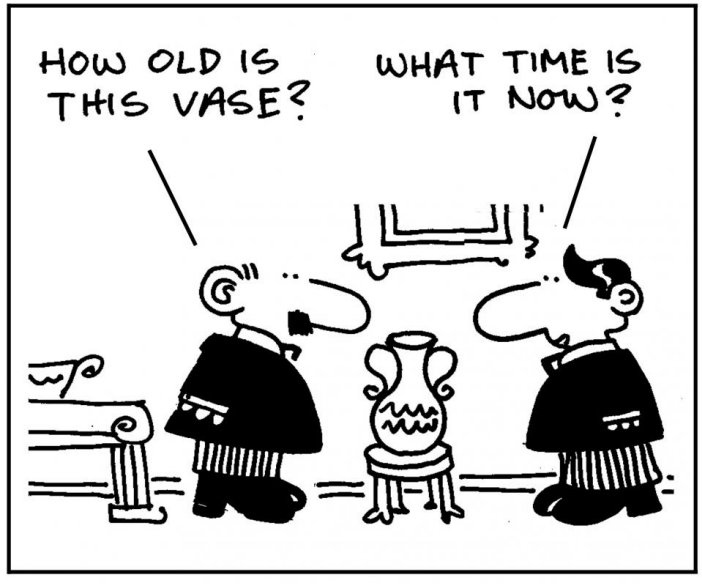
Around the Reproducibility of Scientific Research in the Big Data Era: A Knockoff Filter for Controlling the False Discovery Rate

Emmanuel Candès



Big Data and Computational Scalability, University of Warwick, July 2015

Collaborator

Rina Foygel Barber

The Economist

OCTOBER 19TH–25TH 2013 Economist.com

Washington's lawyer surplus

How to do a nuclear deal with Iran

Investment tips from Nobel economists

Junk bonds are back

The meaning of Sachin Tendulkar

HOW SCIENCE GOES WRONG.

99
Einsteinium

A crisis? The Economist: Oct 19th, 2013

The Economist

Problems with scientific research

How science goes wrong

Scientific research has changed the world. Now it needs to change itself

Oct 19th 2013 | From the print edition

A crisis? The Economist: Oct 19th, 2013

The Economist

Problems with scientific research

How science goes wrong

Scientific research has changed the world. Now it needs to change itself

Oct 19th 2013 | From the print edition

The Economist

Unreliable research

Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition

"I SEE a train wreck looming," warned Daniel Kahneman, an eminent psychologist, in an open letter last year. The premonition concerned research on a phenomenon known as "priming". Priming studies suggest that decisions can be influenced by apparently irrelevant actions or events that took place just before the cusp of choice. They have been a boom area in psychology over the past decade, and some of their insights have already made it out of the lab and into the toolkits of policy wonks keen on "nudging" the populace.

Snippets

The Economist

Unreliable research

Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition

- Systematic attempts to replicate widely cited priming experiments have failed
- Amgen could only replicate 6 of 53 studies they considered landmarks in basic cancer science
- HealthCare could only replicate about 25% of 67 seminal studies

Early report (Kaplan, '08): 50% of Phase III FDA studies ended in failure

New Yorker: December, 2010

ANNALS OF SCIENCE

THE TRUTH WEARS OFF

Is there something wrong with the scientific method?

BY JONAH LEHRER

DECEMBER 13, 2010

Many results that are rigorously proved and accepted start shrinking in later studies.

"Significance chasing"

"Publication bias"

"Selective reporting"

"Why most published research findings are false" (Ioannidis, '05)

The New York Times Jan. 2014

MATTER
Unraveling the Key to a Cold Virus's Effectiveness

Experts Say That Battle on Keystone Pipeline Is Over Politics, Not Facts

Black Holes Inch Ahead to Violent Cosmic Union

GREEN COLUMN
Mapping the World's Problems

SCIENCE

New Truths That Only One Can See

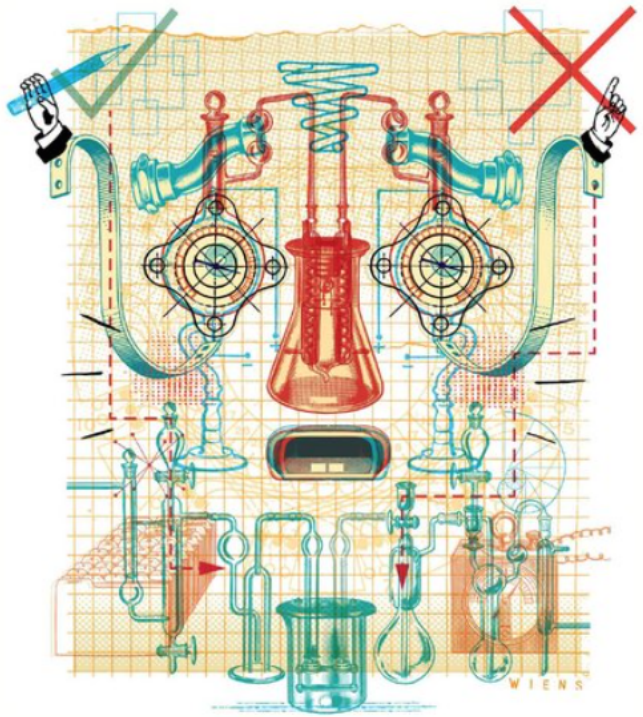
JAN. 20, 2014

George Johnson

RAW DATA

Since 1955, The Journal of Irreproducible Results has offered "spoofs, parodies, whimsies, burlesques, lampoons and satires" about life in the laboratory. Among its greatest hits: "Acoustic Oscillations in Jell-O, With and Without Fruit, Subjected to Varying Levels of Stress" and "Utilizing Infinite Loops to Compute an Approximate Value of Infinity." The good-natured jibes are a backhanded celebration of science. What really goes on in the lab is, by implication, of a loftier, more serious nature.

It has been jarring to learn in recent years that a reproducible result may actually be the rarest of birds. Replication, the ability of another lab to reproduce a finding, is the gold standard of science, reassurance that you have discovered something true. But that is getting harder all the time. With the most accessible truths already discovered, what remains are often subtle effects, some so delicate that they can be conjured up only under ideal circumstances, using highly specialized techniques



Carl Wiens

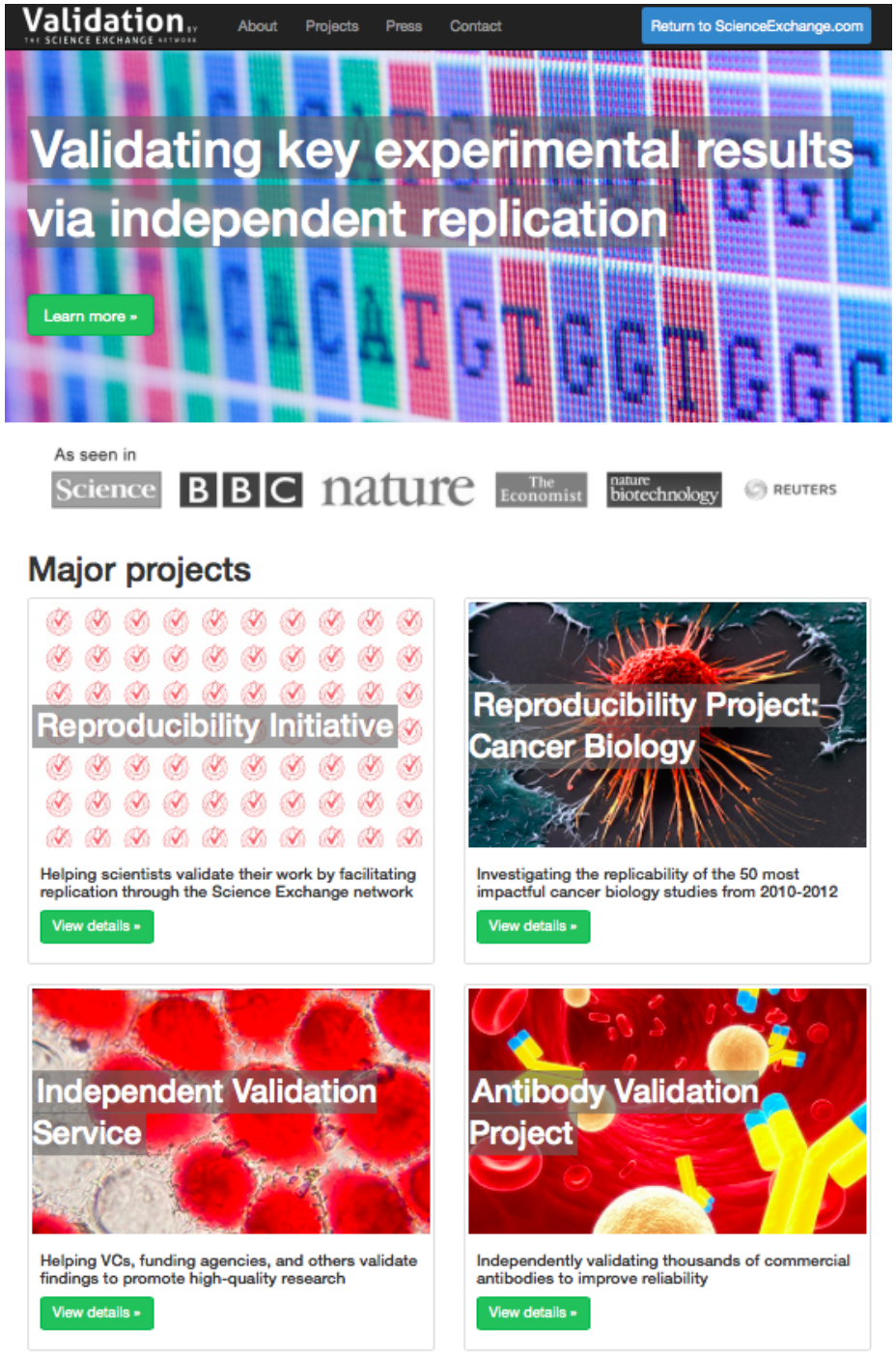
Personal and societal concern

Great danger in seeing erosion of public confidence in science

Personal and societal concern

Great danger in seeing erosion of public confidence in science

Seems like scientific community is responding



Reproducibility Initiative

http://validation.scienceexchange.com/

Nature's 18-point checklist: April 25, 2013

ANNOUNCEMENT

Reducing our irreproducibility

Over the past year, *Nature* has published a string of articles that highlight failures in the reliability and reproducibility of published research (collected and freely available at go.nature.com/huhbyr). The problems arise in laboratories, but journals such as this one compound them when they fail to exert sufficient scrutiny over the results that they publish, and when they do not publish enough information for other researchers to assess results properly.

From next month, *Nature* and the Nature research journals will introduce editorial measures to address the problem by improving the consistency and quality of reporting in life-sciences articles. To ease the interpretation and improve the reliability of published results we will more systematically ensure that key methodological details are reported, and we will give more space to methods sections. We will examine statistics more closely and encourage authors to be transparent, for example by including their raw data.

Central to this initiative is a checklist intended to prompt authors to disclose technical and statistical information in their submissions, and to encourage referees to consider aspects important for research reproducibility (go.nature.com/oloeip). It was developed after discussions with researchers on the problems that lead to irreproducibility, including workshops organized last year by US National Institutes of Health (NIH) institutes. It also draws on published concerns about reporting standards (or the lack of them) and the collective experience of editors at Nature journals.

The checklist is not exhaustive. It focuses on a few experimental and analytical design elements that are crucial for the interpretation of research results but are often reported incompletely. For example, authors will need to describe methodological parameters that can introduce bias or influence robustness, and provide precise characterization of key reagents that may be subject to biological variability, such as cell lines and antibodies. The checklist also consolidates existing policies about data deposition and presentation.

We will also demand more precise descriptions of statistics, and we will commission statisticians as consultants on certain papers, at the editor's discretion and at the referees' suggestion.

We recognize that there is no single way to conduct an experimental study. Exploratory investigations cannot be done with the same level of statistical rigour as hypothesis-testing studies. Few academic laboratories have the means to perform the level of validation required, for example, to translate a finding from the laboratory to the clinic. However, that should not stand in the way of a full report of how a study was designed, conducted and analysed that will allow reviewers and readers to adequately interpret and build on the results.

To allow authors to describe their experimental design and methods in as much detail as necessary, the participating journals, including *Nature*, will abolish space restrictions on the methods section.

To further increase transparency, we will encourage authors to provide tables of the data behind graphs and figures. This builds on our established data-deposition policy for specific experiments and large data sets. The source data will be made available directly from the figure legend, for easy access. We continue to encourage authors to share detailed methods and reagent descriptions by depositing protocols in Protocol Exchange (www.nature.com/protocolexchange), an open resource linked from the primary paper.

Renewed attention to reporting and transparency is a small step. Much bigger underlying issues contribute to the problem, and are beyond the reach of journals alone. Too few biologists receive adequate training in statistics and other quantitative aspects of their subject. Mentoring of young scientists on matters of rigour and transparency is inconsistent at best. In academia, the ever increasing pressures to publish and chase funds provide little incentive to pursue studies and publish results that contradict or confirm previous papers. Those who document the validity or irreproducibility of a published piece of work seldom get a welcome from journals and funders, even as money and effort are wasted on false assumptions.

Tackling these issues is a long-term endeavour that will require the commitment of funders, institutions, researchers and publishers. It is encouraging that NIH institutes have led community discussions on this topic and are considering their own recommendations. We urge others to take note of these and of our initiatives, and do whatever they can to improve research reproducibility. ■

NAS President's address: April 27, 2015

Big data and a new scientific paradigm

Collect data first $\Longrightarrow$ Ask questions later

- Large data sets available prior to formulation of hypotheses
- Need to quantify "reliability" of hypotheses generated by data snooping

Very different from hypothesis-driven research

Big data and a new scientific paradigm

Collect data first $\Longrightarrow$ Ask questions later

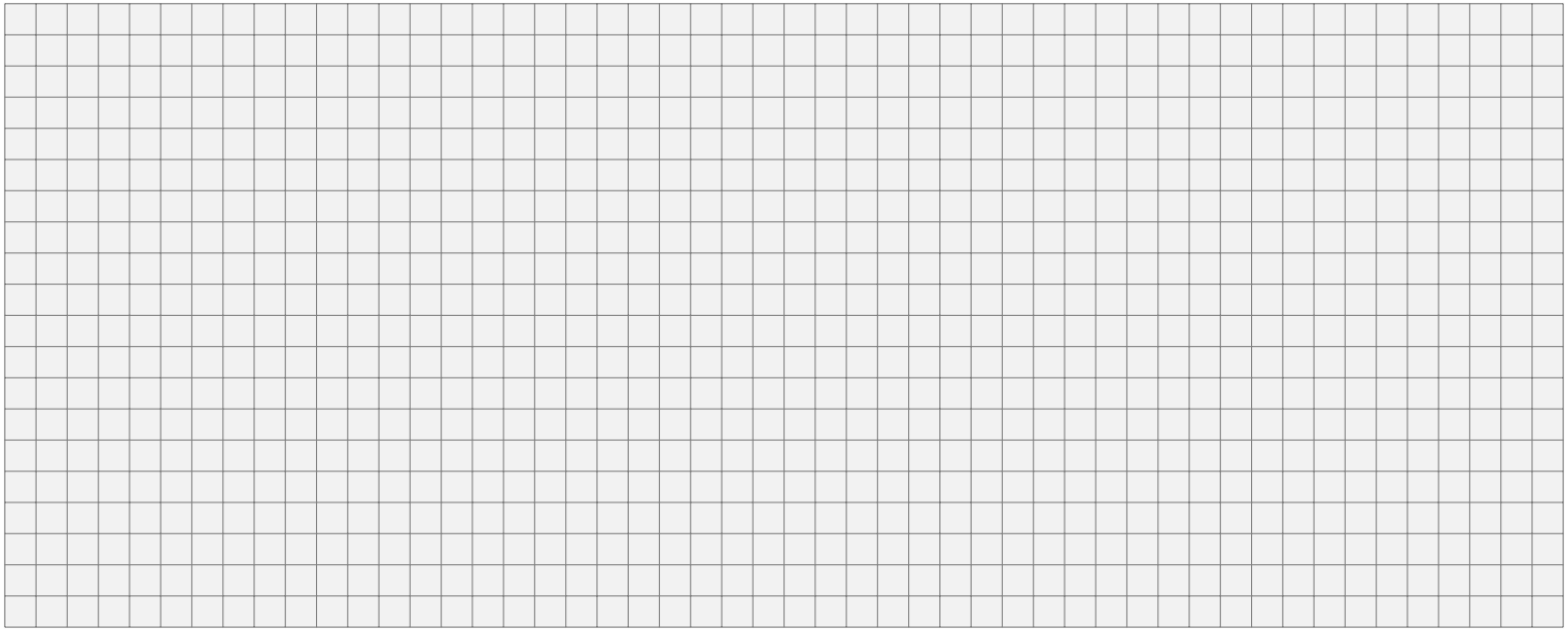
- Large data sets available prior to formulation of hypotheses
- Need to quantify "reliability" of hypotheses generated by data snooping

Very different from hypothesis-driven research

What does statistics have to offer?

- Account for *"look everywhere"* effect
- Understand reliability in the context of all hypotheses that have been explored

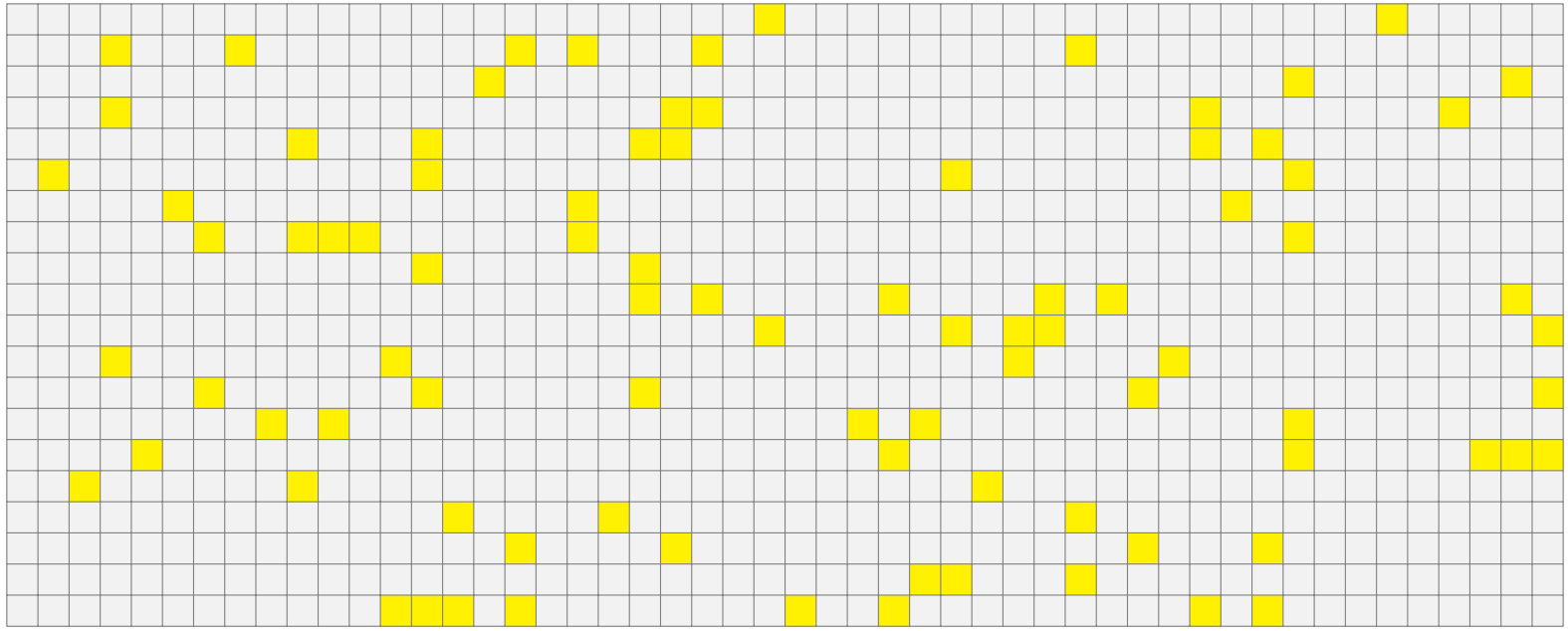
Most discoveries may be false: Sorić ('89)



1000 hypotheses to test

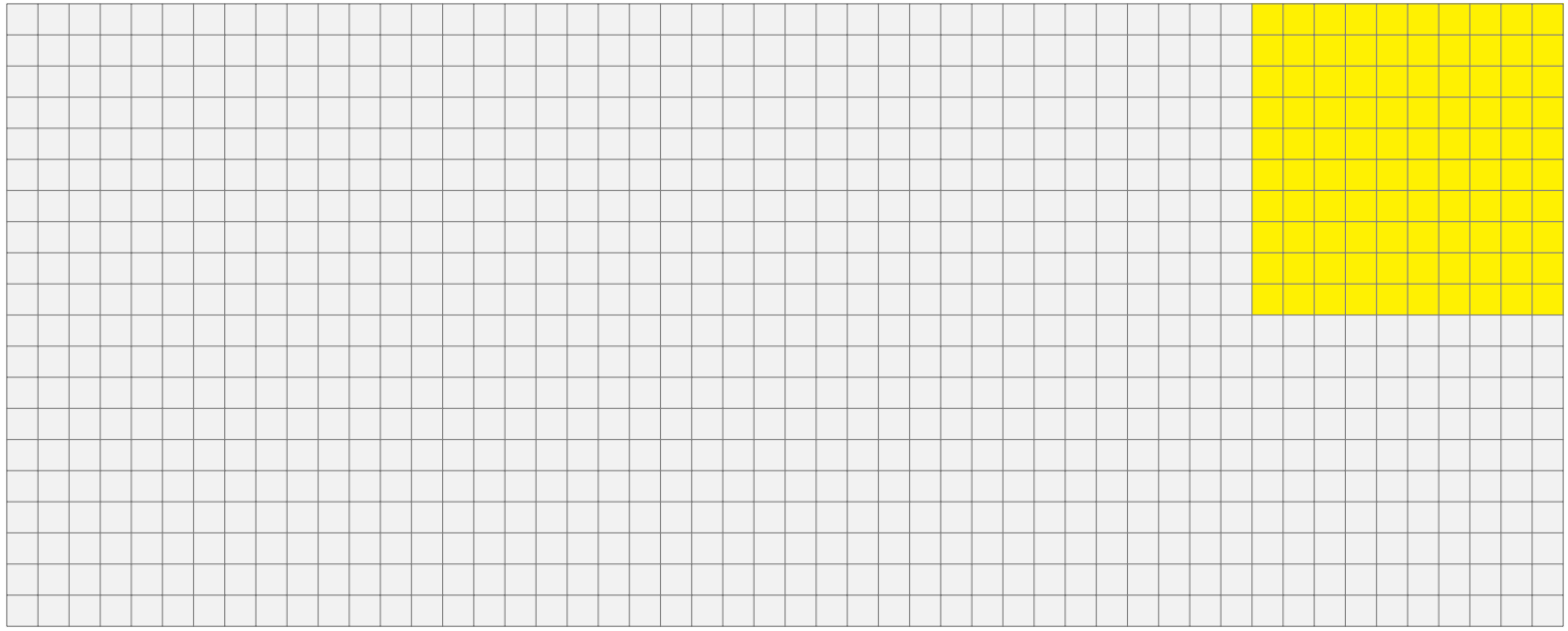
See also *The Economist*

Most discoveries may be false: Sorić ('89)



