt{n}$$

$\hat b = \hat\beta + \frac{1}{n}\hat\Theta X^T(Y - X\hat\beta)$ is the debiased Lasso.

If we choose $\hat\Theta$ s.t. $\|\Delta\|_\infty$ is small we will have

$$\sqrt{n}(\hat b - \beta^0) \approx \tfrac{1}{\sqrt{n}}\hat\Theta X^T\varepsilon$$

By theorem 23, under a compatibility condition and when β^0 is sparse, we know that $\|\beta^0 - \hat\beta\|_1$ is small with high probability. We thus aim to show that the rows of $\hat\Theta\hat\Sigma - I$ have small ℓ_∞-norm, so we can apply Hölder's inequality.

Write $\hat\theta_j$ for the jth row of $\hat\Theta$. Then $\|(\hat\Sigma\hat\Theta^T - I)_j\|_\infty \le \eta$

$\iff |(\hat\Sigma\hat\Theta^T)_{kj}| \le \eta \quad \forall k \ne j$ and $|(\hat\Sigma\hat\Theta^T)_{jj} - 1| \le \eta$

$\iff |(\hat\Sigma_k)^T\hat\theta_j| \le \eta \quad \forall k \ne j \quad \ldots$

$\iff \frac{1}{n}|X_k^TX\hat\theta_j| \le \eta \quad \forall k \ne j$ and $\frac{1}{n}|X_j^TX\hat\theta_j - 1| \le \eta$

$\Leftarrow \frac{1}{n}\|X_{-j}^TX\hat\theta_j\|_\infty \le \eta \quad \ldots$

Let $$\hat\gamma^{(j)} = \operatorname*{arg\,min}_{\gamma \in \mathbb{R}^{p-1}} \left\{\tfrac{1}{2n}\|X_j - X_{-j}\gamma\|_2^2 + \lambda_j\|\gamma\|_1\right\} \quad -(*)$$

Set $$\hat\theta_j = \frac{1}{\hat\tau_j^2}\left(-\hat\gamma_1^{(j)}, -\hat\gamma_2^{(j)}, \ldots, -\hat\gamma_{j-1}^{(j)}, 1, -\hat\gamma_{j+1}^{(j)}, \ldots, -\hat\gamma_p^{(j)}\right)$$

where $\hat\tau_j^2 = \frac{1}{n}X_j^T(X - X_{-j}\hat\gamma^{(j)}) \underset{\text{exercise}}{=} \frac{1}{n}\|X_j - X_{-j}\hat\gamma^{(j)}\|_2^2 + \lambda_j\|\hat\gamma^{(j)}\|_1$

Thus $\frac{1}{n} X_j^T X \hat{\theta}_j = 1$ and by KKT conditions for $(*)$

$$\hat{\varsigma}_j^2 \| X_{-j}^T X \hat{\theta}_j \|_\infty \leq \lambda_j .$$

With this $\hat{\Theta}$ we have (by Hölder)

$$\| \Delta \|_\infty \leq \sqrt{n} \, \| \beta^0 - \hat{\beta} \|_1 \max_j \frac{\lambda_j}{\hat{\varsigma}_j^2}$$

Consider a random design setting where the rows of X are independent and $N_p(0, \Sigma)$ with Σ positive definite. Write $\Omega = \Sigma^{-1}$. From prop 30, we have

$$X_j = X_{-j} \gamma^{(j)} + \varepsilon^{(j)} \quad \text{where} \quad \varepsilon_i^{(j)} | X_{-j} \overset{iid}{\sim} N(0, \Omega_{jj}^{-1})$$ ← see sec on Schur complements

$$\gamma^{(j)} = - \Omega_{jj}^{-1} \Omega_{-j,j} .$$

Let $S = \{ k : \beta_k^0 \neq 0 \}$, $s = |S|$.

$$s_j = \sum_{k \neq j} \mathbb{1}_{\{\Omega_{kj} \neq 0\}} \qquad (\text{sparsity of } \gamma^{(j)})$$

$$s_{max} = \max (s, \max_j s_j)$$