A power and prediction analysis for knockoffs with Lasso statistics

By Weinstein, Barber, Candes

Mengqi Lin

Apr 29, 2021

Berkeley
UNIVERSITY OF CALIFORNIA

Overview

1. Recall the powerful theorem we learned in class.
2. Introduce the knockoff procedure which is aimed for dealing with the Lasso statistics in practice.
3. Show you how the result in 1 can be applied in 2.

Recall the beautiful result we learned in class

Under our ”old friend” setting : Consider linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}$$

where $\mathbf{X} \in \mathbb{R}^{n\times p},\ X_{ij} \sim_{i.i.d.} \mathcal{N}(0, 1/n),\ \ \mathbf{z} \in \mathbb{R}^n,\ z_i \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$,
$\boldsymbol{\beta} \in \mathbb{R}^p,\ \ \beta_j \sim_{i.i.d.} \Pi$, where $\mathbb{P}(\Pi \neq 0) = \epsilon,\ \ \mathbb{P}(\Pi = 0) = 1 - \epsilon$
$n/p \to \delta,\ \ n \to \infty.$
So the expected number of nonzero (nonnull) is ϵp.

The lasso estimator :

$$\hat{\boldsymbol{\beta}}(\lambda) = \underset{\mathbf{b}\in\mathbb{R}^p}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 + \lambda\|\mathbf{b}\|_1$$

Let $V(\lambda) = |\{j : \hat{\beta}_j(\lambda) \neq 0, \beta_j = 0\}|$, $T(\lambda) = |\{j : \hat{\beta}_j(\lambda) \neq 0, \beta_j \neq 0\}|$, $R(\lambda) = |\{j : \hat{\beta}_j(\lambda) \neq 0\}| = V(\lambda) + T(\lambda)$, $k = |\{j : \beta_j \neq 0\}| = \epsilon p$

Lemma (Lec11, 12, AMP)

$$\frac{V(\lambda)}{p} \xrightarrow{\mathbb{P}} 2(1-\epsilon)\Phi(-\alpha),$$
$$\frac{T(\lambda)}{p} \xrightarrow{\mathbb{P}} \mathbb{P}(|\Pi + \tau W| > \alpha\tau, \Pi \neq 0) = \epsilon\mathbb{P}(|\Pi^* + \tau W| > \alpha\tau),$$

where $W \sim \mathcal{N}(0,1)$ is indep of Π, and $\tau,\ \alpha > \max\{\alpha_0, 0\}$ is the unique solution to

$$\begin{aligned} \tau^2 &= \sigma^2 + \frac{1}{\delta}\mathbb{E}(\eta_{\alpha\tau}(\Pi + \tau W) - \Pi)^2 \\ \lambda &= \left(1 - \frac{1}{\delta}\mathbb{P}(|\Pi + \tau W| > \alpha\tau)\right)\alpha\tau. \end{aligned} \tag{1}$$

Berkeley UNIVERSITY OF CALIFORNIA

It follows immediately from Lemma 1 that for a fixed $\lambda > 0$, the limits of FDP and TPP are

$$\mathsf{FDP}(\lambda) = \frac{V(\lambda)}{1 \vee R(\lambda)} \xrightarrow{\mathbb{P}} \frac{2(1-\epsilon)\Phi(-\alpha)}{2(1-\epsilon)\Phi(-\alpha) + \epsilon\mathbb{P}(|\Pi^* + \tau W| > \alpha\tau)} \tag{2}$$

and

$$\mathsf{TPP}(\lambda) = \frac{T(\lambda)}{1 \vee k} \xrightarrow{\mathbb{P}} \mathbb{P}(|\Pi^* + \tau W| > \alpha\tau) \tag{3}$$

philosophy question

We don't know ϵ and Π^* in practice !

In practice

How do scientists decide the threshold λ for the Lasso under FDR-TPP criteria?

Knockoff procedure (Candes, Barber, 2015)

Idea: For each λ, propose an estimator for FDP(λ): $\widehat{\text{FDP}}$ (possibly overestimate FDP), and greedily choose λ so that the procedure stops when $\widehat{\text{FDP}} < q$. But to estimate FDP, we need to estimate $V(\lambda)$, which is unknown, and this is where the knockoff variable plays the role.

