Sorted L-One Penalized Estimation (SLOPE)

Malgorzata Bogdan

University of Wroclaw, Lund University

Warsaw, 02/07/2019

Multiple Testing

Multiple Testing

Model selection in multiple regression

- LASSO
- SLOPE
- FDR control under orthogonal design
- Asymptotic minimaxity
- Extensions and applications
- Adaptive Bayesian Slope

Multiple testing (1)

$$X_i \sim N(\mu_i, \sigma^2), \quad i \in \{1, \ldots, p\}$$

Multiple testing (1)

$$X_i \sim N(\mu_i, \sigma^2), \quad i \in \{1, \ldots, p\}$$

$$H_{0i} : \mu_i = 0 \quad \text{vs} \quad \mu_i \neq 0$$

Multiple testing (1)

$$X_i \sim N(\mu_i, \sigma^2), \quad i \in \{1, \ldots, p\}$$

$$H_{0i} : \mu_i = 0 \quad \text{vs} \quad \mu_i \neq 0$$

$$\text{Reject } H_{0i} \text{ when } |X_i| > c$$

Multiple testing (1)

$$X_i \sim N(\mu_i, \sigma^2), \quad i \in \{1, \ldots, p\}$$

$$H_{0i} : \mu_i = 0 \quad \text{vs} \quad \mu_i \neq 0$$

Reject H_{0i} when $|X_i| > c$

Significance level: $\alpha = P_{H_{0i}}(|X_i| > c) = 2(1 - \Phi(c/\sigma))$

Multiple testing (1)

$$X_i \sim N(\mu_i, \sigma^2), \quad i \in \{1, \dots, p\}$$

$$H_{0i} : \mu_i = 0 \quad \text{vs} \quad \mu_i \neq 0$$

$$\text{Reject } H_{0i} \text{ when } |X_i| > c$$

$$\text{Significance level: } \alpha = P_{H_{0i}}(|X_i| > c) = 2(1 - \Phi(c/\sigma))$$

	H_0 accepted	H_0 rejected	
H_0 true	U	V	p_0
H_0 false	T	S	p_1
	W	R	m

Multiple testing (1)

$$X_i \sim N(\mu_i, \sigma^2), \quad i \in \{1, \ldots, p\}$$

$$H_{0i} : \mu_i = 0 \quad \text{vs} \quad \mu_i \neq 0$$

$$\text{Reject } H_{0i} \text{ when } |X_i| > c$$

$$\text{Significance level: } \alpha = P_{H_{0i}}(|X_i| > c) = 2(1 - \Phi(c/\sigma))$$

	H_0 accepted	H_0 rejected	
H_0 true	U	V	p_0
H_0 false	T	S	p_1
	W	R	m

$$FWER = P(V > 0), \qquad FDR = E\left(\frac{V}{R \vee 1}\right)$$

Multiple testing (1)

$$X_i \sim N(\mu_i, \sigma^2), \quad i \in \{1, \dots, p\}$$

$$H_{0i} : \mu_i = 0 \quad \text{vs} \quad \mu_i \neq 0$$

$$\text{Reject } H_{0i} \text{ when } |X_i| > c$$

$$\text{Significance level: } \alpha = P_{H_{0i}}(|X_i| > c) = 2(1 - \Phi(c/\sigma))$$

	H_0 accepted	H_0 rejected	
H_0 true	U	V	p_0
H_0 false	T	S	p_1
	W	R	m

$$FWER = P(V > 0), \qquad FDR = E\left(\frac{V}{R \vee 1}\right)$$

$$E(V) = \alpha p_0$$

Multiple testing (1)

$$X_i \sim N(\mu_i, \sigma^2), \quad i \in \{1, \dots, p\}$$

$$H_{0i} : \mu_i = 0 \quad \text{vs} \quad \mu_i \neq 0$$

$$\text{Reject } H_{0i} \text{ when } |X_i| > c$$

$$\text{Significance level: } \alpha = P_{H_{0i}}(|X_i| > c) = 2(1 - \Phi(c/\sigma))$$

	H_0 accepted	H_0 rejected	
H_0 true	U	V	p_0
H_0 false	T	S	p_1
	W	R	m

$$FWER = P(V > 0), \qquad FDR = E\left(\frac{V}{R \vee 1}\right)$$

$$E(V) = \alpha p_0$$

$$\alpha = 0.05, p_0 = 5000 \rightarrow E(V) = 250$$

Multiple testing procedures

Bonferroni correction: Use significance level $\frac{\alpha}{p}$.

Multiple testing procedures

Bonferroni correction: Use significance level $\frac{\alpha}{p}$.

Reject H_{0i} if $|X_i| \geq \sigma\Phi^{-1}\left(1 - \frac{\alpha}{2p}\right) = \sigma\sqrt{2\log p}(1 + o(1))$

Multiple testing procedures

Bonferroni correction: Use significance level $\frac{\alpha}{p}$.

Reject H_{0i} if $|X_i| \geq \sigma\Phi^{-1}\left(1 - \frac{\alpha}{2p}\right) = \sigma\sqrt{2\log p}(1 + o(1))$

Benjamini-Hochberg procedure:

(1) $|X|_{(1)} \geq |X|_{(2)} \geq \ldots \geq |X|_{(p)}$

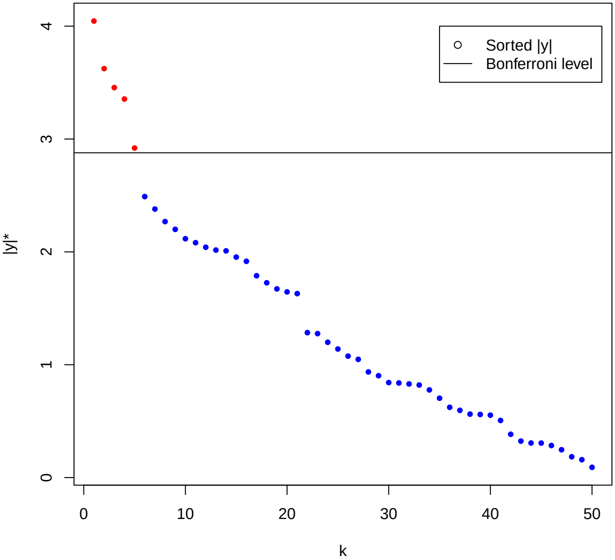
(2) Find the largest index i such that

$$|X|_{(i)} \geq \sigma\Phi^{-1}(1 - \alpha_i), \quad \alpha_i = \alpha\frac{i}{2p}, \tag{1}$$

