Controlling false discovery rate via knockoffs

Rina Foygel Barber & Emmanuel Candès

Jan 21 2015

Code & demos available at `http://web.stanford.edu/~candes/Knockoffs/`
Paper available at `http://arxiv.org/abs/1404.5609`

Setting

An example:
Which mutations in the reverse transcriptase (RT) of HIV-1 determine susceptibility to reverse transcriptase inhibitors (RTIs)?

- $y_i \in \mathbb{R}$ = resistance of virus in sample i to a RTI-type drug
- $X_{ij} \in \{0, 1\}$ indicates if mutation j is present in virus sample i

How can we select mutations that determine drug resistance, in such a way that our answer will replicate in further trials?

Setting

Sparse linear model:

$$y = X \cdot \beta + z, \text{ where } z_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

- n observations, p features
- β is sparse

Setting

Goal: select a set of features X_j
that are likely to be relevant to the response y,
without too many false positives.

One way to measure performance:

$$\underbrace{\mathsf{FDR}}_{\text{False discovery rate}} = \mathbb{E}\left[\underbrace{\frac{\#\text{ false positives}}{\text{total }\#\text{ of features selected}}}_{\text{False discovery proportion}}\right] = \mathbb{E}\left[\frac{|S \cap \mathcal{H}_0|}{|S|}\right].$$

$S =$ set of selected features

$\mathcal{H}_0 =$ "null hypotheses" $= \{j : \beta_j^\star = 0\}$

Sparse regression

$$\text{Lasso:} \quad \beta_\lambda = \underset{\beta \in \mathbb{R}^p}{\arg\min} \left\{ \frac{1}{2} \|y - X \cdot \beta\|_2^2 + \lambda \|\beta\|_1 \right\}$$

Asymptotically, Lasso will select the correct model (at a good λ).

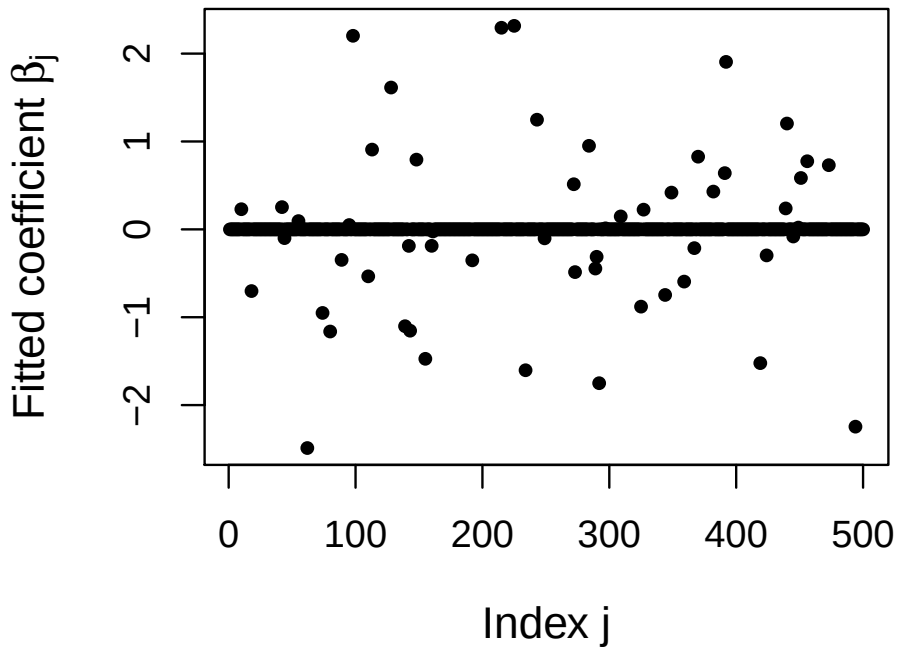
In practice for a finite sample,

- True positives & false positives intermixed along the Lasso path
- How to pick λ to balance FDR vs power?
- Need to account for correlations between X_j & weak signals that may have been missed on the Lasso path.

Sparse regression

Simulated data with $n = 1500$, $p = 500$.

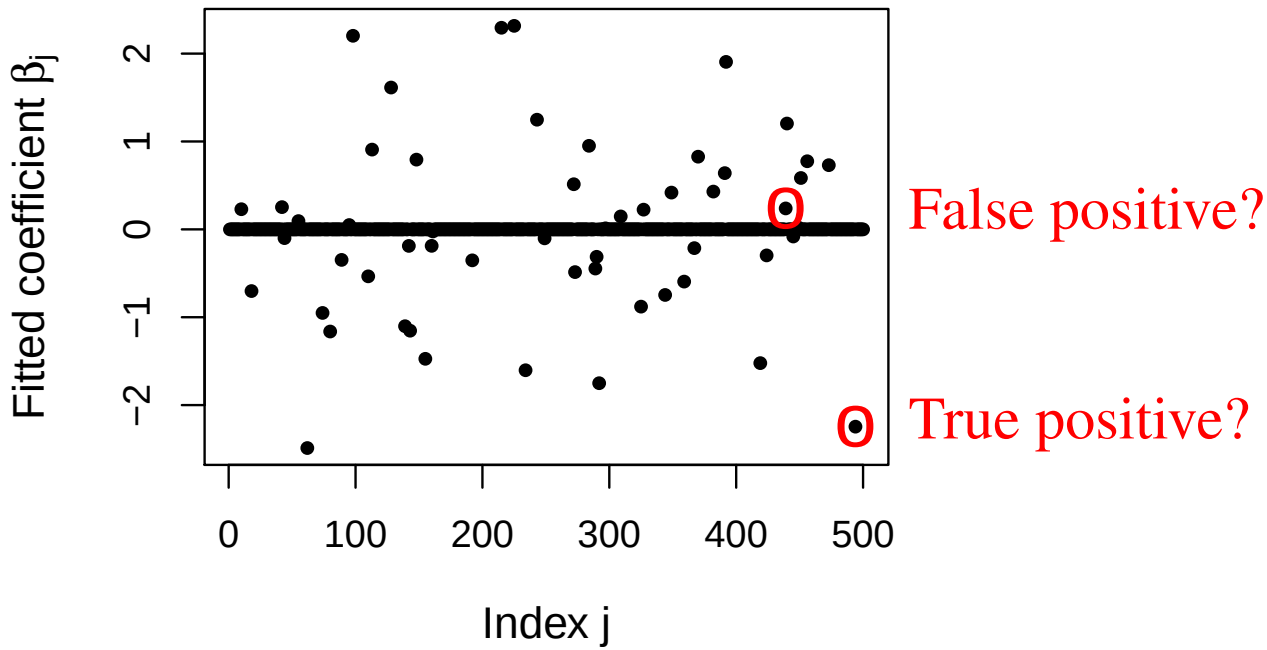
Lasso fitted model for $\lambda = 1.75$:



Sparse regression

Simulated data with $n = 1500$, $p = 500$.

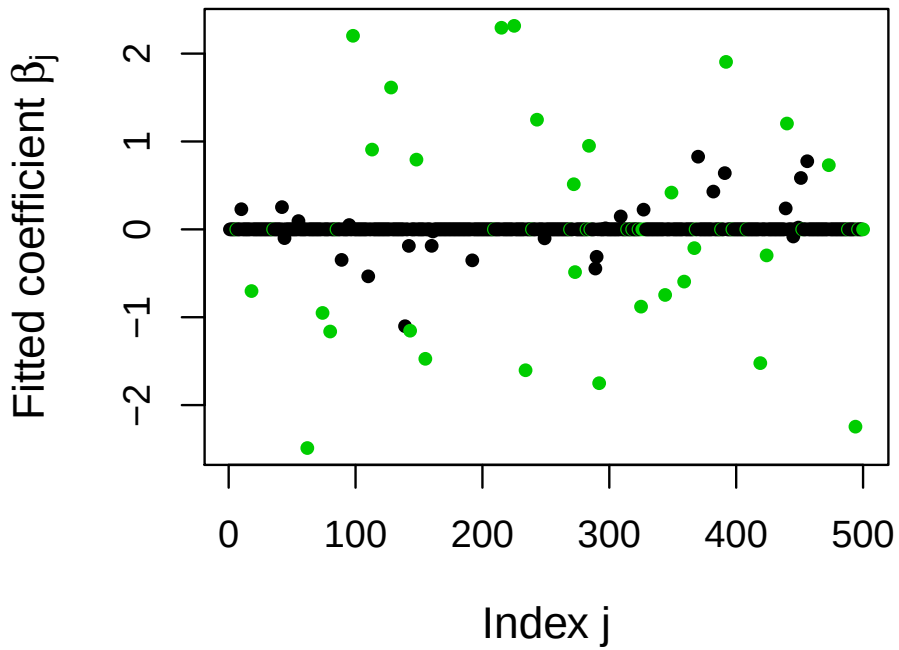
Lasso fitted model for $\lambda = 1.75$:



Sparse regression

Simulated data with $n = 1500$, $p = 500$.

