Sorted L-One Penalized Estimation (SLOPE)

Malgorzata Bogdan

University of Wroclaw

Luminy, 12 July 2017

Outline

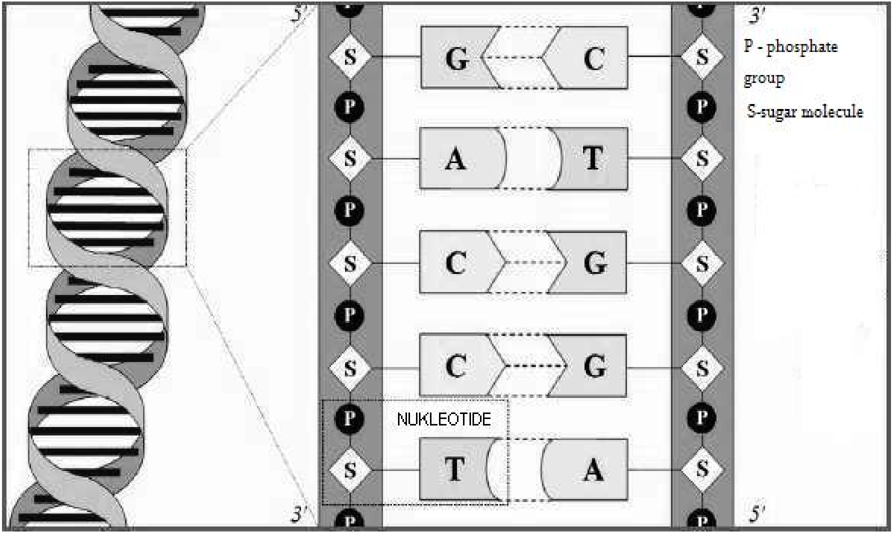
- Genetic motivation

Outline

- Genetic motivation
- Multiple Testing

Outline

- Genetic motivation
- Multiple Testing
- Model selection in multiple regression
  - Noiseless case - Linear programming - transition curve
  - Noisy case - LASSO, SLOPE
  - groupSLOPE
  - further extensions


5'
3'
S
P
G
C
A
T
NUKLEOTIDE
P - phosphate
group
S-sugar molecule
3'
5'

Genetic variability

- About 99,9% of genetic information is the same for all people.
- A ***polymorphism*** is a difference in DNA structure, which is present in at least 1% of population
- A ***Single Nucleotide Polymorphism(SNP)*** is a polymorphism with the difference in the single base:
  - A typical SNP: a position in DNA in which
    - 85% of population has Cytosine(C)
    - 15% has a Thymine(T).
- There are usually two forms of a SNP at a given locus
- three genotypes : AA, Aa, aa.

Main purpose

MAIN PURPOSE: finding the mutations in DNA sequence, that influence the trait of interest.

Main purpose

MAIN PURPOSE: finding the mutations in DNA sequence, that influence the trait of interest.

Y - quantitative trait

Main purpose

MAIN PURPOSE: finding the mutations in DNA sequence, that influence the trait of interest.

Y - quantitative trait

Examples: blood pressure, cholesterol level, gene expression level, response to the treatment

Data structure

$Y = (Y_1, \ldots, Y_n)^T$ - vector of trait values for n individuals

Data structure

$Y = (Y_1, \ldots, Y_n)^T$ - vector of trait values for n individuals

$G_{n \times p}$ - matrix of SNP genotypes

Data structure

$Y = (Y_1, \ldots, Y_n)^T$ - vector of trait values for n individuals

$G_{n \times p}$ - matrix of SNP genotypes

Usual coding for additive effects:

$$X_{ij} = \begin{cases} 0 & \text{if} \quad G_{ij} = aa \\ 1 & \text{if} \quad G_{ij} = Aa \\ 2 & \text{if} \quad G_{ij} = AA \end{cases}$$

Data structure

$Y = (Y_1, \ldots, Y_n)^T$ - vector of trait values for n individuals

$G_{n \times p}$ - matrix of SNP genotypes

Usual coding for additive effects:

$$X_{ij} = \begin{cases} 0 & \text{if} \quad G_{ij} = aa \\ 1 & \text{if} \quad G_{ij} = Aa \\ 2 & \text{if} \quad G_{ij} = AA \end{cases}$$

Usually $n \approx k \times 100$ or $k \times 1000$, $p \approx k \times 10,000$ or $p \approx k \times 100,000$

Data structure

$Y = (Y_1, \ldots, Y_n)^T$ - vector of trait values for n individuals

$G_{n \times p}$ - matrix of SNP genotypes

Usual coding for additive effects:

$$X_{ij} = \begin{cases} 0 & \text{if} \quad G_{ij} = aa \\ 1 & \text{if} \quad G_{ij} = Aa \\ 2 & \text{if} \quad G_{ij} = AA \end{cases}$$

Usually $n \approx k \times 100$ or $k \times 1000$, $p \approx k \times 10,000$ or $p \approx k \times 100,000$

The dependency between SNPs is of a short range - even relatively close SNPs can be modeled as independent random variables.

Multiple testing (1)

Simple regression model for i-th gene:

$$Y_j = \beta_0 + \beta_i X_{ij} + \epsilon_j, \quad \epsilon_j \sim N(0, \tau^2)$$

Multiple testing (1)

Simple regression model for i-th gene:

$$Y_j = \beta_0 + \beta_i X_{ij} + \epsilon_j, \quad \epsilon_j \sim N(0, \tau^2)$$

$$\hat{\beta}_i \text{ - estimated } \beta_i,\ \hat{\beta}_i \sim N(\beta_i, \sigma_i^2), \quad i = 1, \ldots, p$$

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

Multiple testing (1)

Simple regression model for i-th gene:

$$Y_j = \beta_0 + \beta_i X_{ij} + \epsilon_j, \quad \epsilon_j \sim N(0, \tau^2)$$

$$\hat{\beta}_i \text{ - estimated } \beta_i,\ \hat{\beta}_i \sim N(\beta_i, \sigma_i^2), \quad i = 1, \ldots, p$$

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

$$\text{Reject } H_{0i} \text{ when } |\hat{\beta}_i| > c_i$$

Multiple testing (1)

Simple regression model for i-th gene:

$$Y_j = \beta_0 + \beta_i X_{ij} + \epsilon_j, \quad \epsilon_j \sim N(0, \tau^2)$$

$$\hat{\beta}_i \text{ - estimated } \beta_i,\ \hat{\beta}_i \sim N(\beta_i, \sigma_i^2), \quad i = 1, \ldots, p$$

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

$$\text{Reject } H_{0i} \text{ when } |\hat{\beta}_i| > c_i$$

$$\text{Significance level: } \alpha = P_{H_{0i}}(|T_i| > c)$$

Multiple testing (1)