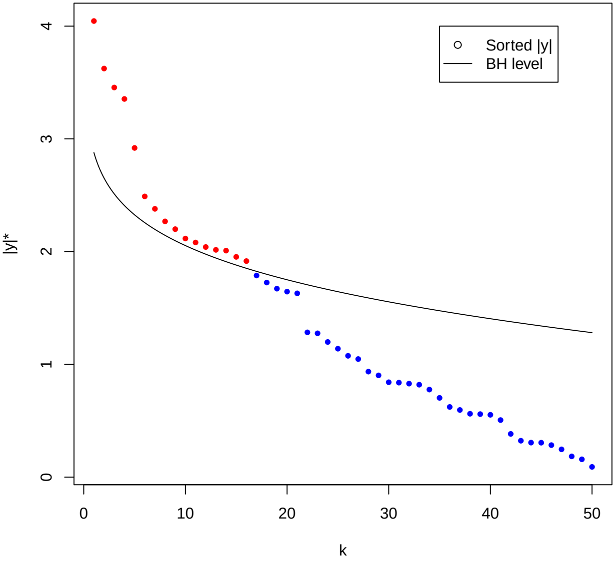
Call this index i_{SU}.

(3) Reject all $H_{(i)}$'s for which $i \leq i_{\mathrm{SU}}$

Bonferroni correction



Benjamini and Hochberg correction



FWER and FDR control

For Bonferroni correction $FWER \leq \alpha$

FWER and FDR control

For Bonferroni correction $FWER \leq \alpha$

(Benjamini,Hochberg, 1995) If $X_1, \ldots, X_p$ are independent then BH controls FDR at:

FWER and FDR control

For Bonferroni correction $FWER \leq \alpha$

(Benjamini,Hochberg, 1995) If $X_1, \ldots, X_p$ are independent then BH controls FDR at:

$$\mathrm{FDR} = \mathbb{E}\left[\frac{V}{R \vee 1}\right] = \alpha \frac{p_0}{p}, \tag{2}$$

where p_0 is the number of true null hypotheses, $p_0 = |\{i : \mu_i = 0\}|$

FWER and FDR control

For Bonferroni correction $FWER \leq \alpha$

(Benjamini,Hochberg, 1995) If $X_1, \ldots, X_p$ are independent then BH controls FDR at:

$$\text{FDR} = \mathbb{E}\left[\frac{V}{R \vee 1}\right] = \alpha\frac{p_0}{p}, \tag{2}$$

where p_0 is the number of true null hypotheses, $p_0 = |\{i : \mu_i = 0\}|$

(Benjamini, Yekutieli, 2001) BH controls FDR at the level $\alpha\frac{p_0}{p}$ if the test statistics are "positively" correlated (Positive Regression Dependency on each from a subset). BH always controls FDR at the level $\alpha\frac{p_0}{p}$ if $|X|_{(i)}$ is compared with $\sigma\Phi^{-1}\left(1 - \frac{i\alpha}{p\sum_{i=1}^{p}\frac{1}{i}}\right)$.

LASSO

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + z_{n\times 1}, \quad z \sim N(0, \sigma I)$$

LASSO

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + z_{n\times 1}, \quad z \sim N(0, \sigma I)$$

LASSO (BPDN) : Convex program: Minimize $||b||_1$ subject to $||Y - Xb||_2^2 \leq \epsilon$

Or alternatively: $min_{b\in R^p} \frac{1}{2}||y - Xb||_2^2 + \lambda||b||_1$

Selection of the tuning parameter for LASSO

- General rule: the reduction of λ_L results in identification of more elements from the true support (true discoveries) but at the same time it produces more falsely identified variables (false discoveries)
- The choice of λ_L is challenging- e.g. crossvalidation typically leads to many false discoveries
- When $X^T X = I$ Lasso selects X_j iff $|\hat{\beta}_j^{LS}| > \lambda$
- Selection $\lambda = \sigma\Phi^{-1}(1 - \alpha/(2p)) \approx \sigma\sqrt{2\log p}$ corresponds to Bonferroni correction and controls FWER.

Sorted L-One Penalized Estimation

M. Bogdan, E. van den Berg, C. Sabatti, W. Su, E. J. Candès, AOAS 2015

SLOPE (2)

SLOPE estimate is given by

$$\min_{b \in \mathbb{R}^p} \quad \frac{1}{2}\|y - Xb\|_{\ell_2}^2 + \lambda_1 |b|_{(1)} + \lambda_2 |b|_{(2)} + \cdots + \lambda_p |b|_{(p)},$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and $|b|_{(1)} \geq |b|_{(2)} \geq \cdots \geq |b|_{(p)}$

SLOPE (2)

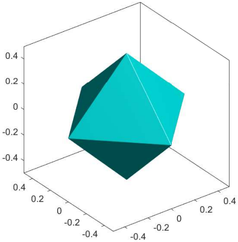
SLOPE estimate is given by

$$\min_{b \in \mathbb{R}^p} \quad \frac{1}{2}\|y - Xb\|_{\ell_2}^2 + \lambda_1 |b|_{(1)} + \lambda_2 |b|_{(2)} + \cdots + \lambda_p |b|_{(p)},$$

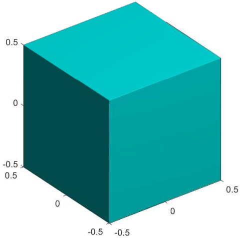
where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and $|b|_{(1)} \geq |b|_{(2)} \geq \cdots \geq |b|_{(p)}$

R package *SLOPE* available on CRAN, by E. Patterson

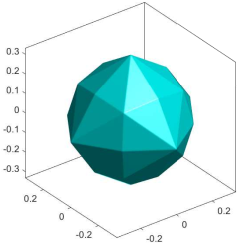
Unit balls for different SLOPE sequences



(a) (2,2,2)



(b) (2,0,0)



(c) (3,2,1)

False discovery rate (FDR) control

- Let $\widetilde{\beta}$ be estimate of β

False discovery rate (FDR) control

- Let $\widetilde{\beta}$ be estimate of β
- We define:

False discovery rate (FDR) control

- Let $\widetilde{\beta}$ be estimate of β
- We define:

- the number of all discoveries, $R := \left|\left\{i : \ \widetilde{\beta}_i \neq 0\right\}\right|$

False discovery rate (FDR) control

- Let $\widetilde{\beta}$ be estimate of β
- We define:

- the number of all discoveries, $R := \left|\left\{i : \ \widetilde{\beta}_i \neq 0\right\}\right|$
- the number of false discoveries, $V := \left|\left\{i : \ \beta_i = 0, \quad \widetilde{\beta}_i \neq 0\right\}\right|$

False discovery rate (FDR) control

- Let $\widetilde{\beta}$ be estimate of β
- We define:

- the number of all discoveries, $R := \left|\left\{i : \ \widetilde{\beta}_i \neq 0\right\}\right|$
- the number of false discoveries, $V := \left|\left\{i : \ \beta_i = 0, \quad \widetilde{\beta}_i \neq 0\right\}\right|$
- false discovery rate, $FDR := \mathbb{E}\left[\frac{V}{\max\{R,1\}}\right]$

FDR control with SLOPE

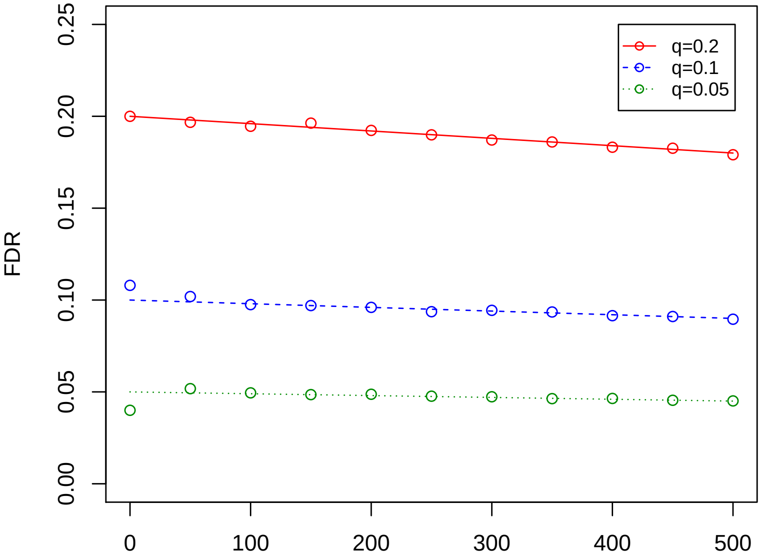
Theorem (B,van den Berg, Sabatti, Su and Candès (2015))

When $X^TX = I$ SLOPE with

$$\lambda_i := \sigma\Phi^{-1}\left(1 - i \cdot \frac{q}{2p}\right)$$

controls FDR at the level $q\frac{p_0}{p}$.

Orthogonal design, $n = p = 5000$



Asymptotic optimality, Su and Candès (Annals of Statistics, 2016)

Theorem

Let $X_{ij} \sim N(0, 1/\sqrt{n})$. *Fix* $0 < q < 1$ *and choose* $\lambda = \sigma(1+\epsilon)\lambda^{BH}(q)$ *for some arbitrary constant* $0 < \epsilon < 1$. *Suppose* $k/p \to 0$ *and* $\frac{k \log p}{n} \to 0$. *Then*

$$\sup_{||\beta_0|| \leq k} P\left(\frac{||\hat{\beta}_{SL} - \beta||^2}{2\sigma^2 k \log(p/k)} > 1 + 3\epsilon\right) \to 0$$

$$\inf_{\hat{\beta}} sup_{||\beta_0|| \leq k} P\left(\frac{||\hat{\beta} - \beta||^2}{2\sigma^2 k \log(p/k)} > 1 - \epsilon\right) \to 1$$

Asymptotic optimality (2)

Minimax estimation/prediction rate $[k \log(p/k)]$ under weighted restricted eigenvalue condition (large collection of random matrices)

Asymptotic optimality (2)

Minimax estimation/prediction rate $[k \log(p/k)]$ under weighted restricted eigenvalue condition (large collection of random matrices)

$\lambda_i = \rho\sqrt{2\log(p/i)}$, ρ is larger than one

Bellec, Lecué, Tsybakov (2018, AOS)

Asymptotic optimality (2)

Minimax estimation/prediction rate $[k \log(p/k)]$ under weighted restricted eigenvalue condition (large collection of random matrices)

$\lambda_i = \rho\sqrt{2\log(p/i)}$, ρ is larger than one

Bellec, Lecué, Tsybakov (2018, AOS)

Extension to GLM by Abramovich and Grinshtein (2018, IEEE Trans. Inf. Theory)

Asymptotic optimality (2)

Minimax estimation/prediction rate $[k \log(p/k)]$ under weighted restricted eigenvalue condition (large collection of random matrices)

$\lambda_i = \rho\sqrt{2\log(p/i)}$, ρ is larger than one

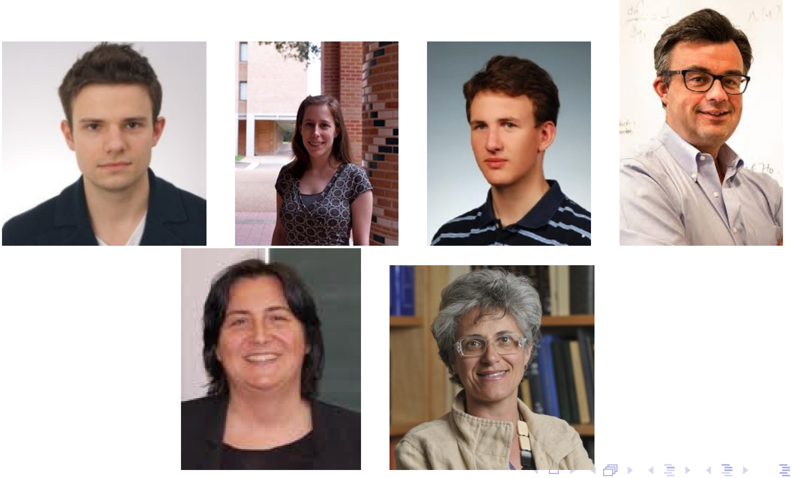
Bellec, Lecué, Tsybakov (2018, AOS)

Extension to GLM by Abramovich and Grinshtein (2018, IEEE Trans. Inf. Theory)

LASSO rate of convergence - $k\log(p)$

Genome Wide Association Studies

D. Brzyski, C.B. Peterson, P.Sobczyk, E.J. Candès, M. Bogdan, C. Sabatti, (Genetics, 2017)
R package *geneSLOPE* available on CRAN, by P. Sobczyk



Computer simulations

Genotypes - the North Finland Birth Cohort (NFBC) dataset.