Lasso fitted model for $\lambda = 1.75$:

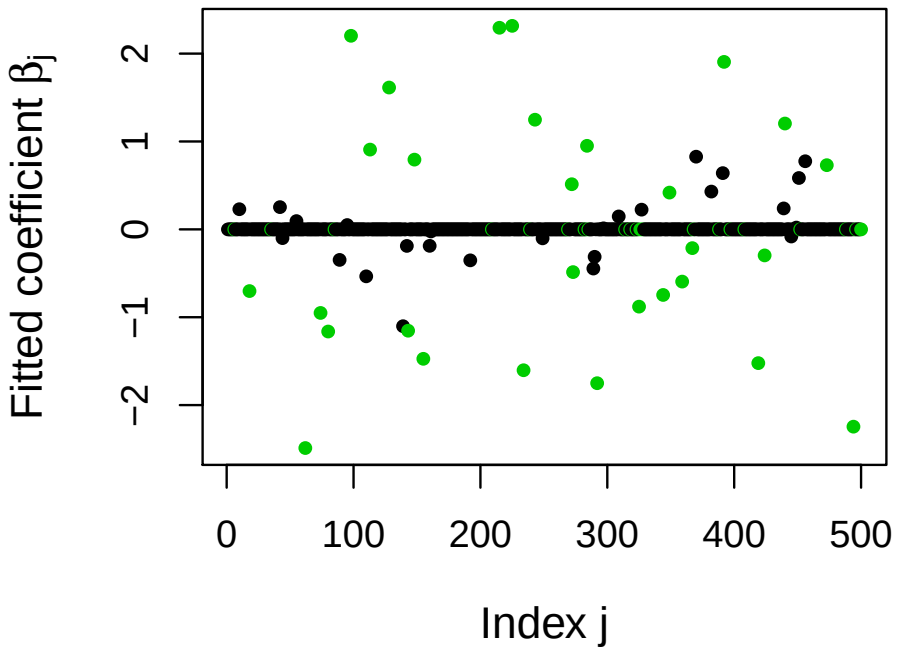


$\text{FDP} = \frac{26}{55} = 47\%$

Sparse regression

Simulated data with $n = 1500$, $p = 500$.

Lasso fitted model for $\lambda = 1.75$:



$\text{FDP} = \frac{26}{55} = 47\%$

To estimate FDP, would need to calculate distribution of β_j^λ for null j (would need to know σ^2, $\beta^\star$, ...). (Donoho et al 2009)

Construct knockoffs

Main idea:

For each feature X_j, construct a knockoff version $\widetilde{X}_j$.
The knockoffs serve as a "control group" $\Rightarrow$ can estimate FDP.

Setting:

- Require $n > p$ (ongoing work for high-dim. setting)
- Don't need to know σ^2
- Don't need any information about $\beta^\star$
- Will get an exact, finite-sample guarantee for FDR

Construct knockoffs

Construction:

- The knockoffs replicate the correlation structure of X:

$$\widetilde{X}_j^\top \widetilde{X}_k = X_j^\top X_k \text{ for all } j, k$$

- Also preserve correlations between knockoffs & originals:

$$\widetilde{X}_j^\top X_k = X_j^\top X_k \text{ for all } j \neq k$$

Augmented design matrix

$$\left[X \;\; \widetilde{X}\right] = (X_1 \;\; X_2 \ldots X_p \;\; \widetilde{X}_1 \;\; \widetilde{X}_2 \ldots \widetilde{X}_p) \in \mathbb{R}^{n \times 2p}$$

Construct knockoffs

How?

Define $\widetilde{X} = X \cdot (\mathbf{I}_p - 2\xi\Sigma^{-1}) + U \cdot C$, where:

$$\Sigma = X^\top X \succeq \xi \mathbf{I}_p$$

$$U = n \times p \text{ orthonormal matrix orthogonal to } X$$

$$C^\top C = 4(\xi \mathbf{I}_p - \xi^2 \Sigma^{-1}) \quad \text{(Cholesky decomposition)}$$

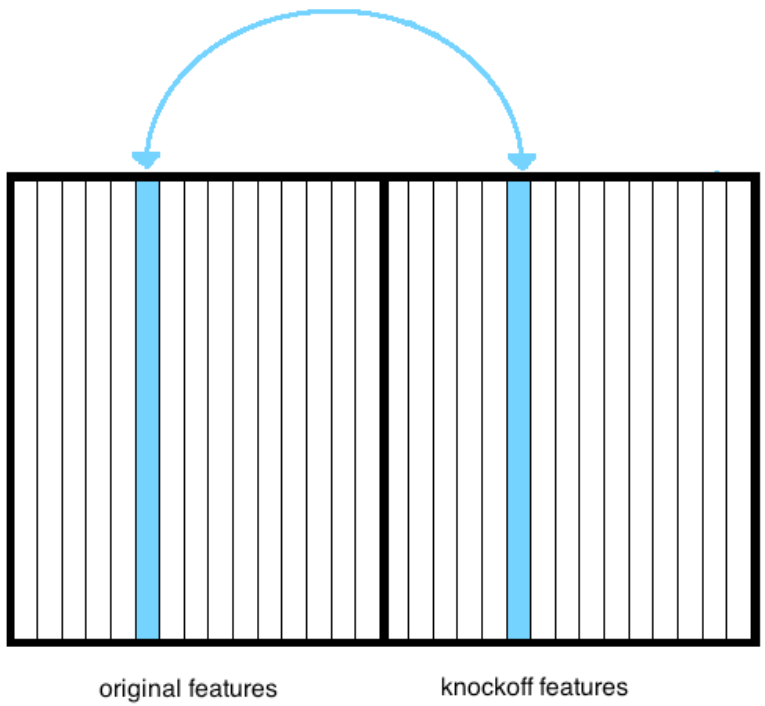
$$\Longrightarrow \quad [X \ \widetilde{X}]^\top [X \ \widetilde{X}] = \begin{pmatrix} \Sigma & \Sigma - 2\xi\mathbf{I}_p \\ \Sigma - 2\xi\mathbf{I}_p & \Sigma \end{pmatrix}$$

Construct knockoffs

Why?

For a null feature X_j,

$$X_j^\top y \;=\; X_j^\top X\beta^\star + X_j^\top z \;\overset{\mathcal{D}}{=}\; \widetilde{X}_j^\top X\beta^\star + \widetilde{X}_j^\top z \;=\; \widetilde{X}_j^\top y$$

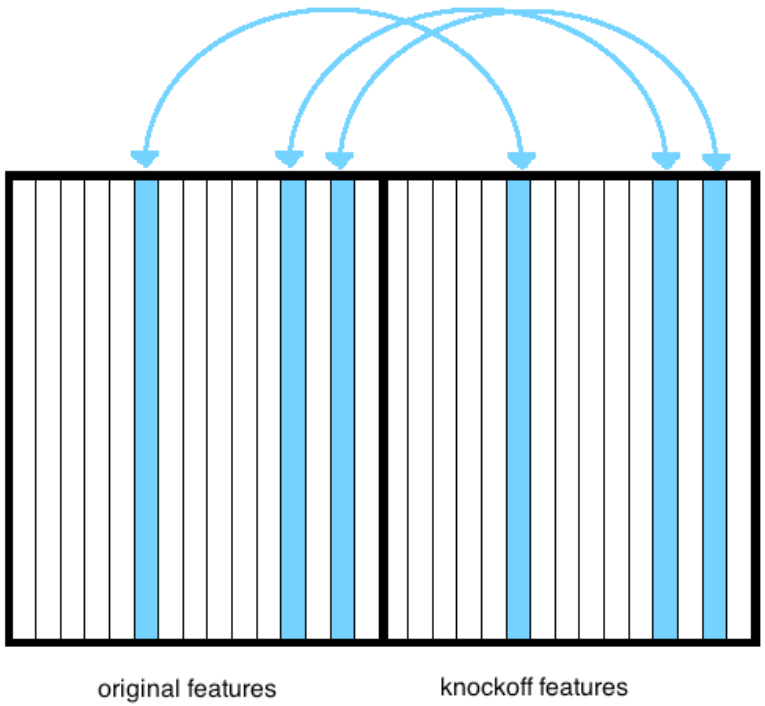


Construct knockoffs

Why?

For a null feature X_j,

$$X_j^\top y \;=\; X_j^\top X\beta^\star + X_j^\top z \;\overset{\mathcal{D}}{=}\; \widetilde{X}_j^\top X\beta^\star + \widetilde{X}_j^\top z \;=\; \widetilde{X}_j^\top y$$



Construct knockoffs

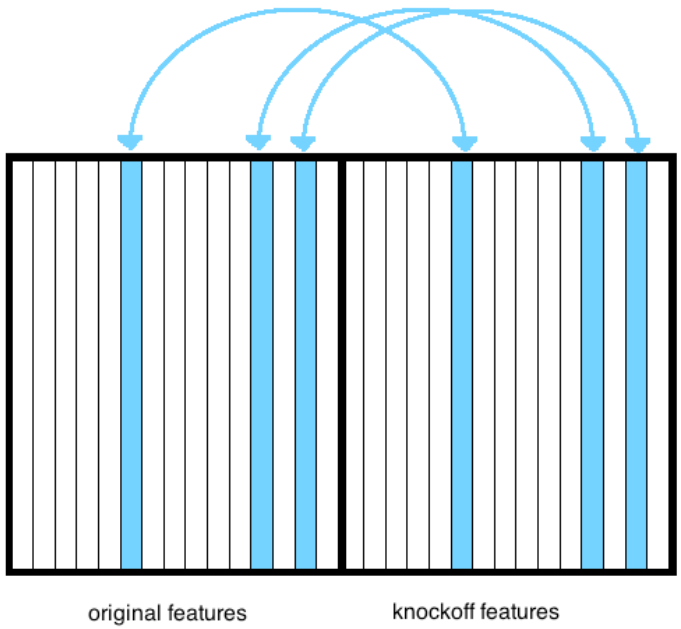
Lemma 1: Pairwise exchangeability property.