Simple regression model for i-th gene:

$$Y_j = \beta_0 + \beta_i X_{ij} + \epsilon_j, \quad \epsilon_j \sim N(0, \tau^2)$$

$$\hat{\beta}_i \text{ - estimated } \beta_i,\ \hat{\beta}_i \sim N(\beta_i, \sigma_i^2), \quad i = 1, \dots, p$$

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

$$\text{Reject } H_{0i} \text{ when } |\hat{\beta}_i| > c_i$$

$$\text{Significance level: } \alpha = P_{H_{0i}}(|T_i| > c)$$

	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)

Simple regression model for i-th gene:

$$Y_j = \beta_0 + \beta_i X_{ij} + \epsilon_j, \quad \epsilon_j \sim N(0, \tau^2)$$

$$\hat{\beta}_i \text{ - estimated } \beta_i,\ \hat{\beta}_i \sim N(\beta_i, \sigma_i^2), \quad i = 1, \ldots, p$$

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

$$\text{Reject } H_{0i} \text{ when } |\hat{\beta}_i| > c_i$$

$$\text{Significance level: } \alpha = P_{H_{0i}}(|T_i| > c)$$

	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)

Simple regression model for i-th gene:

$$Y_j = \beta_0 + \beta_i X_{ij} + \epsilon_j, \quad \epsilon_j \sim N(0, \tau^2)$$

$$\hat{\beta}_i \text{ - estimated } \beta_i, \; \hat{\beta}_i \sim N(\beta_i, \sigma_i^2), \quad i = 1, \ldots, p$$

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

$$\text{Reject } H_{0i} \text{ when } |\hat{\beta}_i| > c_i$$

$$\text{Significance level: } \alpha = P_{H_{0i}}(|T_i| > c)$$

	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)

Simple regression model for i-th gene:

$$Y_j = \beta_0 + \beta_i X_{ij} + \epsilon_j, \quad \epsilon_j \sim N(0, \tau^2)$$

$$\hat{\beta}_i \text{ - estimated } \beta_i,\ \hat{\beta}_i \sim N(\beta_i, \sigma_i^2), \quad i = 1, \ldots, p$$

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

$$\text{Reject } H_{0i} \text{ when } |\hat{\beta}_i| > c_i$$

$$\text{Significance level: } \alpha = P_{H_{0i}}(|T_i| > c)$$

	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

$T_1, \ldots, T_n$ independent $T_i \sim N(\mu_i, 1)$, $H_{0i} : \mu_i = 0$ vs $\mu_i \neq 0$
Bonferroni correction: Use significance level $\frac{\alpha}{p}$.

Multiple testing procedures

$T_1, \ldots, T_n$ independent $T_i \sim N(\mu_i, 1)$, $H_{0i} : \mu_i = 0$ vs $\mu_i \neq 0$
Bonferroni correction: Use significance level $\frac{\alpha}{p}$.

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

Multiple testing procedures

$T_1, \ldots, T_n$ independent $T_i \sim N(\mu_i, 1)$, $H_{0i} : \mu_i = 0$ vs $\mu_i \neq 0$
Bonferroni correction: Use significance level $\frac{\alpha}{p}$.

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

Benjamini-Hochberg procedure:

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

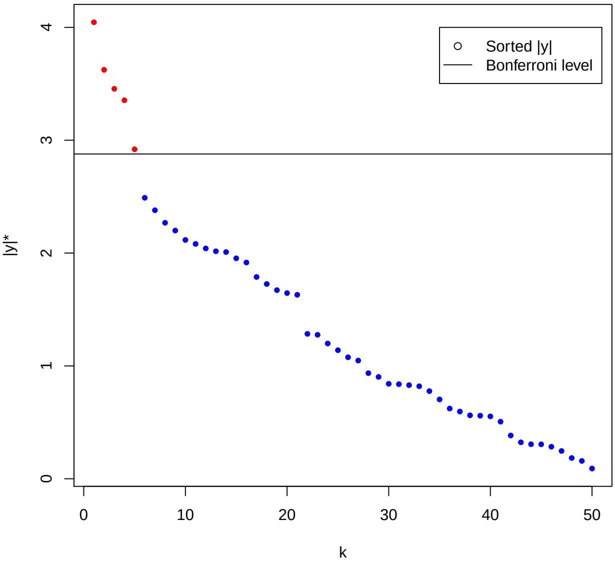
(2) Find the largest index i such that

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

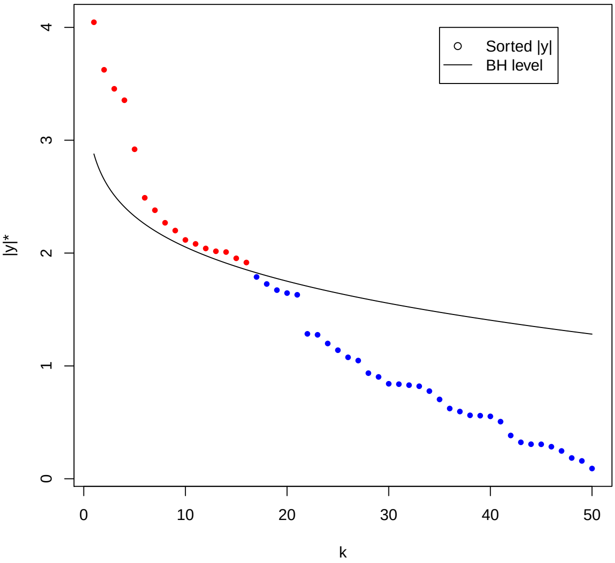
Call this index i_{SU}.

(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 $T_1, \ldots, T_p$ are independent then BH controls FDR at:

FWER and FDR control

For Bonferroni correction $FWER \leq \alpha$

(Benjamini,Hochberg, 1995) If $T_1, \ldots, T_p$ 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, $p_0 = |\{i : \mu_i = 0\}|$

FWER and FDR control

For Bonferroni correction $FWER \leq \alpha$

(Benjamini,Hochberg, 1995) If $T_1, \ldots, T_p$ 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, $p_0 = |\{i : \mu_i = 0\}|$

(Benjamini, Yekutieli, 2001) If test statistics are dependent then BH controls FDR at the level $\alpha \frac{p_0}{p}$ if $|T|_{(i)}$ is compared with $\sigma \Phi^{-1}\left(1 - \frac{i\alpha}{p \sum_{i=1}^{p} \frac{1}{i}}\right)$.