In this section, we work on fixed p, n. And augment the matrix $\mathbf{X}$ with $\widetilde{\mathbf{X}} \in \mathbb{R}^{n\times r}$

$$\mathbb{X} := [\mathbf{X}\,\widetilde{\mathbf{X}}] \in \mathbb{R}^{n\times(p+r)}$$

where the whole $\mathbb{X}_{ij} \sim_{i.i.d.} \mathcal{N}(0, 1/n)$, and the corresponding Lasso estimator

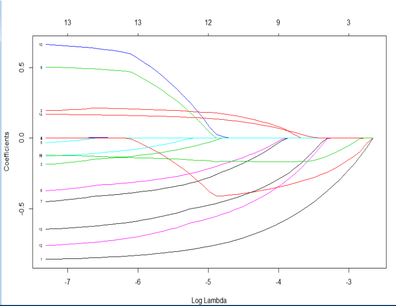
$$\widehat{\boldsymbol{\beta}}(\lambda) = \underset{\mathbf{b}\in\mathbb{R}^{p+r}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y} - \mathbb{X}\mathbf{b}\|^2 + \lambda\|\mathbf{b}\|_1. \tag{4}$$

Denote sets:

$$\mathcal{H} = \{1, ..., p\}, \quad \mathcal{H}_0 = \{j \in \mathcal{H} : \beta_j = 0\}, \quad \mathcal{K}_0 = \{p+1, ..., p+r\}$$

To effectively decide λ, we work on the sufficient statistics for Lasso

$$T_j = \sup\{\lambda : \widehat{\beta}_j(\lambda) \neq 0\}, \;\; j = 1, ..., p+r. \tag{5}$$



Let

$$V_0(\lambda) = |\{j \in \mathcal{H}_0 : T_j \geq \lambda\}|,\ V_1(\lambda) = |\{j \in \mathcal{K}_0 : T_j \geq \lambda\}|,\ R(\lambda) = |\{j \in \mathcal{H} : T_j \geq \lambda\}|$$

$$\widehat{\mathrm{FDP}}(\lambda) = \frac{(1 + V_1(\lambda)) \cdot \frac{|\mathcal{H}|\widehat{\pi}_0}{1+|\mathcal{K}_0|}}{R(\lambda)}$$

And the procedure finds the greedy λ s.t $\widehat{\mathrm{FDP}}(\lambda) \leq q$, i.e

$$\widehat{\lambda} = \inf\left\{\lambda \in \Lambda : \widehat{\mathrm{FDP}}(\lambda) \leq q\right\}$$

and reject $j \in \mathcal{H}$, if $T_j \geq \widehat{\lambda}$.

Berkeley
UNIVERSITY OF CALIFORNIA

Optimality of knockoff procedure

Let's take $r = \rho p$ for a constant $\rho > 0$, we have that $n/(p+r) \to \delta' := \delta/(1+\rho)$, $\epsilon' := \epsilon/(1+\rho)$ as $n, p \to \infty$. Notice that our linear model has become:

$$\mathbb{E}\mathbf{y} = \mathbb{X}\begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{o} \end{bmatrix}$$

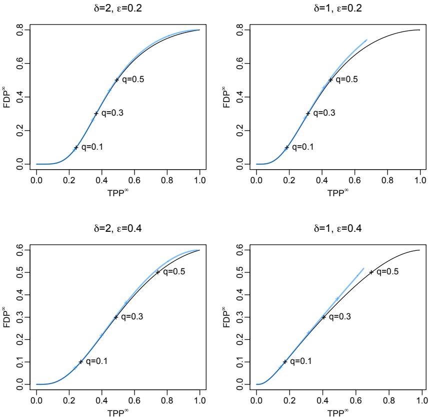
As the similar analysis in section 1, we can derive the oracle FDP versus TPP for each λ in this augment setting with a slight modification. With that being said, when knowing $\epsilon,\ \Pi^*$, we already know that there's an oracle "procedure".

$$\mathrm{FDP}_{\mathrm{aug}}(\lambda) \longrightarrow \frac{2(1-\epsilon)\Phi(-\alpha')}{2(1-\epsilon)\Phi(-\alpha') + \epsilon\mathbb{P}(|\Pi^* + \tau' W| > \alpha'\tau')}$$

$$\mathrm{TPP}_{\mathrm{aug}}(\lambda) \longrightarrow \mathbb{P}(|\Pi^* + \tau' W| > \alpha'\tau')$$

Berkeley
UNIVERSITY OF CALIFORNIA

simulation figure



Berkeley
UNIVERSITY OF CALIFORNIA