For any $N \subset \mathcal{H}_0$,

$$\left(\left[X \ \widetilde{X} \right]_{\text{swap}(N)} \right)^\top y \overset{\mathcal{D}}{=} \left[X \ \widetilde{X} \right]^\top y$$

$\Longrightarrow$ the knockoffs are a "control group" for the nulls

$$\left[X \ \widetilde{X} \right]_{\text{swap}(N)} =$$



Knockoff method

Steps:

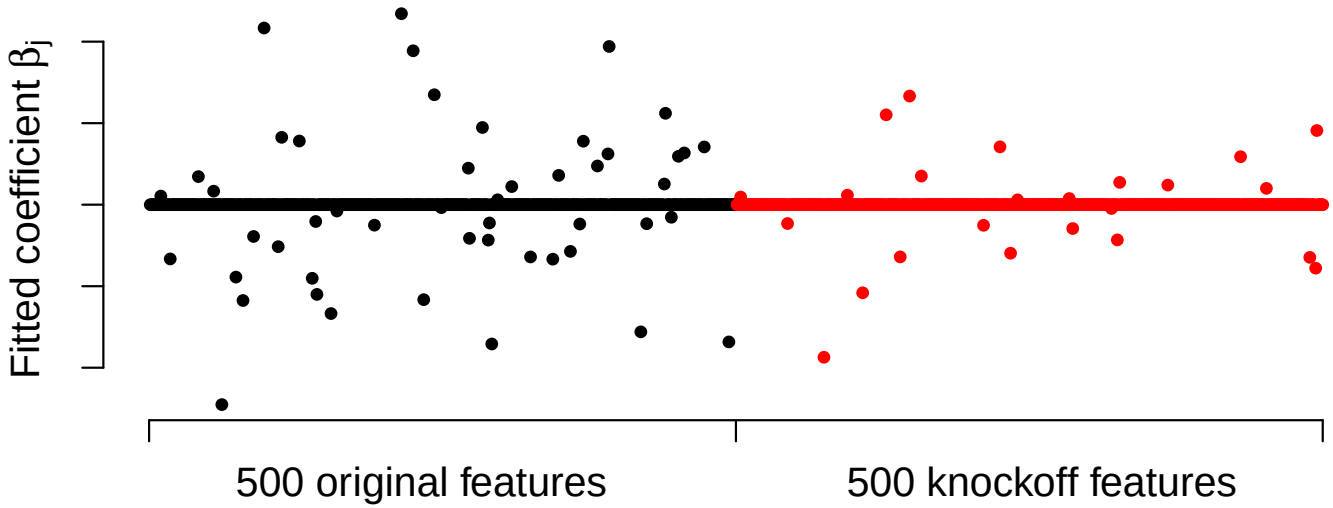
1. Construct knockoffs
2. Compute Lasso with augmented matrix:

$$\beta_\lambda = \underset{\beta \in \mathbb{R}^{2p}}{\arg\min} \left\{ \frac{1}{2} \left\| y - \left[X \;\; \widetilde{X} \right] \cdot \beta \right\|_2^2 + \lambda \left\| \beta \right\|_1 \right\}$$

3. Use $\widetilde{X}_j$ as a "control group" for X_j

Knockoff method

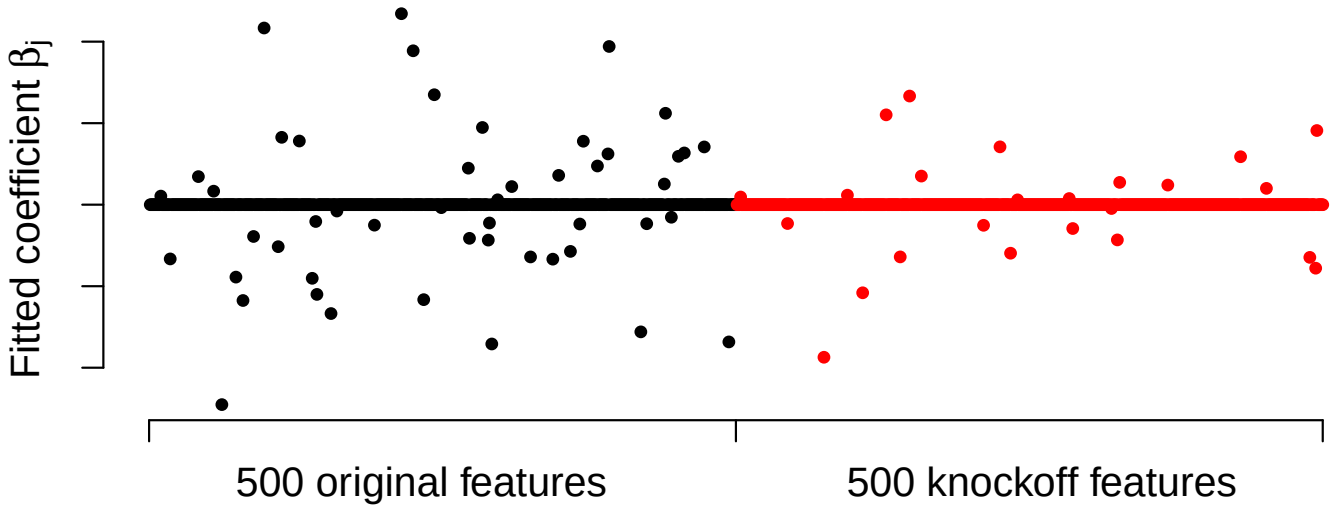
Fitted model for $\lambda = 1.75$ on the simulated dataset:



- Lasso selects 49 original features & 24 knockoff features

Knockoff method

Fitted model for $\lambda = 1.75$ on the simulated dataset:



- Lasso selects 49 original features & 24 knockoff features
- Pairwise exchangeability of the nulls
  $\Longrightarrow$ probably $\approx$ 24 false positives among the 49 original features

Knockoff method

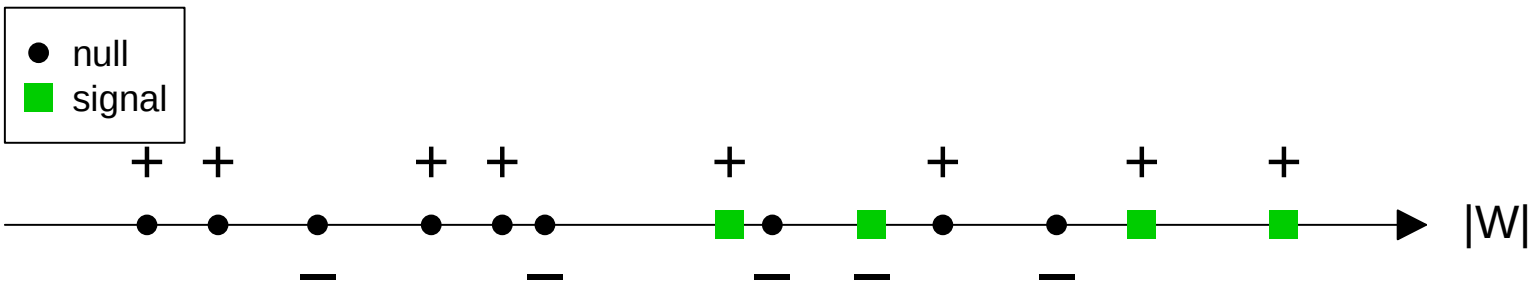
Compute Lasso on the entire path $\lambda \in [0, \infty)$.

$$\lambda_j = \sup\left\{\lambda \ : \ \beta_j^\lambda \neq 0\right\} = \text{ first time } X_j \text{ enters Lasso path}$$
$$\widetilde{\lambda}_j = \sup\left\{\lambda \ : \ \widetilde{\beta}_j^\lambda \neq 0\right\} = \text{ first time } \widetilde{X}_j \text{ enters Lasso path}$$

Then define statistics

$$W_j = \max\{\lambda_j, \widetilde{\lambda}_j\} \cdot \text{sign}(\lambda_j - \widetilde{\lambda}_j)$$