Asymptotic optimality (1)

Abramovich, Benjamini, Donoho and Johnstone, Ann.Statist. 2006

Asymptotic optimality (1)

Abramovich, Benjamini, Donoho and Johnstone, Ann.Statist. 2006

$$T_i \sim N(\beta_i, 1), \text{ independent for } i = 1, \ldots, p$$

Asymptotic optimality (1)

Abramovich, Benjamini, Donoho and Johnstone, Ann.Statist. 2006

$$T_i \sim N(\beta_i, 1), \quad \text{independent for} \quad i = 1, \ldots, p$$

Goal - estimation of $\beta = (\beta_1, \ldots, \beta_p)$

Asymptotic optimality (1)

Abramovich, Benjamini, Donoho and Johnstone, Ann.Statist. 2006

$$T_i \sim N(\beta_i, 1), \quad \text{independent for} \quad i = 1, \ldots, p$$

Goal - estimation of $\beta = (\beta_1, \ldots, \beta_p)$

$$\tilde{\beta}_i = T_i \quad \text{if BH rejects} \quad H_{0i} : \beta_i = 0$$

Asymptotic optimality (1)

Abramovich, Benjamini, Donoho and Johnstone, Ann.Statist. 2006

$$T_i \sim N(\beta_i, 1), \quad \text{independent for} \quad i = 1, \ldots, p$$

Goal - estimation of $\beta = (\beta_1, \ldots, \beta_p)$

$$\tilde{\beta}_i = T_i \quad \text{if BH rejects} \quad H_{0i} : \beta_i = 0$$

$$\tilde{\beta}_i = 0 \quad \text{otherwise}$$

Optimality (2)

$$\theta = \frac{p - p_0}{p}$$ - proportion of alternative hypotheses,

Optimality (2)

$$\theta = \frac{p - p_0}{p}$$ - proportion of alternative hypotheses,

q_p - nominal FDR level

Optimality (2)

$$\theta = \frac{p - p_0}{p}$$ - proportion of alternative hypotheses,

q_p - nominal FDR level

$$R_p = \inf_{\hat{\beta}} \sup_{\beta:\theta<\eta_p} E||\hat{\beta} - \beta||^r$$

Optimality (2)

$$\theta = \frac{p - p_0}{p}$$ - proportion of alternative hypotheses,

q_p - nominal FDR level

$$R_p = \inf_{\hat{\beta}} \sup_{\beta:\theta<\eta_p} E||\hat{\beta} - \beta||^r$$

Theorem

Assumptions:

- $q_p \to c \in [0, 1/2]$ *and* $q_p > \gamma/\log(p)$ *for some* $\gamma > 0$
- $\eta_p \in [p^{-1}\log^5 p, p^{-\delta}], \quad \delta > 0$
- $r \in (0, 2]$

$$\sup_{\beta:\theta<\eta_p} E||\tilde{\beta} - \beta||^r = R_p(1 + o(1))$$

Classification risk

Bogdan, Chakrabarti, Frommlet, Ghosh, Ann.Statist. 2011
Neuvial, Roquain, Ann. Statist. 2012

Classification risk

Bogdan, Chakrabarti, Frommlet, Ghosh, Ann.Statist. 2011
Neuvial, Roquain, Ann. Statist. 2012
Bayes risk, γ_0 - loss for type I error, γ_A - loss for type II error

Classification risk

Bogdan, Chakrabarti, Frommlet, Ghosh, Ann.Statist. 2011
Neuvial, Roquain, Ann. Statist. 2012
Bayes risk, γ_0 - loss for type I error, γ_A - loss for type II error
$\beta_i \sim (1-\theta)\delta_0 + \theta N(0,\tau^2)$

Classification risk

Bogdan, Chakrabarti, Frommlet, Ghosh, Ann.Statist. 2011
Neuvial, Roquain, Ann. Statist. 2012
Bayes risk, γ_0 - loss for type I error, γ_A - loss for type II error
$\beta_i \sim (1-\theta)\delta_0 + \theta N(0,\tau^2)$
$T_i \sim (1-\theta)N(0,1) + \theta N(0,1+\tau^2)$

Classification risk

Bogdan, Chakrabarti, Frommlet, Ghosh, Ann.Statist. 2011
Neuvial, Roquain, Ann. Statist. 2012
Bayes risk, γ_0 - loss for type I error, γ_A - loss for type II error
$\beta_i \sim (1-\theta)\delta_0 + \theta N(0, \tau^2)$
$T_i \sim (1-\theta)N(0,1) + \theta N(0, 1+\tau^2)$
Bayes oracle $\rightarrow$ Bayes classifier

Classification risk

Bogdan, Chakrabarti, Frommlet, Ghosh, Ann.Statist. 2011
Neuvial, Roquain, Ann. Statist. 2012
Bayes risk, γ_0 - loss for type I error, γ_A - loss for type II error
$\beta_i \sim (1-\theta)\delta_0 + \theta N(0, \tau^2)$
$T_i \sim (1-\theta)N(0,1) + \theta N(0, 1+\tau^2)$
Bayes oracle $\rightarrow$ Bayes classifier
The rule is Asymptotically Bayes Optimal under Sparsity (ABOS) if $\lim \frac{R}{R_{opt}} \rightarrow 1$ (as $p \rightarrow \infty$)

Classification risk

Bogdan, Chakrabarti, Frommlet, Ghosh, Ann.Statist. 2011
Neuvial, Roquain, Ann. Statist. 2012
Bayes risk, γ_0 - loss for type I error, γ_A - loss for type II error
$\beta_i \sim (1-\theta)\delta_0 + \theta N(0,\tau^2)$
$T_i \sim (1-\theta)N(0,1) + \theta N(0, 1+\tau^2)$
Bayes oracle $\rightarrow$ Bayes classifier
The rule is Asymptotically Bayes Optimal under Sparsity (ABOS) if
$\lim \frac{R}{R_{opt}} \rightarrow 1$ (as $p \rightarrow \infty$)
BH is ABOS if $\theta \propto p^{-\beta}$, $\beta \in (0,1]$, $\tau \propto \sqrt{2\beta \log p}$

Classification risk

Bogdan, Chakrabarti, Frommlet, Ghosh, Ann.Statist. 2011
Neuvial, Roquain, Ann. Statist. 2012
Bayes risk, γ_0 - loss for type I error, γ_A - loss for type II error
$\beta_i \sim (1-\theta)\delta_0 + \theta N(0,\tau^2)$
$T_i \sim (1-\theta)N(0,1) + \theta N(0,1+\tau^2)$
Bayes oracle $\rightarrow$ Bayes classifier
The rule is Asymptotically Bayes Optimal under Sparsity (ABOS) if $\lim \frac{R}{R_{opt}} \rightarrow 1$ (as $p \rightarrow \infty$)
BH is ABOS if $\theta \propto p^{-\beta}$, $\beta \in (0,1]$, $\tau \propto \sqrt{2\beta \log p}$
Bonferroni correction is ABOS if $\beta = 1$

