Efficient High-dimensional Robust Variable Selection via Rank-based LASSO Methods

A. Bhattacharjee* R. Mondal* R. Vasishtha* S. S. Banerjee*

*Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

April 9, 2022

Contents

Motivation

- Variable selection is a ubiquitous problem while dealing with high-dimensional data, for example gene microarray data.
- Many models make stringent assumptions on the error distribution or the existence of moments - robust methods are required!
- LASSO - very popular variable selection methodology.
- We present fast Rank-based LASSO methods for variable selection that do not make the stringent assumptions and work under high-dimensional settings and multicollinearity.
- **GOAL:** We aim to identify the set of relevant predictors T:

$$T = \{1 \leq j \leq p : \beta_j \neq 0\}.$$

Motivation

- Variable selection is a ubiquitous problem while dealing with high-dimensional data, for example gene microarray data.
- Many models make stringent assumptions on the error distribution or the existence of moments - robust methods are required!
- LASSO - very popular variable selection methodology.
- We present fast Rank-based LASSO methods for variable selection that do not make the stringent assumptions and work under high-dimensional settings and multicollinearity.
- **GOAL:** We aim to identify the set of relevant predictors T:

$$T = \{1 \leq j \leq p : \beta_j \neq 0\}.$$

Motivation

- Variable selection is a ubiquitous problem while dealing with high-dimensional data, for example gene microarray data.
- Many models make stringent assumptions on the error distribution or the existence of moments - robust methods are required!
- LASSO - very popular variable selection methodology.
- We present fast Rank-based LASSO methods for variable selection that do not make the stringent assumptions and work under high-dimensional settings and multicollinearity.
- **GOAL:** We aim to identify the set of relevant predictors T:

$$T = \{1 \leq j \leq p : \beta_j \neq 0\}.$$

Motivation

- Variable selection is a ubiquitous problem while dealing with high-dimensional data, for example gene microarray data.
- Many models make stringent assumptions on the error distribution or the existence of moments - robust methods are required!
- LASSO - very popular variable selection methodology.
- We present fast Rank-based LASSO methods for variable selection that do not make the stringent assumptions and work under high-dimensional settings and multicollinearity.
- **GOAL:** We aim to identify the set of relevant predictors T:

$$T = \{1 \leq j \leq p : \beta_j \neq 0\}.$$

Model

- We consider the model as:

$$Y_i = g(\beta' X_i, \varepsilon_i), \qquad i = 1, \ldots, n. \tag{1}$$

- β is a p-dimensional vector.
- $g(.)$ is an unknown monotonic link function. The covariates influence the response through the link function $g(.)$ of the scalar product $\beta' X_i$.
- No assumptions are made on the form of the link function g or the distribution of the error ε_i.

Model

- We consider the model as:

$$Y_i = g(\beta' X_i, \varepsilon_i), \qquad i = 1, \ldots, n. \tag{1}$$

- β is a p-dimensional vector.
- $g(.)$ is an unknown monotonic link function. The covariates influence the response through the link function $g(.)$ of the scalar product $\beta' X_i$.
- No assumptions are made on the form of the link function g or the distribution of the error ε_i.

Model

- We consider the model as:

$$Y_i = g(\beta' X_i, \varepsilon_i), \qquad i = 1, \ldots, n. \tag{1}$$

- β is a p-dimensional vector.
- $g(.)$ is an unknown monotonic link function. The covariates influence the response through the link function $g(.)$ of the scalar product $\beta' X_i$.
- No assumptions are made on the form of the link function g or the distribution of the error ε_i.

Model

- We consider the model as:

$$Y_i = g(\beta' X_i, \varepsilon_i), \qquad i = 1, \dots, n. \tag{1}$$

- β is a p-dimensional vector.
- $g(.)$ is an unknown monotonic link function. The covariates influence the response through the link function $g(.)$ of the scalar product $\beta' X_i$.
- No assumptions are made on the form of the link function g or the distribution of the error ε_i.