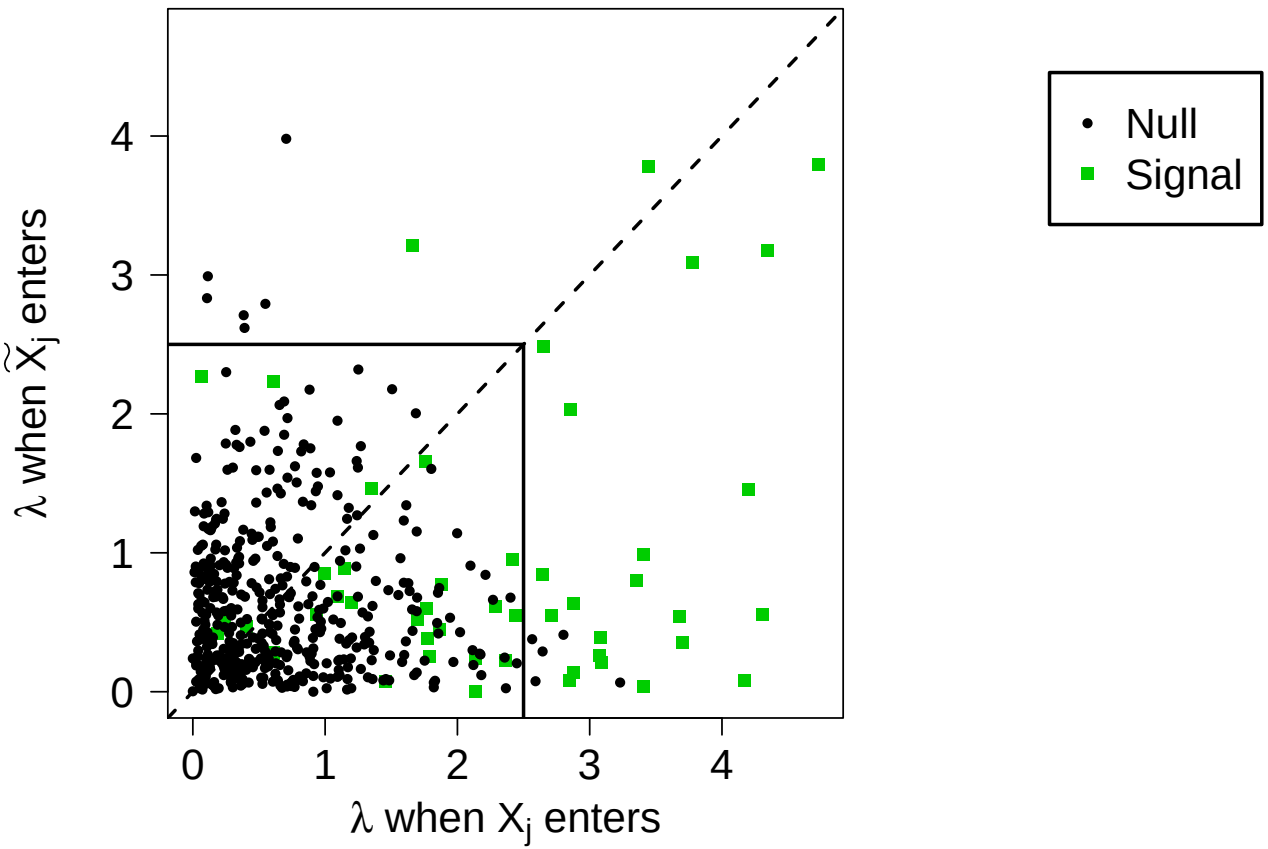
Knockoff method



$\lambda = 0$
variables enter late
(probably not significant)

$\lambda \to \infty$
variables enter early
(likely significant)

Knockoff method

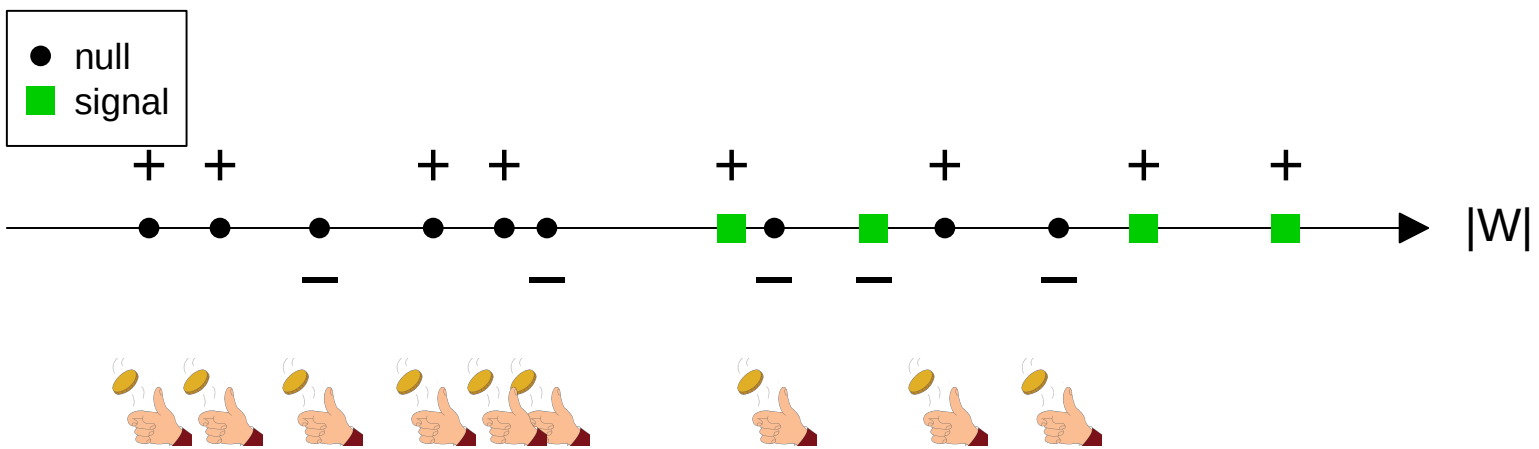


Knockoff method

Lemma 2: Pairwise exchangeability of the nulls $\Longrightarrow$

$$(W_1, W_2, \ldots, W_p) \stackrel{\mathcal{D}}{=} (|W_1| \cdot \epsilon_1, |W_2| \cdot \epsilon_2, \ldots, |W_p| \cdot \epsilon_p)$$

where $\epsilon_j = \text{sign}(W_j)$ for non-nulls and $\epsilon_j \stackrel{\text{iid}}{\sim} \{\pm 1\}$ for nulls.

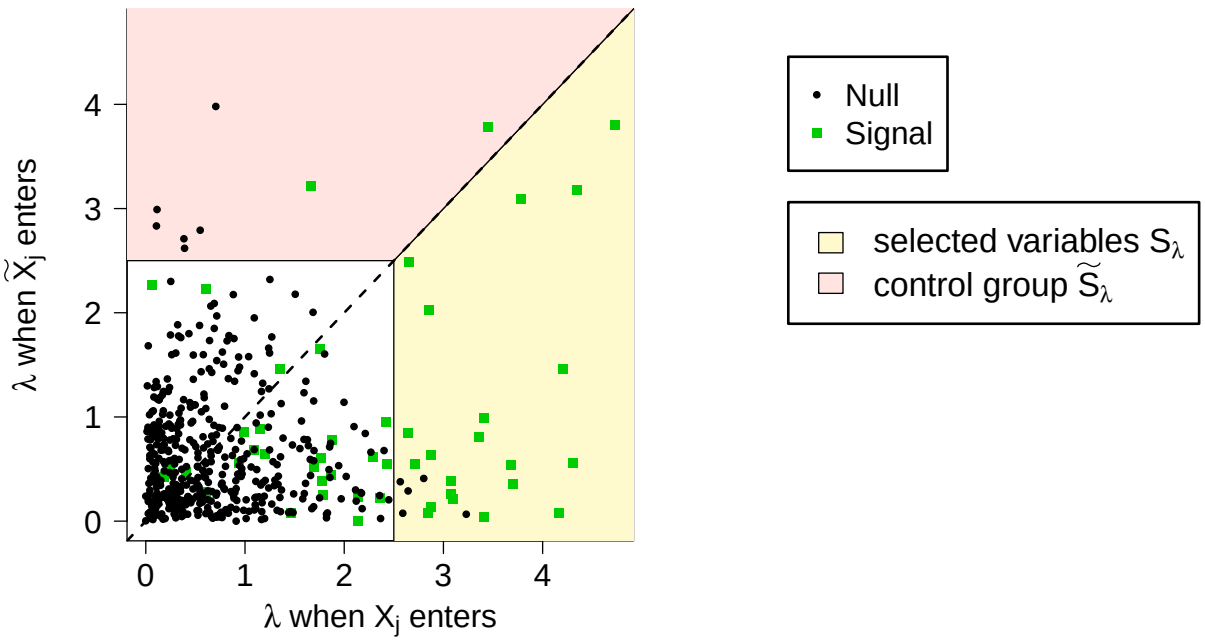


Knockoff method

$$\begin{aligned} \text{Selected variables:} \quad & S_\lambda = \{j : W_j \geq +\lambda\} \\ \text{Control group:} \quad & \widetilde{S}_\lambda = \{j : W_j \leq -\lambda\} \end{aligned} \qquad \rightsquigarrow \qquad \widehat{\mathsf{FDP}}(S_\lambda) := \frac{|\widetilde{S}_\lambda|}{|S_\lambda|}$$