Multiple regression model - Noisless case - Transition curve of Donoho and Tanner (2005)

$$Y = X_{n \times p}\beta$$

Multiple regression model - Noisless case - Transition curve of Donoho and Tanner (2005)

$$Y = X_{n \times p}\beta$$

$k = ||\beta||_0 \leq min(n, p)$ - number of nonzero elements in β

Multiple regression model - Noisless case - Transition curve of Donoho and Tanner (2005)

$$Y = X_{n \times p}\beta$$

$k = ||\beta||_0 \leq min(n, p)$ - number of nonzero elements in β

If $p > n$ minimize $||\beta||_0$ subject to $Y = X\beta$ - problem NP hard

Multiple regression model - Noisless case - Transition curve of Donoho and Tanner (2005)

$$Y = X_{n \times p}\beta$$

$k = ||\beta||_0 \leq min(n, p)$ - number of nonzero elements in β

If $p > n$ minimize $||\beta||_0$ subject to $Y = X\beta$ - problem NP hard

Convex relaxation: minimize $||\beta||_1 = \sum_{i=1}^{n} |\beta_i|$ subject to $Y = X\beta$.

Multiple regression model - Noisless case - Transition curve of Donoho and Tanner (2005)

$$Y = X_{n \times p}\beta$$

$k = ||\beta||_0 \leq min(n, p)$ - number of nonzero elements in β

If $p > n$ minimize $||\beta||_0$ subject to $Y = X\beta$ - problem NP hard

Convex relaxation: minimize $||\beta||_1 = \sum_{i=1}^{n} |\beta_i|$ subject to $Y = X\beta$.

$$\rho = k/p \text{ and } \delta = n/p$$

Multiple regression model - Noisless case - Transition curve of Donoho and Tanner (2005)

$$Y = X_{n \times p} \beta$$

$k = ||\beta||_0 \leq min(n, p)$ - number of nonzero elements in β

If $p > n$ minimize $||\beta||_0$ subject to $Y = X\beta$ - problem NP hard

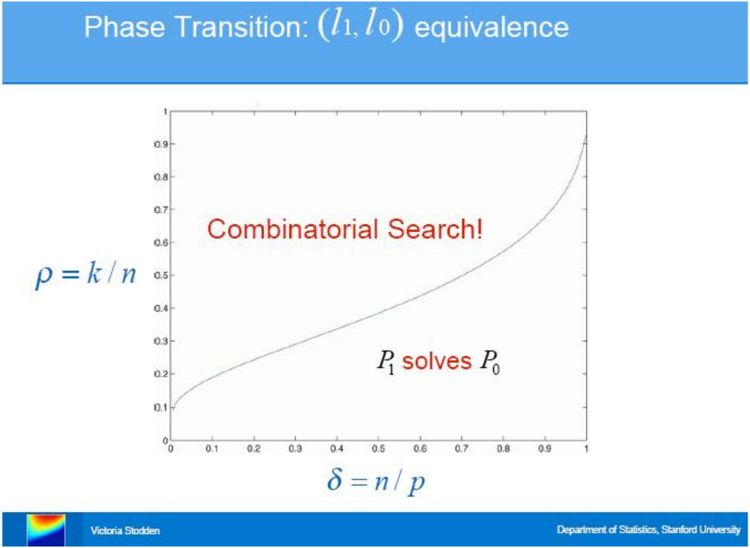
Convex relaxation: minimize $||\beta||_1 = \sum_{i=1}^{n} |\beta_i|$ subject to $Y = X\beta$.

$$\rho = k/p \;\text{ and }\; \delta = n/p$$

Theorem

If X_{ij} are iid $N(0, \sigma^2)$ then with probability converging to 1 the linear program can identify the true solution if $\rho < \rho(\delta)$. If $\rho > \rho(\delta)$ then the probability of recovering the true solution converges to 0.

Transition curve (2)



Noisy case - statistical problem

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

Noisy case - statistical problem

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

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

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

Noisy case - statistical problem

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

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

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

In statistics this procedure is called LASSO (Tibshirani, 1996)

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.

SLOPE (Bogdan,van den Berg, Sabatti, Su and Candès, AOAS, 2015)

SLOPE is an extension of LASSO aimed at control of FDR rather than FWER

- SLOPE is defined as solution to

$$\beta_{SL} = argmin_b \left\{ \frac{1}{2} \|y - Xb\|_2^2 + \sum_{i=1}^{p} \lambda_i |b|_{(i)} \right\}, \qquad \text{(SLOPE)}$$

where $|b|_{(1)} \geq \ldots \geq |b|_{(p)}$ are ordered magnitudes of coefficients of b and $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$ is the sequence of tuning parameters

SLOPE (Bogdan,van den Berg, Sabatti, Su and Candès, AOAS, 2015)

SLOPE is an extension of LASSO aimed at control of FDR rather than FWER

- SLOPE is defined as solution to

$$\beta_{SL} = argmin_b \left\{ \frac{1}{2} \left\| y - Xb \right\|_2^2 + \sum_{i=1}^{p} \lambda_i \left| b \right|_{(i)} \right\}, \qquad \text{(SLOPE)}$$

  where $|b|_{(1)} \geq \ldots \geq |b|_{(p)}$ are ordered magnitudes of coefficients of b and $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$ is the sequence of tuning parameters
- The above optimization problem is convex and can be efficiently solved even for large design matrices

SLOPE (Bogdan,van den Berg, Sabatti, Su and Candès, AOAS, 2015)

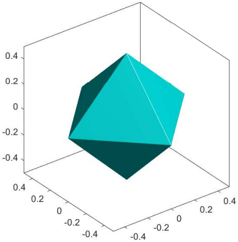
SLOPE is an extension of LASSO aimed at control of FDR rather than FWER

- SLOPE is defined as solution to

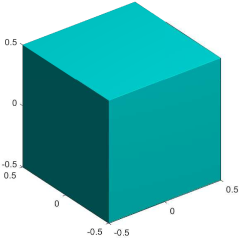
$$\beta_{SL} = argmin_b \left\{ \frac{1}{2} \|y - Xb\|_2^2 + \sum_{i=1}^{p} \lambda_i |b|_{(i)} \right\}, \qquad \text{(SLOPE)}$$

  where $|b|_{(1)} \geq \ldots \geq |b|_{(p)}$ are ordered magnitudes of coefficients of b and $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$ is the sequence of tuning parameters
- The above optimization problem is convex and can be efficiently solved even for large design matrices
- Sorted L-One Norm: $J_\lambda(b) = \sum_{i=1}^{p} \lambda_i |b|_{(i)}$ reduces to $||b||_1$ if $\lambda_1 = \ldots = \lambda_p$ and to $||b||_\infty$ if $\lambda_1 > \lambda_2 = \ldots = \lambda_p = 0$