Rank-LASSO

- We define the rank R_i corresponding to response Y_i as:

$$R_i = \sum_{j=1}^{n} \mathbb{I}(Y_j \leq Y_i), \qquad i = 1, \ldots, n,$$

- The relevant covariates are identified by solving the following rank-based LASSO problem:

$$\textbf{RankLASSO:} \qquad \hat{\theta} = \arg\min_{\theta \in \mathbb{R}^p} Q(\theta) + \lambda |\theta|_1, \tag{2}$$

where

$$Q(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \left(\frac{R_i}{n} - \frac{1}{2} - \theta' X_i \right)^2. \tag{3}$$

Rank-LASSO

- We define the rank R_i corresponding to response Y_i as:

$$R_i = \sum_{j=1}^{n} \mathbb{I}(Y_j \leq Y_i), \qquad i = 1, \ldots, n,$$

- The relevant covariates are identified by solving the following rank-based LASSO problem:

$$\textbf{RankLASSO:} \quad \hat{\theta} = \arg\min_{\theta \in \mathbb{R}^p} Q(\theta) + \lambda |\theta|_1, \tag{2}$$

where

$$Q(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \left(\frac{R_i}{n} - \frac{1}{2} - \theta' X_i \right)^2. \tag{3}$$

Rank-LASSO

- We define the rank R_i corresponding to response Y_i as:

$$R_i = \sum_{j=1}^{n} \mathbb{I}(Y_j \leq Y_i), \qquad i = 1, \ldots, n,$$

- The relevant covariates are identified by solving the following rank-based LASSO problem:

$$\textbf{RankLASSO:} \qquad \hat{\theta} = \arg\min_{\theta \in \mathbb{R}^p} Q(\theta) + \lambda |\theta|_1, \tag{2}$$

where

$$Q(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \left(\frac{R_i}{n} - \frac{1}{2} - \theta' X_i \right)^2. \tag{3}$$

Assumptions

Assumption (A1)

We assume that $(X_1, \varepsilon_1), \ldots, (X_n, \varepsilon_n)$ *are i.i.d. random vectors such that the distribution of* X_1 *is absolutely continuous and* X_1 *is independent of the noise variable* ε_1*. Additionally, we assume that* $\mathbb{E}(X_1) = 0$, $H = \mathbb{E}(X_1 X_1')$ *is positive definite and* $H_{jj} = 1$ *for* $j = 1, \ldots, p$.

Assumption (A2)

We assume that for each $\theta \in \mathbb{R}^p$*, the conditional expectation* $\mathbb{E}(\theta' X_1 | \beta' X_1)$ *exists and* $\mathbb{E}(\theta' X_1 | \beta' X_1) = d_\theta \beta' X_1$ *for a real number* $d_\theta \in \mathbb{R}$.

Assumptions

Assumption (A1)

We assume that $(X_1, \varepsilon_1), \ldots, (X_n, \varepsilon_n)$ are i.i.d. random vectors such that the distribution of X_1 is absolutely continuous and X_1 is independent of the noise variable ε_1. Additionally, we assume that $\mathbb{E}(X_1) = 0$, $H = \mathbb{E}(X_1 X_1')$ is positive definite and $H_{jj} = 1$ for $j = 1, \ldots, p$.

Assumption (A2)

We assume that for each $\theta \in \mathbb{R}^p$, the conditional expectation $\mathbb{E}(\theta' X_1 | \beta' X_1)$ exists and $\mathbb{E}(\theta' X_1 | \beta' X_1) = d_\theta \beta' X_1$ for a real number $d_\theta \in \mathbb{R}$.

Assumptions

Assumption (A3)

We assume that the design matrix and the error term satisfy Assumptions A1 and A2, the cumulative distribution function F of the response variable Y_1 is increasing and g in 1 is increasing with respect to the first argument.

