A very Naive Approach to High-dim Inference (~~Sen~~ Zhao, Witten, Shojaie) 2017.

idea: selection by LASSO is "deterministic" with $p=1$. $\Rightarrow$ data used only once. [selection consistency]

- marginal (& full model) hypothesis testing. (paper of Shi & Qu (2017) falls into this category. (debiased estimator ...)
- submodel inference. (condition on / uniform for selected models). (double dipping).

driven phenomenon: coverage based on $\tilde{\beta}^{(\hat{A}_\lambda)} = (X_{\hat{A}_\lambda}^T X_{\hat{A}_\lambda})^{-1} X_{\hat{A}_\lambda}^T y$ is not too bad.

Main results: Define noiseless LASSO: $\beta_\lambda = \underset{b}{\arg\min} \left\{ \frac{1}{2n} E\|y - Xb\|_2^2 + \lambda \|b\|_1 \right\}$

① $\lim_{n\to\infty} P(\hat{A}_\lambda = A_\lambda) = 1$ under conditions

- (M1) $\sqrt{\frac{\log p}{n}} = o(\lambda)$
- (E) some restriction on eigenvalues of $\hat{\Sigma}$
- (M2) $\lim_{n\to\infty} \frac{\min_{j\in A_\lambda} |\beta_{\lambda,j}|}{\lambda} = \xi$ slow decrease in "weak signal"
  - $\|X_{A_\lambda^c} \beta^*_{A_\lambda^c}\|_2 = O(\sqrt{\log p})$ slow increase in "noise"
- (T) irrepresentable condition on β_λ
  - ↑ selection consistency in this context.
  - $P(A_\lambda \subset \hat{A}_\lambda)$
  - $P(\hat{A}_\lambda \subset A_\lambda)$
- ~~(A)~~ (W). regularity condition for Lindeberg-Feller
- $\frac{\log p}{n - q_\lambda} \to 0$

additionally, λ is not data-driven

② $\frac{\tilde{\beta}_j^{(\hat{A}_\lambda)} - \beta_j^{(\hat{A}_\lambda)}}{\sigma_\varepsilon \sqrt{(X_{\hat{A}_\lambda}^T X_{\hat{A}_\lambda})^{-1}_{jj}}} \xrightarrow{d} N(0,1)$

$\tilde{\beta}_j(\hat{A}_\lambda) = (X_{\hat{A}_\lambda}^T X_{\hat{A}_\lambda})^{-1} X_{\hat{A}_\lambda}^T y$

③ $\frac{1}{n - \hat{q}_\lambda} \|y - X_{\hat{A}_\lambda} \tilde{\beta}^{(\hat{A}_\lambda)}\|_2^2 \xrightarrow{P} \sigma_\varepsilon^2$

$\hat{q}_\lambda = |\hat{A}_\lambda|$

things to notice about simulation part 1:

① data-driven selected λ_{1SE} is worse than data independent λ_{sup}.

② the ~~noise~~ information ratio defined is $\frac{\beta^{*T}\Sigma\beta^*}{\sigma_\varepsilon^2}$, regarding overall information, not weakest signal.

③ the convergence analysis of $\hat{A}_\lambda^b$ is based on "most common" $\hat{A}_\lambda^b$ (? inclusion of weak signals)

④ it is compared with the conditional ~~sele~~ selective inference method.

Naive (marginal) Score test. (based on $\hat{A}_\lambda$)

Define: $S^j = X_j^T(y - \tilde{y}^{(\hat{A}_\lambda \setminus \{j\})}) = x_j^T (I_n - P^{-j}) y$ where $\tilde{y}^{(\hat{A}_\lambda\setminus\{j\})} \equiv X_{\hat{A}_\lambda\setminus\{j\}} \tilde{\beta}^{(\hat{A}_\lambda\setminus\{j\})}$

$P^{-j} \equiv X_{\hat{A}_\lambda\setminus\{j\}} (X_{\hat{A}_\lambda\setminus\{j\}}^T X_{\hat{A}_\lambda\setminus\{j\}})^{-1} X_{\hat{A}_\lambda\setminus\{j\}}^T$

$T_j \equiv \frac{S_j}{\sigma_\varepsilon \sqrt{x_j^T (I - P^{-j}) x_j}}$.

Thm: $T_j \xrightarrow{d} N(0,1)$, $\forall j = 1, \dots, p$. under condition: all above (s.t. $P(\hat{A}_\lambda = A_\lambda) \to 1$) but replace

$M_2 \to M_2^*$: $\|X_{A_\lambda^c} \beta^*_{A_\lambda^c}\|_2 = o(1)$.

$W \to S$: regularity conditions.

Simulation part 2: poor power for weak signals.

(problem with part 1: selection & coverage for really weak signals?)

goal of PoSI? convince people about the model?