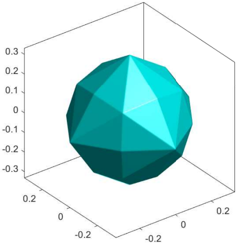
Unit balls for different SLOPE sequences by D.Brzyski



(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

When $X^T X = I$ SLOPE with

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

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

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*

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

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

Random designs - problems with shrinkage

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

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

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

Random designs - problems with shrinkage

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

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

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

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

Random designs - problems with shrinkage

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

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

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

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) and the level of sparsity

Random designs - problems with shrinkage

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

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

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

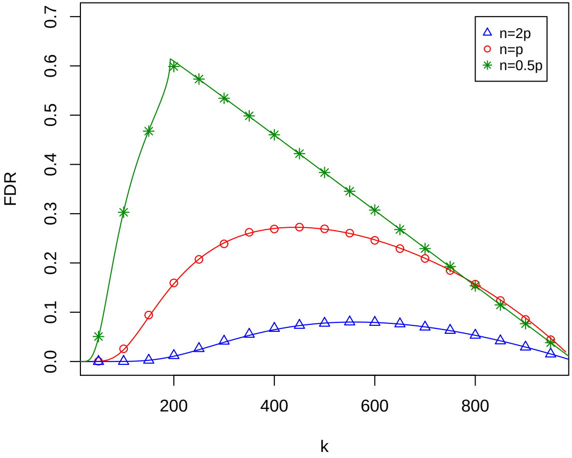
If $X^T X = I$ then $X_i' z = Z_i \sim N(0,1)$, $v_i = 0$ and H_{0i} is rejected if $\beta_i + Z_i > \lambda$

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

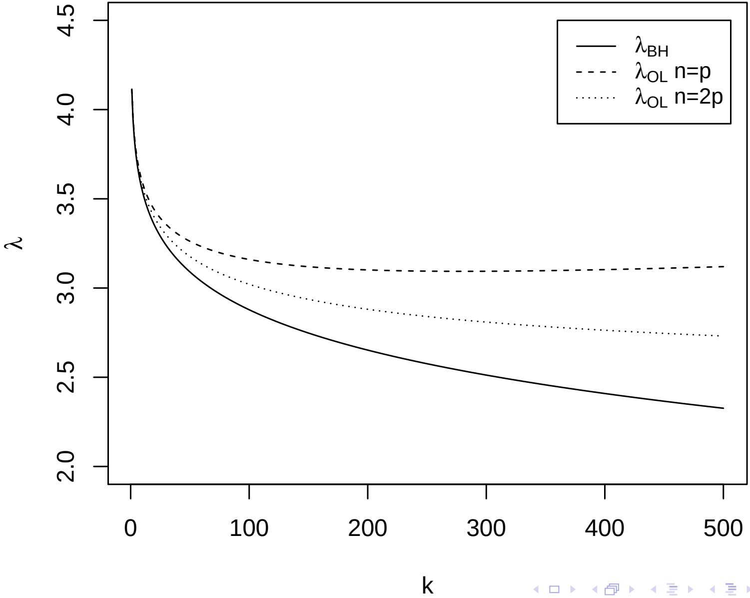
Quantification for LASSO using AMP theory - Bayati and Montanari (IEEE Trans. Infom.Theory, 2011)

Limits on FDR control

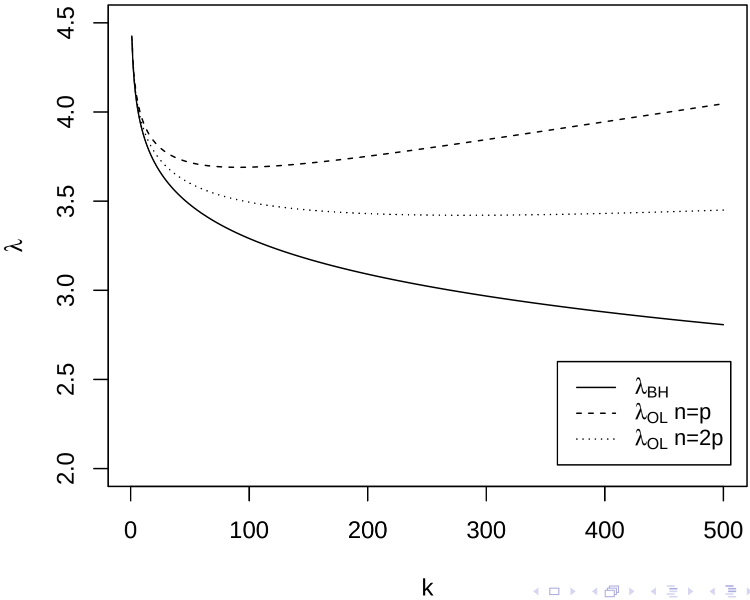
- Bogdan, van den Berg, Su and Candes (2013, arxive)
- Bogdan, Su and Candes (2017, to appear in Ann. Statist)



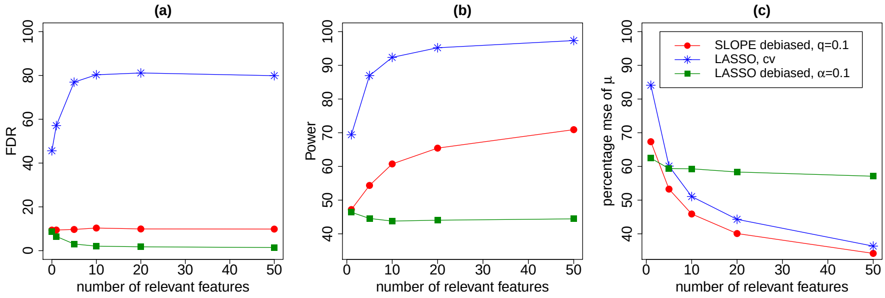
Thresholds, $q = 0.2$, $p = 5000$ (1)



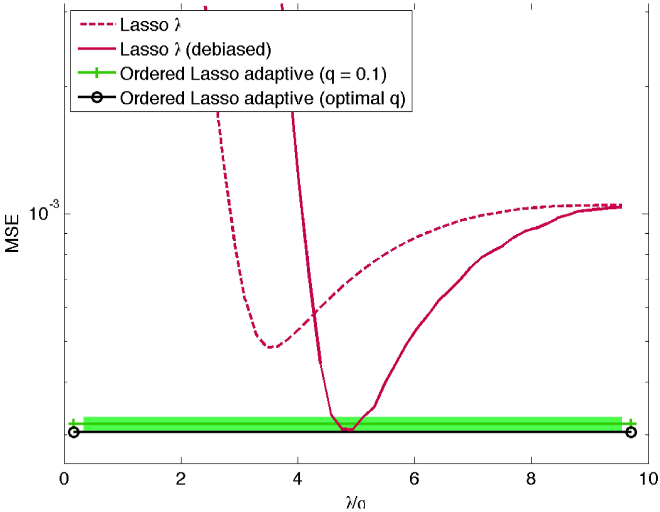
Thresholds, $q = 0.05$, $p = 5000$ (2)



