**HIGH-DIMENSIONAL LINEAR MODEL SELECTION VIA MULTIPLE TESTING**

Konrad Furmańczyk, Wojciech Rejchel

Waraw University of Life Sciences, Institute of Information Technology, Warsaw, Poland, email: konfur@wp.pl

Nicolaus Copernicus University, Torun, Poland, email: w.rejchel@gmail.com

Mathematical Methods of Modern Statistics 2, Limniny 15-19/6/2020

**Introduction** We propose a new model selection algorithm (called „LassoSD”), which combines a stepdown multiple testing approach with penalized minimization. First, we use the Lasso to discard a substantial part of irrelevant predictors. In the next step of the algorithm we consider only the support of the Lasso and select the final model in a way that is motivated by the stepdown multiple testing procedure. We propose to choose thresholds \(t_j\) for predictors, that are smaller in absolute values than the prespecified level \(\gamma\). The Lasso is constructed as follows: if \(t_j < \gamma\), then \(I\) is the empty set. Otherwise, we look for the largest \(\gamma\) satisfying \(t_j \geq \gamma_j\), \(\ldots\), \(t_s \geq \gamma_s\) and \(I\) consists of \(\gamma\)-predictors, which correspond to ordered statistics \(t_j\), \(t_s\), \(\ldots\), \(t_0\).

**Motivated by Multiple Testing**

The third step of the algorithm is motivated by the stepdown multiple testing procedure. We propose to choose thresholds \(\gamma\) for predictors that have a smaller or equal absolute values than the prespecified level \(\gamma\) in the following two ways: for a (small) number \(\alpha > 0\) we consider: the procedure with equal thresholds

\[
\gamma = \sqrt{2 \log \left(\frac{p}{s}\right)} - 2 \log \alpha ,
\]

the procedure with decreasing thresholds

\[
\gamma_j = \sqrt{2 \log \left(\frac{p + 1 - j}{s + 1 - j}\right)} - 2 \log \alpha ,
\]

where \(j = 1, \ldots, s\).

**SCIF (The Sign-Restricted Cone Invertibility Factor)**

For \(x > 0\) and the set \(I\) we define two cones

\[
\mathcal{C}(I) = \{\beta \in \mathbb{R}^p : \|\beta\|_1 \leq \|\beta_{I}\|_1\}, \quad \beta_{I} = \hat{X}_I \hat{X}^{-1}_I \hat{Y}.
\]

and

\[
\mathcal{C}(I, \xi) = \{\beta \in \mathbb{R}^p : \|\beta_{I}\|_1 \leq \xi \|\beta_{I}\|_1\},
\]

where \(I\) is the complement of \(I\). In the case when \(p > n\) three different conditions measuring the degree of correlations between predictors are often used (SCIF),

\[
F(I, \xi) = \inf_{\beta \in \mathcal{C}(I, \xi)} \mathbb{E} \left(\left|\frac{\hat{X}_I \hat{X}^{-1}_I \hat{Y}}{\|\beta_I\|_1}\right|\right),
\]

with \(q > 1\). Note that \(m/J = \|\hat{X}_I \hat{X}^{-1}_I \hat{Y}\|_1\), where \(j \in J\) and \(m_J = \min_{j \in J} \|\beta_j\|_1\).

**Main Result-Theorem** Let \(a \in (0, 1), b > 0, \xi > 1\) be arbitrary. Consider the Lasso SD algorithm with the parameter \(\lambda\) satisfying

\[
\frac{\sigma(\xi)}{\sqrt{b(\xi - 1)}} \leq \lambda \leq \frac{\sigma(\xi + 1)}{\sqrt{b(\xi - 1)}}
\]

\[
\delta = \alpha, \quad \alpha = \left(\frac{1}{2} + \frac{1}{\sqrt{b(\xi - 1)}}\right)
\]

Moreover, suppose that \(K = K_{X}^1 = K_{\hat{X}}^1\),

\[
\frac{(\alpha \sigma)^2}{\alpha^2 m^2} = \frac{16 \log(K)}{N^2},
\]

where

\[
\alpha = \max_{k = 1, \ldots, K_{X}^1} \max_{j \in \mathcal{C}_k} \max_{j \in \mathcal{C}_k} m_{j}\left(I_k \cup J\right).
\]

Then we obtain that

\[
P(\hat{I} \neq I) \leq 2e^{-\exp \left(\frac{(1 - a)(1 - \sqrt{2/\alpha})}{2\sqrt{b(\xi - 1)}}\right)} + a(K - p_0 + 1)\left(\frac{p}{K - p_0 + 1}\right)^{p_0} + K - p_0
\]

**Conclusions** In Furmańczyk and Rejchel, W. (2020) we propose the new algorithm for high-dimensional linear models. It adapts multiple hypothesis testing to the model selection problem. We prove that this procedure is able to find the true model under mild conditions. The experiments confirm that LassoSD performs well in model selection. In fact, it works similarly or better than competitive algorithms based on multsplitting or minimization of information criteria.

**References**

