High-Dimensional Variable Selection in Nonlinear Models that Controls the False Discovery Rate

Lucas Janson

Harvard University Department of Statistics

CMSA Big Data Conference, August 18, 2017

Collaborators: Emmanuel Candès (Stanford), Yingying Fan, Jinchi Lv (USC)

Problem Statement

Controlled Variable Selection

Given:

- Y an outcome of interest (AKA response or dependent variable),
- $X_1, \ldots, X_p$ a set of p potential explanatory variables (AKA covariates, features, or independent variables),

How can we select important explanatory variables with few mistakes?

Controlled Variable Selection

Given:

- Y an outcome of interest (AKA response or dependent variable),
- $X_1, \ldots, X_p$ a set of p potential explanatory variables (AKA covariates, features, or independent variables),

How can we select important explanatory variables with few mistakes?

Applications to:

- Medicine/genetics/health care

Controlled Variable Selection

Given:

- Y an outcome of interest (AKA response or dependent variable),
- $X_1, \ldots, X_p$ a set of p potential explanatory variables (AKA covariates, features, or independent variables),

How can we select important explanatory variables with few mistakes?

Applications to:

- Medicine/genetics/health care
- Economics/political science

Controlled Variable Selection

Given:

- Y an outcome of interest (AKA response or dependent variable),
- $X_1, \ldots, X_p$ a set of p potential explanatory variables (AKA covariates, features, or independent variables),

How can we select important explanatory variables with few mistakes?

Applications to:

- Medicine/genetics/health care
- Economics/political science
- Industry/technology

Controlled Variable Selection (cont'd)

What is an important variable?

Controlled Variable Selection (cont'd)

What is an important variable?

We consider X_j to be unimportant if the conditional distribution of Y given $X_1, \ldots, X_p$ does not depend on X_j. Formally, X_j is unimportant if it is conditionally independent of Y given $X_{\text{-}j}$:

$$Y \perp\!\!\!\perp X_j \,|\, X_{\text{-}j}$$

Controlled Variable Selection (cont'd)

What is an important variable?

We consider X_j to be unimportant if the conditional distribution of Y given $X_1, \ldots, X_p$ does not depend on X_j. Formally, X_j is unimportant if it is conditionally independent of Y given $X_{\text{-}j}$:

$$Y \perp\!\!\!\perp X_j \,|\, X_{\text{-}j}$$

Markov Blanket of Y: smallest set S such that $Y \perp\!\!\!\perp X_{\text{-}S} \,|\, X_S$

Controlled Variable Selection (cont'd)

What is an important variable?

We consider X_j to be unimportant if the conditional distribution of Y given $X_1, \ldots, X_p$ does not depend on X_j. Formally, X_j is unimportant if it is conditionally independent of Y given $X_{\text{-}j}$:

$$Y \perp\!\!\!\perp X_j \,|\, X_{\text{-}j}$$

Markov Blanket of Y: smallest set S such that $Y \perp\!\!\!\perp X_{\text{-}S} \,|\, X_S$

For GLMs with no stochastically redundant covariates, equivalent to $\{j : \beta_j = 0\}$

Controlled Variable Selection (cont'd)

What is an important variable?

We consider X_j to be unimportant if the conditional distribution of Y given $X_1, \ldots, X_p$ does not depend on X_j. Formally, X_j is unimportant if it is conditionally independent of Y given $X_{\text{-}j}$:

$$Y \perp\!\!\!\perp X_j \,|\, X_{\text{-}j}$$

Markov Blanket of Y: smallest set S such that $Y \perp\!\!\!\perp X_{\text{-}S} \,|\, X_S$

For GLMs with no stochastically redundant covariates, equivalent to $\{j \,:\, \beta_j = 0\}$

To make sure we do not make too many mistakes, we seek to select a set $\hat{S}$ to control the **false discovery rate (FDR)**:

$$\text{FDR}(\hat{S}) = \mathbb{E}\left(\frac{\#\{j \text{ in } \hat{S} \,:\, X_j \text{ unimportant}\}}{\#\{j \text{ in } \hat{S}\}}\right) \leq q \;\; \text{(e.g. 10\%)}$$

"Here is a set of variables $\hat{S}$, 90% of which I expect to be important"

Sneak Peak

New interpretation of knockoffs solves the controlled variable selection problem

- Allows any model for Y and $X_1, \ldots, X_p$
- Allows any dimension (including $p > n$)
- Finite-sample control (non-asymptotic) of FDR
- Practical performance on real problems

Sneak Peak

New interpretation of knockoffs solves the controlled variable selection problem

- Allows any model for Y and $X_1, \ldots, X_p$
- Allows any dimension (including $p > n$)
- Finite-sample control (non-asymptotic) of FDR
- Practical performance on real problems

Analysis of the genetic basis of Crohn's Disease (WTCCC, 2007)

- $\approx 5,000$ subjects ($\approx 40\%$ with Crohn's Disease)
- $\approx 375,000$ single nucleotide polymorphisms (SNPs) for each subject

Sneak Peak

New interpretation of knockoffs solves the controlled variable selection problem

- Allows any model for Y and $X_1, \ldots, X_p$
- Allows any dimension (including $p > n$)
- Finite-sample control (non-asymptotic) of FDR
- Practical performance on real problems

Analysis of the genetic basis of Crohn's Disease (WTCCC, 2007)

- $\approx 5,000$ subjects ($\approx 40\%$ with Crohn's Disease)
- $\approx 375,000$ single nucleotide polymorphisms (SNPs) for each subject

Original analysis of the data made **9** discoveries by running marginal tests and selecting p-values to target a FDR of 10%

Sneak Peak

New interpretation of knockoffs solves the controlled variable selection problem

- Allows any model for Y and $X_1, \ldots, X_p$
- Allows any dimension (including $p > n$)
- Finite-sample control (non-asymptotic) of FDR
- Practical performance on real problems

Analysis of the genetic basis of Crohn's Disease (WTCCC, 2007)

- $\approx 5,000$ subjects ($\approx 40\%$ with Crohn's Disease)
- $\approx 375,000$ single nucleotide polymorphisms (SNPs) for each subject

Original analysis of the data made **9** discoveries by running marginal tests and selecting p-values to target a FDR of 10%

Model-free knockoffs used the same FDR of 10% and made **18** discoveries, with many of the new discoveries confirmed by a larger meta-analysis

Review of Methods for Controlled Variable Selection

What is required for valid inference?

	Low dimensions	Model for Y	Asymptopic regime	Sparsity	Random design
OLSp+BHq	Yes	Yes	No	No	No

Review of Methods for Controlled Variable Selection

What is required for valid inference?

	Low dimensions	Model for Y	Asymptopic regime	Sparsity	Random design
OLSp+BHq	Yes	Yes	No	No	No
MLp+BHq	Yes	Yes	Yes	No	No

Review of Methods for Controlled Variable Selection

What is required for valid inference?