Knockoff method

$$\text{Selected variables:} \quad S_\lambda = \{j : W_j \geq +\lambda\} \qquad \rightsquigarrow \qquad \widehat{\text{FDP}}(S_\lambda) := \frac{|\widetilde{S}_\lambda|}{|S_\lambda|}$$
$$\text{Control group:} \quad \widetilde{S}_\lambda = \{j : W_j \leq -\lambda\}$$



$$\text{FDP}(S_\lambda) = \frac{|S_\lambda \cap \mathcal{H}_0|}{|S_\lambda|} \approx \frac{|\widetilde{S}_\lambda \cap \mathcal{H}_0|}{|S_\lambda|} \leq \widehat{\text{FDP}}(S_\lambda)$$

Knockoff method

The knockoff filter: define

$$\widehat{\mathsf{FDP}}(S_\lambda) := \frac{|\widetilde{S}_\lambda|}{|S_\lambda|} = \frac{\#\{j : W_j \le -\lambda\}}{\#\{j : W_j \ge +\lambda\}} ,$$

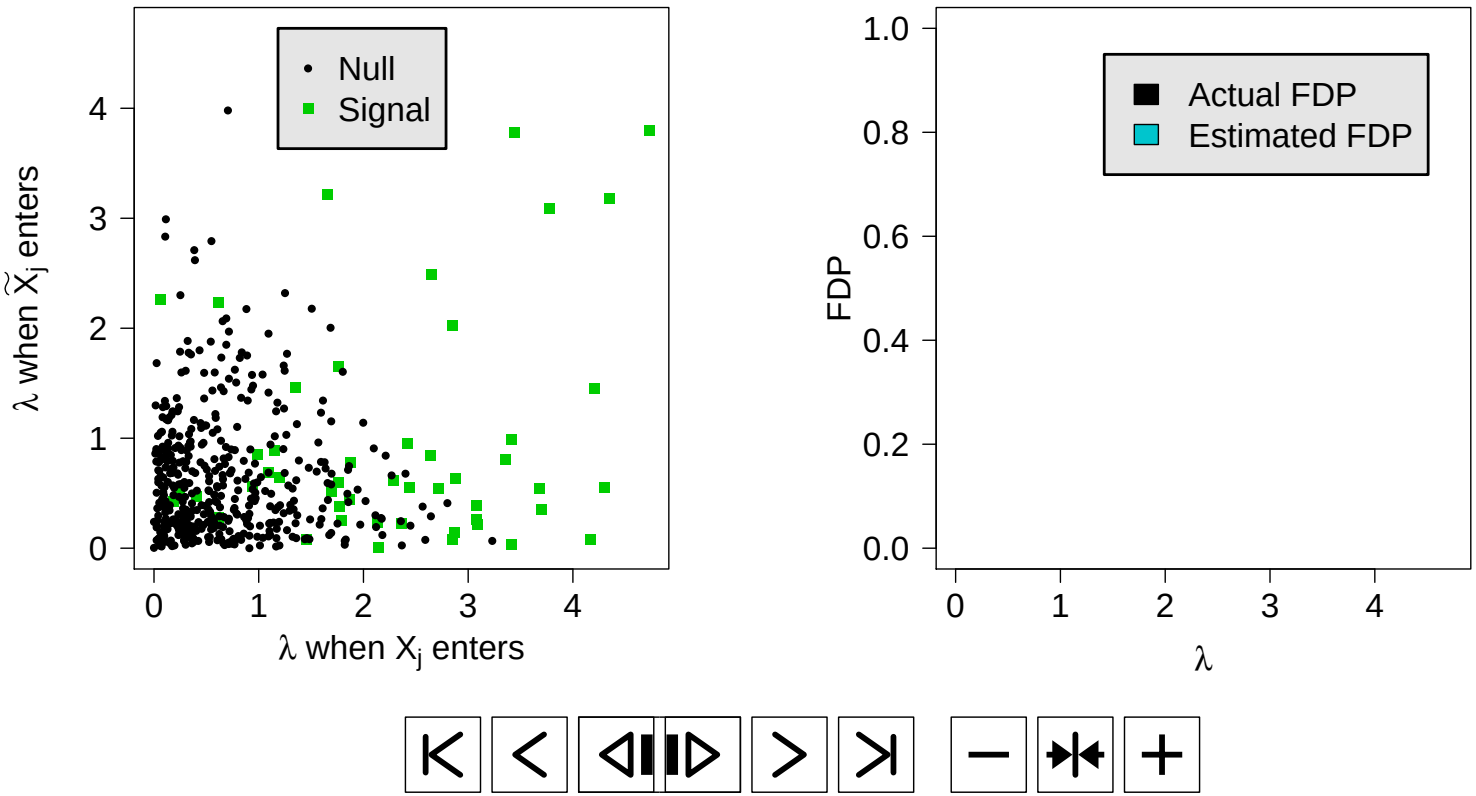
then choose

$$\Lambda = \min\left\{\lambda \ : \ \widehat{\mathsf{FDP}}(S_\lambda) \le q\right\} \quad (\text{or } \lambda = \infty \text{ if empty set})$$

and select the variable set

$$S_\Lambda = \{j : W_j \ge \Lambda\} .$$

Knockoff method



Theoretical guarantees

Theorem 1: For S_Λ chosen by the knockoff filter,

$$\mathbb{E}\left[\mathsf{mFDP}(S_\Lambda)\right] \leq q$$

where the modified FDP is given by

$$\mathsf{mFDP}(S) = \frac{\left|S \cap \mathcal{H}_0\right|}{\left|S\right| + q^{-1}} \ .$$

Theoretical guarantees

The knockoff+ filter: define

$$\widehat{\mathsf{FDP}}_+(S_\lambda) := \frac{|\widetilde{S}_\lambda| + 1}{|S_\lambda|} = \frac{\#\{j : W_j \le -\lambda\} + 1}{\#\{j : W_j \ge +\lambda\}},$$

then choose

$$\Lambda_+ = \min\left\{\lambda \ : \ \widehat{\mathsf{FDP}}_+(S_\lambda) \le q\right\} \quad (\text{or } \lambda = \infty \text{ if empty set})$$

and select the variable set

$$S_{\Lambda_+} = \{j : W_j \ge \Lambda_+\}.$$

Theoretical guarantees

Theorem 2: For S_{Λ_+} chosen by the knockoff+ filter,

$$\mathbb{E}\left[\mathsf{FDP}(S_{\Lambda_+})\right] \leq q \ .$$

Theoretical guarantees

Theorem 2: For S_{Λ_+} chosen by the knockoff+ filter,

$$\mathbb{E}\left[\mathsf{FDP}(S_{\Lambda_+})\right] \leq q\ .$$

Proof sketch:

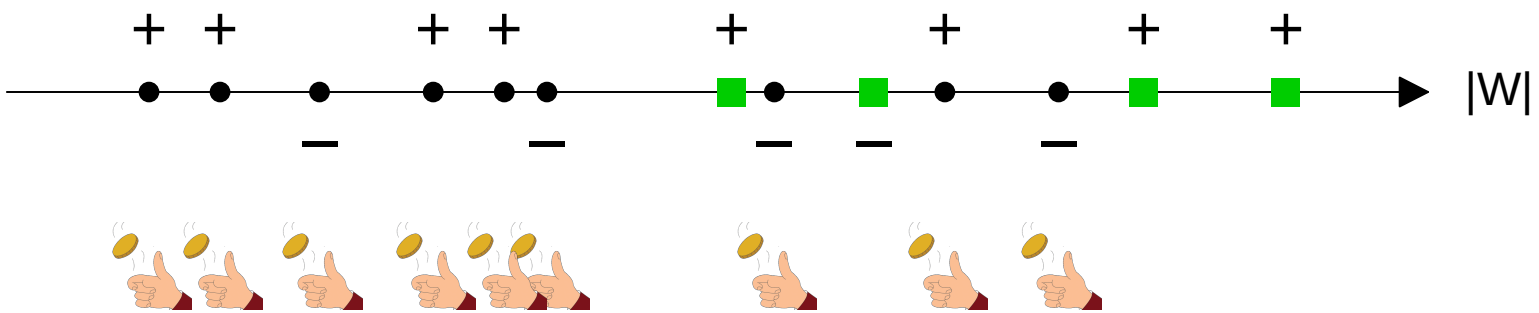
$$\mathsf{FDP}(S_{\Lambda_+}) = \frac{\left|S_{\Lambda_+} \cap \mathcal{H}_0\right|}{\left|S_{\Lambda_+}\right|} = \underbrace{\frac{\left|\widetilde{S}_{\Lambda_+} \cap \mathcal{H}_0\right| + 1}{\left|S_{\Lambda_+}\right|}}_{\leq \widehat{\mathsf{FDP}}_+(S_{\Lambda_+}) \leq q} \cdot \underbrace{\frac{\left|S_{\Lambda_+} \cap \mathcal{H}_0\right|}{\left|\widetilde{S}_{\Lambda_+} \cap \mathcal{H}_0\right| + 1}}_{\text{martingale}}$$

Theoretical guarantees

Proof sketch cont'd:

$$M(\lambda) = \frac{\left|S_\lambda \cap \mathcal{H}_0\right|}{\left|\widetilde{S}_\lambda \cap \mathcal{H}_0\right| + 1}$$ is a supermartingale w.r.t. increasing λ,

and Λ_+ is a stopping time.