1000 hypotheses, 100 potential discoveries

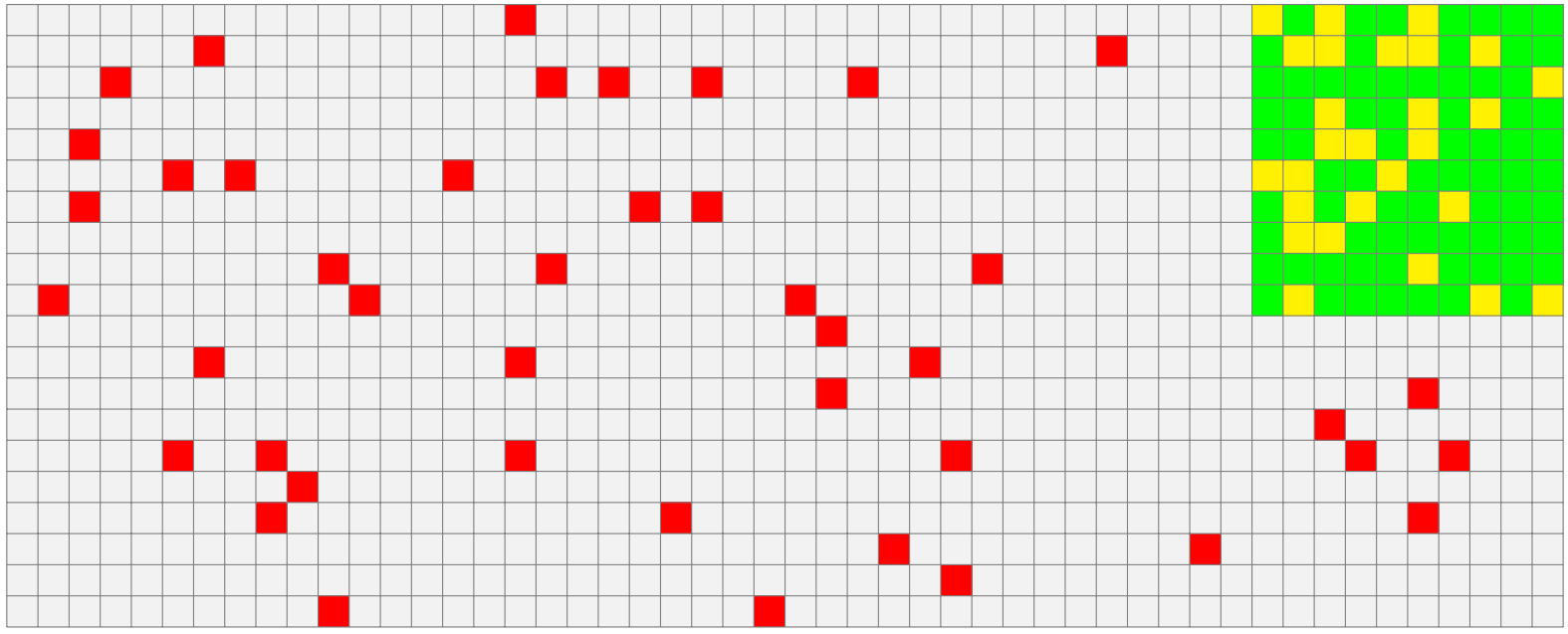
Most discoveries may be false: Sorić ('89)



1000 hypotheses, 100 potential discoveries

Most discoveries may be false: Sorić ('89)

True positives False negatives False positives

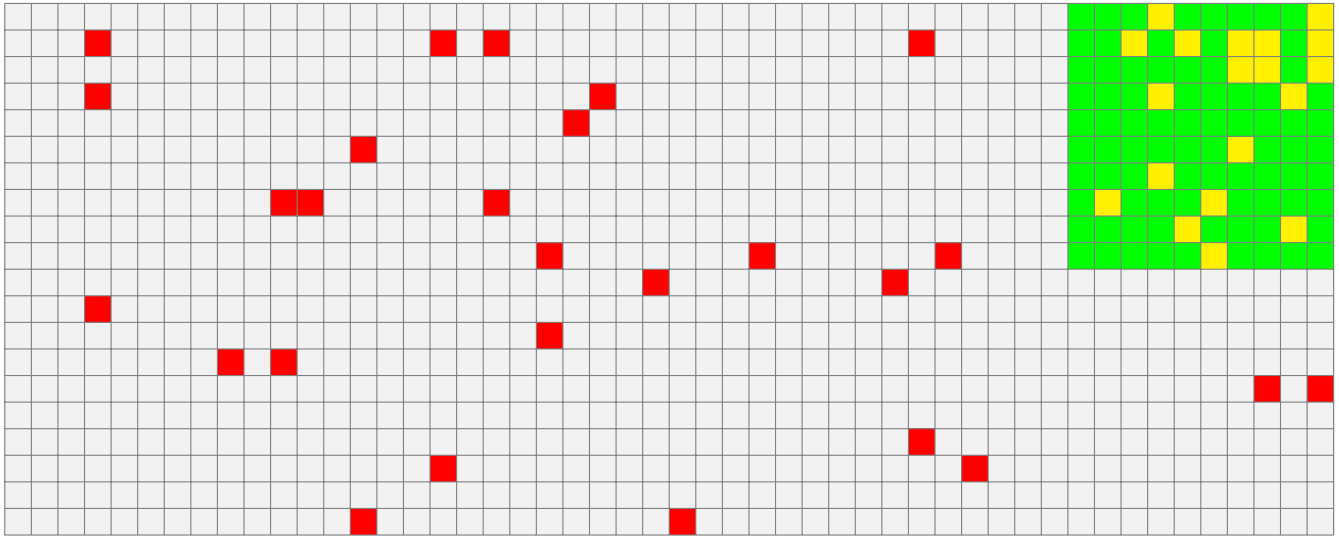


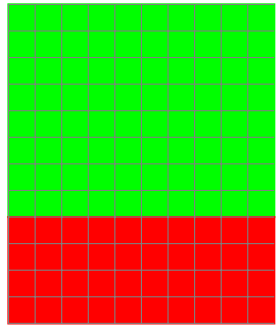
Power $\approx 80\%$ $\longrightarrow$ true positives ≈ 80
False positives (5% level) ≈ 45

$\Longrightarrow$ False discovery rate $\approx 36\%$

Most discoveries may be false: Sorić ('89)

True positives False negatives False positives





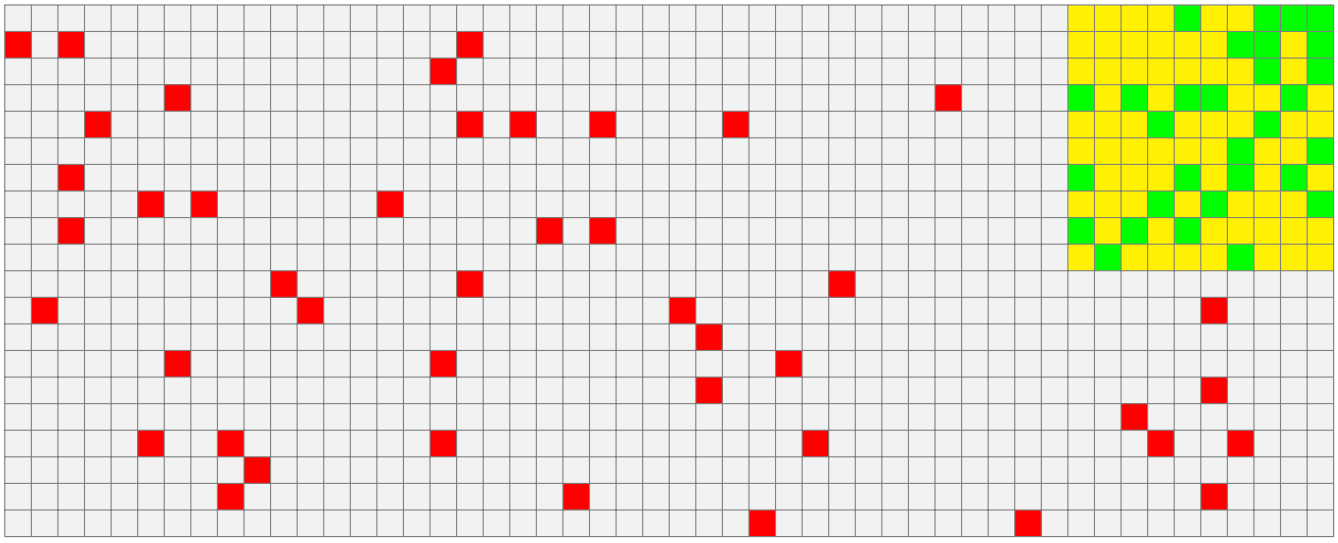
Reported

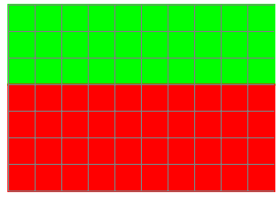
Power $\approx 80\%$ $\longrightarrow$ true positives ≈ 80
False positives (5% level) ≈ 45

$\Longrightarrow$ False discovery rate $\approx 36\%$

Most discoveries may be false: Sorić ('89)

True positives False negatives False positives





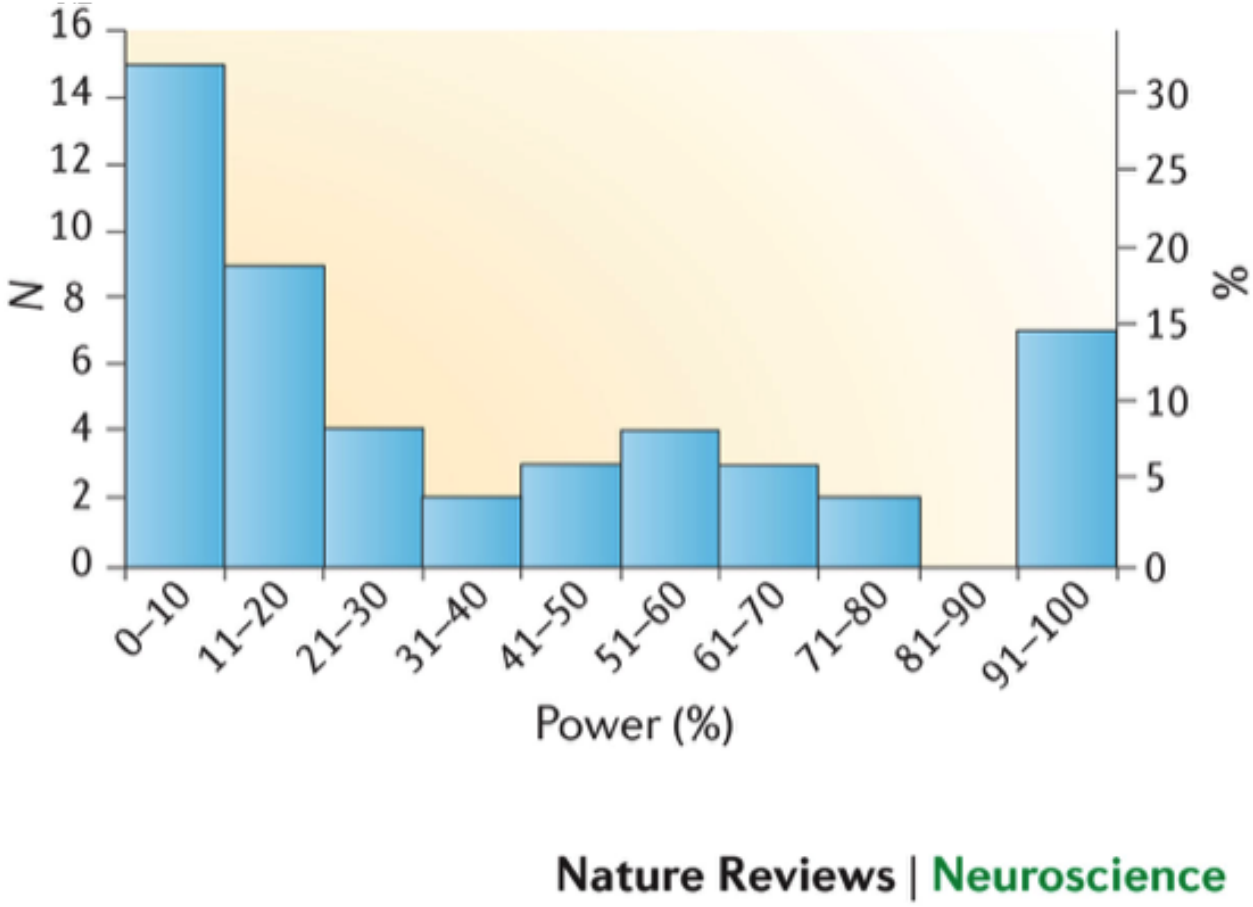
Reported

Power $\approx 30\%$ $\Longrightarrow$ False discovery rate $\approx 60\%$

More false negatives than true positives!

Example: meta-analysis in neuroscience

Button et al. (2013) *Power failure: why small sample size undermines the reliability of neuroscience*



False Discovery Rate (FDR): Benjamini & Hochberg ('95)

$H_1, \dots H_n$ hypotheses subject to some testing procedure

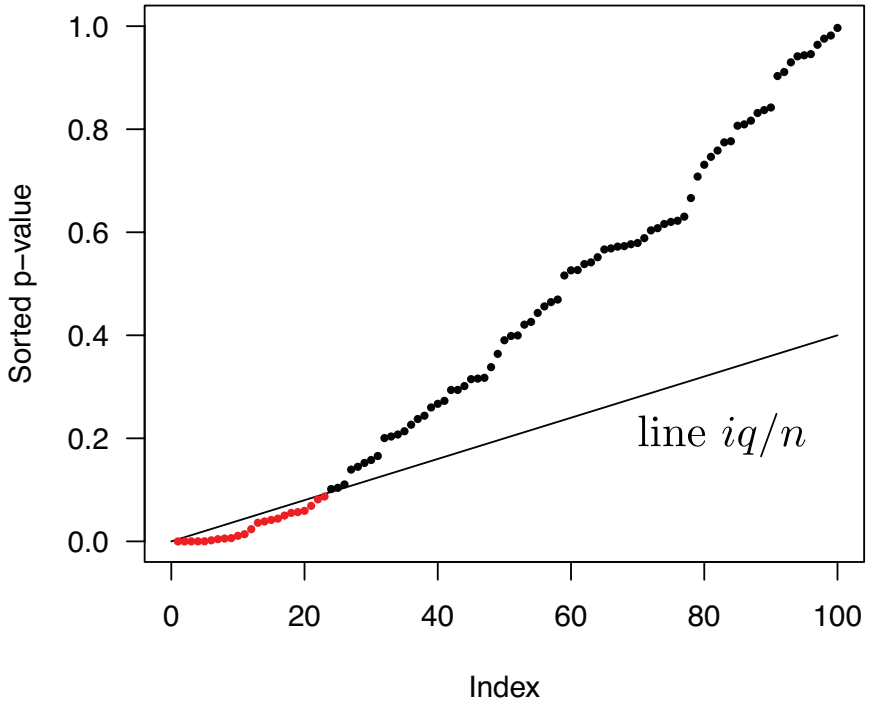
$$\text{FDR} = \mathbb{E}\left[\frac{\#\text{false discoveries}}{\#\text{discoveries}}\right] \qquad \text{`}0/0 = 0\text{'}$$

- Natural type I error
- Under independence (and PRDS) simple rules control FDR (BHq)
- Widely used

FDR control with BHq (under independence)

FDR: expected proportion of false discoveries

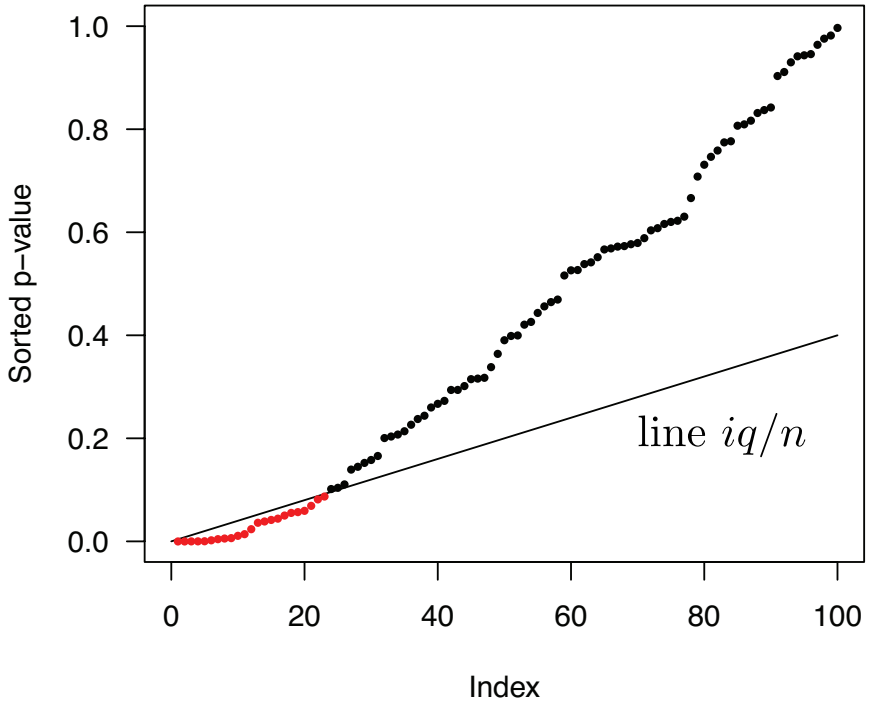
- Sorted p-values: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)}$ (from most to least significant)
- Target FDR q

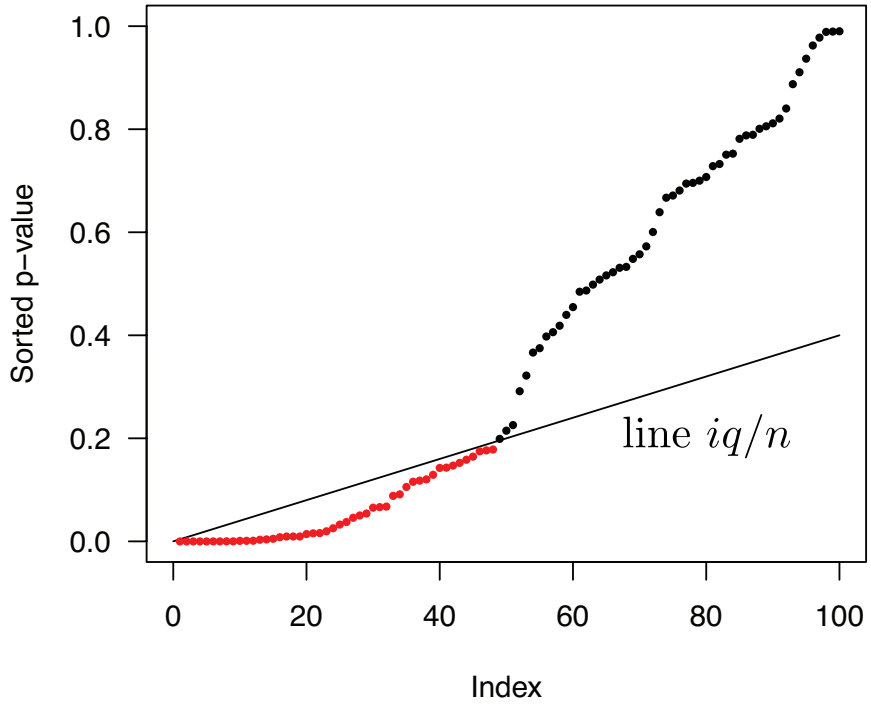


FDR control with BHq (under independence)

FDR: expected proportion of false discoveries

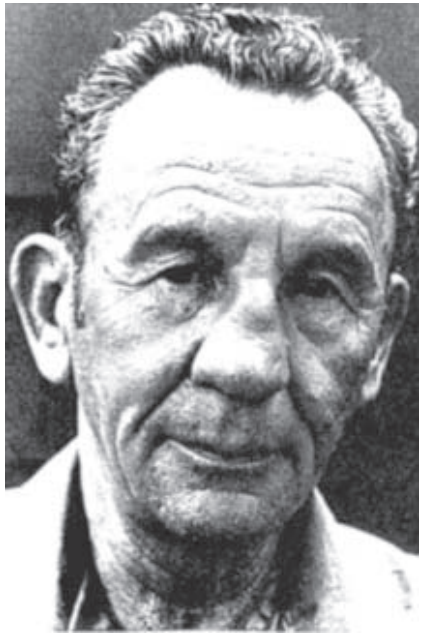
- Sorted p-values: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)}$ (from most to least significant)
- Target FDR q





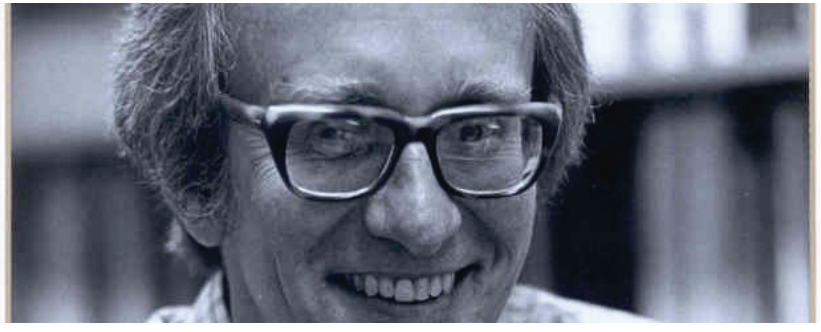
The cut-off is adaptive to number of non-nulls