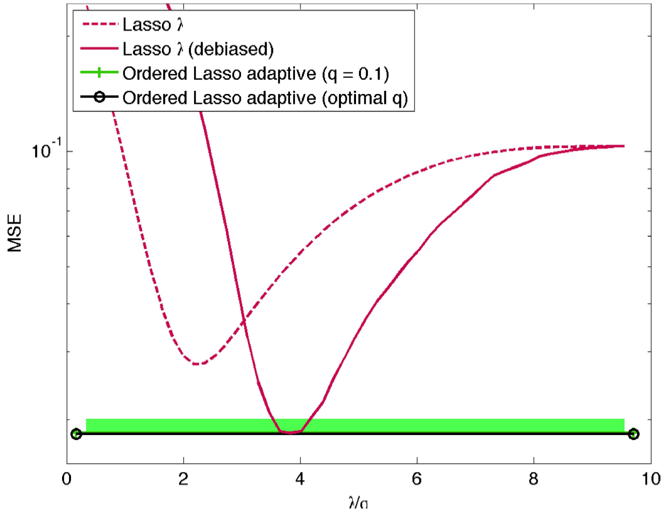
FDR, $p = n = 5000$, Gaussian design



MSE, $p = 262144$, $n = p/2$, $k = 10$, Gaussian design



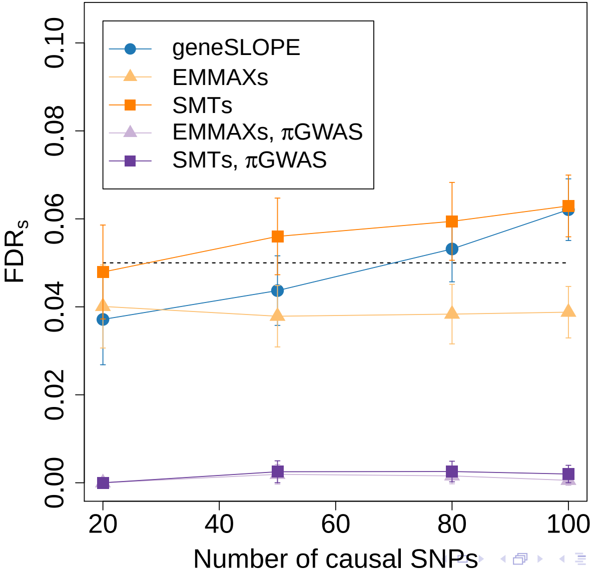
MSE, $p = 262144$, $n = p/2$, $k = 1000$, Gaussian design



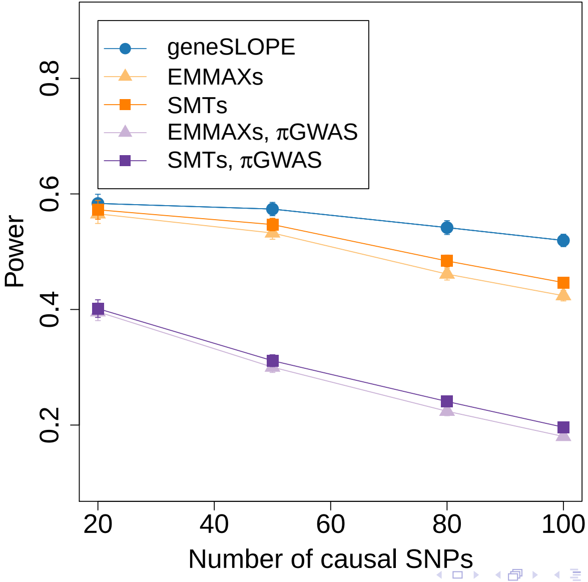
***

geneSLOPE - application for full GWAS data

Brzyski, Peterson, Sobczyk, Bogdan, Candés, Sabatti (Genetics, 2017)



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

Identification of groups of predictors (Brzyski, Gossmann, Su, Bogdan, arxiv 2016, under revision for JASA, ENAR Young Researcher Award for Damian Brzyski, March 2017):

$$[[\beta]]_I := \left(\|X_{I_1}\beta_{I_1}\|_2, \ldots, \|X_{I_m}\beta_{I_m}\|_2\right)^{\mathsf{T}} \quad .$$

$$\beta^{gS} := argmin_b \;\; \left\{\frac{1}{2}\big\|y - Xb\big\|_2^2 + \sigma J_\lambda\big(W[[b]]_I\big)\right\},$$

where W is a diagonal matrix with $W_{i,i} := w_i$, for $i = 1, \ldots, m$.

Group FDR

$$Rg := \left|\left\{i : \ \|X_{I_i}\beta^{gS}_{I_i}\|_2 \neq 0\right\}\right|$$

$$Vg := \left|\left\{i : \ \|X_{I_i}\beta_{I_i}\|_2 = 0, \ \|X_{I_i}\beta^{gS}_{I_i}\|_2 \neq 0\right\}\right|$$

We define the false discovery rate for groups (gFDR) as

$$gFDR := \mathbb{E}\left[\frac{Vg}{\max\{Rg, 1\}}\right]$$

gFDR control

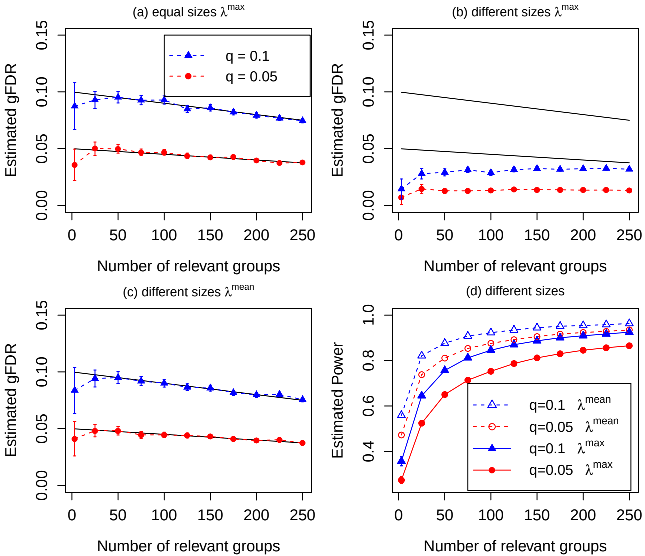
Theorem

Let the design matrix X satisfy $X_{I_i}^\intercal X_{I_j} = 0$, for any $i \neq j$. Denote the number of zero coefficients in $[[\beta]]_I$ by m_0 and let $w_1, \ldots, w_m$ be positive numbers. Moreover, define the sequence of regularizing parameters $\lambda^{\max} = (\lambda_1^{\max}, \ldots, \lambda_m^{\max})^\intercal$, with