	Low dimensions	Model for Y	Asymptopic regime	Sparsity	Random design
OLSp+BHq	Yes	Yes	No	No	No
MLp+BHq	Yes	Yes	Yes	No	No
HDp+BHq	No	Yes	Yes	Yes	Yes

Review of Methods for Controlled Variable Selection

What is required for valid inference?

	Low dimensions	Model for Y	Asymptopic regime	Sparsity	Random design
OLSp+BHq	Yes	Yes	No	No	No
MLp+BHq	Yes	Yes	Yes	No	No
HDp+BHq	No	Yes	Yes	Yes	Yes
Orig KnO	Yes	Yes	No	No	No

Review of Methods for Controlled Variable Selection

What is required for valid inference?

	Low dimensions	Model for Y	Asymptopic regime	Sparsity	Random design
OLSp+BHq	Yes	Yes	No	No	No
MLp+BHq	Yes	Yes	Yes	No	No
HDp+BHq	No	Yes	Yes	Yes	Yes
Orig KnO	Yes	Yes	No	No	No
New KnO	No	No	No	No	Yes*

The Knockoffs Idea

Knockoffs (Barber and Candès, 2015)

$\boldsymbol{y}$ and $\boldsymbol{X}_j$ are $n \times 1$ column vectors of data: n draws from the random variables Y and X_j, respectively; design matrix $\boldsymbol{X} := [\boldsymbol{X}_1 \cdots \boldsymbol{X}_p]$

Knockoffs (Barber and Candès, 2015)

$\boldsymbol{y}$ and $\boldsymbol{X}_j$ are $n \times 1$ column vectors of data: n draws from the random variables Y and X_j, respectively; design matrix $\boldsymbol{X} := [\boldsymbol{X}_1 \cdots \boldsymbol{X}_p]$

(1) **Construct knockoffs**: Knockoffs $\tilde{\boldsymbol{X}}_j$ must satisfy, ($\tilde{\boldsymbol{X}} := [\tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_p]$)

$$[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]^\top [\boldsymbol{X}\ \tilde{\boldsymbol{X}}] = \begin{bmatrix} \boldsymbol{X}^\top \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{X} - \mathrm{diag}\{\boldsymbol{s}\} \\ \boldsymbol{X}^\top \boldsymbol{X} - \mathrm{diag}\{\boldsymbol{s}\} & \boldsymbol{X}^\top \boldsymbol{X} \end{bmatrix}$$

Knockoffs (Barber and Candès, 2015)

$\boldsymbol{y}$ and $\boldsymbol{X}_j$ are $n \times 1$ column vectors of data: n draws from the random variables Y and X_j, respectively; design matrix $\boldsymbol{X} := [\boldsymbol{X}_1 \cdots \boldsymbol{X}_p]$

(1) **Construct knockoffs**: Knockoffs $\tilde{\boldsymbol{X}}_j$ must satisfy, ($\tilde{\boldsymbol{X}} := [\tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_p]$)

$$[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]^\top [\boldsymbol{X}\ \tilde{\boldsymbol{X}}] = \begin{bmatrix} \boldsymbol{X}^\top \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{X} - \operatorname{diag}\{\boldsymbol{s}\} \\ \boldsymbol{X}^\top \boldsymbol{X} - \operatorname{diag}\{\boldsymbol{s}\} & \boldsymbol{X}^\top \boldsymbol{X} \end{bmatrix}$$

(2) **Compute knockoff statistics**:

- Sufficiency: W_j only a function of $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]^\top [\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$ and $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]^\top \boldsymbol{y}$
- Antisymmetry: swapping values of $\boldsymbol{X}_j$ and $\tilde{\boldsymbol{X}}_j$ flips sign of W_j

Knockoffs (Barber and Candès, 2015)

$\boldsymbol{y}$ and $\boldsymbol{X}_j$ are $n \times 1$ column vectors of data: n draws from the random variables Y and X_j, respectively; design matrix $\boldsymbol{X} := [\boldsymbol{X}_1 \cdots \boldsymbol{X}_p]$

(1) **Construct knockoffs**: Knockoffs $\tilde{\boldsymbol{X}}_j$ must satisfy, ($\tilde{\boldsymbol{X}} := [\tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_p]$)

$$[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]^\top [\boldsymbol{X}\ \tilde{\boldsymbol{X}}] = \begin{bmatrix} \boldsymbol{X}^\top \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{X} - \mathrm{diag}\{\boldsymbol{s}\} \\ \boldsymbol{X}^\top \boldsymbol{X} - \mathrm{diag}\{\boldsymbol{s}\} & \boldsymbol{X}^\top \boldsymbol{X} \end{bmatrix}$$

(2) **Compute knockoff statistics**:

- Sufficiency: W_j only a function of $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]^\top [\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$ and $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]^\top \boldsymbol{y}$
- Antisymmetry: swapping values of $\boldsymbol{X}_j$ and $\tilde{\boldsymbol{X}}_j$ flips sign of W_j

(3) **Find the knockoff threshold**:

- Order the variables by decreasing $|W_j|$ and proceed down list
- Select only variables with positive W_j until last time $\frac{\text{negatives}}{\text{positives}} \leq q$

Knockoffs (Barber and Candès, 2015)

$\boldsymbol{y}$ and $\boldsymbol{X}_j$ are $n \times 1$ column vectors of data: n draws from the random variables Y and X_j, respectively; design matrix $\boldsymbol{X} := [\boldsymbol{X}_1 \cdots \boldsymbol{X}_p]$

(1) **Construct knockoffs**: Knockoffs $\tilde{\boldsymbol{X}}_j$ must satisfy, ($\tilde{\boldsymbol{X}} := [\tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_p]$)

$$[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]^\top [\boldsymbol{X}\ \tilde{\boldsymbol{X}}] = \begin{bmatrix} \boldsymbol{X}^\top \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{X} - \mathrm{diag}\{\boldsymbol{s}\} \\ \boldsymbol{X}^\top \boldsymbol{X} - \mathrm{diag}\{\boldsymbol{s}\} & \boldsymbol{X}^\top \boldsymbol{X} \end{bmatrix}$$

(2) **Compute knockoff statistics**:
- Sufficiency: W_j only a function of $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]^\top [\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$ and $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]^\top \boldsymbol{y}$
- Antisymmetry: swapping values of $\boldsymbol{X}_j$ and $\tilde{\boldsymbol{X}}_j$ flips sign of W_j

(3) **Find the knockoff threshold**:
- Order the variables by decreasing $|W_j|$ and proceed down list
- Select only variables with positive W_j until last time $\frac{\text{negatives}}{\text{positives}} \leq q$

Comments:
- Finite-sample FDR control and leverages sparsity for power
- Requires data follow low-dimensional $(n \geq p)$ Gaussian linear model
- Canonical approach: condition on $\boldsymbol{X}$, rely heavily on model for $\boldsymbol{y}$

Generalizing the Knockoffs Procedure

(1) **Construct knockoffs**:

- Artificial versions ("knockoffs") of each variable
- Act as controls for assessing importance of original variables

Generalizing the Knockoffs Procedure

(1) **Construct knockoffs**:
- Artificial versions ("knockoffs") of each variable
- Act as controls for assessing importance of original variables