Relation between Rank-LASSO estimate and β

- RankLASSO does not estimate β, but the vector

$$\theta^0 = \arg\min_{\theta \in \mathbb{R}^p} \mathbb{E} Q(\theta) \qquad (4)$$

- The minimizer θ^0 is given by the formula

$$\theta^0 = \frac{1}{n^2} H^{-1} \left(\mathbb{E} \sum_{i=1}^{n} R_i X_i \right). \qquad (5)$$

- Since

$$\sum_{i=1}^{n} R_i X_i = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{I}(Y_j \leq Y_i) X_i = \sum_{i \neq j} \mathbb{I}(Y_j \leq Y_i) X_i + \sum_{i=1}^{n} X_i$$

and that $\mathbb{E}(X_i) = 0$, we can rewrite (5) as $\theta_0 = \frac{n-1}{n} H^{-1} \mu$ where $\mu = \mathbb{E}[\mathbb{I}(Y_2 \leq Y_1) X_1]$.

Relation between Rank-LASSO estimate and β

- RankLASSO does not estimate β, but the vector

$$\theta^0 = \arg\min_{\theta \in \mathbb{R}^p} \mathbb{E}Q(\theta) \qquad (4)$$

- The minimizer θ^0 is given by the formula

$$\theta^0 = \frac{1}{n^2} H^{-1} \left(\mathbb{E} \sum_{i=1}^{n} R_i X_i \right). \qquad (5)$$

- Since

$$\sum_{i=1}^{n} R_i X_i = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{I}(Y_j \leq Y_i) X_i = \sum_{i \neq j} \mathbb{I}(Y_j \leq Y_i) X_i + \sum_{i=1}^{n} X_i$$

and that $\mathbb{E}(X_i) = 0$, we can rewrite (5) as $\theta_0 = \frac{n-1}{n} H^{-1} \mu$ where $\mu = \mathbb{E}[\mathbb{I}(Y_2 \leq Y_1) X_1]$.

Relation between Rank-LASSO estimate and β

- RankLASSO does not estimate β, but the vector

$$\theta^0 = \arg\min_{\theta \in \mathbb{R}^p} \mathbb{E} Q(\theta) \tag{4}$$

- The minimizer θ^0 is given by the formula

$$\theta^0 = \frac{1}{n^2} H^{-1} \left(\mathbb{E} \sum_{i=1}^{n} R_i X_i \right). \tag{5}$$

- Since

$$\sum_{i=1}^{n} R_i X_i = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{I}(Y_j \leq Y_i) X_i = \sum_{i \neq j} \mathbb{I}(Y_j \leq Y_i) X_i + \sum_{i=1}^{n} X_i$$

and that $\mathbb{E}(X_i) = 0$, we can rewrite (5) as $\theta_0 = \frac{n-1}{n} H^{-1} \mu$ where $\mu = \mathbb{E}[\mathbb{I}(Y_2 \leq Y_1) X_1]$.

Relation between Rank-LASSO estimate and β

Theorem

Consider the model (1). If Assumptions (A1) and (A2) are satisfied, then

$$\theta_0 = \gamma_\beta \beta$$

with

$$\gamma_\beta = \frac{\frac{n-1}{n}\beta'\mu}{\beta' H \beta} = \frac{\frac{n-1}{n} Cov(F(Y_1), \beta' X_1)}{\beta' H \beta}, \quad (6)$$

where F is a cumulative distribution function of a response variable Y_1.
Additionally, if F is increasing and g is increasing with respect to the first argument, then $\gamma_\beta > 0$, so the signs of β coincide with the signs of θ^0 and

$$T = \{j : \beta_j \neq 0\} = \{j : \theta_j^0 \neq 0\}. \quad (7)$$

Relation between Rank-LASSO estimate and β

- Therefore, Rank-LASSO can be used for variable selection from a large number of explanatory variables, as the support of β remains intact through the Rank-based LASSO model.