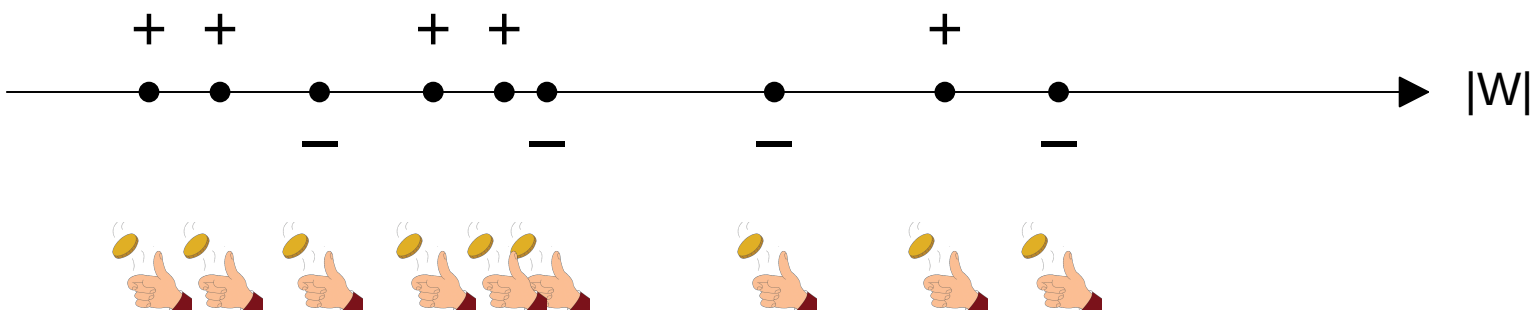


Theoretical guarantees

Proof sketch cont'd:

$$M(\lambda) = \frac{\left|S_\lambda \cap \mathcal{H}_0\right|}{\left|\widetilde{S}_\lambda \cap \mathcal{H}_0\right| + 1}$$ is a supermartingale w.r.t. increasing λ,

and Λ_+ is a stopping time.

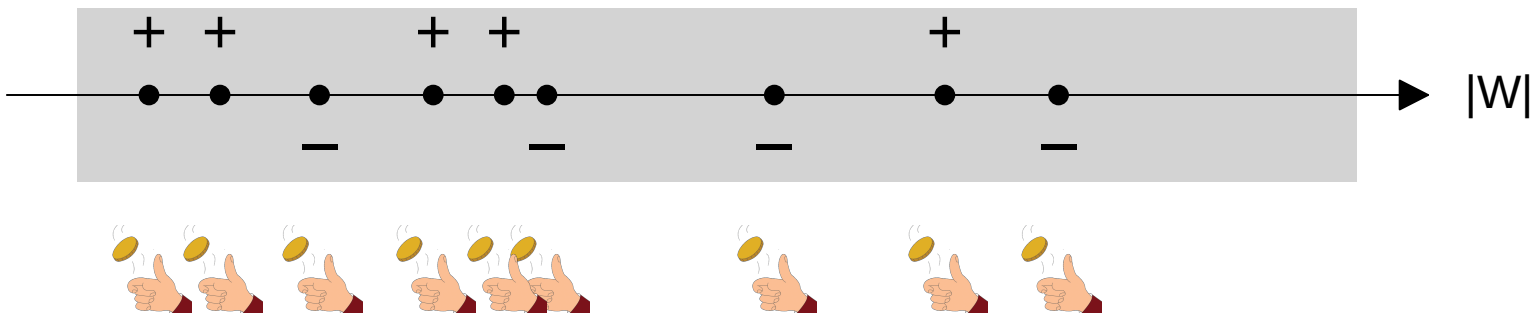


Theoretical guarantees

Proof sketch cont'd:

$$M(\lambda) = \frac{\left|S_\lambda \cap \mathcal{H}_0\right|}{\left|\widetilde{S}_\lambda \cap \mathcal{H}_0\right| + 1}$$ is a supermartingale w.r.t. increasing λ,

and Λ_+ is a stopping time.

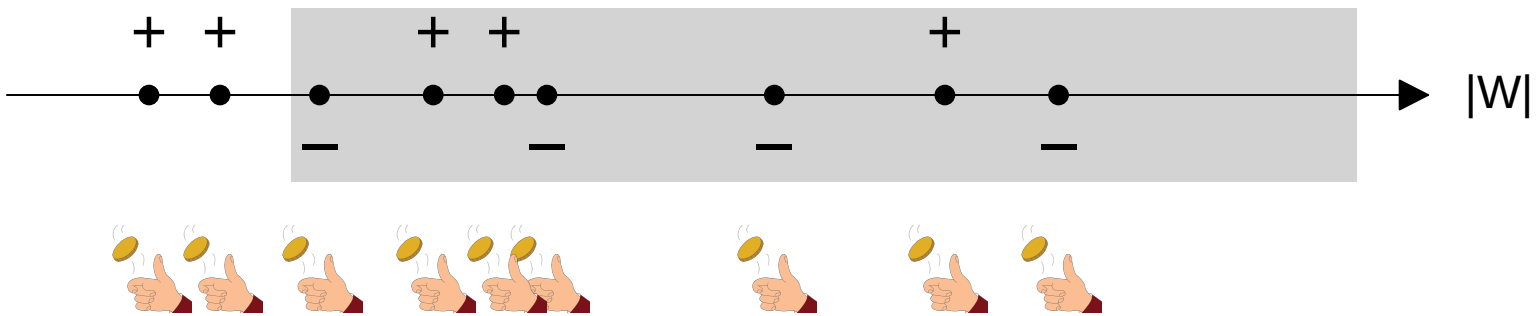


Theoretical guarantees

Proof sketch cont'd:

$$M(\lambda) = \frac{\left|S_\lambda \cap \mathcal{H}_0\right|}{\left|\widetilde{S}_\lambda \cap \mathcal{H}_0\right| + 1}$$ is a supermartingale w.r.t. increasing λ,

and Λ_+ is a stopping time.

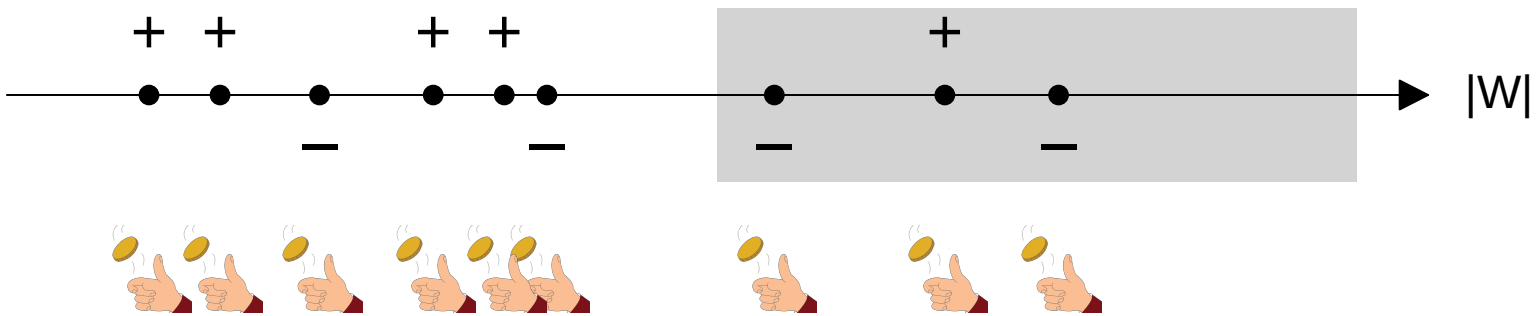


Theoretical guarantees

Proof sketch cont'd:

$$M(\lambda) = \frac{\left|S_\lambda \cap \mathcal{H}_0\right|}{\left|\widetilde{S}_\lambda \cap \mathcal{H}_0\right| + 1}$$ is a supermartingale w.r.t. increasing λ,

and Λ_+ is a stopping time.