(2) **Compute knockoff statistics**:
- Scalar statistic W_j for each variable
- Measures how much more important a variable appears than its knockoff
- Positive W_j denotes original more important, strength measured by magnitude

Generalizing the Knockoffs Procedure

(1) **Construct knockoffs**:
- Artificial versions ("knockoffs") of each variable
- Act as controls for assessing importance of original variables

(2) **Compute knockoff statistics**:
- Scalar statistic W_j for each variable
- Measures how much more important a variable appears than its knockoff
- Positive W_j denotes original more important, strength measured by magnitude

(3) **Find the knockoff threshold**: (same as before)
- Order the variables by decreasing $|W_j|$ and proceed down list
- Select only variables with positive W_j until last time $\frac{\text{negatives}}{\text{positives}} \leq q$

Generalizing the Knockoffs Procedure

(1) **Construct knockoffs**:
- Artificial versions ("knockoffs") of each variable
- Act as controls for assessing importance of original variables

(2) **Compute knockoff statistics**:
- Scalar statistic W_j for each variable
- Measures how much more important a variable appears than its knockoff
- Positive W_j denotes original more important, strength measured by magnitude

(3) **Find the knockoff threshold**: (same as before)
- Order the variables by decreasing $|W_j|$ and proceed down list
- Select only variables with positive W_j until last time $\frac{\text{negatives}}{\text{positives}} \leq q$

Coin-flipping property: The key to knockoffs is that steps (1) and (2) are done specifically to ensure that, conditional on $|W_1|, \ldots, |W_p|$, the signs of the *unimportant/null* W_j are independently ± 1 with probability $1/2$

New Interpretation of Knockoffs

Knockoffs Without a Model for Y (Candès et al., 2016)

Instead of modeling $\boldsymbol{y}$ and conditioning on $\boldsymbol{X}$, condition on $\boldsymbol{y}$ and model $\boldsymbol{X}$
(shifts the *burden of knowledge* from $\boldsymbol{y}$ onto $\boldsymbol{X}$)

Knockoffs Without a Model for Y (Candès et al., 2016)

Instead of modeling $\boldsymbol{y}$ and conditioning on $\boldsymbol{X}$, condition on $\boldsymbol{y}$ and model $\boldsymbol{X}$
(shifts the *burden of knowledge* from $\boldsymbol{y}$ onto $\boldsymbol{X}$)

Explicitly,

$$\text{rows of } \boldsymbol{X} = (X_{i,1}, \ldots, X_{i,p}) \overset{\text{iid}}{\sim} G$$

where G can be arbitrary but is assumed known

Knockoffs Without a Model for Y (Candès et al., 2016)

Instead of modeling $\boldsymbol{y}$ and conditioning on $\boldsymbol{X}$, condition on $\boldsymbol{y}$ and model $\boldsymbol{X}$ (shifts the *burden of knowledge* from $\boldsymbol{y}$ onto $\boldsymbol{X}$)

Explicitly,

$$\text{rows of } \boldsymbol{X} = (X_{i,1}, \ldots, X_{i,p}) \overset{\text{iid}}{\sim} G$$

where G can be arbitrary but is assumed known

- As compared to original knockoffs, removes
  - Restriction on dimension
  - Linear model requirement for $Y \mid X_1, \ldots, X_p$
  - "Sufficiency" constraint for W_j

Knockoffs Without a Model for Y (Candès et al., 2016)

Instead of modeling $\boldsymbol{y}$ and conditioning on $\boldsymbol{X}$, condition on $\boldsymbol{y}$ and model $\boldsymbol{X}$ (shifts the *burden of knowledge* from $\boldsymbol{y}$ onto $\boldsymbol{X}$)

Explicitly,

$$\text{rows of } \boldsymbol{X} = (X_{i,1}, \dots, X_{i,p}) \overset{\text{iid}}{\sim} G$$

where G can be arbitrary but is assumed known

- As compared to original knockoffs, removes
  - Restriction on dimension
  - Linear model requirement for $Y \mid X_1, \dots, X_p$
  - "Sufficiency" constraint for W_j
- The rows of $\boldsymbol{X}$ must be i.i.d., *not* the columns (covariates)

Knockoffs Without a Model for Y (Candès et al., 2016)

Instead of modeling $\boldsymbol{y}$ and conditioning on $\boldsymbol{X}$, condition on $\boldsymbol{y}$ and model $\boldsymbol{X}$ (shifts the *burden of knowledge* from $\boldsymbol{y}$ onto $\boldsymbol{X}$)

Explicitly,

$$\text{rows of } \boldsymbol{X} = (X_{i,1}, \ldots, X_{i,p}) \overset{\text{iid}}{\sim} G$$

where G can be arbitrary but is assumed known

- As compared to original knockoffs, removes
  - Restriction on dimension
  - Linear model requirement for $Y \mid X_1, \ldots, X_p$
  - "Sufficiency" constraint for W_j
- The rows of $\boldsymbol{X}$ must be i.i.d., *not* the columns (covariates)
- Nothing about $\boldsymbol{y}$'s distribution is assumed or need be known

Knockoffs Without a Model for Y (Candès et al., 2016)

Instead of modeling $\boldsymbol{y}$ and conditioning on $\boldsymbol{X}$, condition on $\boldsymbol{y}$ and model $\boldsymbol{X}$ (shifts the *burden of knowledge* from $\boldsymbol{y}$ onto $\boldsymbol{X}$)

Explicitly,

$$\text{rows of } \boldsymbol{X} = (X_{i,1}, \ldots, X_{i,p}) \overset{\text{iid}}{\sim} G$$

where G can be arbitrary but is assumed known

- As compared to original knockoffs, removes
  - Restriction on dimension
  - Linear model requirement for $Y \mid X_1, \ldots, X_p$
  - "Sufficiency" constraint for W_j
- The rows of $\boldsymbol{X}$ must be i.i.d., *not* the columns (covariates)
- Nothing about $\boldsymbol{y}$'s distribution is assumed or need be known
- Robust to overfitting $\boldsymbol{X}$'s distribution in preliminary experiments