Motivation

- Presenting the important properties of Rank-LASSO via **non-asymptotic** results.
- Ensuring applicability of the method for high-dimensional scenario especially for $p >> n$.

Motivation

- Presenting the important properties of Rank-LASSO via **non-asymptotic** results.
- Ensuring applicability of the method for high-dimensional scenario especially for $p >> n$.

Assumption 4

Assumption

Let $(X_1)_T$ be the vector of significant predictors and suppose that it is subgaussian with coefficients $\tau_0 > 0$ i.e for each $u \in \mathbb{R}^{p_0}$ we have $\mathbb{E}exp(u^T(X_1)_T) \leq exp(\tau_0^2 u^T u/2)$. Also we have, the insignificant predictors are univariate subgaussian, i.e for each $a \in \mathbb{R}$ and $j \notin T$, $\mathbb{E}(aX_{1j}) \leq exp(\tau_j^2 a^2/2)$, for $\tau_j > 0$. Denote, $\tau = max(\tau_0, \tau_j, j \notin T)$.

Characteristics measuring the potential for consistent estimation of model parameters

- Let T be the set of indices corresponding to the support of true vector β.
- Suppose that θ_T and $\theta_{T'}$ be the restrictions of the vector $\theta \in \mathbb{R}^p$ to indices of the indices from T and T', respectively.
- For, $\zeta > 1$, a cone can be considered,

$$C(\zeta) = \{\theta \in \mathbb{R}^p : \ |\theta_{T'}|_1 \le \zeta \, |\theta_T|_1\}$$

Characteristics measuring the potential for consistent estimation of model parameters

- Let T be the set of indices corresponding to the support of true vector β.
- Suppose that θ_T and $\theta_{T'}$ be the restrictions of the vector $\theta \in \mathbb{R}^p$ to indices of the indices from T and T', respectively.
- For, $\zeta > 1$, a cone can be considered,

$$C(\zeta) = \{\theta \in \mathbb{R}^p : |\theta_{T'}|_1 \leq \zeta |\theta_T|_1\}$$

Characteristics measuring the potential for consistent estimation of model parameters

- Let T be the set of indices corresponding to the support of true vector β.
- Suppose that θ_T and $\theta_{T'}$ be the restrictions of the vector $\theta \in \mathbb{R}^p$ to indices of the indices from T and T', respectively.
- For, $\zeta > 1$, a cone can be considered,

$$C(\zeta) = \{\theta \in \mathbb{R}^p : \; |\theta_{T'}|_1 \leq \zeta \, |\theta_T|_1\}$$

Characteristics measuring potential for consistent estimations of model parameters

- Restricted Eigen Value (Bickel et al. (2009)):

$$RE(\zeta) = \inf_{0 \neq \theta \in C(\zeta)} \frac{\theta^T X^T X \theta}{n \, |\theta_T|_2^2}$$

- Compatibility Factor (Van de Geer (2008)):

$$K(\zeta) = \inf_{0 \neq \theta \in C(\zeta)} \frac{p_0 \theta^T X^T X \theta}{n \, |\theta_T|_1^2}$$

- Cone Invertibility Factor(CIF, Ye and Zhang (2010)):

$$\bar{F}_q(\zeta) = \inf_{0 \neq \theta \in C(\zeta)} \frac{p_0^{1/q} \, |X^T X \theta|_\infty}{n \, |\theta_T|_q}$$

Characteristics measuring potential for consistent estimations of model parameters

- Population version of CIF is given by,

$$F_q(\zeta) = \inf_{0 \neq \theta \in C(\zeta)} \frac{p_0^{1/q} \, |H\theta\,|_\infty}{n \, |\theta_T\,|_q},$$

where $H = E(X^T X)$.

- In this report the CIF will be used as it allows formulation of convergence results for any l_q norm, for $q \geq 1$.

Characteristics measuring potential for consistent estimations of model parameters

- Population version of CIF is given by,

$$F_q(\zeta) = \inf_{0 \neq \theta \in C(\zeta)} \frac{p_0^{1/q} |H\theta|_\infty}{n |\theta_T|_q},$$

where $H = E(X^T X)$.