Earlier work on multiple comparisons

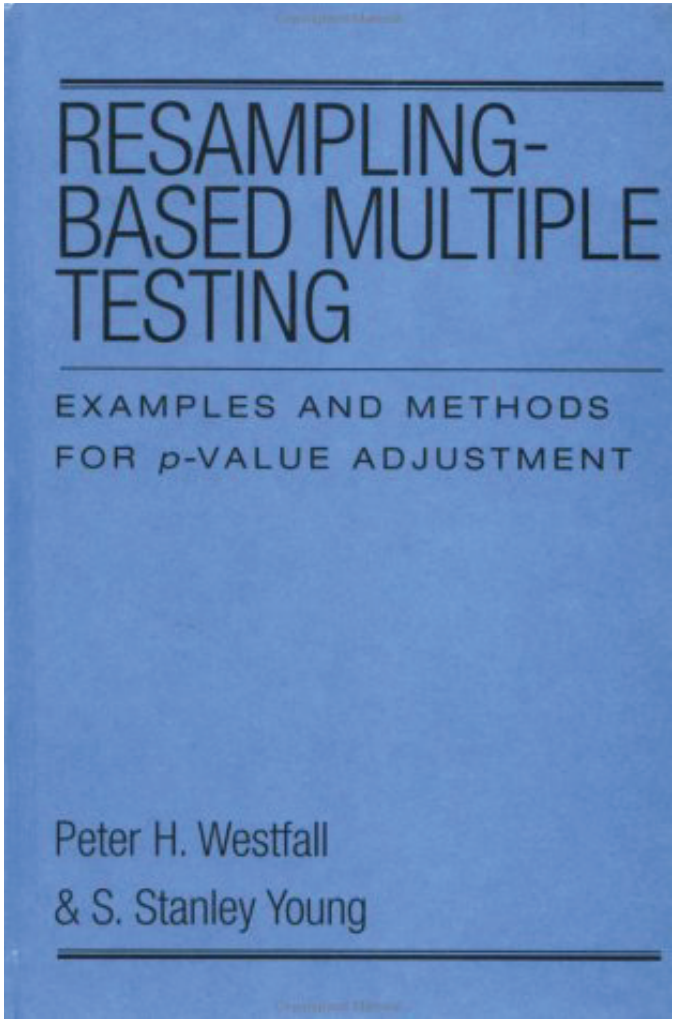


Henry Scheffe

John Tukey



Rupert Miller

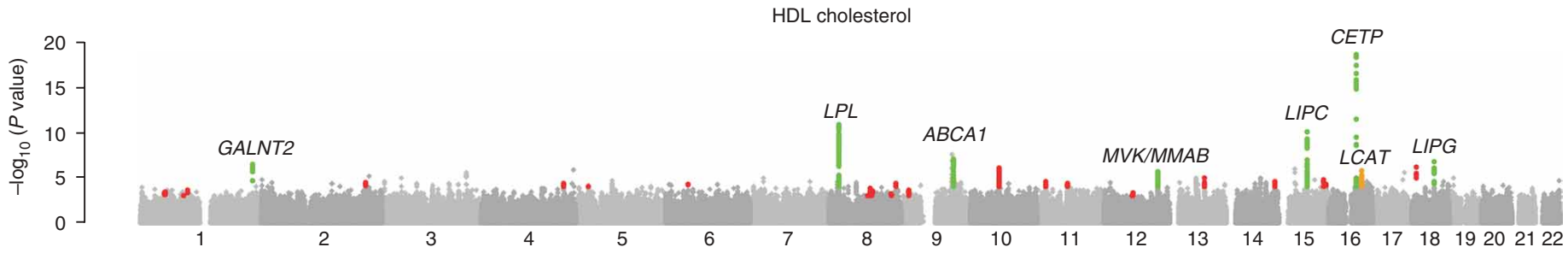


Westfall and Young ('93)

Controlled Variable Selection

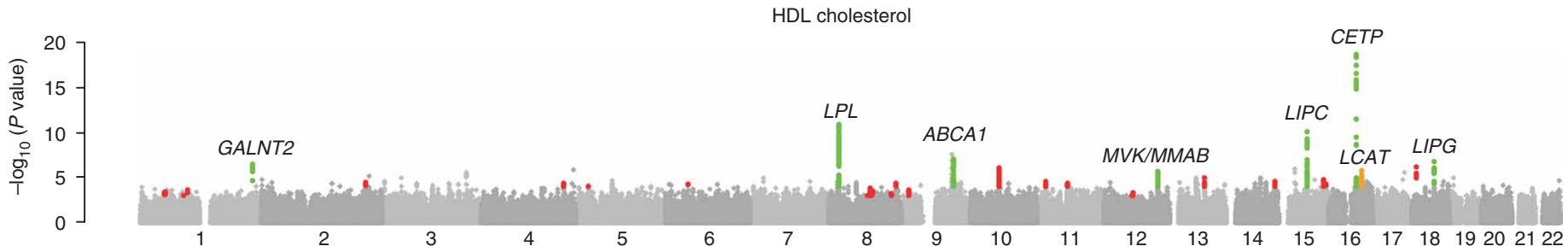
Contemporary problem: inference for sparse regression

Find locations on the genome that influence a trait: e.g. cholesterol level



Contemporary problem: inference for sparse regression

Find locations on the genome that influence a trait: e.g. cholesterol level



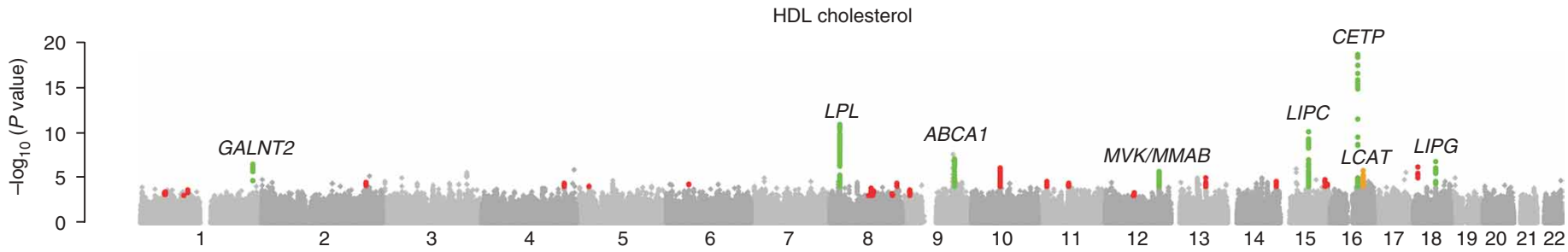
Linear model ($y = X\beta + z$) + e.g. Lasso fit

$$\min \quad \frac{1}{2}\|y - X\hat{\beta}\|_2^2 + \lambda\|\hat{\beta}\|_1$$

- y : cholesterol level of patients
- X : genotype matrix; e.g. $X_{i,j}$ # of alleles of recessive type at location j
- z environmental factors (not accounted by genetics)

Contemporary problem: inference for sparse regression

Find locations on the genome that influence a trait: e.g. cholesterol level



Linear model ($y = X\beta + z$) + e.g. Lasso fit

$$\text{min} \quad \tfrac{1}{2}\|y - X\hat{\beta}\|_2^2 + \lambda\|\hat{\beta}\|_1$$

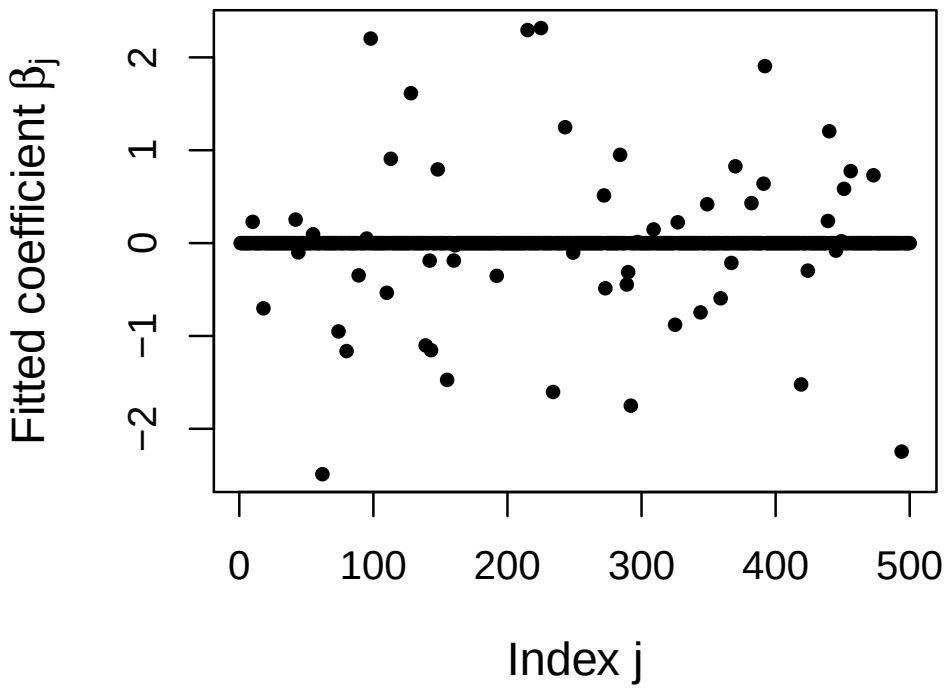
- y : cholesterol level of patients
- X : genotype matrix; e.g. $X_{i,j}$ # of alleles of recessive type at location j
- z environmental factors (not accounted by genetics)

How do we control the FDR of selected variables $\{i : \hat{\beta}_i \neq 0\}$?

Sparse regression

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

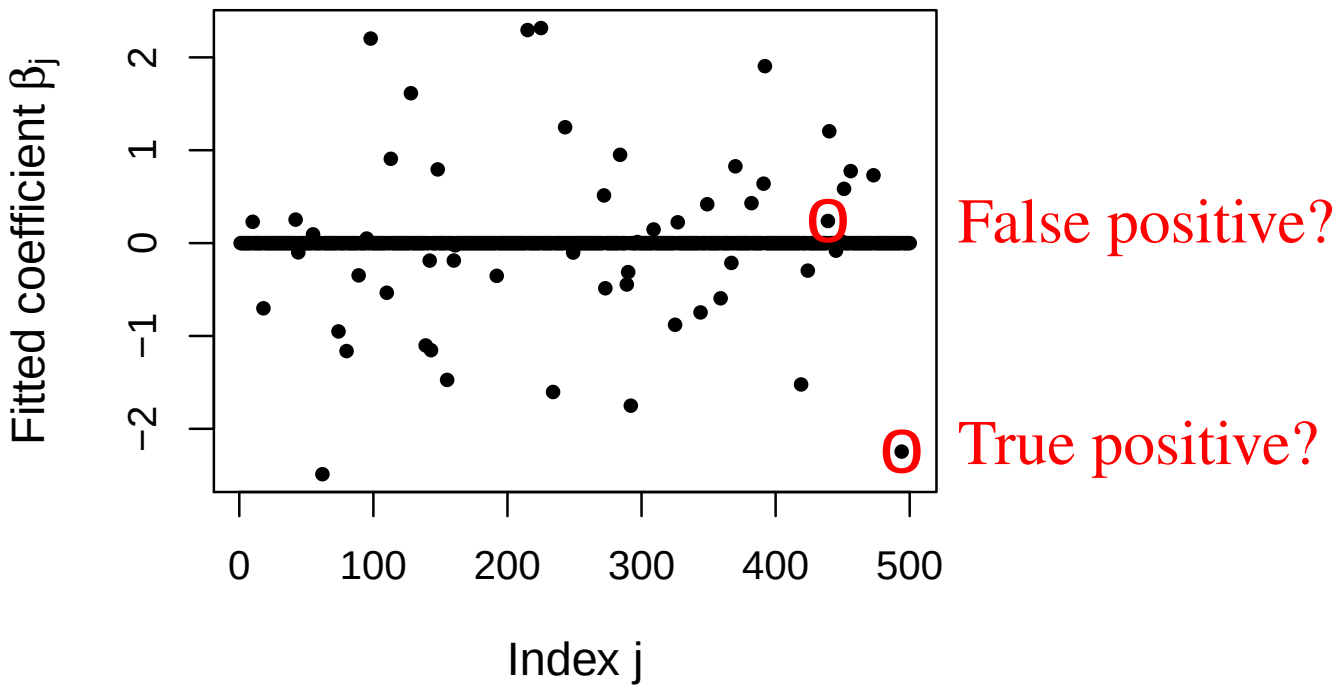
Lasso model with $\lambda = 1.75$



Sparse regression

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

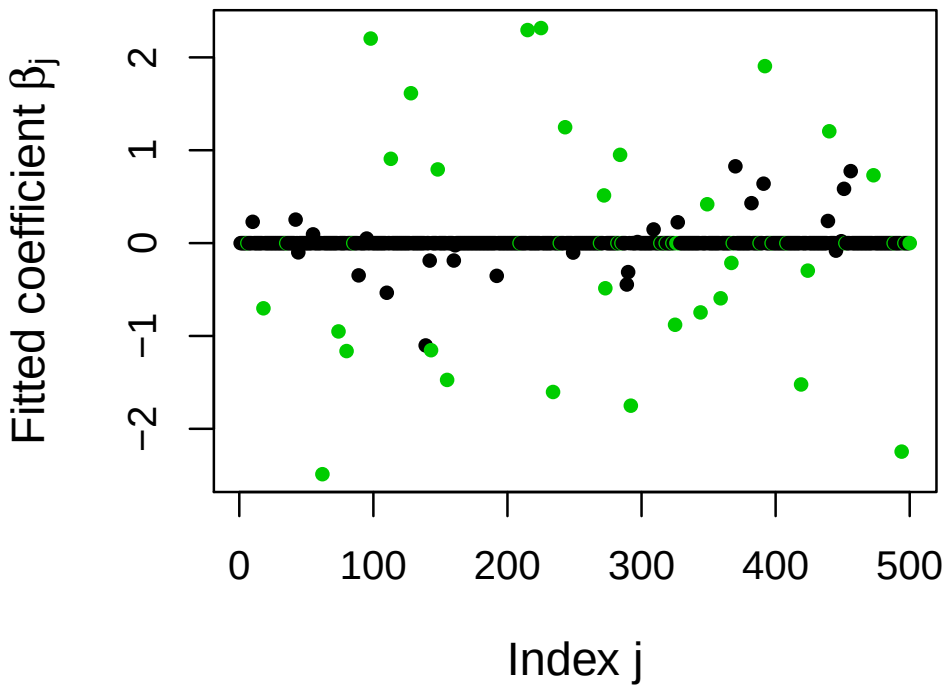
Lasso model with $\lambda = 1.75$



Sparse regression

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

Lasso model with $\lambda = 1.75$

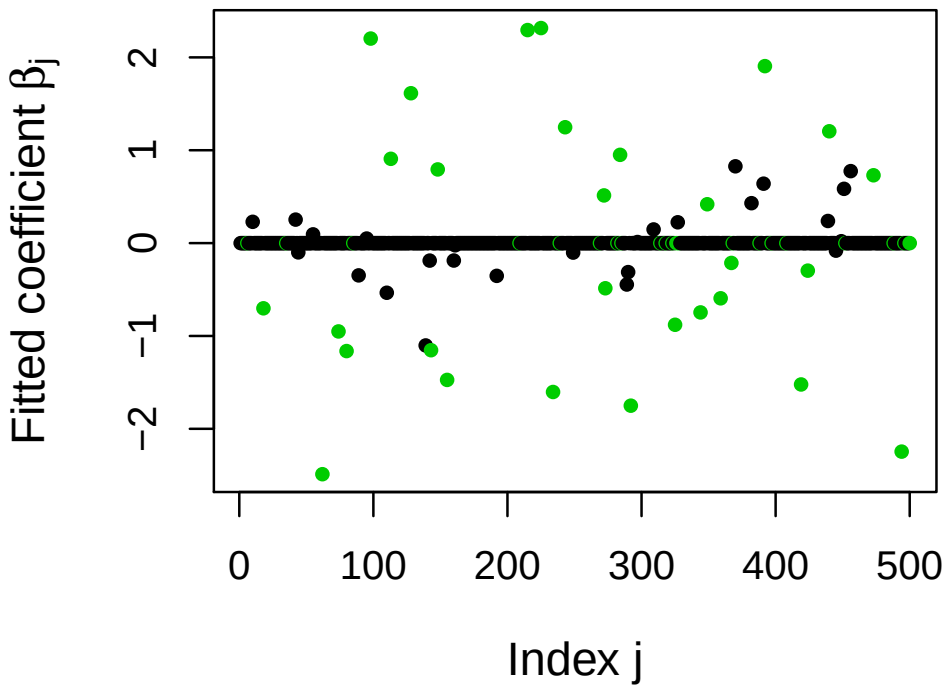


FDP $= \frac{26}{55} = 47\%$

Sparse regression

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

Lasso model with $\lambda = 1.75$



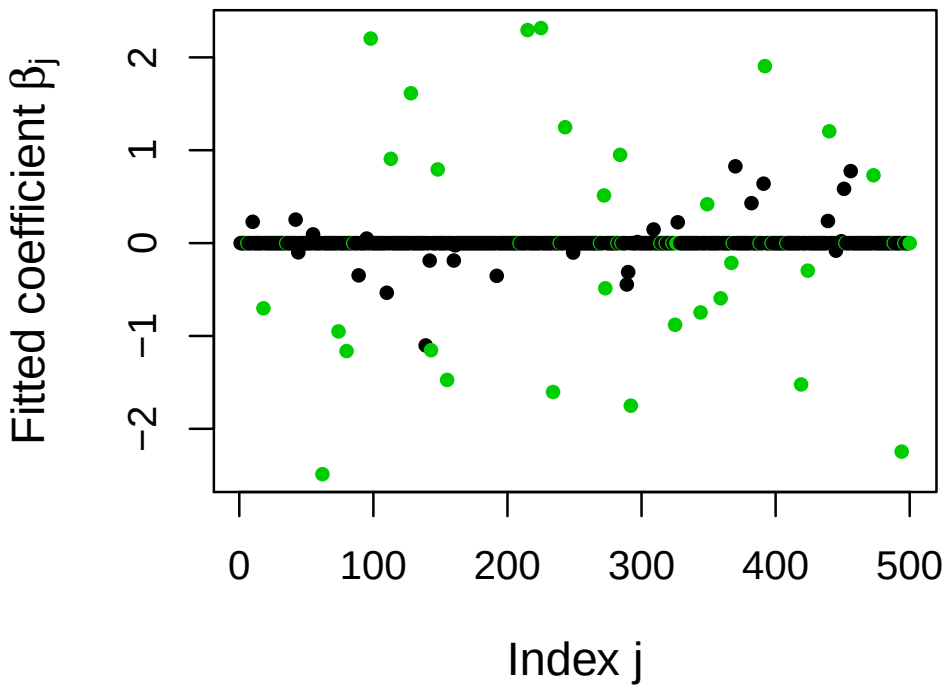
FDP $= \frac{26}{55} = 47\%$

Estimate FDP?

Sparse regression

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

Lasso model with $\lambda = 1.75$



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

Estimate FDP? Forget it...

Controlled variable selection

$$\underset{n\times 1}{y} = \overbrace{\underset{n\times p}{X}\underset{p\times 1}{\beta}}^{\sum_j \beta_j X_j} + \underset{n\times 1}{z} \qquad y \sim \mathcal{N}(X\beta, \sigma^2 I)$$

Controlled variable selection

$$\underset{n \times 1}{y} = \overbrace{\underset{n \times p}{X} \underset{p \times 1}{\beta}}^{\sum_j \beta_j X_j} + \underset{n \times 1}{z} \qquad y \sim \mathcal{N}(X\beta, \sigma^2 I)$$

Goal: select set of features X_j without too many false positives

$$\underbrace{\text{FDR}}_{\text{False discovery rate}} = \mathbb{E}\Big[\underbrace{\frac{\#\text{ false positives}}{\#\text{ features selected}}}_{\text{False discovery proportion}}\Big] \qquad \text{`}0/0 = 0\text{'}$$

Controlled variable selection

$$\underset{n\times 1}{y} = \overbrace{\underset{n\times p}{X}\underset{p\times 1}{\beta}}^{\sum_j \beta_j X_j} + \underset{n\times 1}{z} \qquad y \sim \mathcal{N}(X\beta, \sigma^2 I)$$

Goal: select set of features X_j without too many false positives

$$\underbrace{\text{FDR}}_{\text{False discovery rate}} = \mathbb{E}\Big[\underbrace{\frac{\#\text{ false positives}}{\#\text{ features selected}}}_{\text{False discovery proportion}}\Big] \qquad \text{`}0/0 = 0\text{'}$$

Context of multiple testing (with possibly very many irrelevant variables)

$$H_j : \beta_j = 0 \qquad j = 1, \ldots, p$$

This lecture

Novel procedure for variable selection controlling FDR in finite sample settings

- Requires $n \geq p$ (and full column rank) for identifiability

$$n < p \quad \Longrightarrow \quad X\beta = X\beta' \text{ and } \beta \neq \beta'$$

- Works under any design X
- Does not require any knowledge of noise level σ

Ongoing research: $n < p$

The Knockoff Filter

The knockoff filter

It's not a name brand bag, just a cheap knockoff

Thesaurize.com

For each feature X_j, construct knockoff version $\tilde{X}_j$
Knockoffs serve as control group $\Longrightarrow$ can estimate FDP

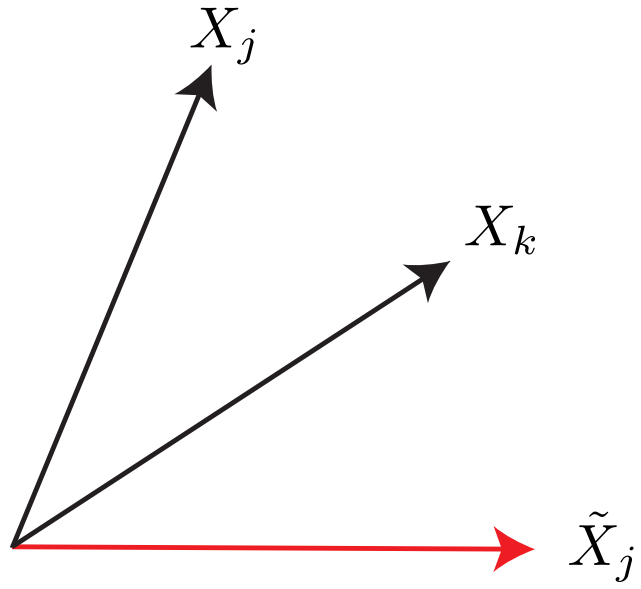
(1) Construct fake variables (knockoffs)

(2) Calculate statistics for each original/knockoff pair

(3) Calculate a data-dependent threshold for the statistics

1. Knockoff features $\tilde{X}_j$

$$\tilde{X}_j'\tilde{X}_k = X_j'X_k \qquad \text{for all } j,k$$
$$\tilde{X}_j'X_k = X_j'X_k \qquad \text{for all } j \neq k$$



Would like knockoffs as uncorrelated to features as possible

How?