Robustness

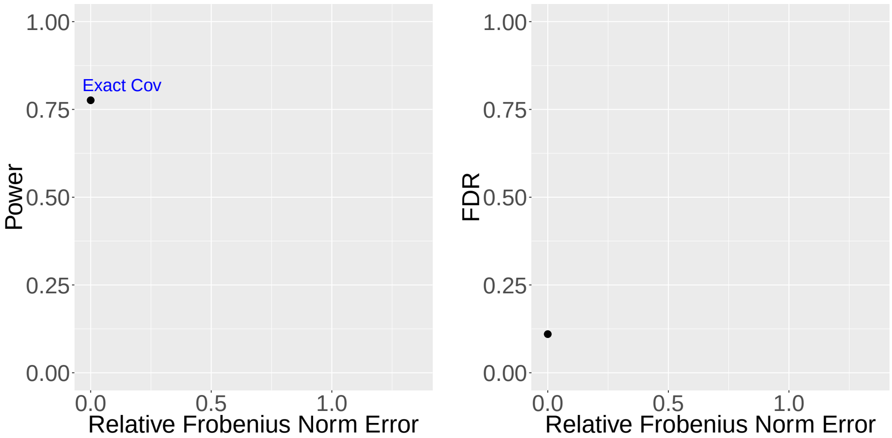


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

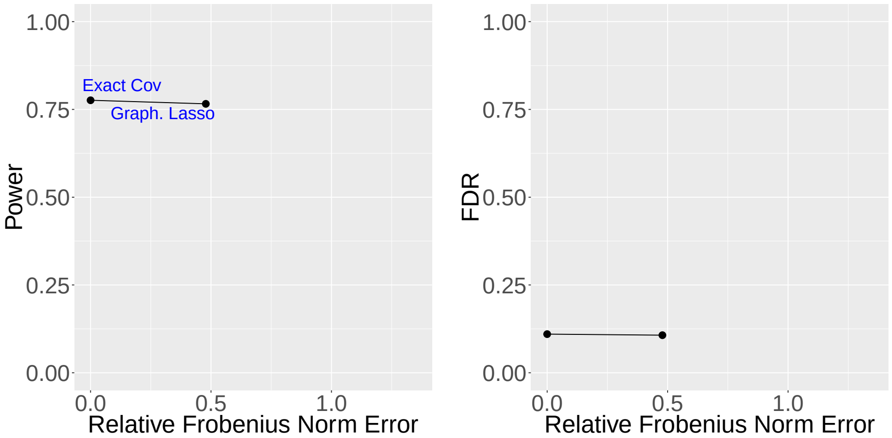


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

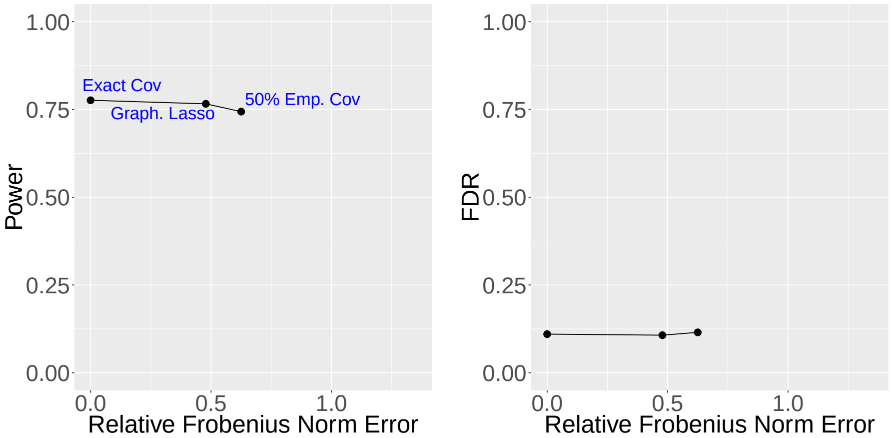


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

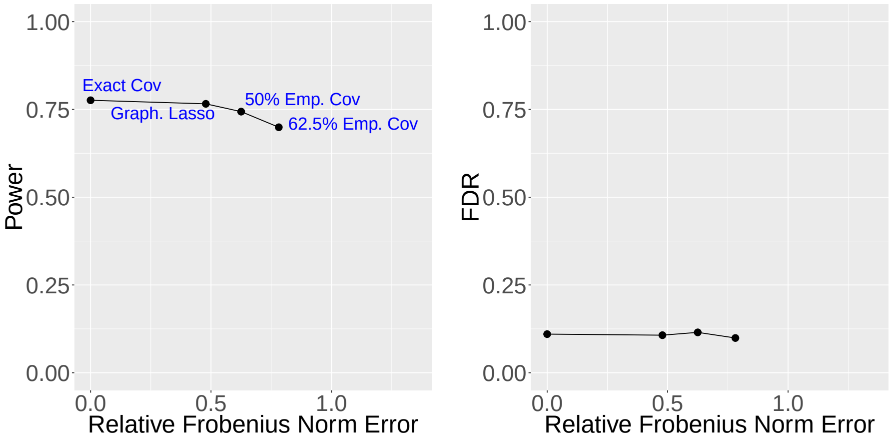


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

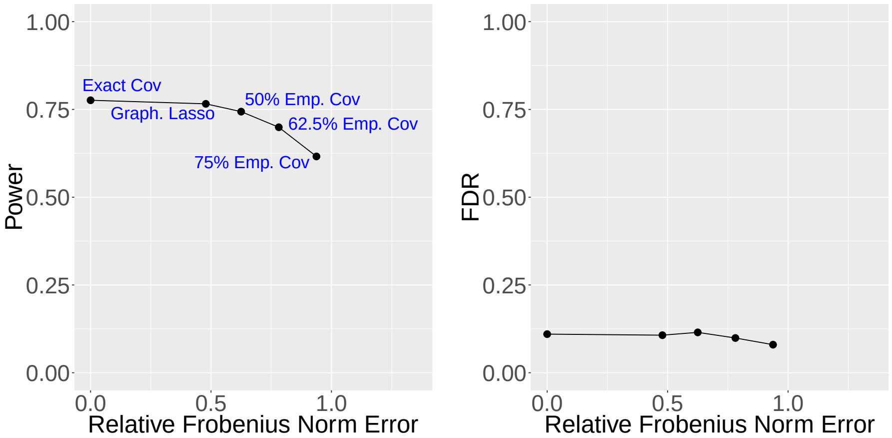


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

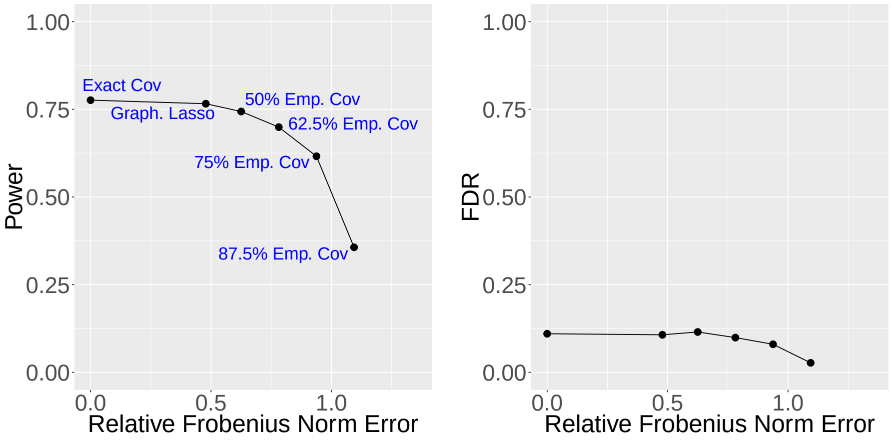


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

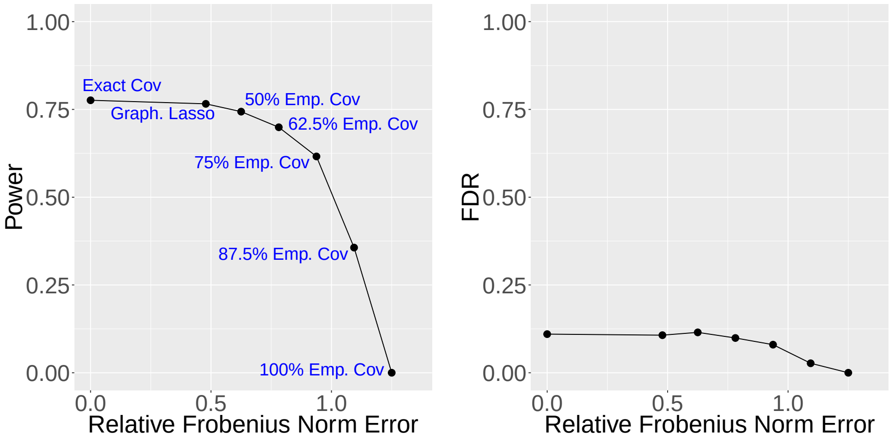


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Shifting the Burden of Knowledge

When is it appropriate?