- In this report the CIF will be used as it allows formulation of convergence results for any l_q norm, for $q \geq 1$.

Estimation Accuracy of Rank-LASSO

Theorem

Let $a \in (0,1)$, $q \geq 1$ and $\zeta \geq 1$ be arbitrary. Suppose that the assumptions A3 and A4 are satisfied. Also,

$$n \geq \frac{K_1 p_0^2 \tau^4 (1+\zeta)^2 log(p/a)}{F_q^2(\zeta)}$$

$$\lambda \geq K_2 \frac{\zeta+1}{\zeta-1} \tau^2 \sqrt{\frac{log(p/a)}{kn}}$$

where K_1, K_2 are universal constants and k is the smallest eigen value of the correlation matrix between true predictor $H_T = (H_{i,j})_{j,k \in T}$.

Estimation Accuracy of Rank-LASSO

Theorem

Then there exists a universal constant K_3 such that,

$$|\hat{\theta} - \theta^0|_q \leq \frac{4\zeta p_0^{1/q}\lambda}{(\zeta+1)F_q(\zeta)}$$

with probability at least $1 - K_3 a$

Estimation Accuracy of Rank-LASSO

- This theorem provides bound to the estimation error.
- It does not require n to be very large. It allows p to increase exponentially as a function of n.
- By replacing a by a sequence a_n, that does not decreases too fast and replacing λ by corresponding sequence λ_n based on a_n, the consistency conditions can be presented.
- The consistency holds even when number of predictors is significantly. larger than sample size.

Estimation Accuracy of Rank-LASSO

- This theorem provides bound to the estimation error.
- It does not require n to be very large. It allows p to increase exponentially as a function of n.
- By replacing a by a sequence a_n, that does not decreases too fast and replacing λ by corresponding sequence λ_n based on a_n, the consistency conditions can be presented.
- The consistency holds even when number of predictors is significantly. larger than sample size.

Estimation Accuracy of Rank-LASSO

- This theorem provides bound to the estimation error.
- It does not require n to be very large. It allows p to increase exponentially as a function of n.
- By replacing a by a sequence a_n, that does not decreases too fast and replacing λ by corresponding sequence λ_n based on a_n, the consistency conditions can be presented.
- The consistency holds even when number of predictors is significantly. larger than sample size.

Estimation Accuracy of Rank-LASSO

- This theorem provides bound to the estimation error.
- It does not require n to be very large. It allows p to increase exponentially as a function of n.
- By replacing a by a sequence a_n, that does not decreases too fast and replacing λ by corresponding sequence λ_n based on a_n, the consistency conditions can be presented.
- The consistency holds even when number of predictors is significantly. larger than sample size.

Separability of Rank-LASSO

Corollary

If the conditions of Theorem 2 are satisfied for $q = \infty$, then for $\theta^0_{min} \geq \frac{8\zeta\lambda}{(\zeta+1)F_\infty(\zeta)}$ we have,

$$P(\forall_{j\in T, k\notin T} \, |\hat{\theta}_j| > |\hat{\theta}_k|) \geq 1 - K_3 a$$

where $\theta^0_{min} = \min_{j\in T} |\theta^0_j|$

- θ^0_{min} can not be too small.
- As, $\theta^0 = \gamma_\beta \beta$, according to Corollary 4 , $\min_{j\in T} |\beta_j| \geq \frac{8\zeta\lambda}{\gamma_\beta(\zeta+1)F_\infty(\zeta)}$.