Compute knockoffs via matrix computations and/or numerical optimization (later)

$$[X \quad \tilde{X}]' [X \quad \tilde{X}] = \begin{bmatrix} \Sigma & \Sigma - \text{diag}\{s\} \\ \Sigma - \text{diag}\{s\} & \Sigma \end{bmatrix} \succeq 0 \qquad s \in \mathbb{R}^p$$

How?

Compute knockoffs via matrix computations and/or numerical optimization (later)

$$[X \;\; \tilde{X}]' [X \;\; \tilde{X}] = \begin{bmatrix} \Sigma & \Sigma - \mathrm{diag}\{s\} \\ \Sigma - \mathrm{diag}\{s\} & \Sigma \end{bmatrix} \succeq 0 \qquad s \in \mathbb{R}^p$$

$$\tilde{X} = X(I - \Sigma^{-1}\,\mathrm{diag}\{s\}) + \tilde{U}C$$

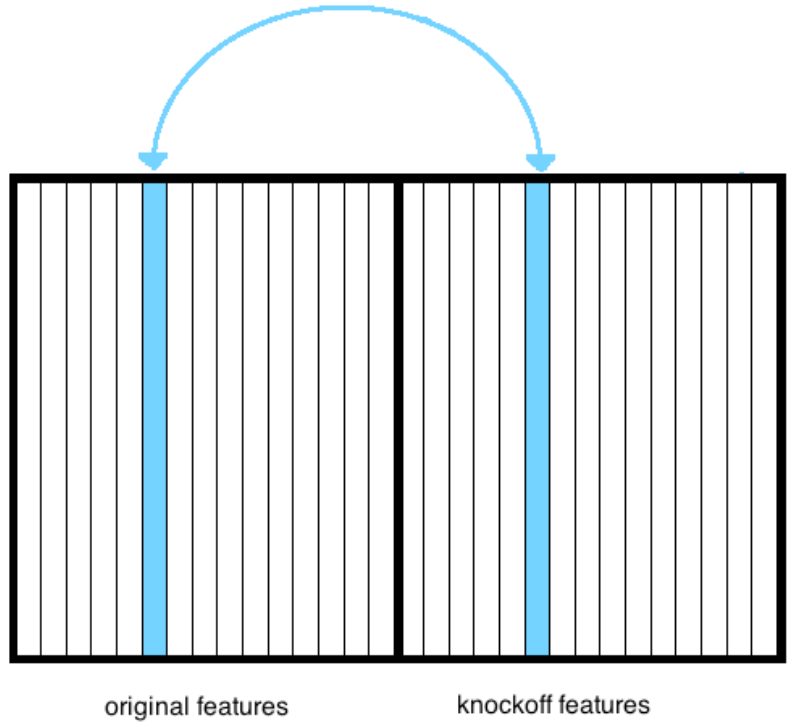
- $\tilde{U} \in \mathbb{R}^{n \times p}$ with col. space orthogonal to that of X
- $C'C$ Cholevsky factorization of $2\,\mathrm{diag}\{s\} - \mathrm{diag}\{s\}\Sigma^{-1}\,\mathrm{diag}\{s\} \succeq 0$

No need for new data or experiment

Why?

For null feature X_j

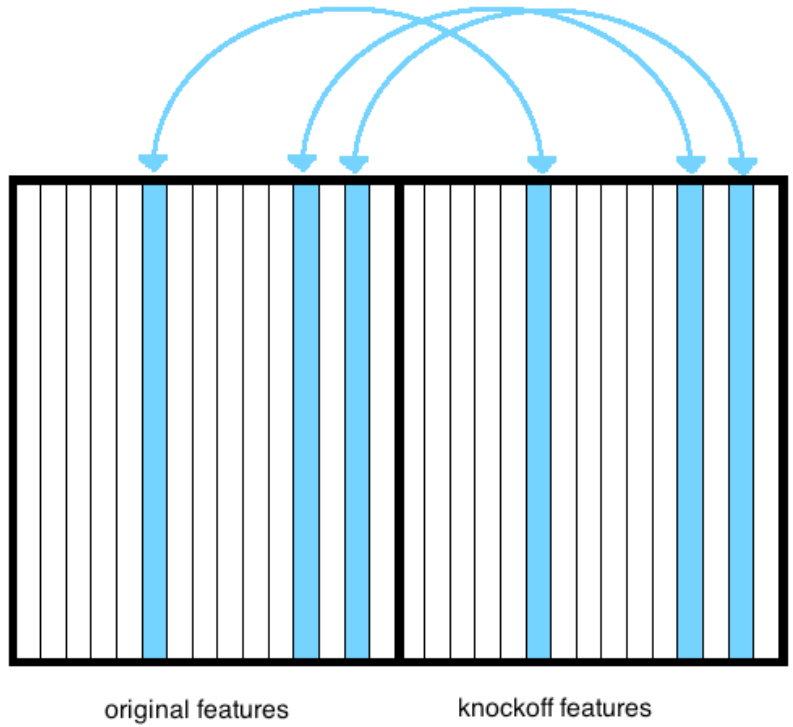
$$X_j'y = X_j'X\beta + X_j'z \stackrel{d}{=} \tilde{X}_j'X\beta + \tilde{X}_j'z = \tilde{X}_j'y$$



Why?

For null feature X_j

$$X_j'y = X_j'X\beta + X_j'z \overset{d}{=} \tilde{X}_j'X\beta + \tilde{X}_j'z = \tilde{X}_j'y$$



Why?

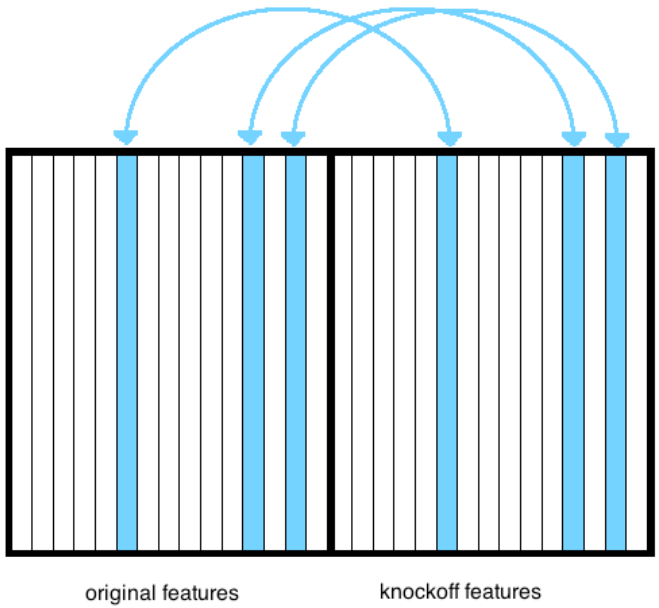
Lemma

Pairwise exchangeability property. For any subset of nulls N

$$[X\ \tilde{X}]'_{\text{swap}(N)}\ y \overset{d}{=} [X\ \tilde{X}]'\ y$$

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

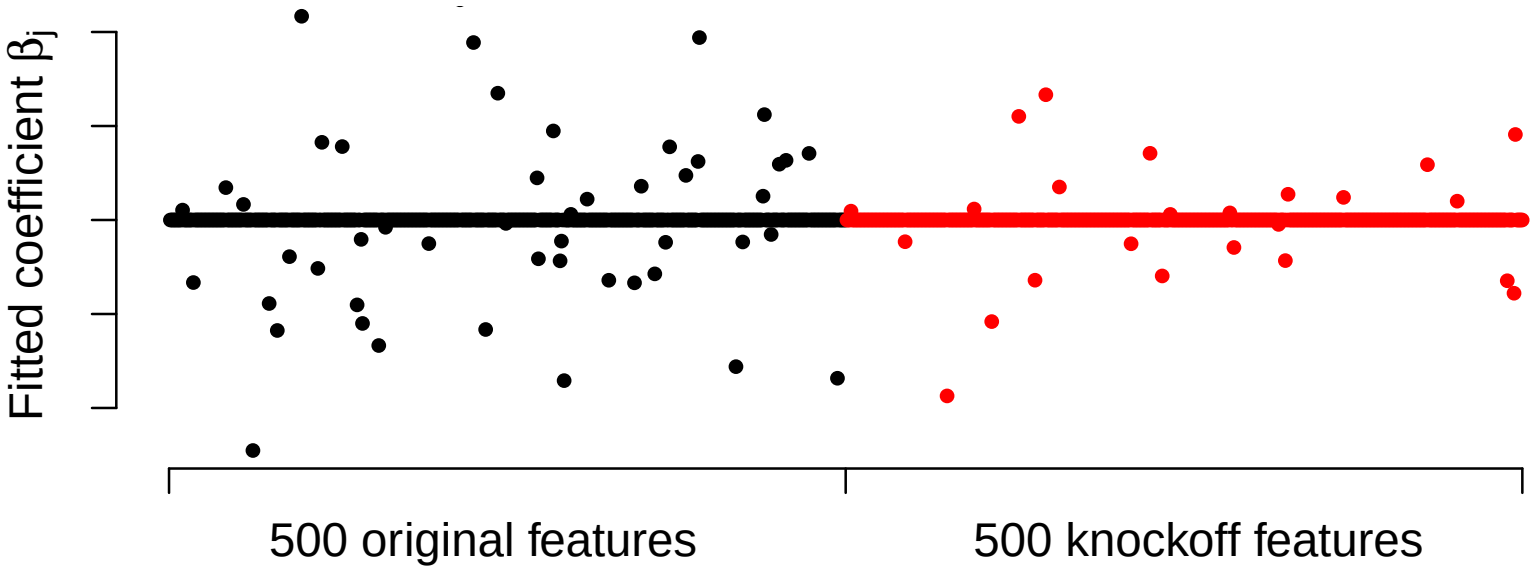
$$[X\ \tilde{X}]'_{\text{swap}(N)} =$$



Knockoff method

Compute Lasso with augmented matrix:

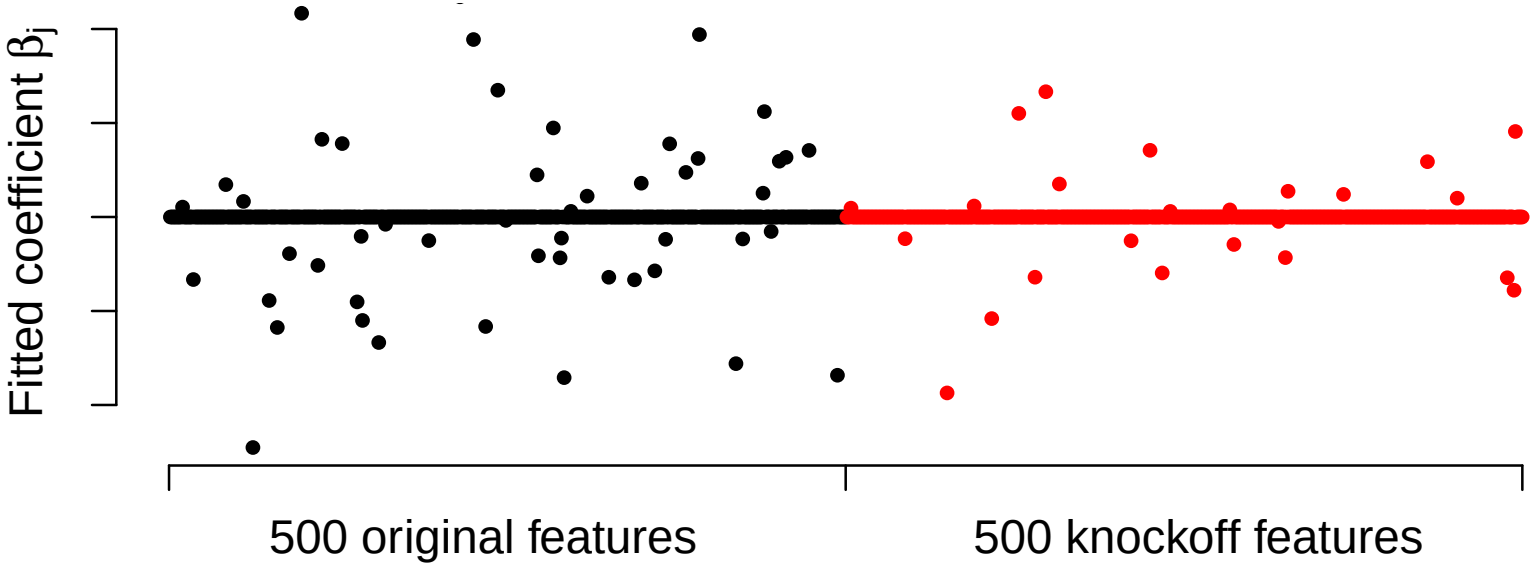
$$\hat{\beta}(\lambda) = \text{argmin}_{b \in \mathbb{R}^{2p}} \ \frac{1}{2} \left\| y - [X \ \tilde{X}] \cdot b \right\|^2 + \lambda \|b\|_1$$



Knockoff method

Compute Lasso with augmented matrix:

$$\hat{\beta}(\lambda) = \text{argmin}_{b \in \mathbb{R}^{2p}} \ \frac{1}{2} \left\| y - [X \ \tilde{X}] \cdot b \right\|^2 + \lambda \|b\|_1$$

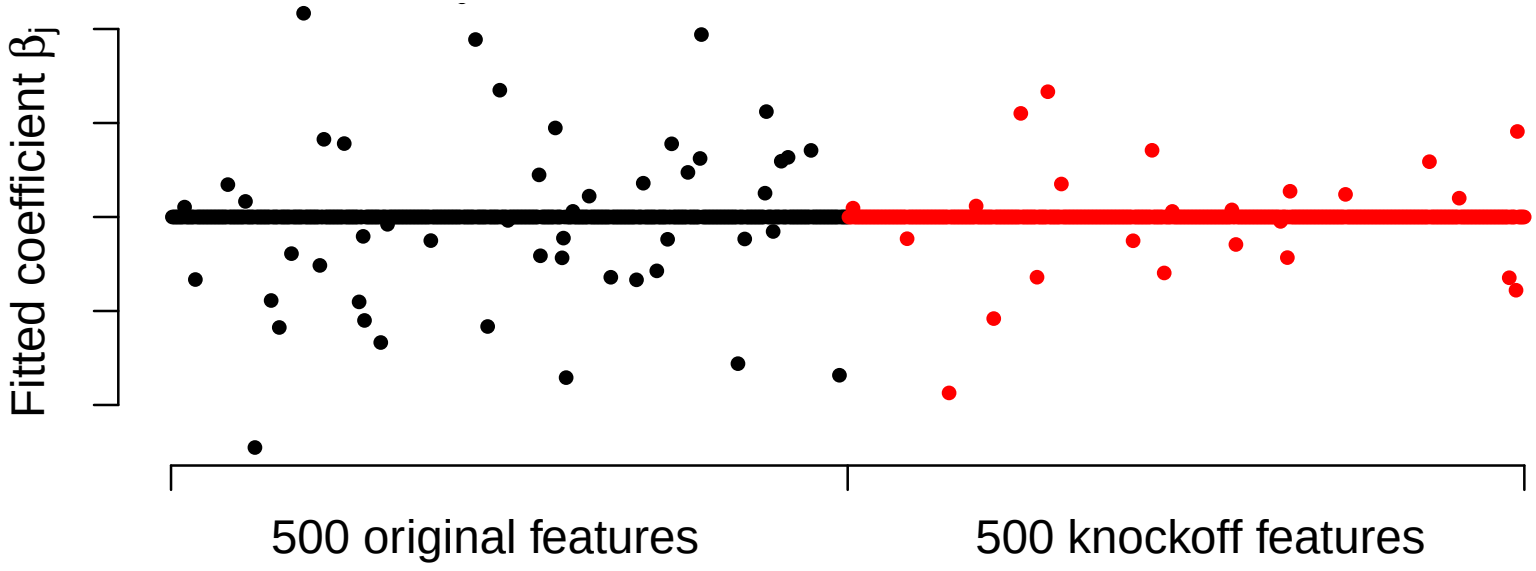


- Lasso selects 49 original features & 24 knockoff features

Knockoff method

Compute Lasso with augmented matrix:

$$\hat{\beta}(\lambda) = \text{argmin}_{b \in \mathbb{R}^{2p}} \ \frac{1}{2} \left\| y - [X \ \tilde{X}] \cdot b \right\|^2 + \lambda \|b\|_1$$



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

2. Statistics

$$Z_j = \sup\{\lambda : b_j(\lambda) \neq 0\} \qquad \text{first time } X_j \text{ enters model}$$
$$\tilde{Z}_j = \sup\{\lambda : \tilde{b}_j(\lambda) \neq 0\} \qquad \text{first time } \tilde{X}_j \text{ enters model}$$

Test statistic W_j for feature j

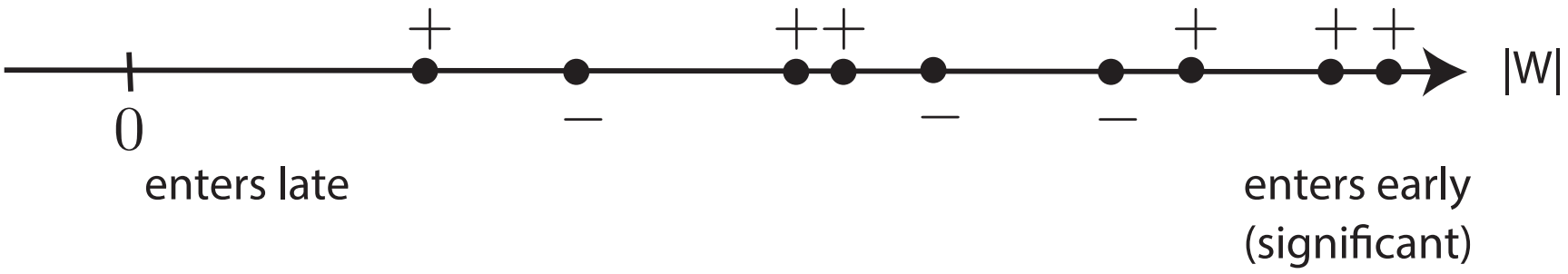
$$W_j = \max(Z_j, \tilde{Z}_j) \cdot \begin{cases} +1 & Z_j > \tilde{Z}_j \\ -1 & Z_j < \tilde{Z}_j \end{cases}$$

2. Statistics

$$Z_j = \sup\{\lambda : b_j(\lambda) \neq 0\} \qquad \text{first time } X_j \text{ enters model}$$
$$\tilde{Z}_j = \sup\{\lambda : \tilde{b}_j(\lambda) \neq 0\} \qquad \text{first time } \tilde{X}_j \text{ enters model}$$

Test statistic W_j for feature j

$$W_j = \max(Z_j, \tilde{Z}_j) \cdot \begin{cases} +1 & Z_j > \tilde{Z}_j \\ -1 & Z_j < \tilde{Z}_j \end{cases}$$

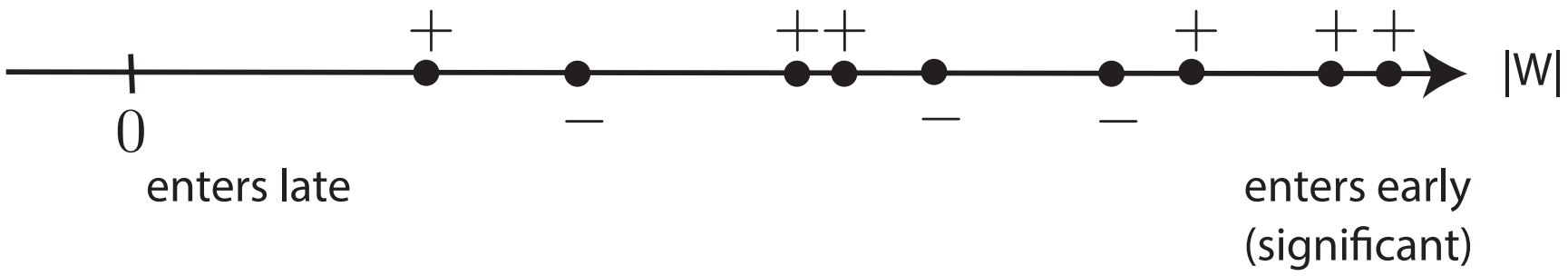


2. Statistics

$$Z_j = \sup\{\lambda : b_j(\lambda) \neq 0\} \qquad \text{first time } X_j \text{ enters model}$$
$$\tilde{Z}_j = \sup\{\lambda : \tilde{b}_j(\lambda) \neq 0\} \qquad \text{first time } \tilde{X}_j \text{ enters model}$$

Test statistic W_j for feature j

$$W_j = \max(Z_j, \tilde{Z}_j) \cdot \begin{cases} +1 & Z_j > \tilde{Z}_j \\ -1 & Z_j < \tilde{Z}_j \end{cases}$$

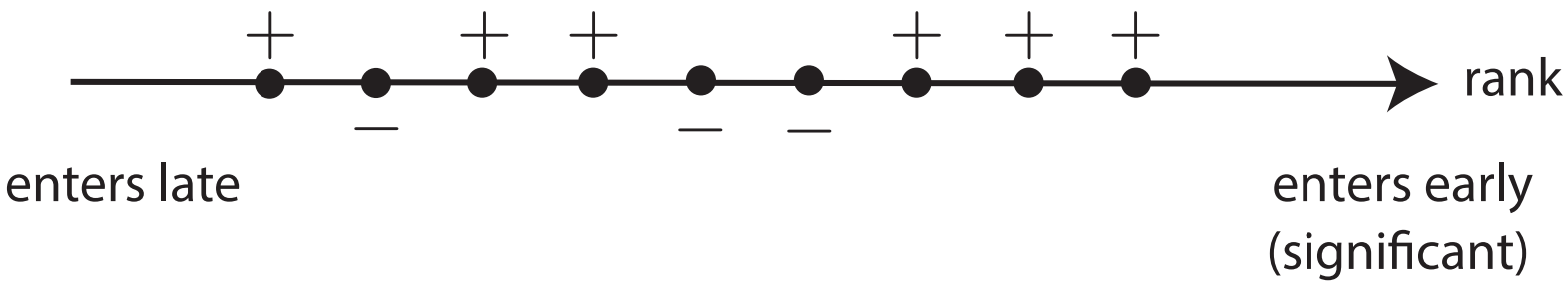


Many other choices (later)