1. Subjects sampled from a population, and

2a. X_j highly structured, well-studied, or well-understood, OR

Shifting the Burden of Knowledge

When is it appropriate?

1. Subjects sampled from a population, and

2a. X_j highly structured, well-studied, or well-understood, OR

2b. Large set of unsupervised X data (without Y's)

Shifting the Burden of Knowledge

When is it appropriate?

1. Subjects sampled from a population, and

2a. X_j highly structured, well-studied, or well-understood, OR

2b. Large set of unsupervised X data (without Y's)

For instance, many **genome-wide association studies** satisfy all conditions:

1. Subjects sampled from a population (oversampling cases still valid)

Shifting the Burden of Knowledge

When is it appropriate?

1. Subjects sampled from a population, and

2a. X_j highly structured, well-studied, or well-understood, OR

2b. Large set of unsupervised X data (without Y's)

For instance, many **genome-wide association studies** satisfy all conditions:

1. Subjects sampled from a population (oversampling cases still valid)

2a. Strong spatial structure: linkage disequilibrium models, e.g., Markov chains, are well-studied and work well

Shifting the Burden of Knowledge

When is it appropriate?

1. Subjects sampled from a population, and

2a. X_j highly structured, well-studied, or well-understood, OR

2b. Large set of unsupervised X data (without Y's)

For instance, many **genome-wide association studies** satisfy all conditions:

1. Subjects sampled from a population (oversampling cases still valid)

2a. Strong spatial structure: linkage disequilibrium models, e.g., Markov chains, are well-studied and work well

2b. Other studies have collected same or similar SNP arrays on different subjects

The New Knockoffs Procedure

(1) **Construct knockoffs**: Exchangeability

$$[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \overset{\mathcal{D}}{=} [\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]$$

The New Knockoffs Procedure

(1) **Construct knockoffs**: Exchangeability

$$[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \overset{\mathcal{D}}{=} [\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]$$

(2) **Compute knockoff statistics**:

- Variable importance measure Z
- Antisymmetric function $f_j : \mathbb{R}^2 \to \mathbb{R}$, i.e.,

$$f_j(z_1, z_2) = -f_j(z_2, z_1)$$

- $W_j = f_j(Z_j, \widetilde{Z}_j)$, where Z_j and $\widetilde{Z}_j$ are the variable importances of $\boldsymbol{X}_j$ and $\tilde{\boldsymbol{X}}_j$, respectively

The New Knockoffs Procedure

(1) **Construct knockoffs**: Exchangeability

$$[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \overset{\mathcal{D}}{=} [\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]$$

(2) **Compute knockoff statistics**:

- Variable importance measure Z
- Antisymmetric function $f_j : \mathbb{R}^2 \to \mathbb{R}$, i.e.,

$$f_j(z_1, z_2) = -f_j(z_2, z_1)$$

- $W_j = f_j(Z_j, \widetilde{Z}_j)$, where Z_j and $\widetilde{Z}_j$ are the variable importances of $\boldsymbol{X}_j$ and $\tilde{\boldsymbol{X}}_j$, respectively

(3) **Find the knockoff threshold**: (same as before)

- Order the variables by decreasing $|W_j|$ and proceed down list
- Select only variables with positive W_j until last time $\frac{\text{negatives}}{\text{positives}} \leq q$

Step (1): Construct Knockoffs

Knockoff Construction

Proof that valid knockoff variables can be generated for any X distribution

Knockoff Construction

Proof that valid knockoff variables can be generated for any X distribution

If $(X_1, \ldots, X_p)$ multivariate Gaussian, exchangeability reduces to matching first and second moments when X_j, $\tilde{X}_j$ swapped

For $\mathrm{Cov}(X_1, \ldots, X_p) = \boldsymbol{\Sigma}$:

$$\mathrm{Cov}(X_1, \ldots, X_p, \tilde{X}_1, \ldots, \tilde{X}_p) = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} \\ \boldsymbol{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} & \boldsymbol{\Sigma} \end{bmatrix}$$

For non-Gaussian X, still second-order-correct approximate knockoffs

Knockoff Construction

Proof that valid knockoff variables can be generated for any X distribution

If $(X_1, \ldots, X_p)$ multivariate Gaussian, exchangeability reduces to matching first and second moments when X_j, $\tilde{X}_j$ swapped

For $\mathrm{Cov}(X_1, \ldots, X_p) = \boldsymbol{\Sigma}$:

$$\mathrm{Cov}(X_1, \ldots, X_p, \tilde{X}_1, \ldots, \tilde{X}_p) = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} \\ \boldsymbol{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} & \boldsymbol{\Sigma} \end{bmatrix}$$

For non-Gaussian X, still second-order-correct approximate knockoffs

- Linear algebra and semidefinite programming to find good $\boldsymbol{s}$
- Recently: construction for Markov chains and HMMs (Sesia et al., 2017)
- Constructions also possible for grouped variables (Dai and Barber, 2016)

Step (2): Compute Knockoff Statistics

Strategy for Choosing Knockoff Statistics

Recall W_j an antisymmetric function f_j of Z_j and $\widetilde{Z}_j$ (the variable importances of $\boldsymbol{X}_j$ and $\tilde{\boldsymbol{X}}_j$, respectively):

$$W_j = f_j(Z_j, \widetilde{Z}_j) = -f_j(\widetilde{Z}_j, Z_j)$$

Strategy for Choosing Knockoff Statistics

Recall W_j an antisymmetric function f_j of Z_j and $\widetilde{Z}_j$ (the variable importances of $\boldsymbol{X}_j$ and $\tilde{\boldsymbol{X}}_j$, respectively):

$$W_j = f_j(Z_j, \widetilde{Z}_j) = -f_j(\widetilde{Z}_j, Z_j)$$

For example,

- Z is magnitude of fitted coefficient β from a lasso regression of $\boldsymbol{y}$ on $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$
- $f_j(z_1, z_2) = z_1 - z_2$