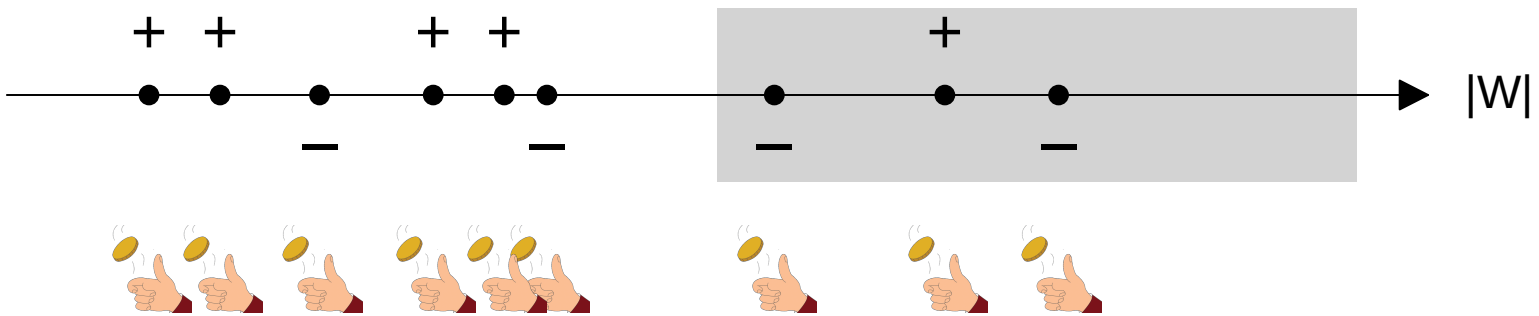


Theoretical guarantees

Proof sketch cont'd:

$$M(\lambda) = \frac{|S_\lambda \cap \mathcal{H}_0|}{|\widetilde{S}_\lambda \cap \mathcal{H}_0| + 1}$$ is a supermartingale w.r.t. increasing λ,

and Λ_+ is a stopping time.



$$\mathbb{E}\left[M(\Lambda_+)\right] \leq \mathbb{E}\left[M(0)\right] = \mathbb{E}\left[\frac{C}{|\mathcal{H}_0| - C + 1}\right] \leq 1,$$

for C = # of + coin flips $\sim \mathsf{Bin}(|\mathcal{H}_0|, 0.5)$

Simulations

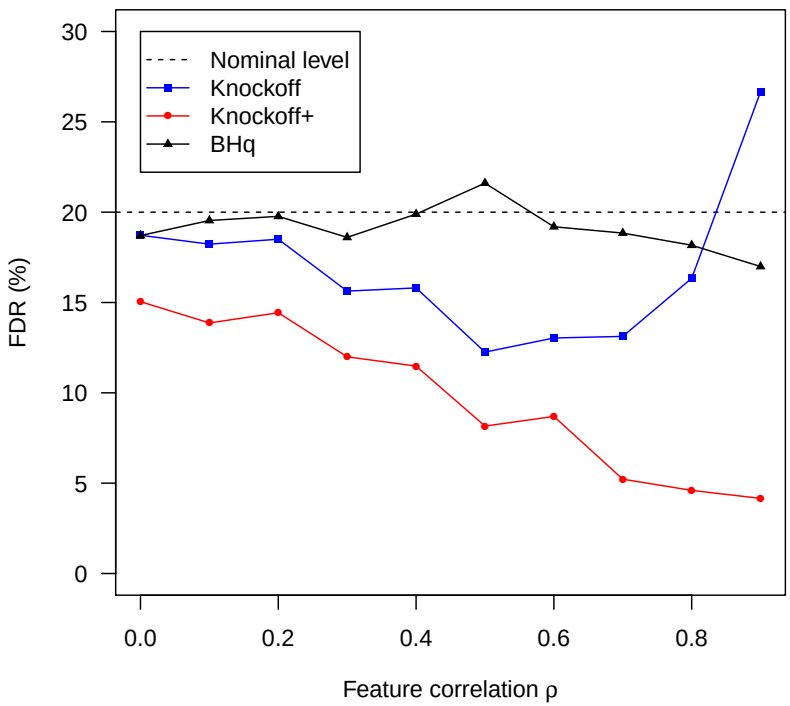
Setup:

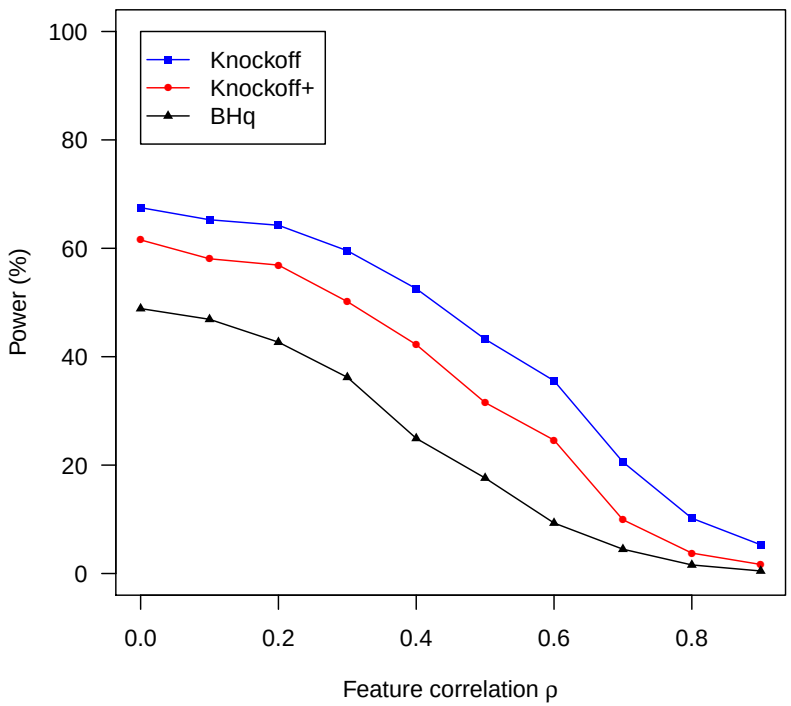
- $n = 3000$, $p = 1000$, sparsity level k
- Features X_j are random unit vectors with correlation level ρ
- For signals j, $\beta_j^\star \overset{\text{iid}}{\sim} \{\pm A\}$ for amplitude level A
- $y = X\beta + N(0, \mathbf{I}_n)$

Compare knockoff, knockoff+, & Benjamini-Hochberg (BH).

Simulations

- Fix amplitude $A = 3.5$ & sparsity level $k = 30$
- Vary feature correlation ρ from 0 to 0.9 (set $\mathbb{E}[X_j^\top X_k] = \rho^{|j-k|}$)





HIV data

Which mutations in the RT or protease of HIV-1
determine susceptibility to RT inhibitors or protease inhibitors?

Data:
Genotypic predictors of HIV type 1 drug resistance, Rhee et al (2006)
Available at `hivdb.stanford.edu` (Stanford HIV Drug Resistance Database)

- Each drug analysed separately
- Response y = resistance to the drug
- Features X = which mutations are present in the RT or in the protease

HIV data

The data set:

Drug type	# drugs	Sample size	# protease or RT positions genotyped	# mutations appearing ≥ 3 times in sample
PI	6	848	99	209
NRTI	6	639	240	294
NNRTI	3	747	240	319

HIV data

The data set:

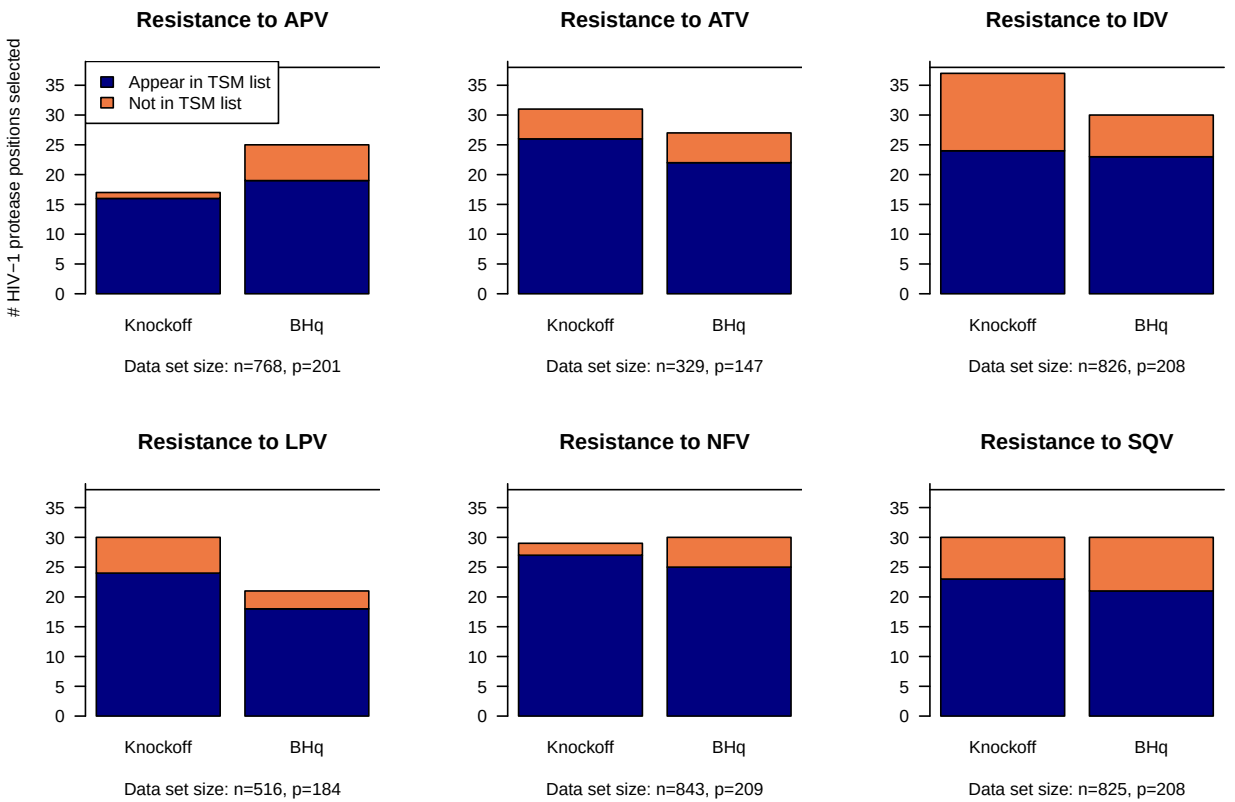
Drug type	# drugs	Sample size	# protease or RT positions genotyped	# mutations appearing ≥ 3 times in sample
PI	6	848	99	209
NRTI	6	639	240	294
NNRTI	3	747	240	319