$n = 5\ 402$ and $p = 334\ 103$

Missing values replaced by the column means.

X - centered and normalized matrix of genotypes.

Computer simulations

Genotypes - the North Finland Birth Cohort (NFBC) dataset.

$n = 5\ 402$ and $p = 334\ 103$

Missing values replaced by the column means.

X - centered and normalized matrix of genotypes.
Signals are evenly distributed in the interval $[0.6\sqrt{2\log p}, 1.4\sqrt{2\log p}]$ ($h^2 \approx 5\%$ for $k = 10$ and 32% for $k = 100$).

Computer simulations

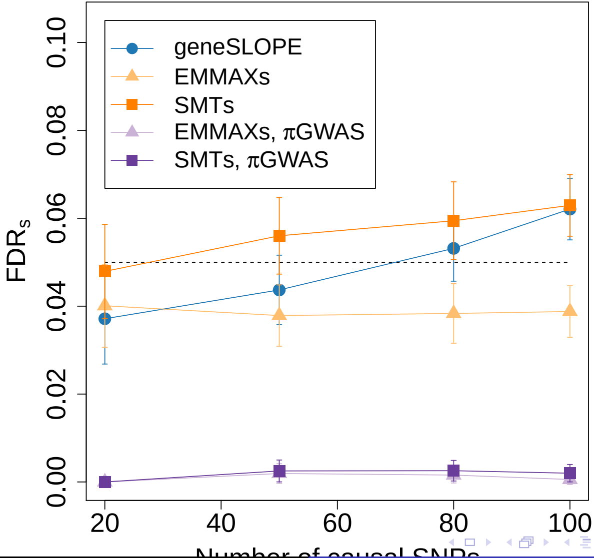
Genotypes - the North Finland Birth Cohort (NFBC) dataset.

$n = 5\,402$ and $p = 334\,103$

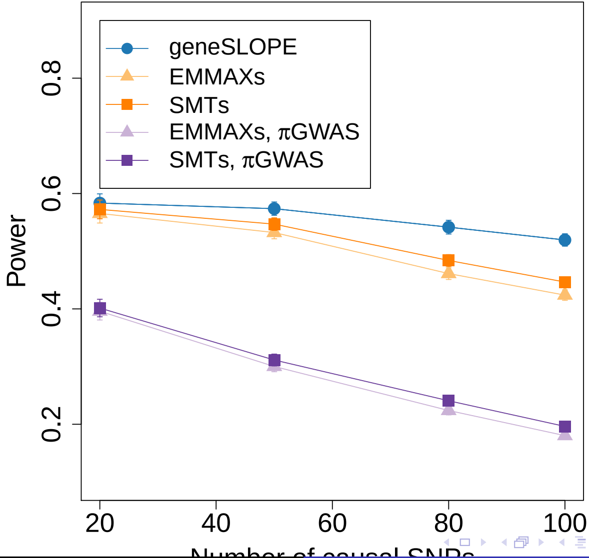
Missing values replaced by the column means.

X - centered and normalized matrix of genotypes.
Signals are evenly distributed in the interval $[0.6\sqrt{2\log p}, 1.4\sqrt{2\log p}]$ ($h^2 \approx 5\%$ for $k = 10$ and 32% for $k = 100$).
Support randomly selected for every replication from the set of well separated SNPs

FDR


geneSLOPE
EMMAXs
SMTs
EMMAXs, πGWAS
SMTs, πGWAS
FDR_s
0.00
0.02
0.04
0.06
0.08
0.10
20
40
60
80
100
Number of causal SNPs

Power


geneSLOPE
EMMAXs
SMTs
EMMAXs, πGWAS
SMTs, πGWAS
Power
0.2
0.4
0.6
0.8
20
40
60
80
100
Number of causal SNPs

Group SLOPE, (Brzyski, Gossmann, Su and Bogdan, JASA, 2019)

GroupSLOPE (2)

ENAR (East North American Region of the International Biometric Society) award for the best young scientist for Damian Brzyski

GroupSLOPE (2)

ENAR (East North American Region of the International Biometric Society) award for the best young scientist for Damian Brzyski

R package *grpSLOPE* available on CRAN by A. Gossmann

GroupSLOPE (2)

ENAR (East North American Region of the International Biometric Society) award for the best young scientist for Damian Brzyski

R package *grpSLOPE* available on CRAN by A. Gossmann

A. Gossmann, S. Cao, D. Brzyski, L. Zhao, H. Deng, and Y. Wang, ”A sparse regression method for group-wise feature selection with false discovery rate control",*IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2017

GroupSLOPE (2)

ENAR (East North American Region of the International Biometric Society) award for the best young scientist for Damian Brzyski

R package *grpSLOPE* available on CRAN by A. Gossmann

A. Gossmann, S. Cao, D. Brzyski, L. Zhao, H. Deng, and Y. Wang, "A sparse regression method for group-wise feature selection with false discovery rate control",*IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2017

group - additive and dominance effect of a given SNP

Applications for GWAS

$n = 5402$, $p = 26233$ - roughly independent SNPs

Applications for GWAS

$n = 5402$, $p = 26233$ - roughly independent SNPs

Scenario 1: $Y = X\beta + z$ - additive model

$$X_{ij} = \begin{cases} -1 & \text{for} \quad aa \\ 0 & \text{for} \quad aA \\ 1 & \text{for} \quad AA \end{cases}, \tag{3}$$

Applications for GWAS

$n = 5402$, $p = 26233$ - roughly independent SNPs

Scenario 1: $Y = X\beta + z$ - additive model

$$X_{ij} = \begin{cases} -1 & \text{for} \quad aa \\ 0 & \text{for} \quad aA \\ 1 & \text{for} \quad AA \end{cases}, \tag{3}$$