Strategy for Choosing Knockoff Statistics

Recall W_j an antisymmetric function f_j of Z_j and $\widetilde{Z}_j$ (the variable importances of $\boldsymbol{X}_j$ and $\tilde{\boldsymbol{X}}_j$, respectively):

$$W_j = f_j(Z_j, \widetilde{Z}_j) = -f_j(\widetilde{Z}_j, Z_j)$$

For example,

- Z is magnitude of fitted coefficient β from a lasso regression of $\boldsymbol{y}$ on $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$
- $f_j(z_1, z_2) = z_1 - z_2$

Lasso Coefficient Difference (LCD) statistic:

$$W_j = |\beta_j| - |\tilde{\beta}_j|$$

Exchangeability Endows Coin-Flipping

Recall exchangeability property: for any j,

$$[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \overset{\mathcal{D}}{=} [\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]$$

Exchangeability Endows Coin-Flipping

Recall exchangeability property: for any j,

$$[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \stackrel{\mathcal{D}}{=} [\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]$$

Coin-flipping property for W_j:

Exchangeability Endows Coin-Flipping

Recall exchangeability property: for any j,

$$[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \stackrel{\mathcal{D}}{=} [\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]$$

Coin-flipping property for W_j: for any *unimportant* variable j,

$$\left(Z_j, \widetilde{Z}_j\right) := \left(Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right), \ \ \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right)$$

Exchangeability Endows Coin-Flipping

Recall exchangeability property: for any j,

$$[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \overset{\mathcal{D}}{=} [\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]$$

Coin-flipping property for W_j: for any *unimportant* variable j,

$$\left(Z_j, \widetilde{Z}_j\right) := \left(Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right), \; \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right) \overset{\mathcal{D}}{=} \left(Z_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right), \; \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right)\right)$$

Exchangeability Endows Coin-Flipping

Recall exchangeability property: for any j,

$$[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \overset{\mathcal{D}}{=} [\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]$$

Coin-flipping property for W_j: for any *unimportant* variable j,

$$\begin{aligned}
\left(Z_j, \widetilde{Z}_j\right) &:= \left(Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right),\ \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right) \\
&\overset{\mathcal{D}}{=} \left(Z_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right),\ \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right)\right) \\
&= \left(\widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right),\ Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right)
\end{aligned}$$

Exchangeability Endows Coin-Flipping

Recall exchangeability property: for any j,

$$\begin{aligned}&[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \\ \overset{\mathcal{D}}{=} \, &[\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]\end{aligned}$$

Coin-flipping property for W_j: for any *unimportant* variable j,

$$\begin{aligned}\left(Z_j, \widetilde{Z}_j\right) &:= \left(Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right), \;\; \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right) \\ &\overset{\mathcal{D}}{=} \left(Z_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right), \;\; \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right)\right) \\ &= \left(\widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right), \;\; Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right) \\ &= \left(\widetilde{Z}_j, Z_j\right)\end{aligned}$$

Exchangeability Endows Coin-Flipping

Recall exchangeability property: for any j,

$$[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p\, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \stackrel{\mathcal{D}}{=} [\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p\, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]$$

Coin-flipping property for W_j: for any *unimportant* variable j,

$$\begin{aligned}
\left(Z_j, \widetilde{Z}_j\right) &:= \left(Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right),\ \ \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right) \\
&\stackrel{\mathcal{D}}{=} \left(Z_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right),\ \ \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right)\right) \\
&= \left(\widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right),\ \ Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right) \\
&= \left(\widetilde{Z}_j, Z_j\right)
\end{aligned}$$

$$W_j = f_j(Z_j, \widetilde{Z}_j) \stackrel{\mathcal{D}}{=} f_j(\widetilde{Z}_j, Z_j)$$

Exchangeability Endows Coin-Flipping

Recall exchangeability property: for any j,

$$[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p]$$

$$\stackrel{\mathcal{D}}{=} [\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]$$

Coin-flipping property for W_j: for any *unimportant* variable j,

$$\begin{aligned}
\left(Z_j, \widetilde{Z}_j\right) &:= \left(Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right), \;\; \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right) \\
&\stackrel{\mathcal{D}}{=} \left(Z_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right), \;\; \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right)\right) \\
&= \left(\widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right), \;\; Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right) \\
&= \left(\widetilde{Z}_j, Z_j\right)
\end{aligned}$$

$$W_j = f_j(Z_j, \widetilde{Z}_j) \stackrel{\mathcal{D}}{=} f_j(\widetilde{Z}_j, Z_j) = -f_j(Z_j, \widetilde{Z}_j) = -W_j$$

Exchangeability Endows Coin-Flipping

Recall exchangeability property: for any j,

$$\begin{aligned}&[\boldsymbol{X}_1 \cdots \boldsymbol{X}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \tilde{\boldsymbol{X}}_p] \\ \overset{\mathcal{D}}{=} &[\boldsymbol{X}_1 \cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_p \, \tilde{\boldsymbol{X}}_1 \cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_p]\end{aligned}$$

Coin-flipping property for W_j: for any *unimportant* variable j,

$$\begin{aligned}\left(Z_j, \widetilde{Z}_j\right) &:= \left(Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right), \; \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right) \\ &\overset{\mathcal{D}}{=} \left(Z_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right), \; \widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \tilde{\boldsymbol{X}}_j \cdots \boldsymbol{X}_j \cdots\right]\right)\right) \\ &= \left(\widetilde{Z}_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right), \; Z_j\left(\boldsymbol{y}, \left[\cdots \boldsymbol{X}_j \cdots \tilde{\boldsymbol{X}}_j \cdots\right]\right)\right) \\ &= \left(\widetilde{Z}_j, Z_j\right)\end{aligned}$$

$$W_j \overset{\mathcal{D}}{=} -W_j$$

Adaptivity and Prior Information in W_j

Recall LCD: $W_j = |\beta_j| - |\tilde{\beta}_j|$, where β_j, $\tilde{\beta}_j$ come from ℓ_1-penalized regression

Adaptivity

- Cross-validation (on $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$) to choose the penalty parameter in LCD

Adaptivity and Prior Information in W_j

Recall LCD: $W_j = |\beta_j| - |\tilde{\beta}_j|$, where β_j, $\tilde{\beta}_j$ come from ℓ_1-penalized regression

Adaptivity

- Cross-validation (on $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$) to choose the penalty parameter in LCD
- Higher-level adaptivity: CV to choose best-fitting model for inference

Adaptivity and Prior Information in W_j

Recall LCD: $W_j = |\beta_j| - |\tilde{\beta}_j|$, where β_j, $\tilde{\beta}_j$ come from ℓ_1-penalized regression

Adaptivity

- Cross-validation (on $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$) to choose the penalty parameter in LCD
- Higher-level adaptivity: CV to choose best-fitting model for inference
  - E.g., fit random forest and ℓ_1-penalized regression; derive feature importance from whichever has lower CV error—**still strict FDR control**

Adaptivity and Prior Information in W_j

Recall LCD: $W_j = |\beta_j| - |\tilde{\beta}_j|$, where β_j, $\tilde{\beta}_j$ come from ℓ_1-penalized regression