Estimation Accuracy of Rank-LASSO

Corollary

Let $a \in (0,1)$ be arbitrary and Assumptions A3 and A4 are satisfied. Suppose that, there exists $\zeta_0 > 1, C_1 > 0$ and $C_2 < \infty$ such that $k \geq C_1, F_\infty(\zeta_0) \geq C_1$ and $\tau \leq C_2$. Then for,
$n\& \geq K_1 p_0^2 log(p/a), \lambda \geq K_2 \sqrt{\frac{log(p/a)}{n}}$
we have ,

$$P(\,|\hat{\theta} - \theta^0\,|_\infty \leq 4\lambda/C_1) \geq 1 - K_3 a \tag{8}$$

where K_1, K_2 depend only on ζ_0, C_1, C_2 and K_3 is a universal constant as mentioned in Theorem 2.

The above corollary is a simplified version of Theorem 2.

Extensions to Rank-LASSO technique

- Main drawback of Rank-LASSO that it can recover true model only if irrepresentable condition is satisfied.
- If the condition does not hold, then we need to add a large number of irrelevant predictors for the process to yield true model.
- We discuss following techniques by which this problem can be solved.

Extensions to Rank-LASSO technique

- Main drawback of Rank-LASSO that it can recover true model only if irrepresentable condition is satisfied.
- If the condition does not hold, then we need to add a large number of irrelevant predictors for the process to yield true model.
- We discuss following techniques by which this problem can be solved.

Extensions to Rank-LASSO technique

- Main drawback of Rank-LASSO that it can recover true model only if irrepresentable condition is satisfied.
- If the condition does not hold, then we need to add a large number of irrelevant predictors for the process to yield true model.
- We discuss following techniques by which this problem can be solved.

Threshold Rank-LASSO

We consider thresholded RankLASSO, denoted by $\hat{\theta}^{th}$ and defined as

$$\hat{\theta}_j^{th} = \hat{\theta}_j \mathbb{I}(|\hat{\theta}_j| \geq \delta), \ j = 1, \ldots, p$$

where $\hat{\theta}$ is the RankLASSO esimator and δ is a threshold.

Theorem

Assuming Cor. 5 holds, and selecting the sample size and tuning parameter accordingly, if $\theta^0_{min}/2 \geq \delta > K_4 \lambda$, ($K_4$ defined in Cor. 5, then

$$P(\hat{T}^{th} = T) \geq 1 - K_3 a$$

where, $\hat{T}^{th} = \{1 \leq j \leq p : \hat{\theta}_j^{th} \neq 0\}$ is the estimated estimated set of relevant predictors by thresholded RankLASSO.

Thresholded RankLASSO

- This suggests that the thresholded RankLASSO has potential for identifying the support of β under milder regularity conditions.
- This also suggests that under the conditions, the sequence of models based on ranking provided by RankLASSO estimates contain the true model.

Weighted RankLASSO

We redefine our objective function as follows:

$$Q(\theta) + \lambda_a \sum_{j=1}^{p} w_j |\theta_j|$$

where $\lambda_a > 0$, with weights defined as follows: For an arbitrary number $K > 0$ and the RankLASSO estimator $\hat{\theta}$,

$$w_j = |\hat{\theta}_j|^{-1}, \text{ for } |\hat{\theta}_j| \leq \lambda_a \text{ and } w_j \leq K \text{ otherwise}$$

Weighted RankLASSO

Theorem

Assuming Cor. 5 holds, let $\lambda_a = K_4\lambda$, if $\theta^0_{min}/2 > \lambda_a$ and $p_0\lambda \leq K_5$, with K_5 being sufficiently small, then there exists a global minimizer $\hat{\theta}^a$, such that $\hat{\theta}_{T'} = 0$ and

$$P[|\hat{\theta}^a_T - \theta^0_T|_1 \leq K_7 p_0 \lambda] \geq 1 - K_6 a$$

Advantage of these modifications

- Absolute value loss function is robust with respect to distribution of noise variable.
- However, it requires that that the density of the noise is continuous in a neighbourhood of 0.
- The modifications suggested do not require such restrictions and the procedures work well in single index models.