Scenario 2: modeling dominance

$$Z_{ij} = \begin{cases} -1 & \text{for} \quad aa, AA \\ 1 & \text{for} \quad aA \end{cases}, \tag{4}$$

$$y = [X, Z][\beta'_X, \beta'_Z]' + \epsilon \quad .$$

Simulations in the genetic context

$n = 5402$, $p = 26233$ - roughly independent SNPs

Simulations in the genetic context

$n = 5402$, $p = 26233$ - roughly independent SNPs

Scenario 1: $Y = X\beta + z$ - additive model

Simulations in the genetic context

$n = 5402$, $p = 26233$ - roughly independent SNPs

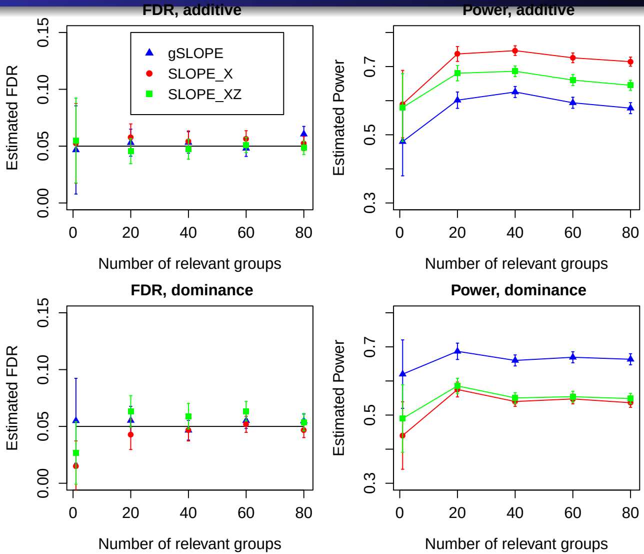
Scenario 1: $Y = X\beta + z$ - additive model

Scenario 2: modeling dominance

$$\tilde{z}_{ij} = \begin{cases} -1 & \text{for} \quad aa, AA \\ 1 & \text{for} \quad aA \end{cases} , \tag{5}$$

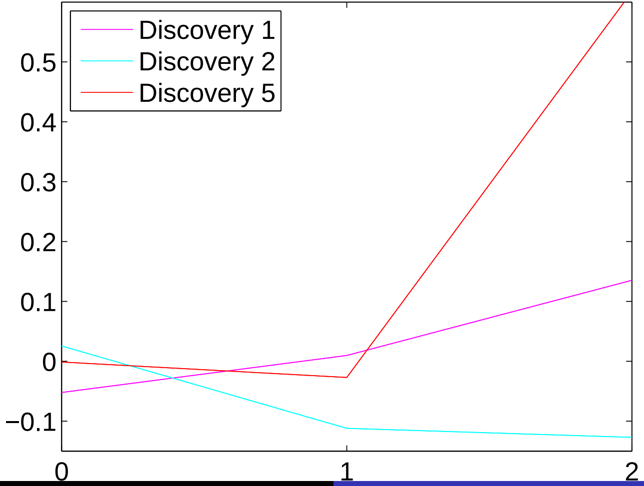
$$y = [X, Z][\beta_X', \beta_Z']' + \epsilon \quad .$$

Simulation results



Genes Influencing Level of Triglicerides

5 new discoveries with group SLOPE - recessive rare genetic variants. Discovery 5 - 37 rare homozygotes



Robust regression with SLOPE

A.Virouleau, A.Guilloux, S.Gaiffas, M.Bogdan (arxive, 2017)

Robust regression (2)

Mean shift model: Candes and Randall (2006), Gannaz (2006) and McCann and Welsch (CSDA, 2007) ,

$$y = X\beta + I\mu + \varepsilon \tag{6}$$

$\mu \in R^n$ is the sparse vector of "outliers" and $\varepsilon \sim N(0, \sigma^2 I)$

Robust regression (2)

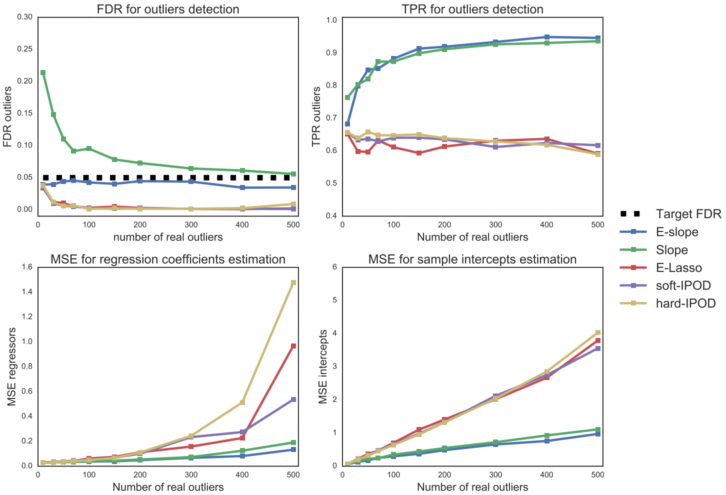
Mean shift model: Candes and Randall (2006), Gannaz (2006) and McCann and Welsch (CSDA, 2007) ,

$$y = X\beta + I\mu + \varepsilon \tag{6}$$

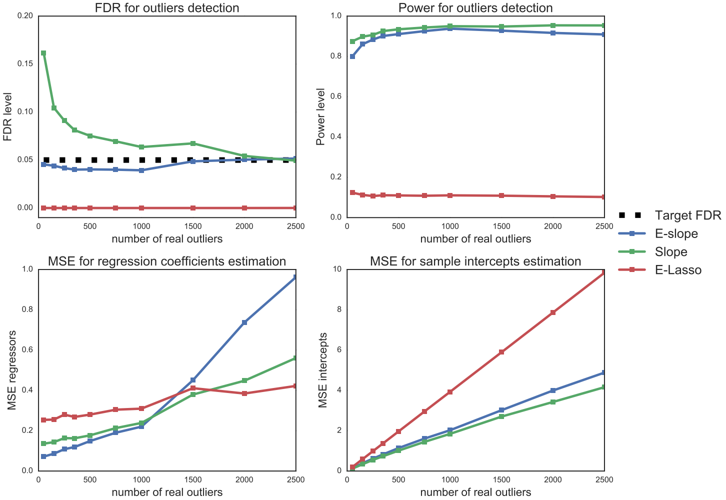
$\mu \in R^n$ is the sparse vector of "outliers" and $\varepsilon \sim N(0, \sigma^2 I)$

$$\min_{\beta \in \mathbb{R}^p, \mu \in \mathbb{R}^n} \left\{ \|y - X\beta - \mu\|_2^2 + 2\rho_1 J_{\tilde{\lambda}}(\beta) + 2\rho_2 J_\lambda(\mu) \right\}$$