$$\lambda_i^{\max} := \max_{j=1,\ldots,m} \left\{ \frac{1}{w_j} F_{\chi_{l_j}}^{-1} \left(1 - \frac{q \cdot i}{m} \right) \right\}, \tag{3}$$

where $F_{\chi_{l_j}}$ is a cumulative distribution function of χ distribution with l_j degrees of freedom. Then any solution, $\beta^{\mathrm{ES}(gS)}$, to gSLOPE generates the same vector $[[\beta^{gS}]]_I$ and it holds

$$gFDR = \mathbb{E}\left[\frac{Vg}{\max\{Rg, 1\}} \right] \leq q \cdot \frac{m_0}{m}.$$



(a) equal sizes λ^max
(b) different sizes λ^max
(c) different sizes λ^mean
(d) different sizes
Estimated gFDR
Estimated Power
Number of relevant groups
q = 0.1
q = 0.05
q=0.1 λ^mean
q=0.05 λ^mean
q=0.1 λ^max
q=0.05 λ^max

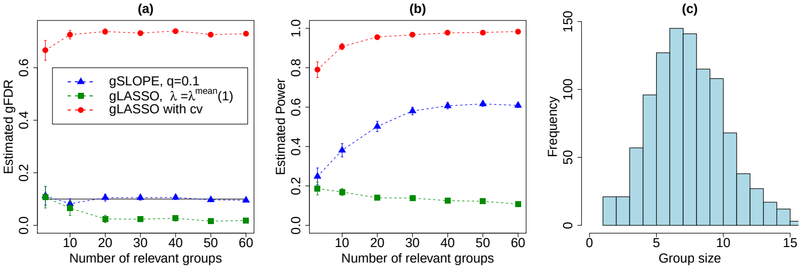
***

Averaged λ

$$\lambda_r^{mean} := \overline{F}^{-1}\left(1 - \frac{qr}{m}\right) \quad \text{for} \quad \overline{F}(x) := \frac{1}{m}\sum_{i=1}^{m} F_{w_i^{-1}\chi_{l_i}}(x) \ , \qquad (4)$$

where $F_{w_i^{-1}\chi_{l_i}}$ is the cumulative distribution function of scaled chi distribution with l_i degrees of freedom and scale $\mathcal{S} = w_i^{-1}$.
Gossmann, Brzyski et al. (2017) - Proof of FDR control in case when averaging is with respect to the distribution generating groups of different sizes and the signal is randomly placed between groups.

Simulations under Gaussian design, $n = 5000$, $m = 1000$, $p = 7917$



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

Applications for GWAS

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

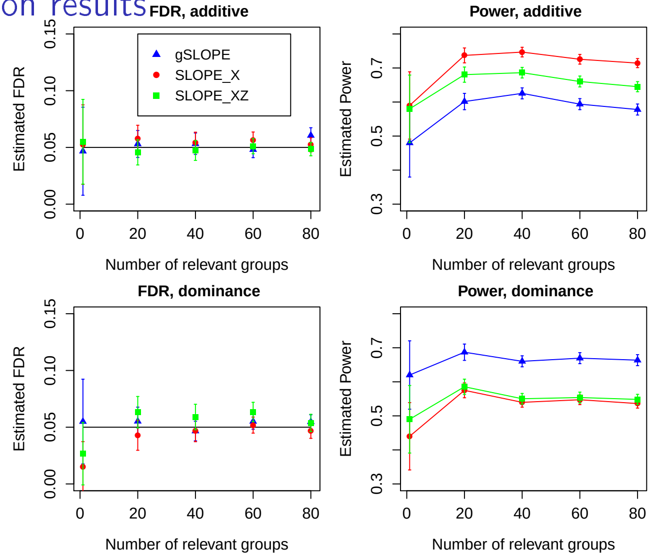
Scenario 1: $Y = X\beta + z$ - additive model

Scenario 2: modeling dominance

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

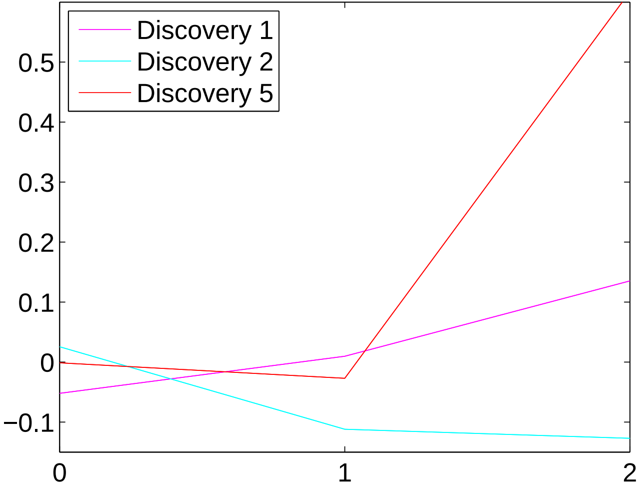
$$y = [X, Z][\beta_X', \beta_Z']' + \epsilon \quad .$$

Simulation results



Genes Influencing Level of Triglicerides

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



Logistic SLOPE (Michal Kos, Sangkyun Lee, M.Bogdan

$l(b)$ - log-likelihood

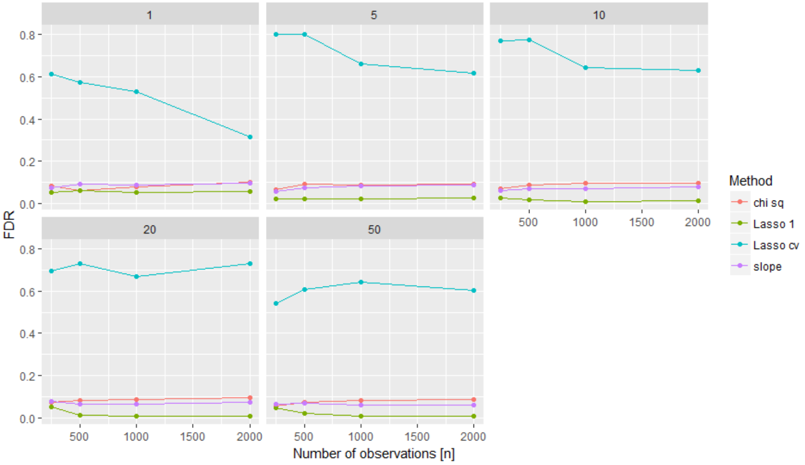
$$\min_b \left(-l(b) + \sum \lambda_i |b_{(i)}|\right)$$