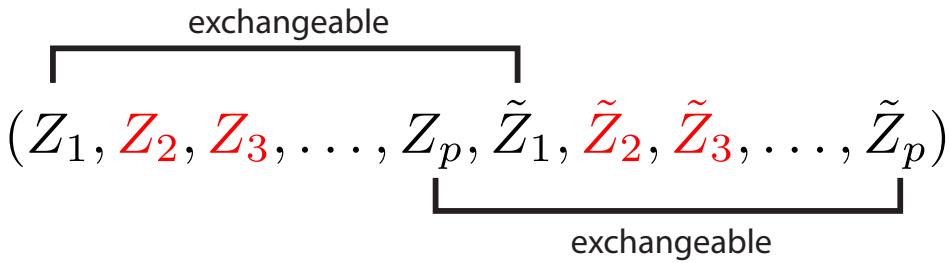
Variation

Forward selection on augmented design $\begin{bmatrix} X & \tilde{X} \end{bmatrix}$

- First time (rank) either original or knockoff enters
- Tag '+' if original comes before its knockoff, '-' otherwise



Pairwise exchangeability of the nulls

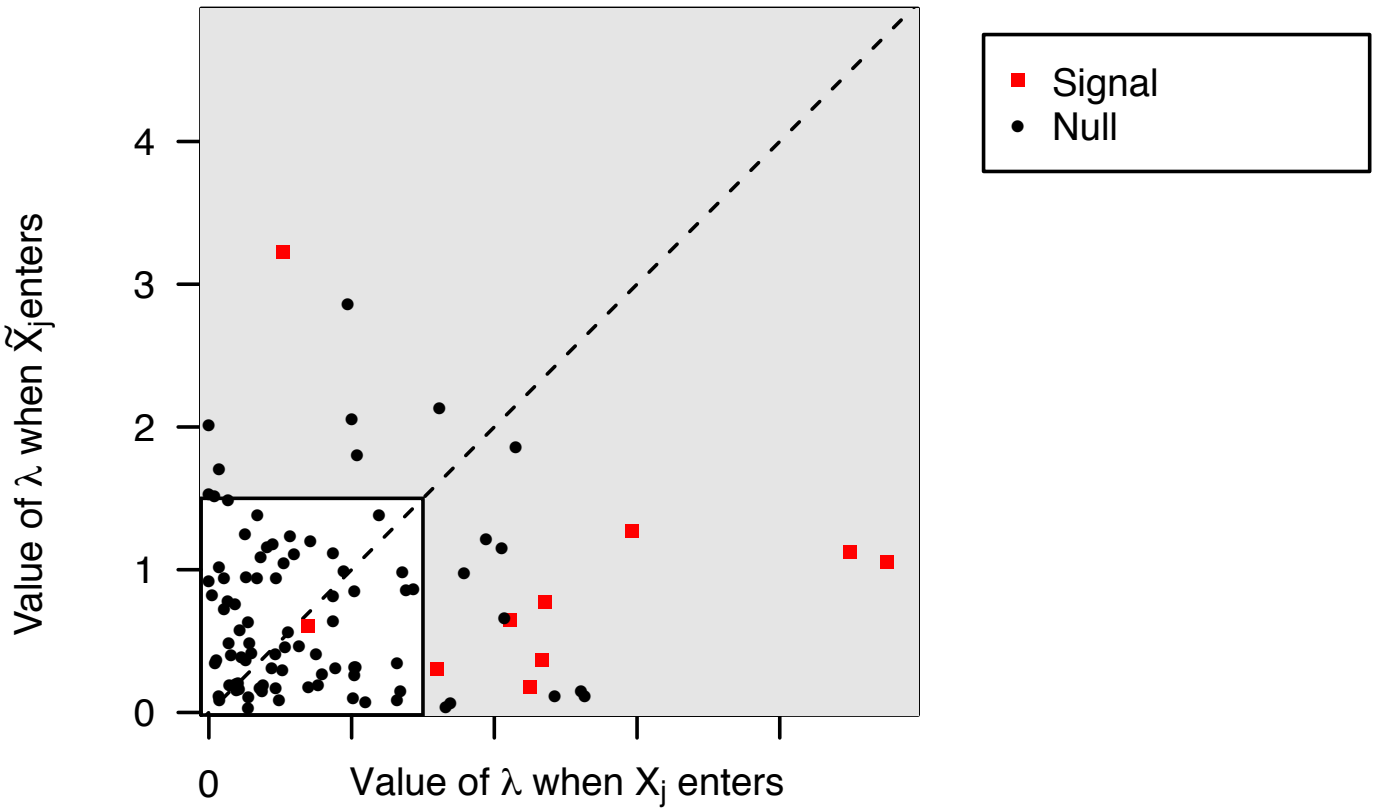


Pairwise exchangeability of the nulls

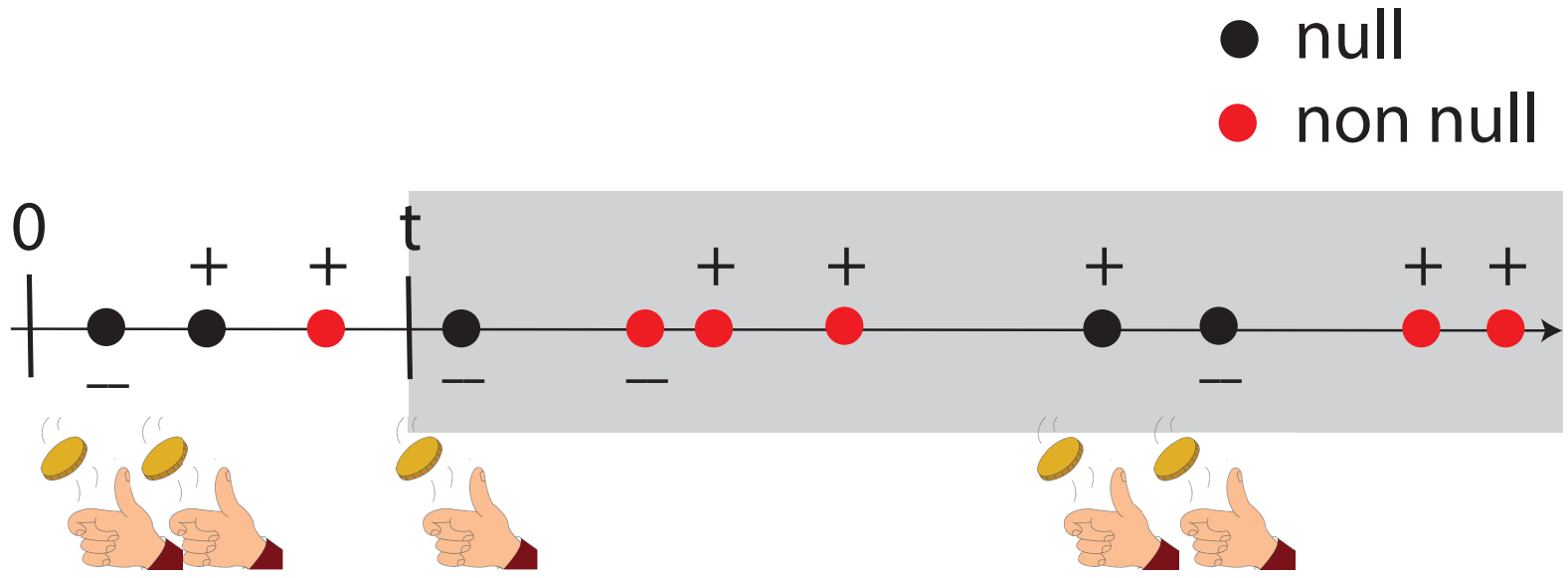
exchangeable

$$(Z_1, Z_2, Z_3, \ldots, Z_p, \tilde{Z}_1, \tilde{Z}_2, \tilde{Z}_3, \ldots, \tilde{Z}_p)$$

exchangeable



Consequence of exchangeability



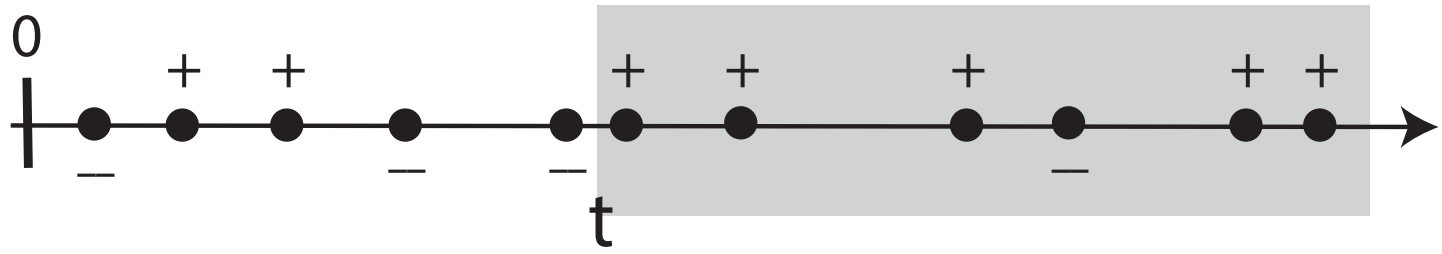
Signs of nulls iid ± 1 indep. of $|W|$ (ordering)

$$(W_1, \ldots, W_p) \stackrel{d}{=} (W_1 \cdot \epsilon_1, \ldots, W_p \cdot \epsilon_p)$$

Sign seq. $\{\epsilon_j\}$ indep. of W, $\epsilon_j = +1$ for all non-null j and $\epsilon_j \overset{\text{i.i.d.}}{\sim} \{\pm 1\}$ for null j

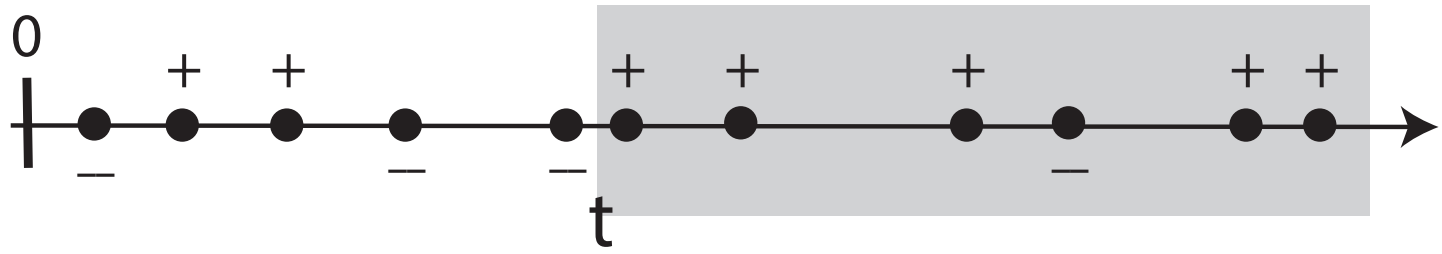
Signs $\rightarrow$ 1-bit p-values

Knockoff estimate of FDR



$$\mathsf{FDP}(t) = \frac{\#\{j \text{ null } : W_j \geq t\}}{\#\{j : W_j \geq t\} \vee 1} \approx \frac{\#\{j \text{ null } : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1}$$

Knockoff estimate of FDR



$$\mathsf{FDP}(t) = \frac{\#\{j \text{ null} \ : W_j \geq t\}}{\#\{j : W_j \geq t\} \vee 1} \approx \frac{\#\{j \text{ null} \ : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1}$$

$$\leq \frac{\#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1} := \widehat{\mathsf{FDP}}(t)$$

3. Selection and FDR control

Select features with large and positive statistics $\{W_j \geq T\}$

$$T = \min\left\{t \in \mathcal{W} : \frac{\#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1} \leq q\right\} \qquad \textbf{Knockoff}$$

Theorem (Knockoff)

$$\mathbb{E}\left[\frac{V}{R + q^{-1}}\right] \leq q$$

V : # false positives
R : total # of selections

3. Selection and FDR control

Select features with large and positive statistics $\{W_j \geq T\}$

$$T = \min\left\{t \in \mathcal{W} : \frac{\#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1} \leq q\right\} \qquad \textbf{Knockoff}$$

Theorem (Knockoff)

$$\mathbb{E}\left[\frac{V}{R + q^{-1}}\right] \leq q$$

V : # false positives
R : total # of selections

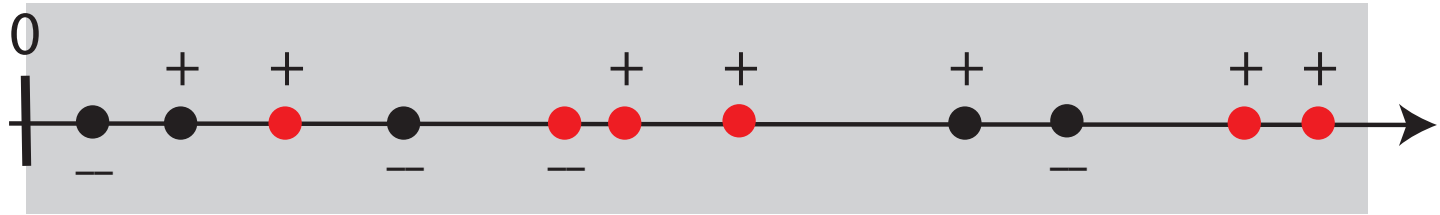
$$T = \min\left\{t \in \mathcal{W} : \frac{1 + \#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1} \leq q\right\} \qquad \textbf{Knockoff+}$$

Theorem (Knockoff+)

$$\mathbb{E}\left[\frac{V}{R \vee 1}\right] \leq q$$

Stopping rule

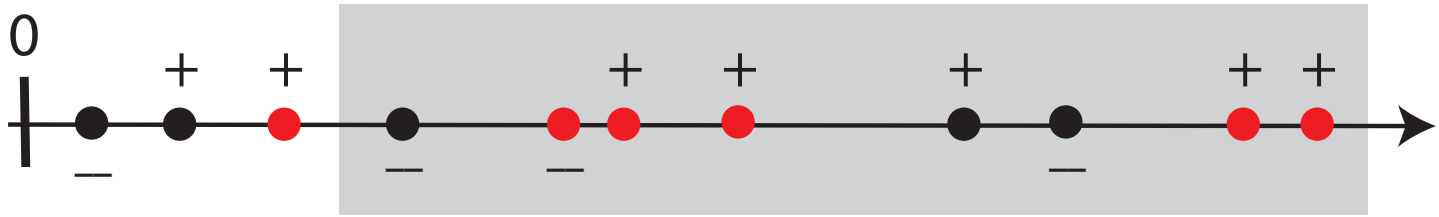
$$T = \min\left\{t : \frac{0/1 + \#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1} \leq q\right\}$$



Stop first time ratio between # negatives and # positives below q

Stopping rule

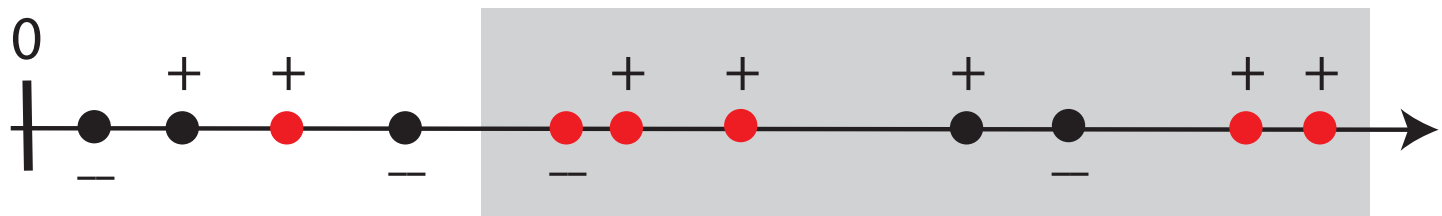
$$T = \min\left\{t : \frac{0/1 + \#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1} \leq q\right\}$$



Stop first time ratio between # negatives and # positives below q

Stopping rule

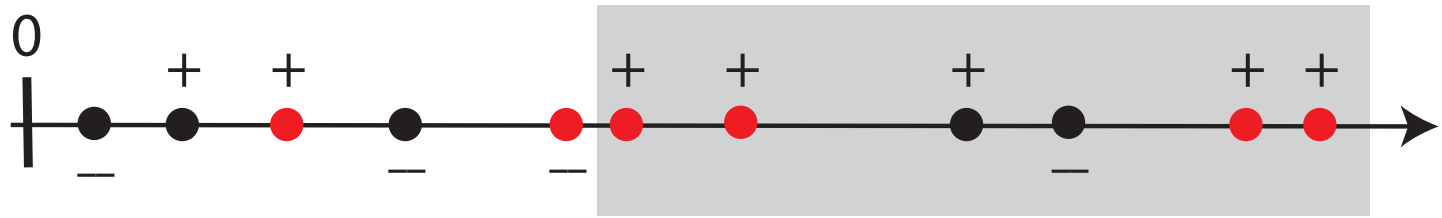
$$T = \min\left\{t : \frac{0/1 + \#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1} \leq q\right\}$$



Stop first time ratio between # negatives and # positives below q

Stopping rule

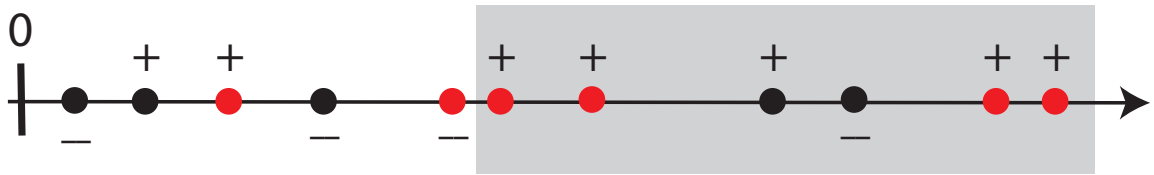
$$T = \min\left\{t : \frac{0/1 + \#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1} \leq q\right\}$$



Stop first time ratio between # negatives and # positives below q

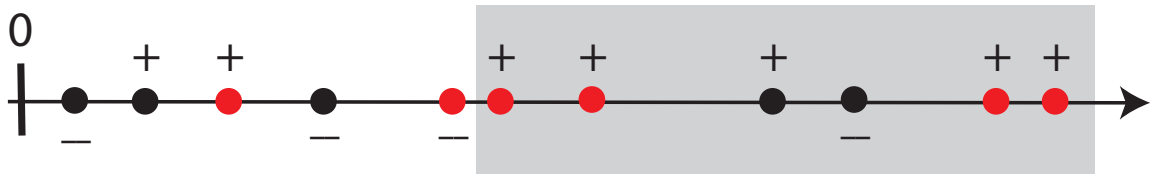
Why does all this work?

$$T = \min\left\{t : \frac{1+\#\{j : W_j \le -t\}}{\#\{j : W_j \ge t\} \vee 1} \le q\right\}$$



Why does all this work?