Low dimensional set-up; small outliers



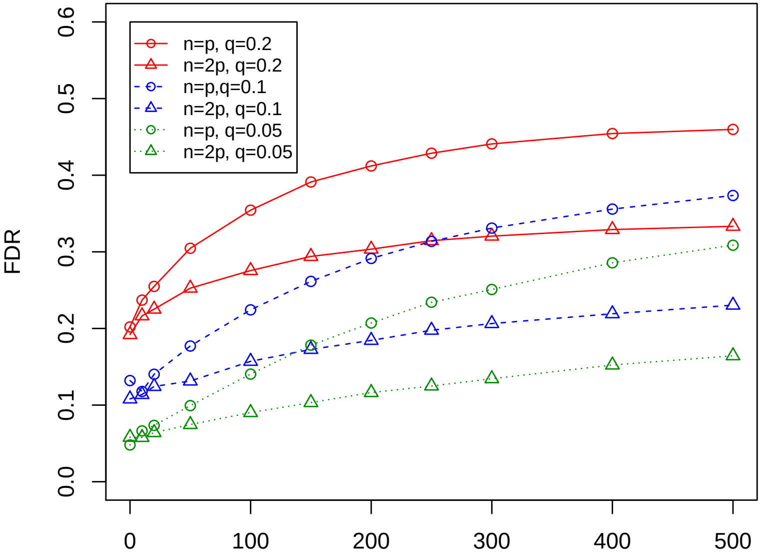
High dimensional set-up; small outliers



Adaptive SLOPE

W. Jiang, B. Miasojedow, V. Rockova, J. Josse, M. Bogdan (in progress)

Gaussian design (1), $n = p = 5000$



Problems with FDR control

Similar problems occur for LASSO: Precise FDR-Power Tradeoff under asymptotic asymptotic assumptions of AMP theory provided in (Su, B, Candès, AOS 2017)

Intuitive explanation:

$$\hat{\beta} = \eta_\lambda(\beta_i + X_i' z + v_i)$$

$$v_i = \langle X_i, \sum_{j \neq i} X_j(\beta_j - \hat{\beta}_j) \rangle$$

$$\eta_\lambda(t) = sign(t)(|t| - \lambda)_+, \quad \text{applied componentwise}$$

Problems with FDR control

Similar problems occur for LASSO: Precise FDR-Power Tradeoff under asymptotic asymptotic assumptions of AMP theory provided in (Su, B, Candès, AOS 2017)

Intuitive explanation:

$$\hat{\beta} = \eta_\lambda(\beta_i + X_i' z + v_i)$$

$$v_i = \langle X_i, \sum_{j \neq i} X_j(\beta_j - \hat{\beta}_j) \rangle$$

$$\eta_\lambda(t) = sign(t)(|t| - \lambda)_+, \quad \text{applied componentwise}$$

If $X^T X = I$ then $X_i' z = Z_i \sim N(0,1)$, $v_i = 0$ and H_{0i} is rejected if $\beta_i + Z_i > \lambda$

Problems with FDR control

Similar problems occur for LASSO: Precise FDR-Power Tradeoff under asymptotic asymptotic assumptions of AMP theory provided in (Su, B, Candès, AOS 2017)

Intuitive explanation:

$$\hat{\beta} = \eta_\lambda(\beta_i + X_i' z + v_i)$$

$$v_i = \langle X_i, \sum_{j \neq i} X_j(\beta_j - \hat{\beta}_j) \rangle$$

$$\eta_\lambda(t) = sign(t)(|t| - \lambda)_+, \quad \text{applied componentwise}$$

If $X^T X = I$ then $X_i' z = Z_i \sim N(0,1)$, $v_i = 0$ and H_{0i} is rejected if $\beta_i + Z_i > \lambda$

When the design is not orthogonal: $v_i \neq 0$ - additional noise, dependent on λ (level of shrinkage), the level of sparsity and magnitude of true signals

Adaptive LASSO and Bayesian Interpretation of LASSO

Adaptive LASSO [Zou, *JASA* 2006], [Candès, Wakin and Boyd, *J. Fourier Anal. Appl.* 2008]

$$\beta_{aL} = argmin_b \left\{ \frac{1}{2} \|y - Xb\|_2^2 + \lambda \sum_{i=1}^{p} w_i |b|_i \right\}, \qquad (7)$$

where $w_i = \frac{1}{\hat{\beta}_i}$, and $\hat{\beta}_i$ is some consistent estimator of β_i.

Adaptive LASSO and Bayesian Interpretation of LASSO

Adaptive LASSO [Zou, *JASA* 2006], [Candès, Wakin and Boyd, *J. Fourier Anal. Appl.* 2008]

$$\beta_{aL} = argmin_b \left\{ \frac{1}{2} \|y - Xb\|_2^2 + \lambda \sum_{i=1}^{p} w_i |b|_i \right\}, \qquad (7)$$

where $w_i = \frac{1}{\hat{\beta}_i}$, and $\hat{\beta}_i$ is some consistent estimator of β_i.

LASSO has a Bayesian interpretation as a posterior mode under the prior

$$\pi(\beta) = C(\lambda) \prod_{i=1}^{n} \mathrm{e}^{-|\beta_i|\lambda}$$

Spike and Slab LASSO (Rockova, George, 2018)

Model for SSL: $\gamma_i = 1$ if i^{th} variable is a signal
Prior for β is given by

$$\pi(\beta|\lambda,\gamma) \propto c^{\sum_{i=1}^{p} 1(\gamma_i=1)} \prod_{i=1}^{p} \mathrm{e}^{-w_i|\beta_i|\lambda_0},$$

where $w_i = 1$ if $\gamma_i = 0$ and $w_i = c \in (0,1)$ if $\gamma_i = 1$.