Advantage of these modifications

- Absolute value loss function is robust with respect to distribution of noise variable.
- However, it requires that that the density of the noise is continuous in a neighbourhood of 0.
- The modifications suggested do not require such restrictions and the procedures work well in single index models.

Advantage of these modifications

- Absolute value loss function is robust with respect to distribution of noise variable.
- However, it requires that that the density of the noise is continuous in a neighbourhood of 0.
- The modifications suggested do not require such restrictions and the procedures work well in single index models.

Simulation Scenarios

- $$Y_i = \beta' X_i + \varepsilon_i$$
- $X_i \sim N(0, \Sigma)$
- $\Sigma = I$ or $\Sigma_{jj} = 1, \Sigma_{jk} = 0.3$
- $\varepsilon \sim$ Cauchy distribution
- $$\beta = (\underbrace{3, \ldots, 3}_{p_0}, \underbrace{0, \ldots, 0}_{p - p_0})$$
- $p_0 \in \{3, 10, 20\}$
- $n \in \{100, 200, 300, 400\}$
- $p \in \{100, 400, 900, 1600\}$

Simulation Scenarios

- We also simulate the genotypes of p independent Single Nucleotide Polymorphisms (SNPs)
- Explanatory variables can take only three values: 0, 1 and 2.
- Given the frequency π_j for j-th SNP, the explanatory variable X_{ij} has the distribution:

$$P(X_{ij}=0)=\pi_j^2, P(X_{ij}=1)=2\pi_j(1-\pi_j) and P(X_{ij}=2)=(1-\pi_j)^2.$$

  Here, $\pi_j \sim$ U(0,1,0.5).
- $Y_i = \beta' X_i + \varepsilon_i$

Simulation Scenarios

- $$Y_i = \exp\left(1 + 0.05\beta' X_i\right) + \varepsilon_i$$
- $X_i \sim N(0, \Sigma)$
- $\Sigma_{jk} = 0.3$
- $\varepsilon \sim$ Cauchy distribution
- $$\beta = (\underbrace{3, \ldots, 3}_{p_0}, \underbrace{0, \ldots, 0}_{p-p_0})$$
- $p_0 \in \{3, 10, 20\}$
- $n \in \{100, 200, 300, 400\}$
- $p \in \{100, 400, 900, 1600\}$

Methods

- RankLasso (rL)
- adaptive RankLasso (arL)
- thresholded RankLasso (thrL)
- Lasso with cross-validation (cv)

- NMP - average number of misclassified predictors

Results

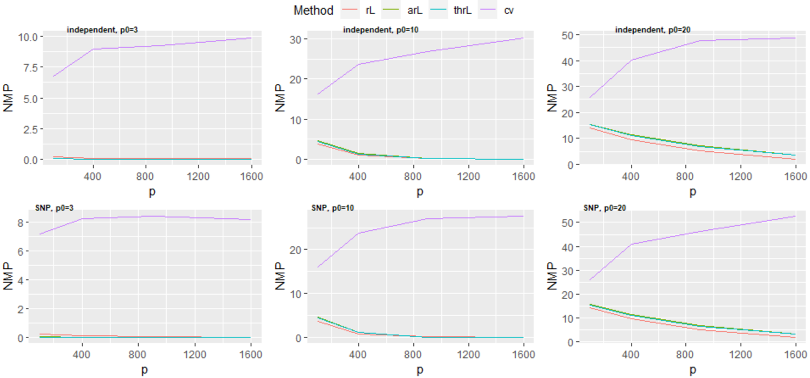

Figure: Plots of NMP (average number of misclassified predictors) as the function of p.