$\lambda_i = 0.5\Phi^{-1}\left(1 - \frac{qi}{2p}\right)$

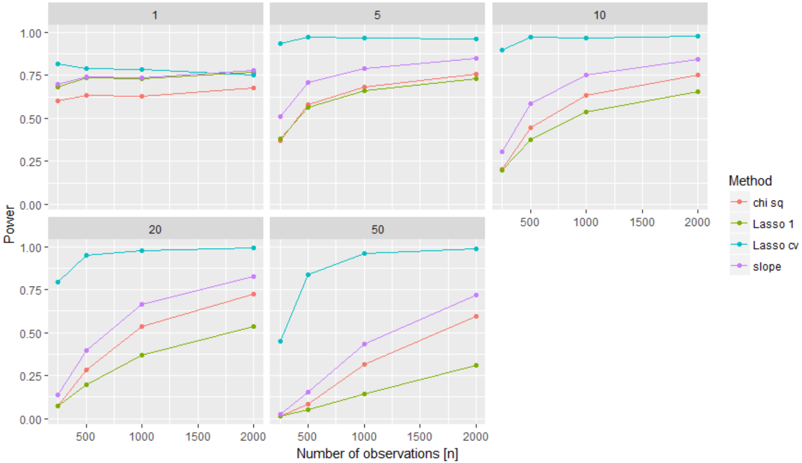
Asymptotic FDR control when X is a random design such that

- for all i, j: x_{ij} are independent and $Ex_{ij} = 0$
- elements in each column of design matrix X are i.i.d.
- $\mathbb{E}[X^T X] = I$ ($\Rightarrow var(x_{ij}) = n^{-1}$)
- there exist constant M that for all i and j:
  $$\frac{|x_{ij}|}{\sqrt{var(x_{ij})}} = x_{ij}\sqrt{n} \leqslant M$$
- p and β are fixed, $n \to \infty$

FDR

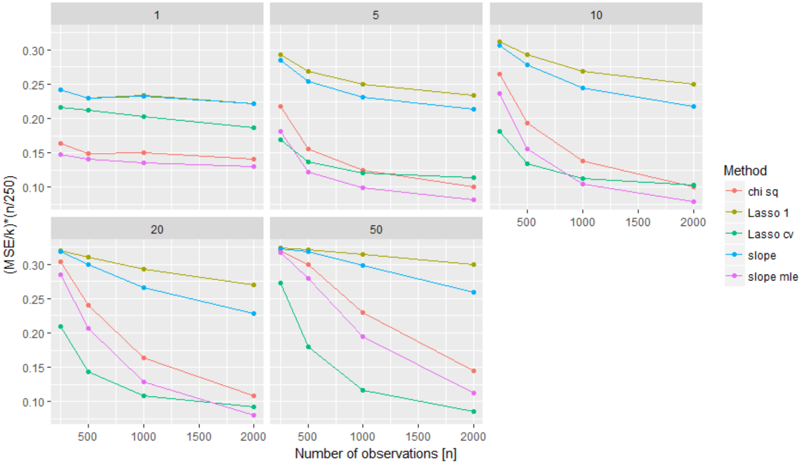

1
5
10
20
50
0.8
0.6
0.4
0.2
0.0
FDR
500
1000
1500
2000
Number of observations [n]
Method
chi sq
Lasso 1
Lasso cv
slope

Power



1
5
10
20
50
Power
1.00
0.75
0.50
0.25
0.00
500
1000
1500
2000
Number of observations [n]
Method
chi sq
Lasso 1
Lasso cv
slope

MSE



1
5
10
20
50
0.30
0.25
0.20
0.15
0.10
500
1000
1500
2000
(MSE/k)*(n/250)
Number of observations [n]
Method
chi sq
Lasso 1
Lasso cv
slope
slope mle

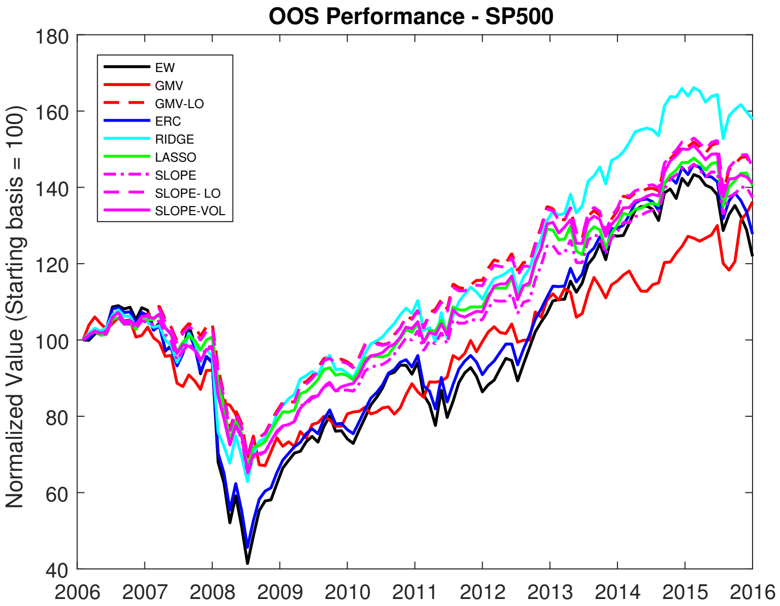
Portfolio Optimization, (Philipp Kremmer, Sangkyun Lee, M. Bogdan, Sandra Paterlini)

$R_{t\times k} = (R_1, \ldots, R_k)$ - asset returns,
$E(R) = \mu$, $R = F_{t\times r}B_{r\times k}$, $r << k$

$$\min_{w\in\mathbb{R}^k} \frac{\phi}{2} w'\Sigma w - \mu' w + J_\lambda(w) \tag{6}$$

$$\text{s.t.} \sum_{i=1}^{k} w_i = 1 \tag{7}$$

Evolution of Portfolio



Other results/current work

1. Brzyski, Gossmann, Su, Bogdan - Group SLOPE, under revision for JASA
2. Lee, Brzyski, Bogdan - Ordered Dantzig Selector, AISTAT 2016
3. Kos, Lee, Bogdan - Logistic regression - asymptotic FDR control under random designs (p fixed, $n \to \infty$).
4. Lee, Sobczyk, Bogdan - Graphical models - FDR control at the block level
4. Kremmer, Lee, Bogdan, Paterlini - Portfolio optimization - prediction under factorial (correlated) designs