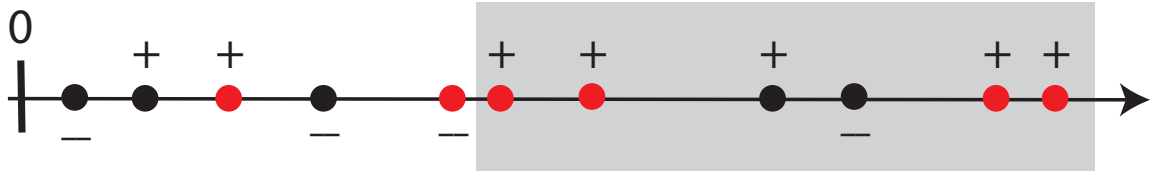
$$T = \mathsf{min}\left\{t : \frac{1+\#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\} \vee 1} \leq q\right\}$$



$$\mathsf{FDP}(T) = \frac{\#\{j \text{ null} : W_j \geq T\}}{\#\{j : W_j \geq T\} \vee 1} \cdot \frac{1 + \#\{j \text{ null} : W_j \leq -T\}}{1 + \#\{j \text{ null} : W_j \leq -T\}}$$

Why does all this work?

$$T = \min\left\{t : \frac{1+\#\{j : W_j \le -t\}}{\#\{j : W_j \ge t\} \vee 1} \le q\right\}$$

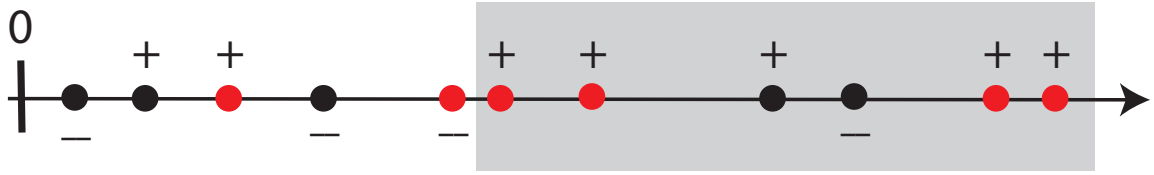


$$\mathsf{FDP}(T) = \frac{\#\{j \text{ null} : W_j \ge T\}}{\#\{j : W_j \ge T\} \vee 1} \cdot \frac{1 + \#\{j \text{ null} : W_j \le -T\}}{1 + \#\{j \text{ null} : W_j \le -T\}}$$

$$\le q \cdot \frac{\overbrace{\#\{j \text{ null} : W_j \ge T\}}^{V^+(T)}}{1 + \underbrace{\#\{j \text{ null} : W_j \le -T\}}_{V^-(T)}}$$

Why does all this work?

$$T = \min\left\{t : \frac{1+\#\{j : W_j \le -t\}}{\#\{j : W_j \ge t\} \vee 1} \le q\right\}$$

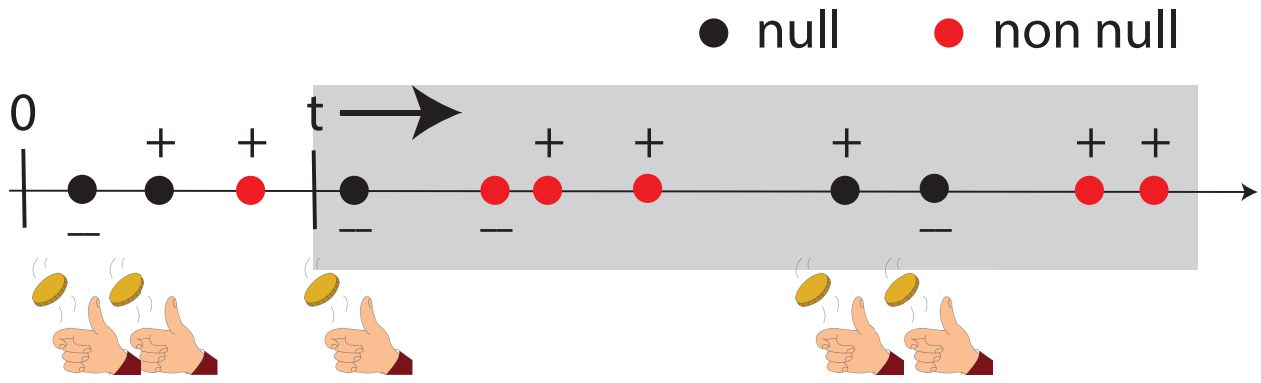


$$\mathrm{FDP}(T) = \frac{\#\{j \text{ null} : W_j \ge T\}}{\#\{j : W_j \ge T\} \vee 1} \cdot \frac{1 + \#\{j \text{ null} : W_j \le -T\}}{1 + \#\{j \text{ null} : W_j \le -T\}}$$

$$\le q \cdot \frac{\overbrace{\#\{j \text{ null} : W_j \ge T\}}^{V^+(T)}}{1 + \underbrace{\#\{j \text{ null} : W_j \le -T\}}_{V^-(T)}}$$

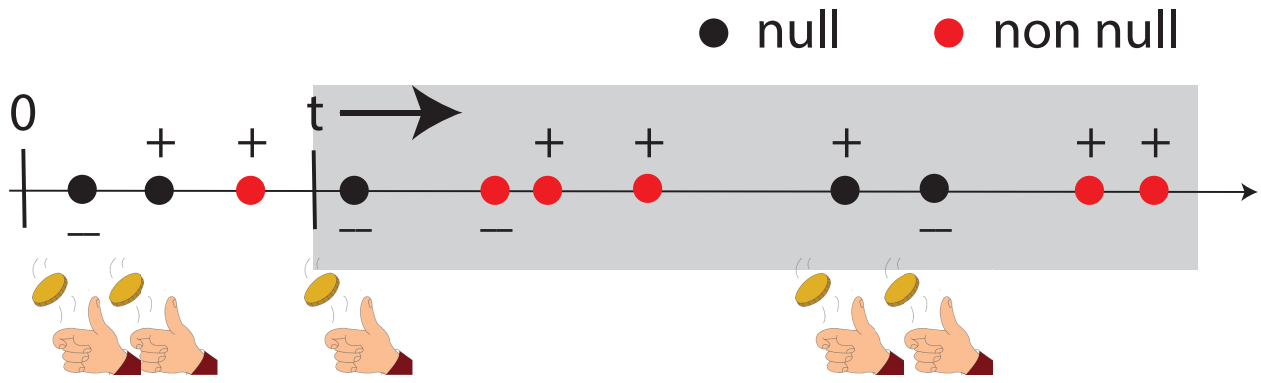
- $V^+(t)/(1+V^-(t))$ is a super-martingale w.r.t. well defined filtration
- T is stopping time

Optional stopping time theorem



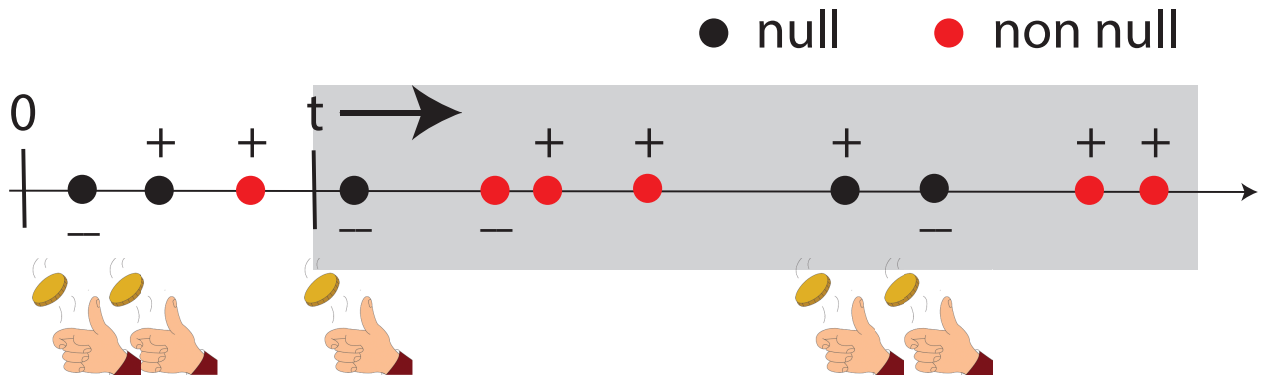
$$\text{FDR} \leq q\, \mathbb{E}\left[\frac{V^+(T)}{1+V^-(T)}\right]$$

Optional stopping time theorem



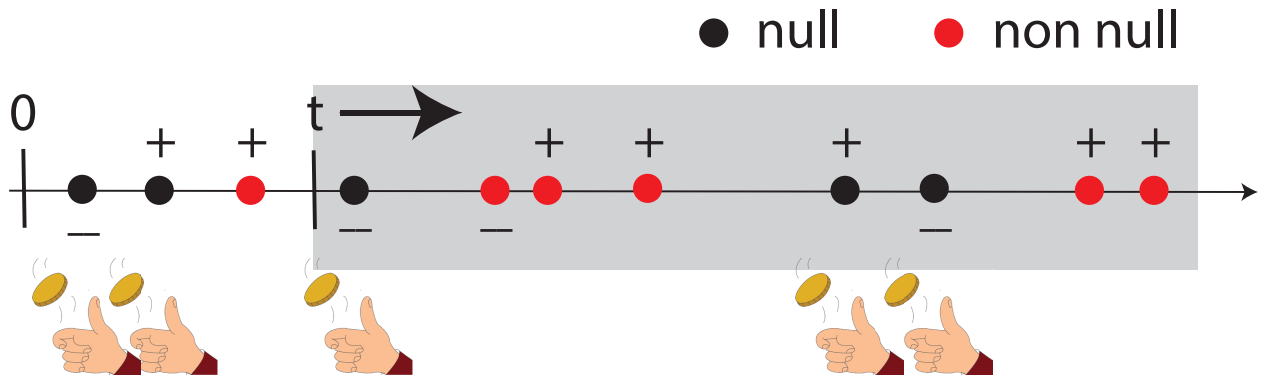
$$\text{FDR} \leq q\ \mathbb{E}\left[\frac{V^+(T)}{1+V^-(T)}\right] \leq q\ \mathbb{E}\left[\frac{V^+(0)}{1+V^-(0)}\right]$$

Optional stopping time theorem



$$\text{FDR} \leq q\ \mathbb{E}\left[\frac{V^+(T)}{1+V^-(T)}\right] \leq q\ \mathbb{E}\left[\frac{V^+(0)}{1+V^-(0)}\right] = q\ \mathbb{E}\left[\frac{\overbrace{V^+(0)}^{\text{Ber}(\#\text{nulls},1/2)}}{1+\#\text{nulls}-V^+(0)}\right]$$

Optional stopping time theorem



$$\text{FDR} \leq q\ \mathbb{E}\left[\frac{V^+(T)}{1+V^-(T)}\right] \leq q\ \mathbb{E}\left[\frac{V^+(0)}{1+V^-(0)}\right] = q\ \mathbb{E}\left[\frac{\overbrace{V^+(0)}^{\text{Ber}(\#\text{nulls},1/2)}}{1+\#\text{nulls}-V^+(0)}\right] \leq q$$

Comparison with other methods

Permutation methods

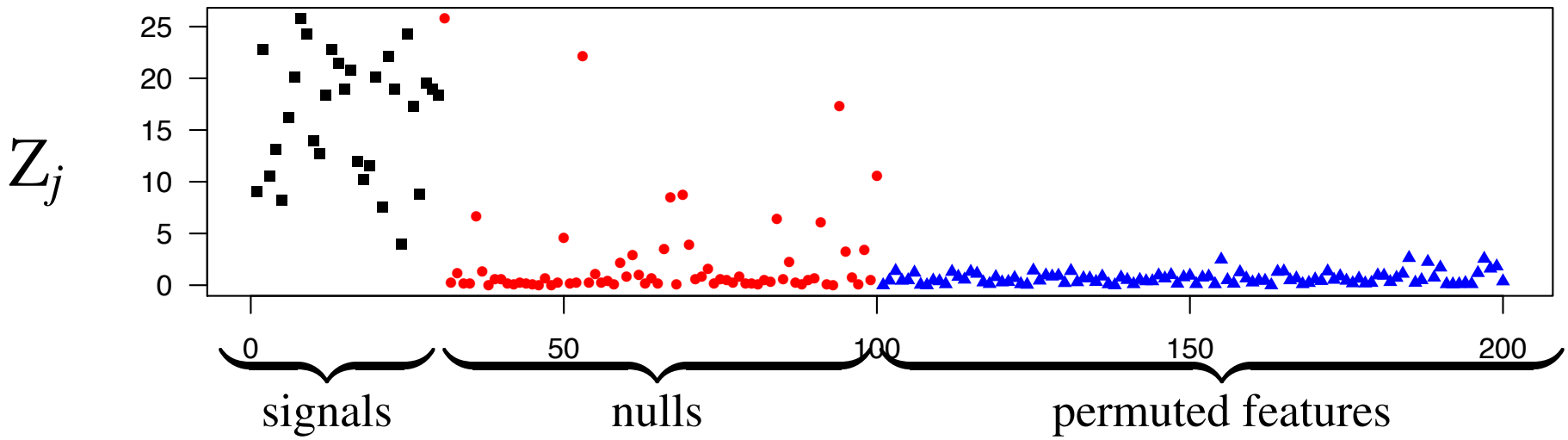
Let $X^\pi = X$ with rows randomly permuted

$$\begin{bmatrix} X & X^\pi \end{bmatrix}' \begin{bmatrix} X & X^\pi \end{bmatrix} \approx \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma \end{bmatrix}$$

Permutation methods

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

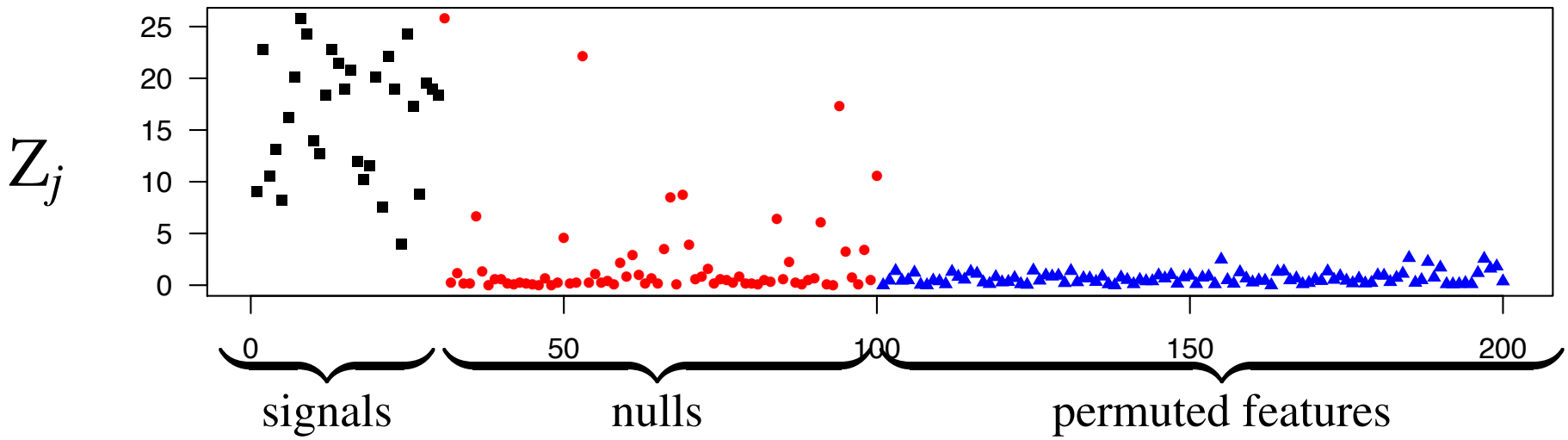
$$\begin{bmatrix} X & X^\pi \end{bmatrix}' \begin{bmatrix} X & X^\pi \end{bmatrix} \approx \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma \end{bmatrix}$$



Permutation methods

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

$$\begin{bmatrix} X & X^\pi \end{bmatrix}' \begin{bmatrix} X & X^\pi \end{bmatrix} \approx \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma \end{bmatrix}$$



	FDR over 1000 trials (nominal level $q = 20\%$)
Knockoff method	12.29%
Permutation method	45.61%

Other methods

- Benjamini-Hochberg (BHq)
- BHq + log factor correction
- BHq with whitened noise

Other methods

- Benjamini-Hochberg (BHq)
- BHq + log factor correction
- BHq with whitened noise

$$y \sim \mathcal{N}(X\beta, \sigma^2 I) \quad \Longleftrightarrow \quad \hat{\beta}^{\mathsf{LS}} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1})$$

Apply BHq to

$$Z_j = \frac{\hat{\beta}_j^{\mathsf{LS}}}{\sigma \sqrt{(\Sigma^{-1})_{jj}}}$$

Not known to control FDR $\rightarrow$ log factor correction (Benjamini Yekutieli)

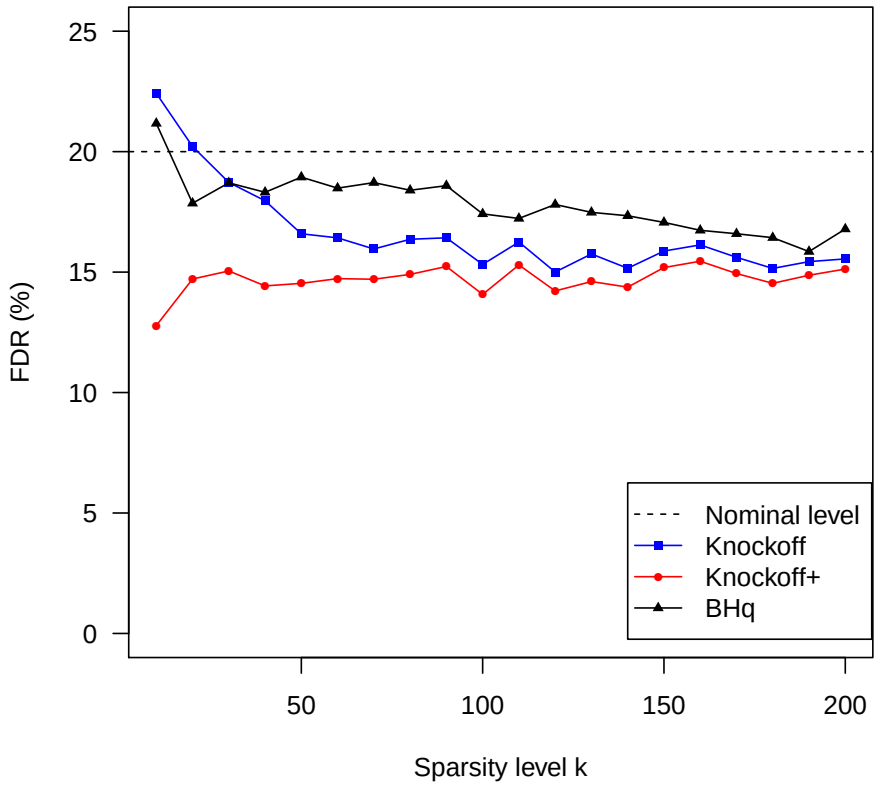
Empirical results

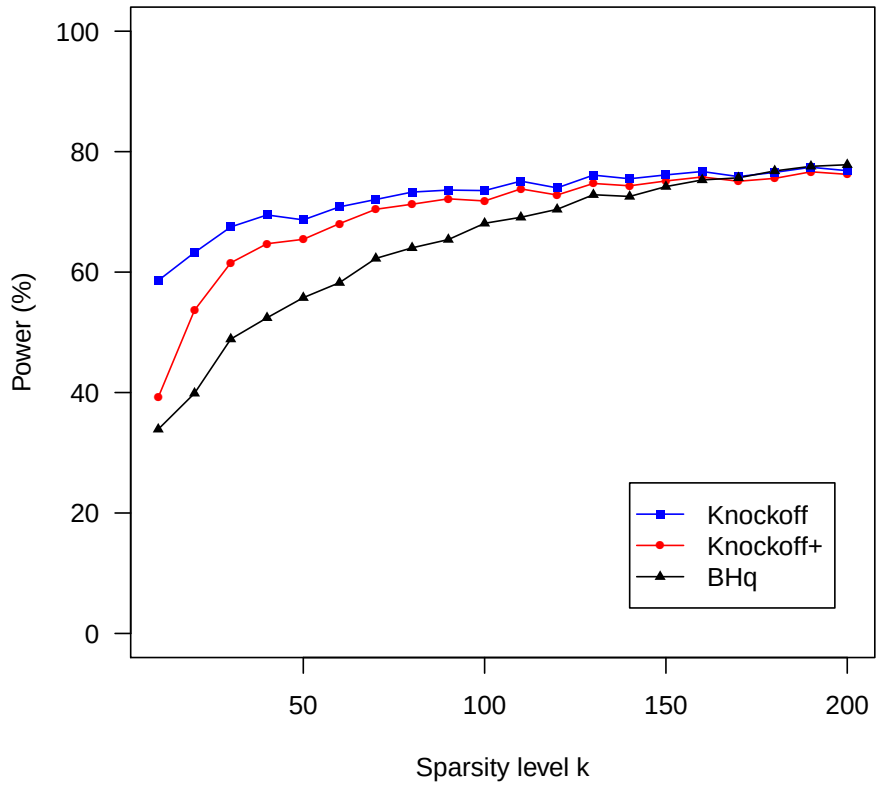
- Features $\mathcal{N}(0, I_n)$, $n = 3000$, $p = 1000$
- $k = 30$ variables with regression coefficients of magnitude 3.5

Method	FDR (%) (nominal level $q = 20\%$)	Power (%)	Theor. FDR control?
Knockoff+ (equivariant)	**14.40**	**60.99**	**Yes**
Knockoff (equivariant)	17.82	66.73	No
Knockoff+ (SDP)	**15.05**	**61.54**	**Yes**
Knockoff (SDP)	18.72	67.50	No
BHq	18.70	48.88	No
BHq + log-factor correction	**2.20**	**19.09**	**Yes**
BHq with whitened noise	**18.79**	**2.33**	**Yes**

Effect of sparsity level

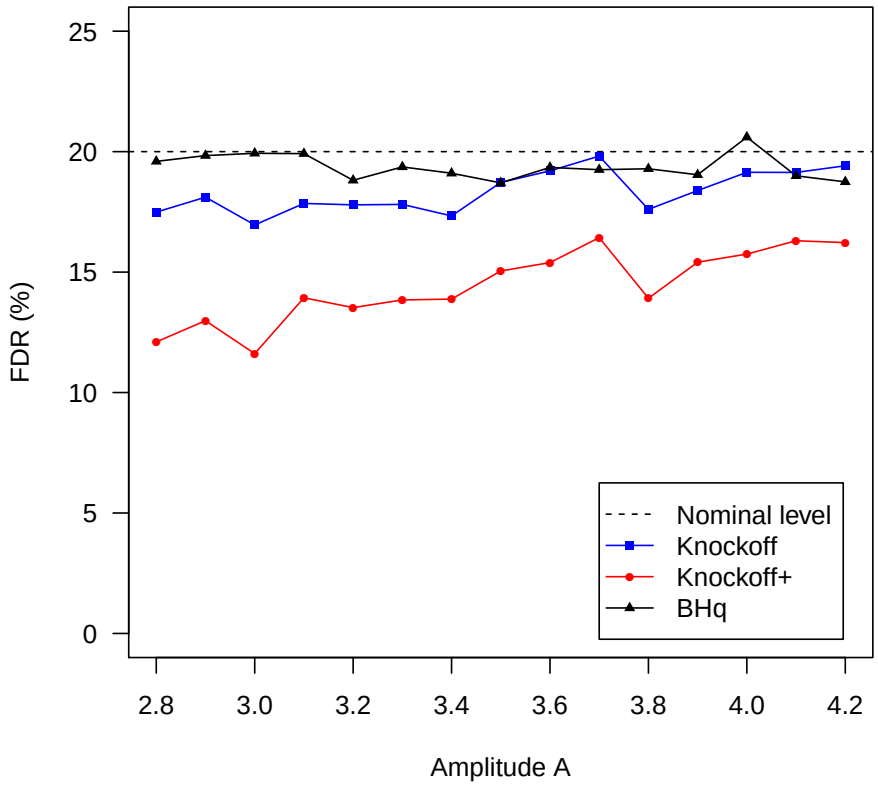
Same setup with amplitudes set to 3.5 ($q = 0.2$)

