LASSO in sub-samples

November 27, 2020

1 Sub-samples and Weighted Least Square

1.1 Notations

Let $y_i = \boldsymbol{x}_i^{'}\boldsymbol{\beta}+\epsilon_i$, where $\epsilon_1, ..., \epsilon_n$ are i.i.d. with mean 0 and variance σ^2. Denote the corresponding sub-sampling probabilities for each observation as $\pi_1, ..., \pi_n$.

To save notations:

$$\begin{aligned}
\boldsymbol{X} &= (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)' \\
\boldsymbol{y} &= (y_1, ..., y_n)' \\
\boldsymbol{\epsilon} &= (\epsilon_1, ..., y_n)' \\
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}
\end{aligned}$$

Further, define the weight matrix as:

$$\boldsymbol{W} = \begin{bmatrix} \frac{1}{\pi_1} & & \\ & \ddots & \\ & & \frac{1}{\pi_n} \end{bmatrix} = \begin{bmatrix} w_1 & & \\ & \ddots & \\ & & w_n \end{bmatrix}$$

Follow the notations in Knight and Fu (2000), the l1-penalized weighted least squares (WLS) criterion:

$$\sum_{i=1}^{n} w_i(y_i - \boldsymbol{x}_i^{'}\boldsymbol{\Phi})^2 + \lambda_n \sum_{j=1}^{p} |\phi_j| \tag{1}$$

For a given λ_n, the estimator minimizing (1) by $\hat{\boldsymbol{\beta}}_n$. When $\lambda_n = 0$, the WLS estimator is $\hat{\boldsymbol{\beta}}_n^{(0)}$.

1.2 Asymptotic Properties for WLS

Denote

$$C_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i\boldsymbol{x}_i^{'}}{\pi_i} = \frac{\boldsymbol{X}^{'}\boldsymbol{W}\boldsymbol{X}}{n}$$

and assume $C_n \to C$, which is n.n.d. Also denote

$$D_n = \frac{1}{n}\sum_{i=1}^{n} \frac{\boldsymbol{x}_i \boldsymbol{x}_i^{'}}{\pi_i^2} = \frac{\boldsymbol{X}^{'}\boldsymbol{W^2}\boldsymbol{X}}{n}$$

and assume $D_n \to D$, which is also n.n.d. Note that for uniform sub-samples, $\boldsymbol{W} = \boldsymbol{I}$ and therefore $C_n = D_n$ and $C = D$

By these assumptions and notations, we can derive the asymptotic distribution for $\hat{\boldsymbol{\beta}}_n^{(0)}$ as follows:

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(0)} - \boldsymbol{\beta}) = \sqrt{n}(\boldsymbol{X}^{'}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^{'}\boldsymbol{W}\boldsymbol{\epsilon} = (\frac{\boldsymbol{X}^{'}\boldsymbol{W}\boldsymbol{X}}{n})^{-1}\frac{\boldsymbol{X}^{'}\boldsymbol{W}\boldsymbol{\epsilon}}{\sqrt{n}} = C_n^{-1}\frac{\boldsymbol{X}^{'}\boldsymbol{W}\boldsymbol{\epsilon}}{\sqrt{n}}$$

Since

$$C_n \to C$$
$$D_n \to D$$
$$\frac{\boldsymbol{X}^{'}\boldsymbol{W}\boldsymbol{\epsilon}}{\sqrt{n}} = \sqrt{n}\frac{\boldsymbol{X}^{'}\boldsymbol{W}\boldsymbol{\epsilon}}{n} \xrightarrow{d} N(0, \sigma^2 D)$$

Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(0)} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2 C^{-1} D C^{-1})$$

2 Asymptotic Analysis for LASSO

Follow the rationale of Knight and Fu (2000), the limiting distributions for WLS can be easily derived. Define te random function

$$Z_n(\boldsymbol{\Phi}) = \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\Phi})^{'}\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\Phi})}{n} + \frac{\lambda_n}{n}\sum_{j=1}^{p}|\phi_j|$$

Then we can get the Theorem 1.

Theorem 1 *If $\lambda_n/n \to \lambda_0 \geq 0$, then $\hat{\boldsymbol{\beta}}_n \xrightarrow{p} argmin(Z)$ where*

$$Z(\Phi) = (\boldsymbol{\Phi} - \boldsymbol{\beta})'C(\boldsymbol{\Phi} - \boldsymbol{\beta}) + \lambda_0 \sum_{j=1}^{p}|\phi_j|$$

Further, define another random function

$$\begin{aligned} V_n(\boldsymbol{u}) &= \sum_{i=1}^{n}\{(\frac{\epsilon_i}{\sqrt{\pi_i}} - \frac{\boldsymbol{u}^{'}\boldsymbol{x}_i}{\sqrt{n\pi_i}})^2 - (\frac{\epsilon_i}{\sqrt{\pi_i}})^2\} + \lambda_n \sum_{j=1}^{p}\{|\beta_j + u_j/\sqrt{n}| - |\beta_j|\} \\ &= -2\frac{\boldsymbol{u}'\boldsymbol{X}'\boldsymbol{W}\boldsymbol{\epsilon}}{\sqrt{n}} + \frac{\boldsymbol{u}'\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\boldsymbol{u}}{n} + \lambda_n \sum_{j=1}^{p}\{|\beta_j + u_j/\sqrt{n}| - |\beta_j|\} \end{aligned}$$

and notice that

$$-2\frac{\boldsymbol{u}'\boldsymbol{X}'\boldsymbol{W}\boldsymbol{\epsilon}}{\sqrt{n}} + \frac{\boldsymbol{u}'\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\boldsymbol{u}}{n} \xrightarrow{d} -2\boldsymbol{u}'\boldsymbol{M} + \boldsymbol{u}'C\boldsymbol{u}$$

where $\boldsymbol{M} \sim N(\boldsymbol{0}, \sigma^2 D)$.Then we can get the Theorem 2.

Theorem 2 *If $\lambda_n/\sqrt{n} \to \lambda_0 \geq 0$, then*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} argmin(V)$$

where

$$V(\boldsymbol{u}) = -2\boldsymbol{u}'\boldsymbol{M} + \boldsymbol{u}'C\boldsymbol{u} + \lambda_0 \sum_{j=1}^{p} \{u_j sin(\beta_j) I(\beta_j \neq 0) + |u_j| I(\beta_j = 0)\}$$

and $\boldsymbol{M} \sim N(\boldsymbol{0}, \sigma^2 D)$