Results

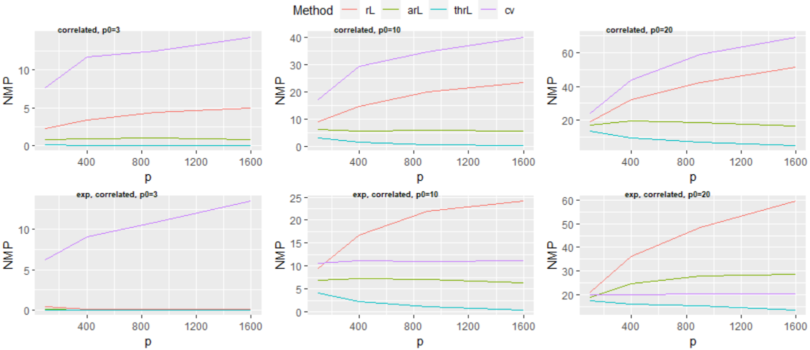

Figure: Plots of NMP (average number of misclassified predictors) as the function of p.

Final Remarks

- The methodology described does not require knowledge of the distribution of the covariates or make moment assumptions on the error distribution.
- The RankLASSO is essentially a convex optimization problem. Hence, it is computationally fast, even when $p >> n$ or in presence of multicollinearity.
- Under certain assumptions, the support of θ_0 coincides with that of β.
- Our simulations illustrate that the thresholded and adaptive versions of RankLasso can properly identify the predictors even when the link function is non-linear, predictors are correlated and the error comes from the Cauchy distribution.
- Some open questions: selection of optimal λ, δ and w_j's.

Thank You!

Final Remarks

- The methodology described does not require knowledge of the distribution of the covariates or make moment assumptions on the error distribution.
- The RankLASSO is essentially a convex optimization problem. Hence, it is computationally fast, even when $p >> n$ or in presence of multicollinearity.
- Under certain assumptions, the support of θ_0 coincides with that of β.
- Our simulations illustrate that the thresholded and adaptive versions of RankLasso can properly identify the predictors even when the link function is non-linear, predictors are correlated and the error comes from the Cauchy distribution.
- Some open questions: selection of optimal λ, δ and w_j's.

Thank You!

Final Remarks

- The methodology described does not require knowledge of the distribution of the covariates or make moment assumptions on the error distribution.
- The RankLASSO is essentially a convex optimization problem. Hence, it is computationally fast, even when $p > > n$ or in presence of multicollinearity.
- Under certain assumptions, the support of θ_0 coincides with that of β.
- Our simulations illustrate that the thresholded and adaptive versions of RankLasso can properly identify the predictors even when the link function is non-linear, predictors are correlated and the error comes from the Cauchy distribution.
- Some open questions: selection of optimal λ, δ and w_j's.

Thank You!

Final Remarks

- The methodology described does not require knowledge of the distribution of the covariates or make moment assumptions on the error distribution.
- The RankLASSO is essentially a convex optimization problem. Hence, it is computationally fast, even when $p > > n$ or in presence of multicollinearity.
- Under certain assumptions, the support of θ_0 coincides with that of β.
- Our simulations illustrate that the thresholded and adaptive versions of RankLasso can properly identify the predictors even when the link function is non-linear, predictors are correlated and the error comes from the Cauchy distribution.
- Some open questions: selection of optimal λ, δ and w_j's.

Thank You!

Final Remarks

- The methodology described does not require knowledge of the distribution of the covariates or make moment assumptions on the error distribution.
- The RankLASSO is essentially a convex optimization problem. Hence, it is computationally fast, even when $p > > n$ or in presence of multicollinearity.
- Under certain assumptions, the support of θ_0 coincides with that of β.
- Our simulations illustrate that the thresholded and adaptive versions of RankLasso can properly identify the predictors even when the link function is non-linear, predictors are correlated and the error comes from the Cauchy distribution.
- Some open questions: selection of optimal λ, δ and w_j's.

Thank You!

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of statistics*, 37(4):1705–1732.

Van de Geer, S. A. (2008). High-dimensional generalized linear models and the lasso. *The Annals of Statistics*, 36(2):614–645.

Ye, F. and Zhang, C.-H. (2010). Rate minimaxity of the lasso and dantzig selector for the lq loss in lr balls. *The Journal of Machine Learning Research*, 11:3519–3540.