To validate results:

- Treatment-selected mutation (TSM) panel:
  A separate study identifies mutations frequently present in patients who have been treated with each type of drug

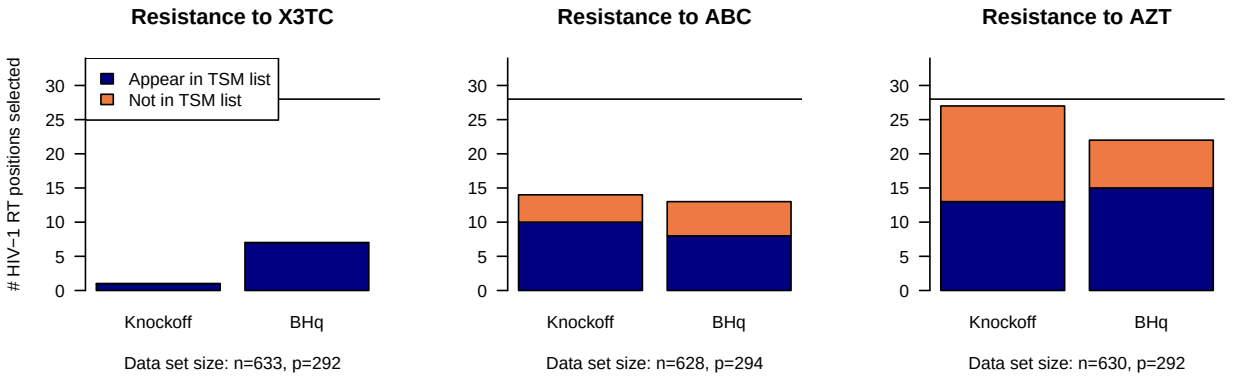
HIV data

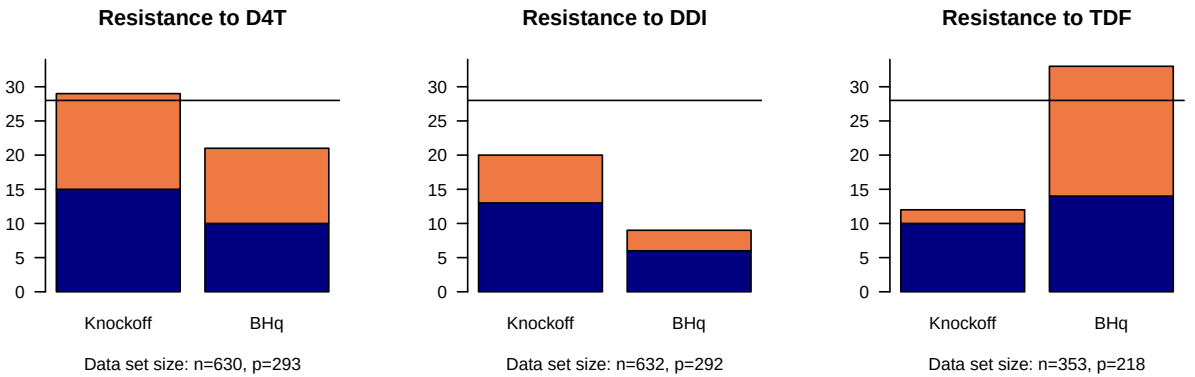
Results for
PI type drugs



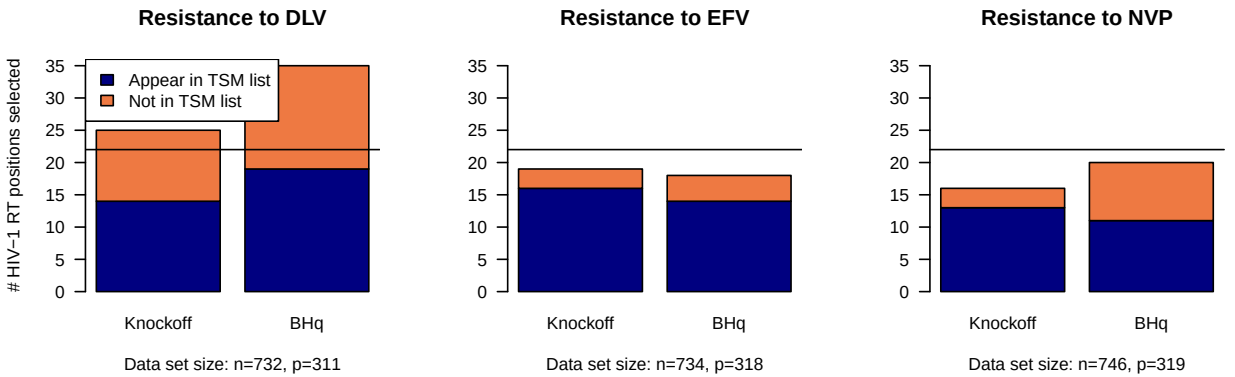
HIV data

Results for NRTI type drugs





Results for NNRTI type drugs



Can knockoffs be replaced by permutations?

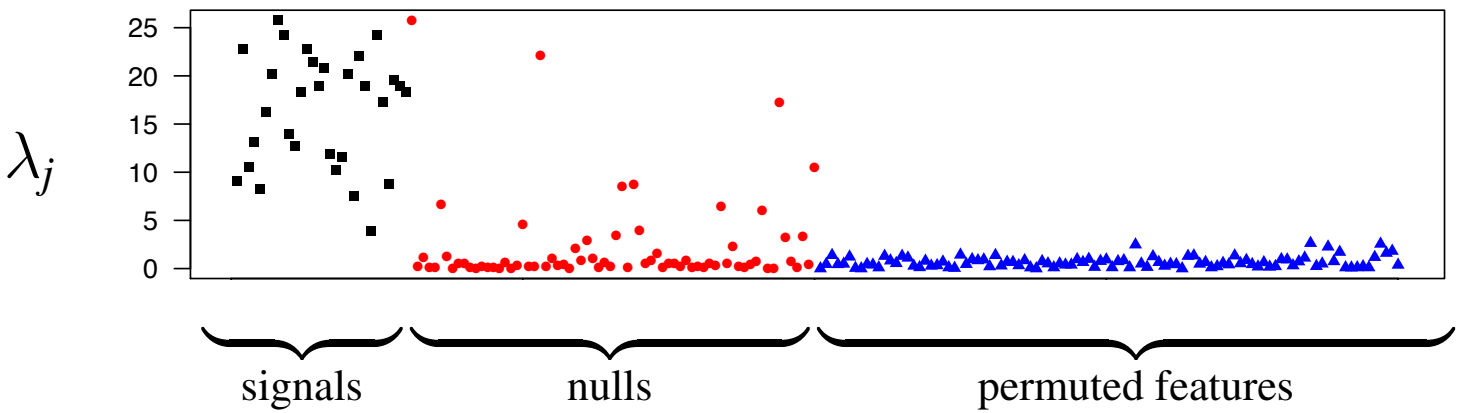
Let $X^\pi = X$ with rows randomly permuted. Then

$$[X \;\; X^\pi]^\top [X \;\; X^\pi] \approx \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix}$$

Can knockoffs be replaced by permutations?

Let $X^\pi = X$ with rows randomly permuted. Then

$$[X \;\; X^\pi]^\top [X \;\; X^\pi] \approx \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix}$$



	FDR (target level $q = 20\%$)
Knockoff method	12.29%
Permutation method	45.61%

Summary

The knockoff filter for inference in a sparse linear model:

- Creates a "control group" for any type of statistic
- Handles any type of feature correlation
- Unknown noise level & sparsity level
- Finite-sample FDR guarantees

Summary

Future work:

1. How to move to high-dimensional setting?
2. Extend to GLMs or other regression models?
3. Similar principles for other problems, e.g. graphical models?

Summary

Thank you!

- Code & demos available at `http://web.stanford.edu/~candes/Knockoffs/`
- Paper available at `http://arxiv.org/abs/1404.5609`

- Joint work with Emmanuel Candès @ Stanford
- R. F. B. was partially supported by NSF award DMS-1203762. E. C. is partially supported by AFOSR under grant FA9550-09-1-0643, by NSF via grant CCF-0963835 and by the Math + X Award from the Simons Foundation.