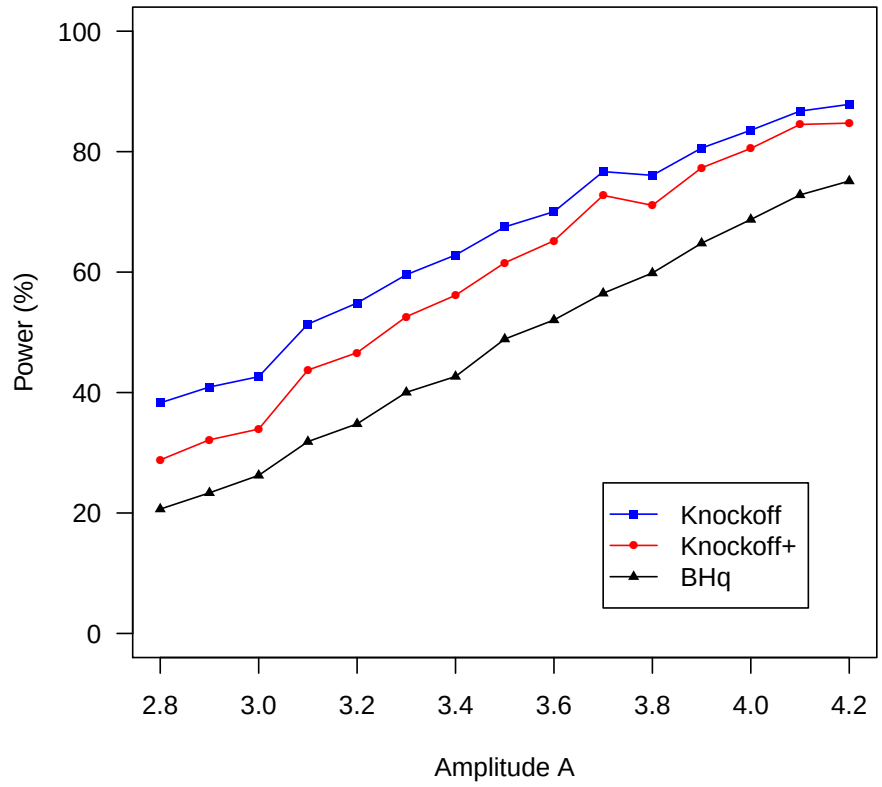




Effect of signal amplitude

Same setup with $k = 30$ $(q = 0.2)$

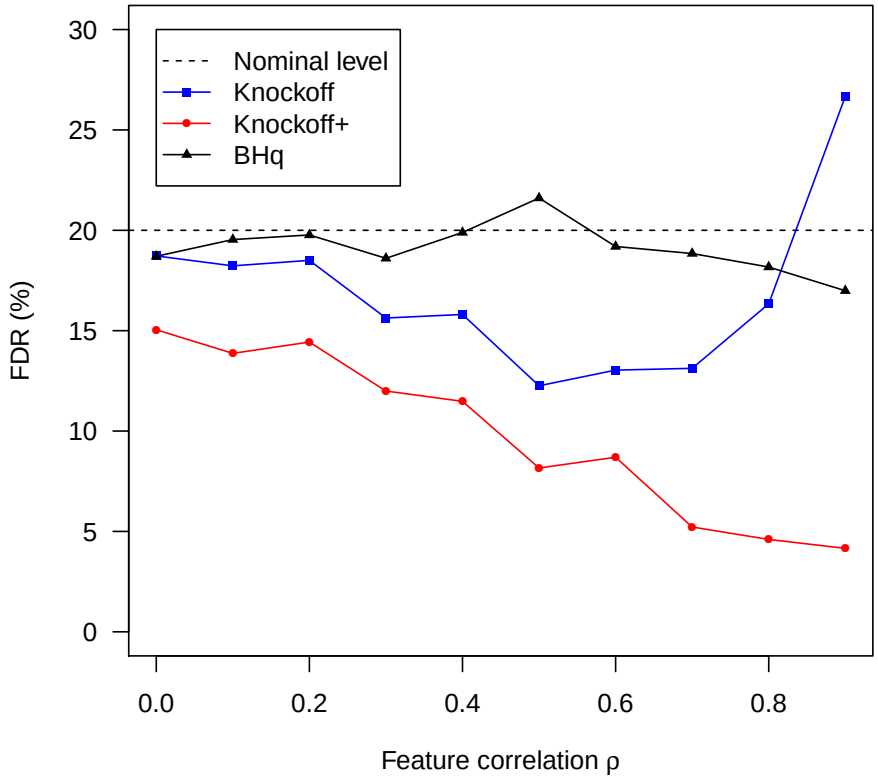


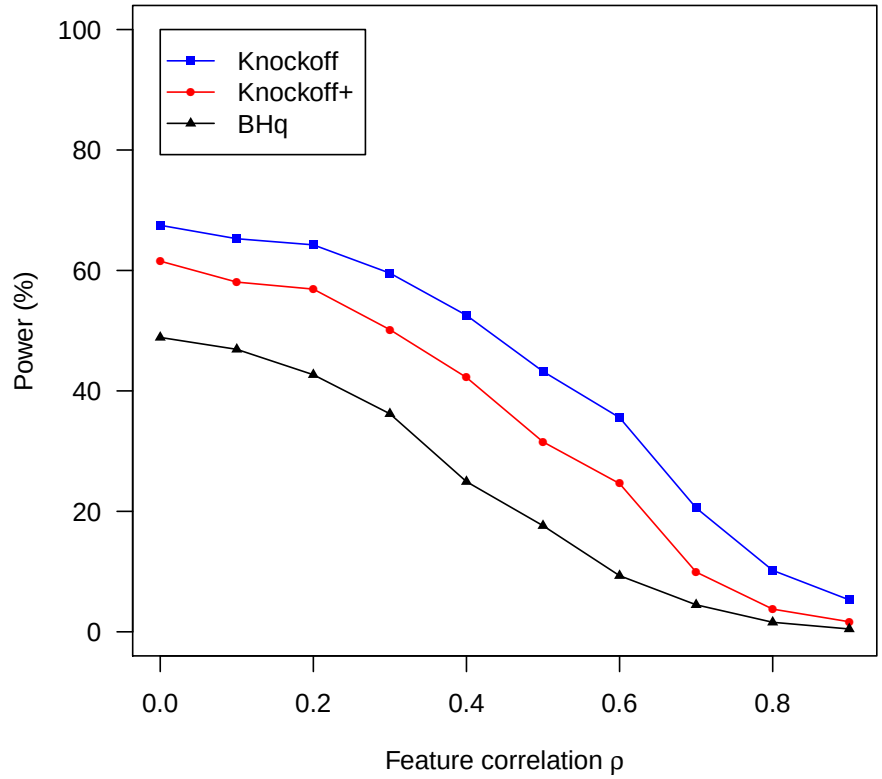


Effect of feature correlation

$$\text{Features} \sim \mathcal{N}(0, \Theta) \qquad \Theta_{jk} = \rho^{|j-k|}$$

$n = 3000$, $p = 1000$, and $k = 30$ and amplitude $= 3.5$





Application to real HIV data

HIV drug resistance

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

- response y: log-fold-increase of lab-tested drug resistance in
- covariate X_j: presence or absence of mutation $\#j$

Data from R. Shafer (Stanford) available at:

`http://hivdb.stanford.edu/pages/published_analysis/genophenoPNAS2006/`

PI-type drug resistance

TSM list: mutations associated with the PI class of drugs in general, and is not specialized to the individual drugs in the class

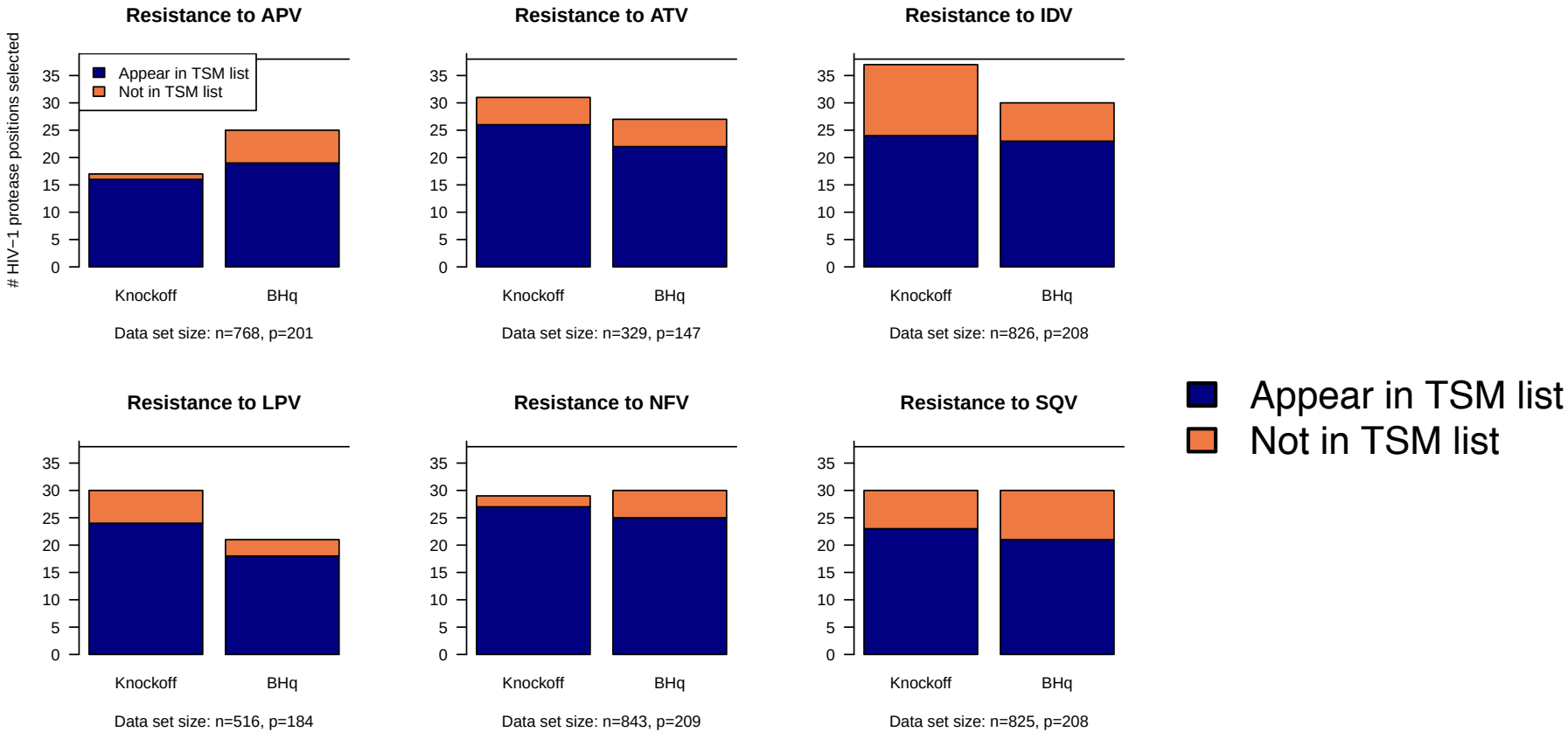


Figure: $q = 0.2$. # positions on the HIV-1 protease where mutations were selected. Horizontal line = # HIV-1 protease positions in TSM list.

NRTI-type drug resistance

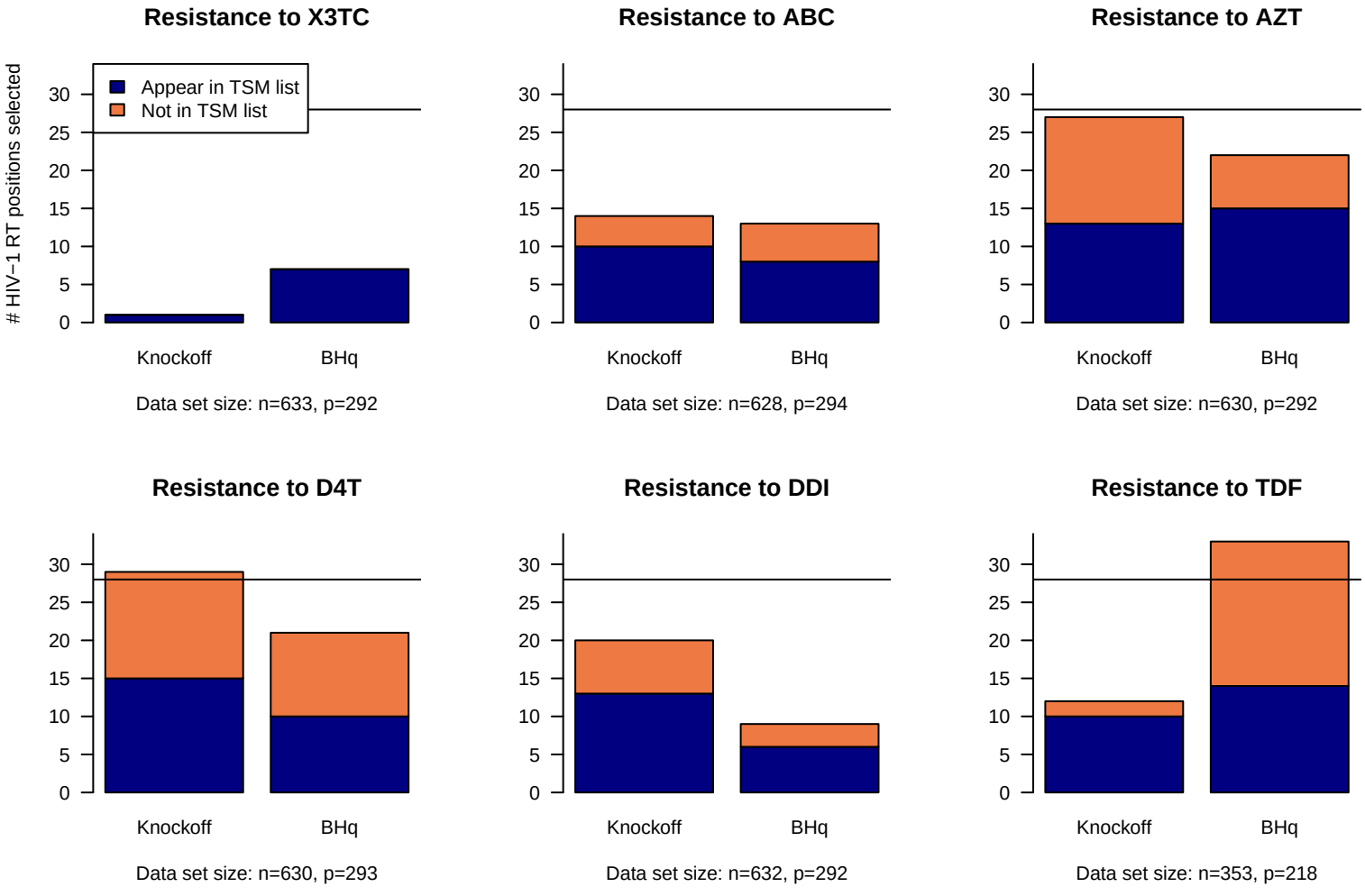


Figure: Validation against the treatment-selected mutation (TSM) panel

NNRTI-type drug resistance

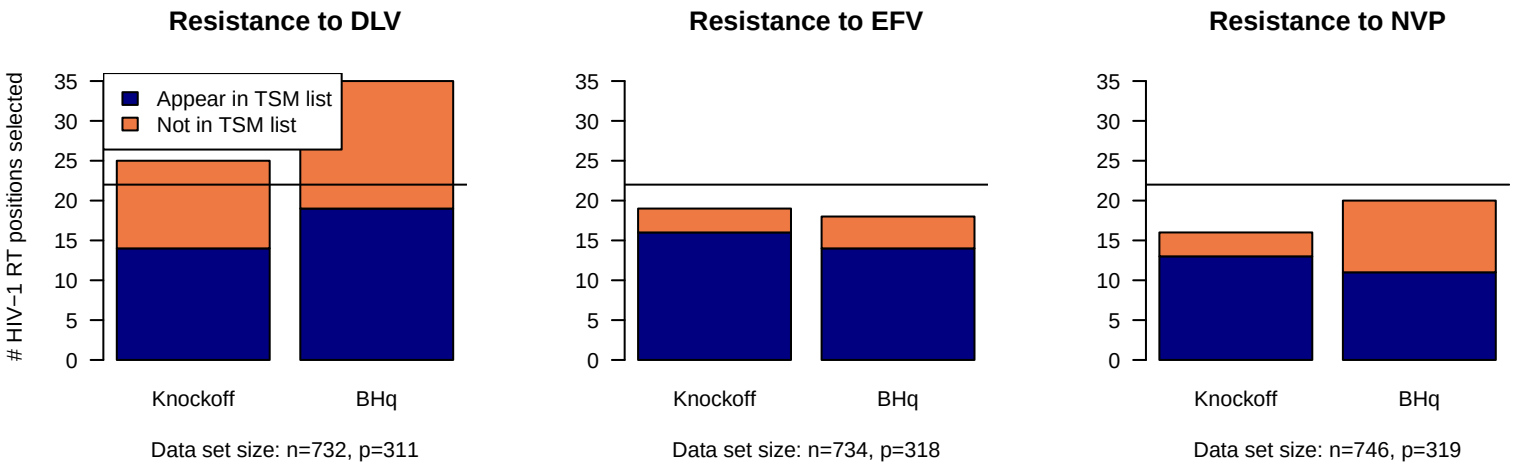


Figure: Validation against the treatment-selected mutation (TSM) panel

Heavy-tailed noise

- Design as in HIV example (sparse matrix)
- Errors as residuals from HIV example
- Regression coefficients entered manually

	FDR ($q = 20\%$)	Power
Knockoff+	20.31%	60.67%
BHq	25.47%	69.42%

Details

Knockoff constructions ($n \geq 2p$)

$$\tilde{X} = X(I - \Sigma^{-1}\,\mathrm{diag}\{s\}) + \tilde{U}C$$

Knockoff constructions ($n \geq 2p$)

$$\tilde{X} = X(I - \Sigma^{-1}\,\mathrm{diag}\{s\}) + \tilde{U}C$$

$$[X \quad \tilde{X}]'\,[X \quad \tilde{X}] = \begin{bmatrix} \Sigma & \Sigma - \mathrm{diag}\{s\} \\ \Sigma - \mathrm{diag}\{s\} & \Sigma \end{bmatrix} := G \succeq 0$$

Knockoff constructions ($n \geq 2p$)

$$\tilde{X} = X(I - \Sigma^{-1}\,\mathrm{diag}\{s\}) + \tilde{U}C$$

$$[X \quad \tilde{X}]'\,[X \quad \tilde{X}] = \begin{bmatrix} \Sigma & \Sigma - \mathrm{diag}\{s\} \\ \Sigma - \mathrm{diag}\{s\} & \Sigma \end{bmatrix} := G \succeq 0$$

$$G \succeq 0 \quad \Longleftrightarrow \quad \begin{array}{l} \mathrm{diag}\{s\} \succeq 0 \\ 2\Sigma - \mathrm{diag}\{s\} \succeq 0 \end{array}$$

Knockoff constructions($n \geq 2p$)

- *Equi-correlated knockoffs:* $s_j = 2\lambda_{\min}(\Sigma) \wedge 1$

$$\langle X_j, \tilde{X}_j \rangle = 1 - 2\lambda_{\min}(\Sigma) \wedge 1$$

Under equivariance, minimizes the value of $|\langle X_j, \tilde{X}_j \rangle|$

Knockoff constructions($n \geq 2p$)

- *Equi-correlated knockoffs:* $s_j = 2\lambda_{\min}(\Sigma) \wedge 1$

$$\langle X_j, \tilde{X}_j \rangle = 1 - 2\lambda_{\min}(\Sigma) \wedge 1$$

Under equivariance, minimizes the value of $|\langle X_j, \tilde{X}_j \rangle|$

- *SDP knockoffs:*

$$\begin{array}{lll} \text{minimize} & \sum_j |1 - s_j| & \\ \text{subject to} & s_j \geq 0 & \iff \\ & \mathrm{diag}\{s\} \preceq 2\Sigma & \end{array} \qquad \begin{array}{lll} \text{minimize} & \sum_j (1 - s_j) & \\ \text{subject to} & s_j \geq 0 & \\ & \mathrm{diag}\{s\} \preceq 2\Sigma & \end{array}$$

Highly structured semidefinite program (SDP)

- Other possibilities

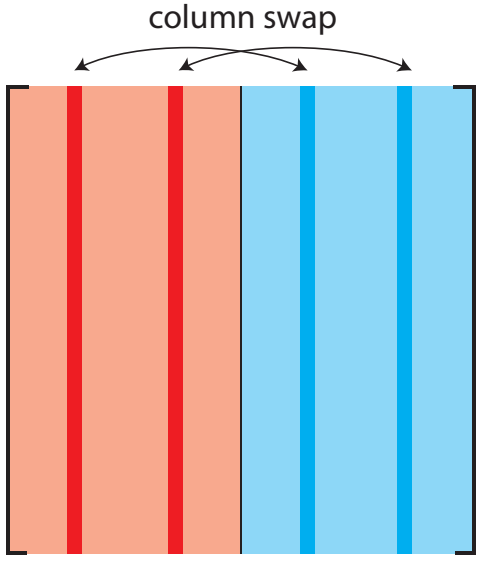
Symmetric statistics: $W([X \;\; \tilde{X}], y)$

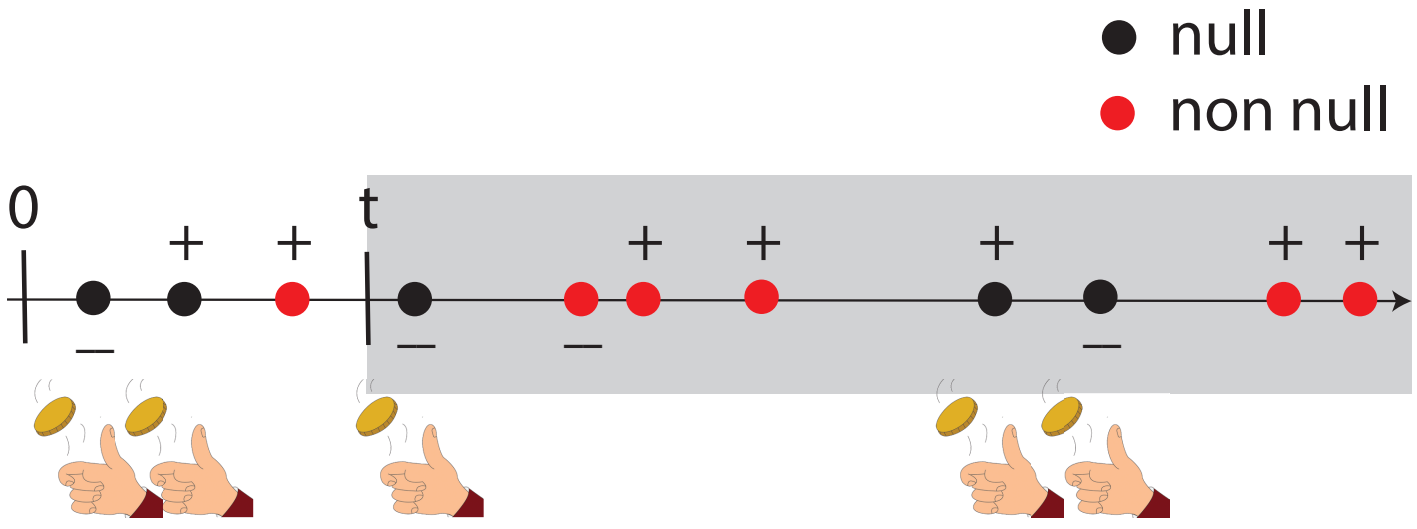
- *Sufficiency property:*

$$W = f\left([X \;\; \tilde{X}]'[X \;\; \tilde{X}], [X \;\; \tilde{X}]' y\right)$$