Spike and Slab LASSO (2)

The maximum aposteriori rule is given by reweighted LASSO

$$\hat{\beta}(\gamma) = argmin_{b \in R^p} \frac{1}{2} ||y - Xb||_2^2 + \lambda \sum_{i=1}^{p} w_i |b_i|$$

$$w_i = c\gamma_i + (1 - \gamma_i)$$

Spike and Slab LASSO (2)

The maximum aposteriori rule is given by reweighted LASSO

$$\hat{\beta}(\gamma) = argmin_{b \in R^p} \frac{1}{2} ||y - Xb||_2^2 + \lambda \sum_{i=1}^{p} w_i |b_i|$$

$$w_i = c\gamma_i + (1 - \gamma_i)$$

Prior for γ: $\gamma_1, \ldots, \gamma_p$ are iid such that

$$P(\gamma_i = 1) = \theta = 1 - P(\gamma_i = 0)$$

Spike and Slab LASSO (2)

The maximum aposteriori rule is given by reweighted LASSO

$$\hat{\beta}(\gamma) = argmin_{b \in R^p} \frac{1}{2} ||y - Xb||_2^2 + \lambda \sum_{i=1}^{p} w_i |b_i|$$

$$w_i = c\gamma_i + (1 - \gamma_i)$$

Prior for γ: $\gamma_1, \ldots, \gamma_p$ are iid such that

$$P(\gamma_i = 1) = \theta = 1 - P(\gamma_i = 0)$$

In consecutive iterations of SSL γ_i is replaced with

$$\pi_i^t = P(\gamma_i = 1|\beta^t, c) = \frac{c\theta e^{-c|\beta_i^t|\lambda_0}}{c\theta e^{-c|\beta_i^t|\lambda_0} + (1-\theta)e^{-|\beta_i^t|\lambda_0}}$$

and then a new estimate $\hat{\beta}^{t+1}$ is calculated by solving reweighted LASSO with the vector γ replaced with the vector π^t.

Adaptive SLOPE with missing values

Prior for β is given by

$$\pi(\beta|\gamma, c, \sigma^2) \propto c^{\sum_{i=1}^{n} 1(\gamma_i=1)} \prod_{i=1}^{n} \mathrm{e}^{-w_i|\beta_i|\lambda_{r(W\beta,i)}},$$

where W is the diagonal matrix with $W_{ii} = w_i$ and $\lambda = \lambda^{BH}$

Adaptive SLOPE with missing values

Prior for β is given by

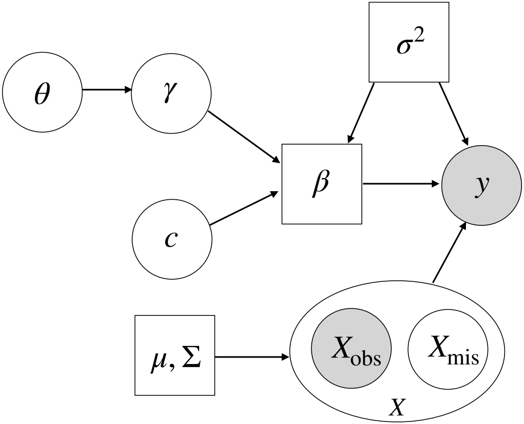
$$\pi(\beta|\gamma, c, \sigma^2) \propto c^{\sum_{i=1}^{n} 1(\gamma_i=1)} \prod_{i=1}^{n} \mathrm{e}^{-w_i|\beta_i|\lambda_{r(W\beta,i)}},$$

where W is the diagonal matrix with $W_{ii} = w_i$ and $\lambda = \lambda^{BH}$

Missing at Random (MAR) mechanism under assumption $X_i = (X_{i1}, \ldots, X_{ip})$ is normally distributed:

$$X_i \ \text{iid} \ \mathcal{N}_p(\mu, \Sigma), \quad i = 1, \cdots, n.$$

Graphical model of ABSLOPE



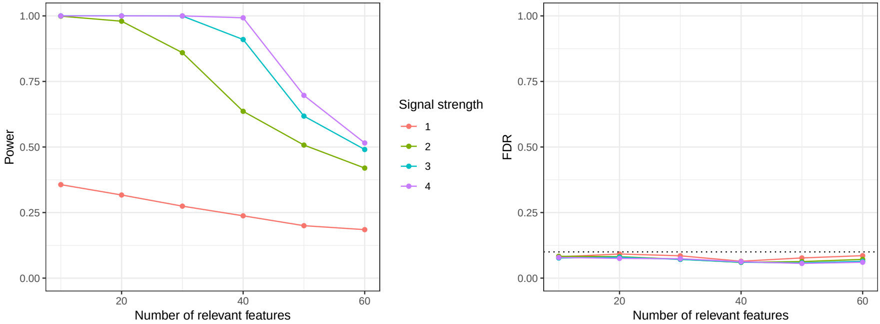
Stochastic approximation EM algorithm

- $\pi(\theta)$ - B(a,b), $\pi(c)$ - $U(0,1)$
- Gibbs sampling of latent variables : $\theta, c, \gamma, c, X_{mis}$
- Estimate parameters $\beta, \sigma, \mu, \Sigma$ by maximizing the complete-data likelihood with sampled values for the latent variables
- When $p > n$, Σ is estimated using the shrinkage estimator of Ledoit and Wolf (2004)
- Approximation of SAEM: ψ,

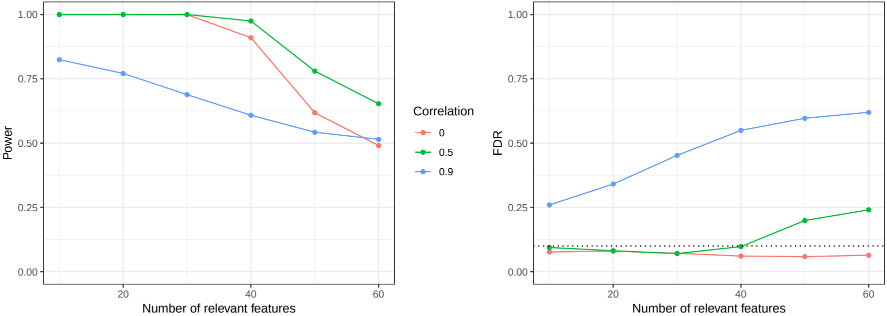
$$\psi^{t+1} = \psi^t + \eta_t \left[\hat{\psi}^t_{MLE} - \psi^t\right],$$