- *Anti-symmetry property:* swapping changes signs

$$W_j\left([X \;\; \tilde{X}]_{\text{swap}(S)}, y\right) = W_j\left([X \;\; \tilde{X}], y\right) \cdot \begin{cases} +1 & j \notin S \\ -1 & j \in S \end{cases}$$





Examples of statistics

- $Z_j = \sup\{\lambda : \hat{\beta}_j(\lambda) \neq 0\}$, $j = 1, \ldots, 2p$, and $\hat{\beta}(\lambda)$ sol. to augmented Lasso

$$W_j = (Z_j \vee Z_{j+p}) \cdot \operatorname{sign}(Z_j - Z_{j+p}) \qquad W_j = Z_j - Z_{j+p}$$

Examples of statistics

- $Z_j = \sup\{\lambda : \hat{\beta}_j(\lambda) \neq 0\}$, $j = 1, \ldots, 2p$, and $\hat{\beta}(\lambda)$ sol. to augmented Lasso

$$W_j = (Z_j \vee Z_{j+p}) \cdot \operatorname{sign}(Z_j - Z_{j+p}) \qquad W_j = Z_j - Z_{j+p}$$

- Same with penalized estimate

$$\min \quad \tfrac{1}{2}\|y - Xb\|_2^2 + \lambda P(b)$$

Examples of statistics

- $Z_j = \sup\{\lambda : \hat{\beta}_j(\lambda) \neq 0\}$, $j = 1, \ldots, 2p$, and $\hat{\beta}(\lambda)$ sol. to augmented Lasso

$$W_j = (Z_j \vee Z_{j+p}) \cdot \operatorname{sign}(Z_j - Z_{j+p}) \qquad W_j = Z_j - Z_{j+p}$$

- Same with penalized estimate

$$\min \quad \tfrac{1}{2}\|y - Xb\|_2^2 + \lambda P(b)$$

- Forward selection/orthogonal matching pursuit: $Z_1, \cdots, Z_{2p}$ (reverse) order in which $2p$ variables entered model

$$W_j = (Z_j \vee Z_{j+p}) \cdot \operatorname{sign}(Z_j - Z_{j+p})$$

Examples of statistics

- $Z_j = \sup\{\lambda : \hat{\beta}_j(\lambda) \neq 0\}$, $j = 1, \ldots, 2p$, and $\hat{\beta}(\lambda)$ sol. to augmented Lasso

$$W_j = (Z_j \vee Z_{j+p}) \cdot \operatorname{sign}(Z_j - Z_{j+p}) \qquad W_j = Z_j - Z_{j+p}$$

- Same with penalized estimate

$$\min \quad \tfrac{1}{2}\|y - Xb\|_2^2 + \lambda P(b)$$

- Forward selection/orthogonal matching pursuit: $Z_1, \cdots, Z_{2p}$ (reverse) order in which $2p$ variables entered model

$$W_j = (Z_j \vee Z_{j+p}) \cdot \operatorname{sign}(Z_j - Z_{j+p})$$

- Statistics based on LS estimates

$$W_j = |\hat{\beta}_j^{\mathsf{LS}}|^2 - |\hat{\beta}_{j+p}^{\mathsf{LS}}|^2 \qquad |\hat{\beta}_j^{\mathsf{LS}}| - |\hat{\beta}_{j+p}^{\mathsf{LS}}|$$

Examples of statistics

- $Z_j = \sup\{\lambda : \hat{\beta}_j(\lambda) \neq 0\}$, $j = 1, \ldots, 2p$, and $\hat{\beta}(\lambda)$ sol. to augmented Lasso

$$W_j = (Z_j \vee Z_{j+p}) \cdot \operatorname{sign}(Z_j - Z_{j+p}) \qquad W_j = Z_j - Z_{j+p}$$

- Same with penalized estimate

$$\min \quad \tfrac{1}{2}\|y - Xb\|_2^2 + \lambda P(b)$$

- Forward selection/orthogonal matching pursuit: $Z_1, \cdots, Z_{2p}$ (reverse) order in which $2p$ variables entered model

$$W_j = (Z_j \vee Z_{j+p}) \cdot \operatorname{sign}(Z_j - Z_{j+p})$$

- Statistics based on LS estimates

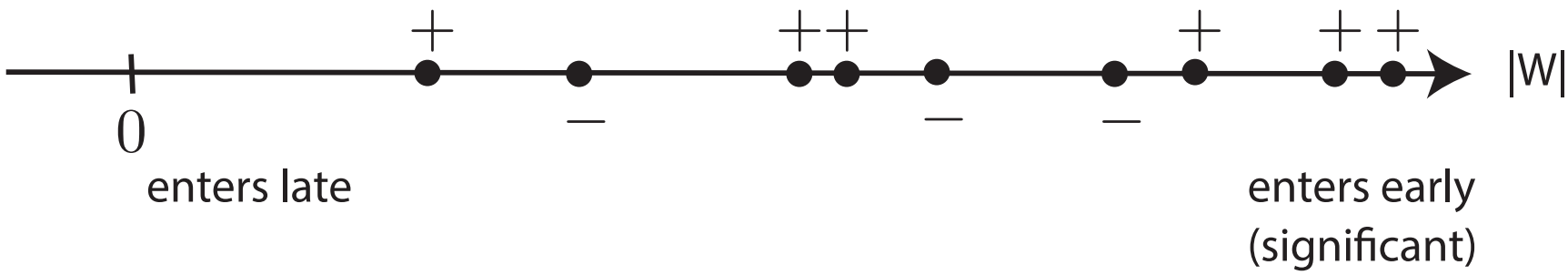
$$W_j = |\hat{\beta}_j^{\mathsf{LS}}|^2 - |\hat{\beta}_{j+p}^{\mathsf{LS}}|^2 \qquad |\hat{\beta}_j^{\mathsf{LS}}| - |\hat{\beta}_{j+p}^{\mathsf{LS}}|$$

- ... (endless possibilities)

Extensions

Other type-I errors

Can control familywise error rate (FWER), k-FWER, ...: Janson and Su ('15)



- T: time at which m knockoffs have entered before originals appear before
- Reject hypotheses with '+'

Expected number of false discoveries

$$\mathbb{E}\, V \leq m$$

Other models

Suppose we wish to test for groups

$$y = \sum_{g \in G} X_g \beta_g + z \qquad H_g : \beta_g = 0$$

- Group lasso

$$\text{min } \tfrac{1}{2} \|y - \textstyle\sum_g X_g \hat{\beta}_g\|_2^2 + \lambda \textstyle\sum_g \|\hat{\beta}_g\|_2$$

- Forward group selection
- ...

Other models

Suppose we wish to test for groups

$$y = \sum_{g \in G} X_g \beta_g + z \qquad H_g : \beta_g = 0$$

- Group lasso

$$\text{min } \tfrac{1}{2} \|y - \textstyle\sum_g X_g \hat{\beta}_g\|_2^2 + \lambda \textstyle\sum_g \|\hat{\beta}_g\|_2$$

- Forward group selection
- ...

- Construct group knockoffs for exchangeability

Other models

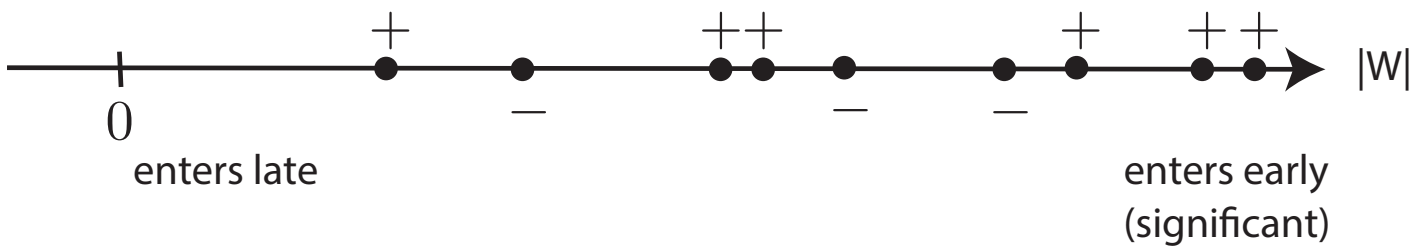
Suppose we wish to test for groups

$$y = \sum_{g \in G} X_g \beta_g + z \qquad H_g : \beta_g = 0$$

- Group lasso

$$\min \tfrac{1}{2} \|y - \textstyle\sum_g X_g \hat{\beta}_g\|_2^2 + \lambda \textstyle\sum_g \|\hat{\beta}_g\|_2$$

- Forward group selection
- ...



- Construct group knockoffs for exchangeability
- Calculate statistics; e.g. signed reversed ranks

Other models

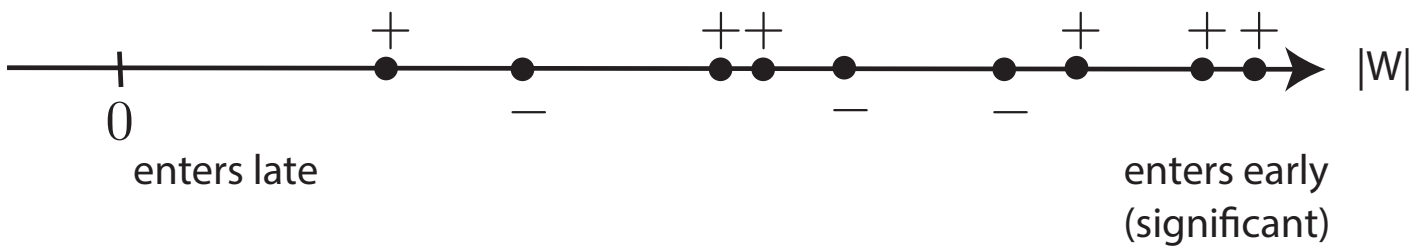
Suppose we wish to test for groups

$$y = \sum_{g \in G} X_g \beta_g + z \qquad H_g : \beta_g = 0$$

- Group lasso

$$\min \tfrac{1}{2} \|y - \textstyle\sum_g X_g \hat{\beta}_g\|_2^2 + \lambda \textstyle\sum_g \|\hat{\beta}_g\|_2$$

- Forward group selection
- ...



- Construct group knockoffs for exchangeability
- Calculate statistics; e.g. signed reversed ranks
- Compute threshold as before

Other models

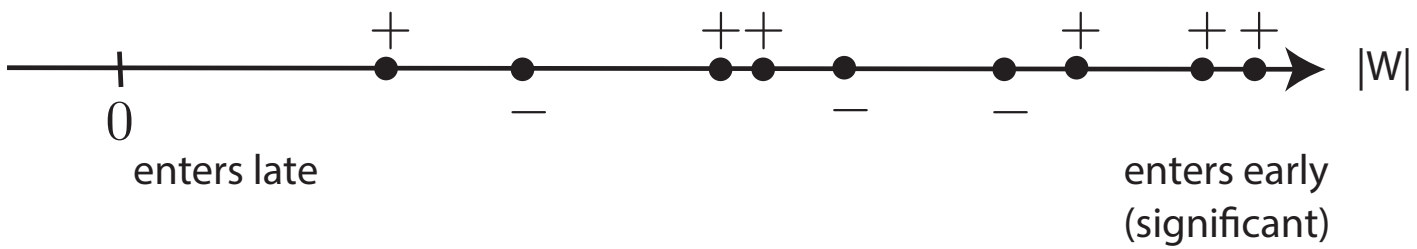
Suppose we wish to test for groups

$$y = \sum_{g \in G} X_g \beta_g + z \qquad H_g : \beta_g = 0$$

- Group lasso

$$\min \tfrac{1}{2} \|y - \textstyle\sum_g X_g \hat{\beta}_g\|_2^2 + \lambda \textstyle\sum_g \|\hat{\beta}_g\|_2$$

- Forward group selection
- ...



- Construct group knockoffs for exchangeability
- Calculate statistics; e.g. signed reversed ranks
- Compute threshold as before

Provides FDR control

Summary

Knockoff filter = inference machine

You design the statistics, knockoffs take care of inference

- Works under any design X (handles arb. correlations)
- Does not require any knowledge of noise level σ
- Very powerful when sparse effects

Open research: $n < p$...

FDR control is an extremely useful concept even away from counting errors and successes

BHq

$$y \sim \mathcal{N}(X\beta, \sigma^2 I) \quad \Longleftrightarrow \quad \hat{\beta}^{\mathsf{LS}} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1})$$

Statistics are independent iff $X'X$ is diagonal (orthogonal design)

BHq

Orthogonal model: $y \sim \mathcal{N}(\beta, I)$ ($\sigma = 1$ is known)

$$T_{\mathsf{BH}} = \min \left\{ t : \frac{p \cdot \mathbb{P}\{|\mathcal{N}(0,1)| \geq t\}}{\#\{j : |y_j| = |\beta_j + z_j| \geq t\}} \leq q \right\}$$

BHq

Orthogonal model: $y \sim \mathcal{N}(\beta, I)$ ($\sigma = 1$ is known)

$$T_{\mathsf{BH}} = \min \left\{ t : \frac{p \cdot \mathbb{P}\{|\mathcal{N}(0,1)| \geq t\}}{\#\{j : |y_j| = |\beta_j + z_j| \geq t\}} \leq q \right\}$$

Knockoff procedure (with lasso) is quite different:

- Make control group

$$[X \quad \tilde{X}] = \begin{bmatrix} I_p & 0 \\ 0 & I_p \end{bmatrix} \quad \Longrightarrow \quad \text{statistics} = \begin{bmatrix} y \\ z' \end{bmatrix}$$

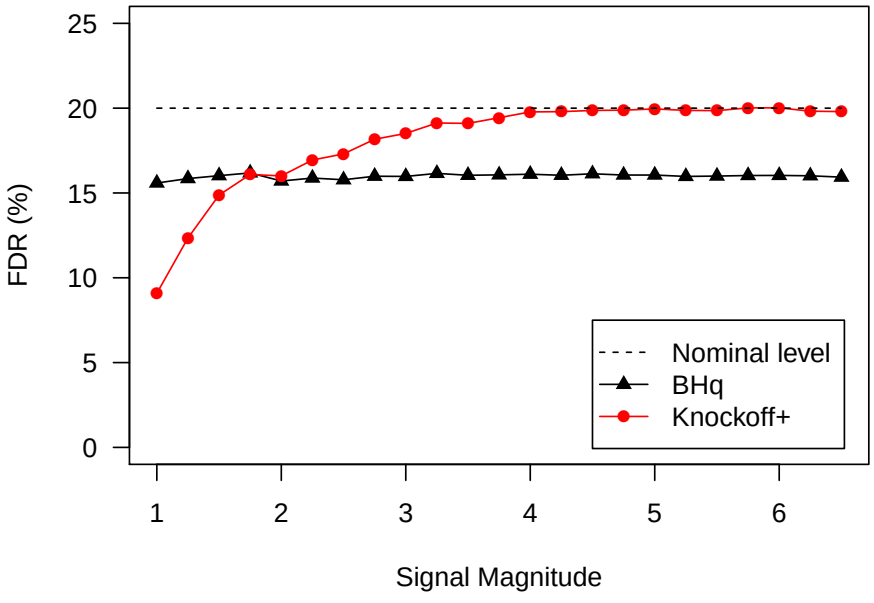
  $y \sim \mathcal{N}(\beta, I)$ indep. from $z' \sim \mathcal{N}(0, I)$

- Compute threshold (knockoff+) via

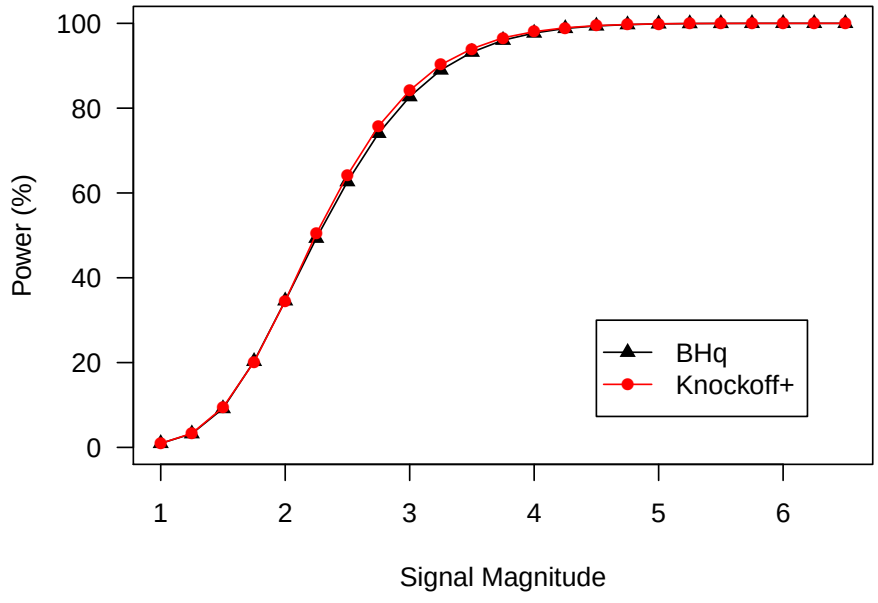
$$T = \min \left\{ t : \widehat{\mathsf{FDP}}(t) = \frac{\#\{j : |z'_j| \geq t \text{ and } |z'_j| > |y_j|\}}{\#\{j : |y_j| \geq t \text{ and } |y_j| > |z'_j|\}} \leq q \right\}$$

Empirical comparison

- $p = 1000$
- # true signals is $200 \rightarrow$ fraction of nulls is 0.8



FDR



Power

FDR and power of the BHq and knockoff+ methods vs. size A of the regression coefficients (signal magnitude)

Some Closing Remarks: How About Prediction/Estimation?

FDR for estimation: $y = X\beta + z$

- Goal: predict response from explanatory variables
- Sparse (modern) setup: p large and only few important variables

FDR for estimation: $y = X\beta + z$

- Goal: predict response from explanatory variables
- Sparse (modern) setup: p large and only few important variables

Model selection via C_p

$$C_p(S) = \underbrace{\|y - X\hat{\beta}[S]\|^2}_{\text{RSS}} + 2\sigma^2 \underbrace{|S|}_{\text{model dim.}} \qquad S \subset \{1, \ldots, p\}$$

- $\hat{\beta}[S]$: fitted LS coefficients from variables in S
- $C_p(S)$ unbiased for prediction error of model S

FDR for estimation: $y = X\beta + z$

- Goal: predict response from explanatory variables
- Sparse (modern) setup: p large and only few important variables

	Low bias	Low variance
C_p	✓	✗
FWER	✗	✓
FDR	✓	✓

FDR thresholding (Abramovich and Benjamini ('96))

- Orthogonal design: $X^T X = I_p \Longrightarrow X^T y \sim \mathcal{N}(\beta, \sigma^2 I_p)$
- Select nominal level q and perform BH(q) testing

$$\hat{\beta}_i = \begin{cases} X_i^T y & |X_i^T y| \geq t_{\text{FDR}} \\ 0 & \text{otherwise} \end{cases}$$

FDR thresholding (Abramovich and Benjamini ('96))

- Orthogonal design: $X^T X = I_p \Longrightarrow X^T y \sim \mathcal{N}(\beta, \sigma^2 I_p)$
- Select nominal level q and perform BH(q) testing

$$\hat{\beta}_i = \begin{cases} X_i^T y & |X_i^T y| \geq t_{\mathsf{FDR}} \\ 0 & \text{otherwise} \end{cases}$$

Theorem (Abramovich, Benjamini, Donoho, Johnstone ('05))

- *Sparsity class* $\ell_0(k) = \{\beta : \|\beta\|_0 \leq k\}$
- *Minimax risk*

$$R(k) = \inf_{\hat{\beta}} \sup_{\beta \in \ell_0(k)} \mathbb{E} \|\hat{\beta} - \beta\|_2^2$$

FDR thresholding (Abramovich and Benjamini ('96))

- Orthogonal design: $X^T X = I_p \Longrightarrow X^T y \sim \mathcal{N}(\beta, \sigma^2 I_p)$
- Select nominal level q and perform BH(q) testing

$$\hat{\beta}_i = \begin{cases} X_i^T y & |X_i^T y| \geq t_{\mathsf{FDR}} \\ 0 & \text{otherwise} \end{cases}$$

Theorem (Abramovich, Benjamini, Donoho, Johnstone ('05))

- *Sparsity class* $\ell_0(k) = \{\beta : \|\beta\|_0 \leq k\}$
- *Minimax risk*

$$R(k) = \inf_{\hat{\beta}} \sup_{\beta \in \ell_0(k)} \mathbb{E} \|\hat{\beta} - \beta\|_2^2$$

Set $q < 1/2$. *Then asymptotically, as* $p \to \infty$ *and* $k \in [(\log p)^5, p^{1-\delta}]$

$$\mathbb{E} \|\hat{\beta}_{\mathsf{FDR}} - \beta\|_2^2 = \mathbb{E} \|X\hat{\beta}_{\mathsf{FDR}} - X\beta\|_2^2 = R(k)(1 + o(1))$$

Other connections

SLOPE: Bogdan, van den Berg, Sabatti, Su and Candès ('13)

$$\text{min} \quad \frac{1}{2}\|y - X\hat{\beta}\|_2^2 + \lambda_1|\hat{\beta}|_{(1)} + \lambda_2|\hat{\beta}|_{(2)} + \ldots + \lambda_p|\hat{\beta}|_{(p)}$$

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$$

$$|\hat{\beta}|_{(1)} \geq |\hat{\beta}|_{(2)} \geq \cdots \geq |\hat{\beta}|_{(p)}$$

[order statistic]

Other connections

SLOPE: Bogdan, van den Berg, Sabatti, Su and Candès ('13)

$$\text{min} \quad \frac{1}{2}\|y - X\hat{\beta}\|_2^2 + \lambda_1|\hat{\beta}|_{(1)} + \lambda_2|\hat{\beta}|_{(2)} + \ldots + \lambda_p|\hat{\beta}|_{(p)}$$

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$$

$$|\hat{\beta}|_{(1)} \geq |\hat{\beta}|_{(2)} \geq \cdots \geq |\hat{\beta}|_{(p)}$$
[order statistic]

Adaptive minimaxity: C. and Su ('15)

- Sparse class $\ell_0(k) = \{\beta : \|\beta\|_0 \leq k\}$
- Fix $q \in (0, 1]$ and set $\lambda_i = \sigma \cdot \Phi^{-1}(1 - iq/2p)$ (BHq)

For some linear models, adaptive minimax estimation

$$\sup_{\beta \in \ell_0(k)} \mathbb{E}\,\|X\hat{\beta}_{\text{SLOPE}} - X\beta\|_2^2 = R(k)(1 + o(1))$$