$\eta^t = 1$ for $t \in \{1, \dots, t_0\}$ and $\eta^t = \frac{1}{t-t_0}$ for $t > t_0$

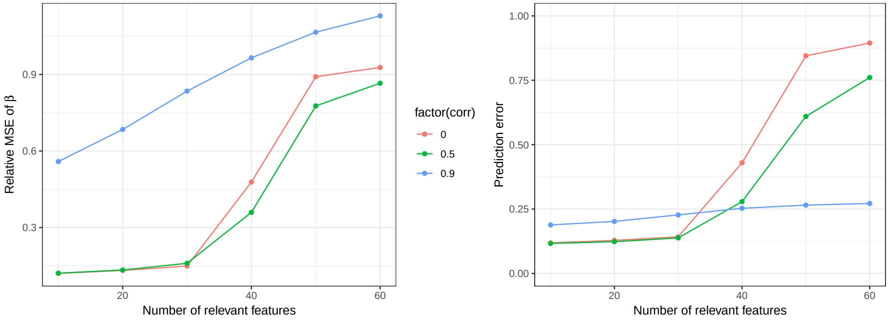
$n = p = 500$, $\rho = 0$, $Na = 10\%$, Power and FDR



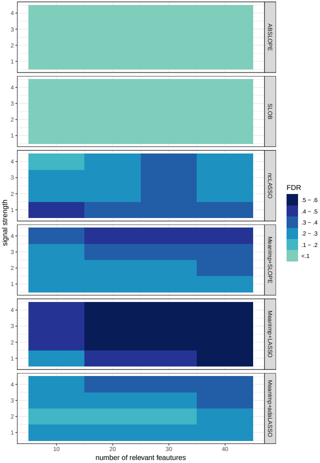
$n = p = 500$, $NA = 10\%$, strong signal, Power and FDR



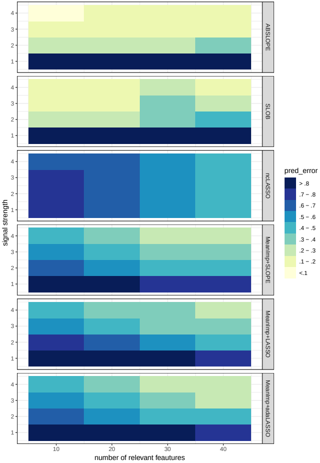
$n = p = 500$, $NA = 10\%$, strong signal, $MSE(\hat{\beta})$, $MSE(X\hat{\beta})$



500×500 dataset, 10% missingness, $\rho = 0.5$



(j) FDR



(k) Prediction error

Variables in the TraumaBase data set (APHP)

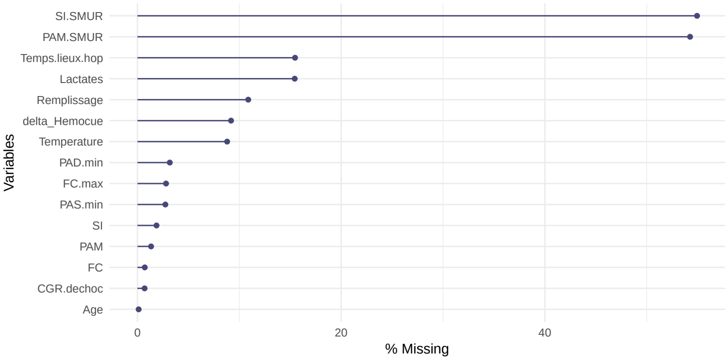
Goal - quick prediction of the level of platelets

- *Age:* Age
- *SI:* Shock index indicates level of occult shock based on heart rate (FC) and systolic blood pressure (PAS). $SI = \frac{FC}{PAS}$. Evaluated on arrival of hospital.
- *PAM:* Mean arterial pressure is an average blood pressure in an individual during a single cardiac cycle, based on systolic blood pressure (PAS) and diastolic blood pressure (PAD). $PAM = \frac{2PAD+PAS}{3}$. Evaluated on arrival of hospital.
- *delta_Hemocue:* The difference between the hemoglobin on arrival at hospital and that in the ambulance.
- *Temps.lieux.hop:* Time spent in hospital *i.e.*, medicalization time, in minutes.
- *Lactates:* The conjugate base of lactic acid.
- *Temperature:* Patient's body temperature.

Variables

- *FC:* heart rate measured on arrival of hospital.
- *Remplissage:* A volume expander is a type of intravenous therapy that has the function of providing volume for the circulatory system.
- *CGR.dechoc:* A binary index which indicates whether the transfusion of Red Blood Cells Concentrates is performed.
- *SI.SMUR:* Shock index measured on ambulance.
- *PAM.SMUR:* Mean arterial pressure measured in the ambulance.
- *FC.max:* Maximum value of measured heart rate in the ambulance.
- *PAS.min:* Minimum value of measured systolic blood pressure in the ambulance.
- *PAD.min:* Minimum value of measured diastolic blood pressure in the ambulance.

Percentage of missing values



Rysunek – Percentage of missing values in each pre-selected variable from TraumaBase.

Selected variables

Method	Variables selected
ABSLOPE	Age $*$ PAM.SMUR, delta_Hemocue $*$ Lactates, Lactates $*$ CGR.dechoc, FC $*$ PAS.min
LASSO	CGR.dechoc, PAS.min, Age $*$ Lactates, Age $*$ Remplissage, delta_Hemocue $*$ Lactates, delta_Hemocue $*$ Remplissage, Lactates $*$ Remplissage, Lactates $*$ CGR.dechoc
adaLASSO	Age $*$ Temps.lieux.hop, Age $*$ FC, Age $*$ PAM.SMUR, Age $*$ PAS.min, PAM $*$ FC, delta_Hemocue $*$ Remplissage, Lactates $*$ Remplissage, FC $*$ FC.max, FC $*$ PAS.min, Remplissage $*$ CGR.dechoc

Prediction error