Adaptivity

- Cross-validation (on $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$) to choose the penalty parameter in LCD
- Higher-level adaptivity: CV to choose best-fitting model for inference
  - E.g., fit random forest and ℓ_1-penalized regression; derive feature importance from whichever has lower CV error—**still strict FDR control**
- Can even let analyst look at (masked version of) data to choose Z function

Adaptivity and Prior Information in W_j

Recall LCD: $W_j = |\beta_j| - |\tilde{\beta}_j|$, where β_j, $\tilde{\beta}_j$ come from ℓ_1-penalized regression

Adaptivity

- Cross-validation (on $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$) to choose the penalty parameter in LCD
- Higher-level adaptivity: CV to choose best-fitting model for inference
  - E.g., fit random forest and ℓ_1-penalized regression; derive feature importance from whichever has lower CV error—**still strict FDR control**
- Can even let analyst look at (masked version of) data to choose Z function

Prior information

- **Bayesian approach**: choose prior and model, and Z_j could be the posterior probability that X_j contributes to the model

Adaptivity and Prior Information in W_j

Recall LCD: $W_j = |\beta_j| - |\tilde{\beta}_j|$, where β_j, $\tilde{\beta}_j$ come from ℓ_1-penalized regression

Adaptivity

- Cross-validation (on $[\boldsymbol{X}\ \tilde{\boldsymbol{X}}]$) to choose the penalty parameter in LCD
- Higher-level adaptivity: CV to choose best-fitting model for inference
  - E.g., fit random forest and ℓ_1-penalized regression; derive feature importance from whichever has lower CV error—**still strict FDR control**
- Can even let analyst look at (masked version of) data to choose Z function

Prior information

- **Bayesian approach**: choose prior and model, and Z_j could be the posterior probability that X_j contributes to the model
- Still strict FDR control, **even if wrong prior or MCMC has not converged**

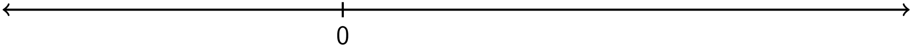
Step (3): Find the Knockoff Threshold

Find the Knockoff Threshold

Example with $p = 10$ and $q = 20\% = 1/5$:

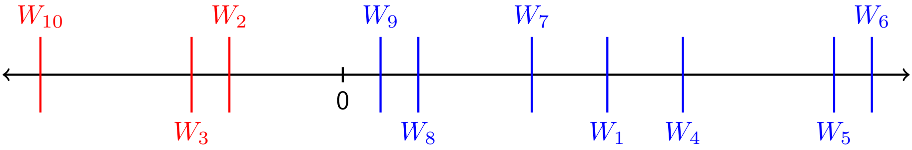
Find the Knockoff Threshold

Example with $p = 10$ and $q = 20\% = 1/5$:



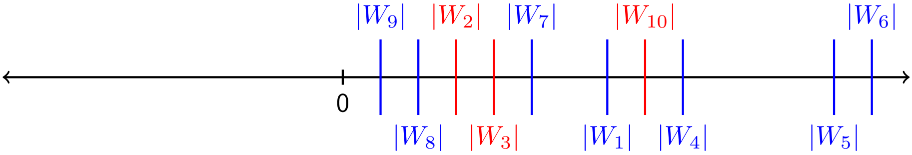
Find the Knockoff Threshold

Example with $p = 10$ and $q = 20\% = 1/5$:



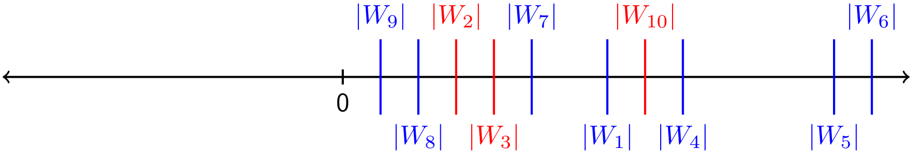
Find the Knockoff Threshold

Example with $p = 10$ and $q = 20\% = 1/5$:



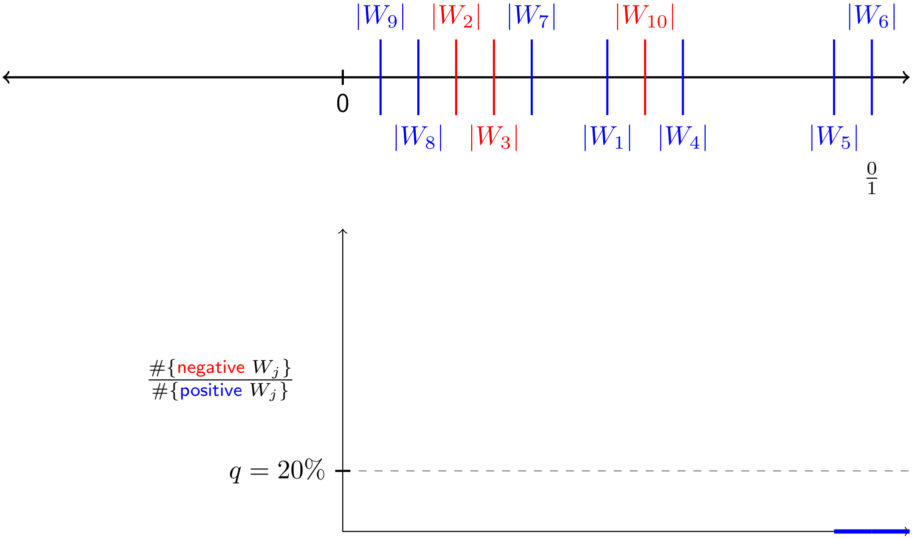
Find the Knockoff Threshold

Example with $p = 10$ and $q = 20\% = 1/5$:



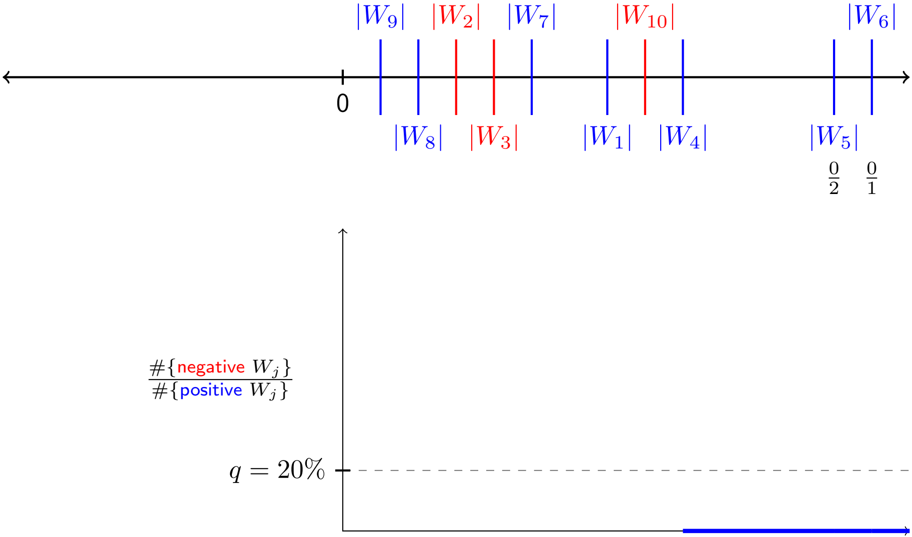
Find the Knockoff Threshold

Example with $p = 10$ and $q = 20\% = 1/5$:



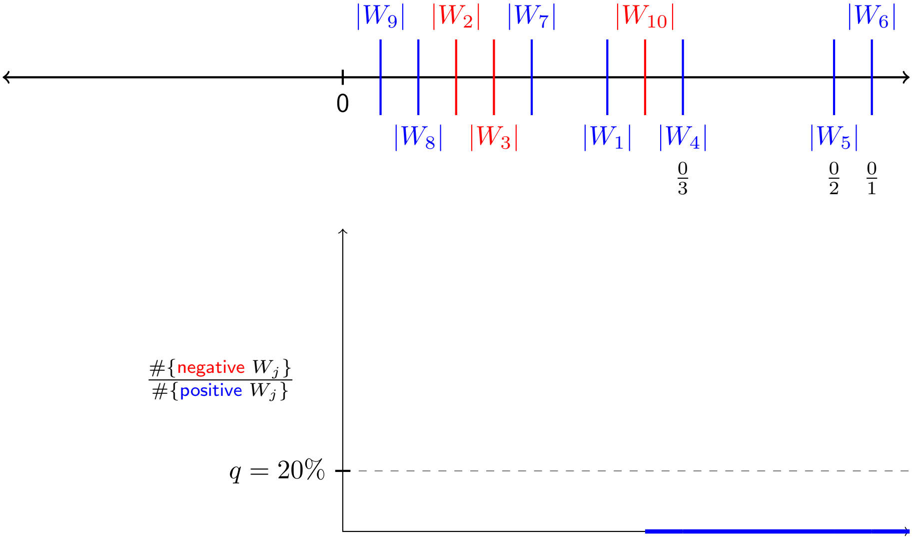
Find the Knockoff Threshold

Example with $p = 10$ and $q = 20\% = 1/5$:



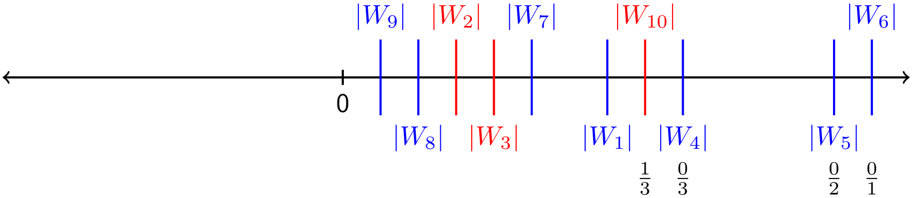
Find the Knockoff Threshold

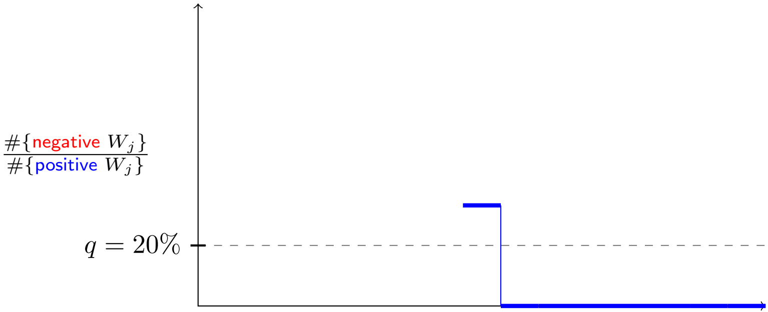
Example with $p = 10$ and $q = 20\% = 1/5$:



Find the Knockoff Threshold

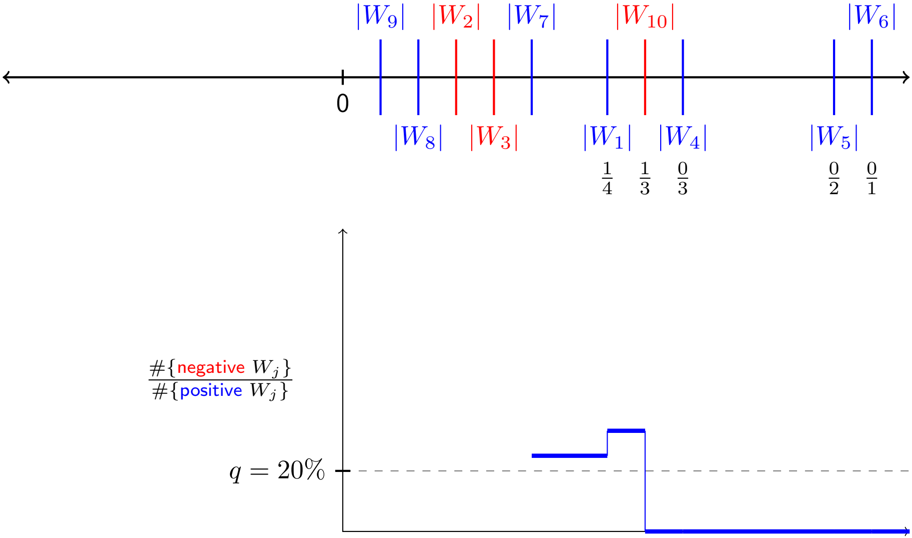
Example with $p = 10$ and $q = 20\% = 1/5$:





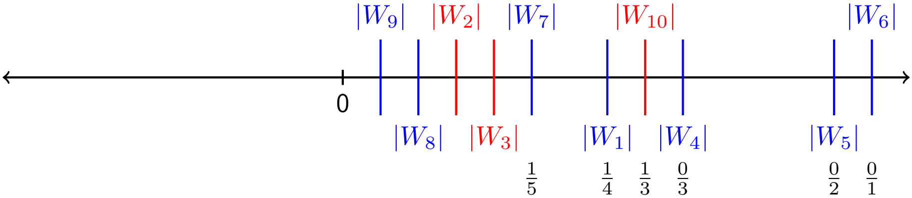
Find the Knockoff Threshold

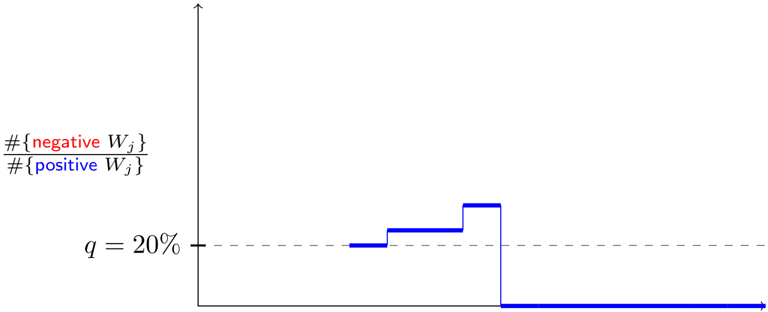
Example with $p = 10$ and $q = 20\% = 1/5$:



Find the Knockoff Threshold

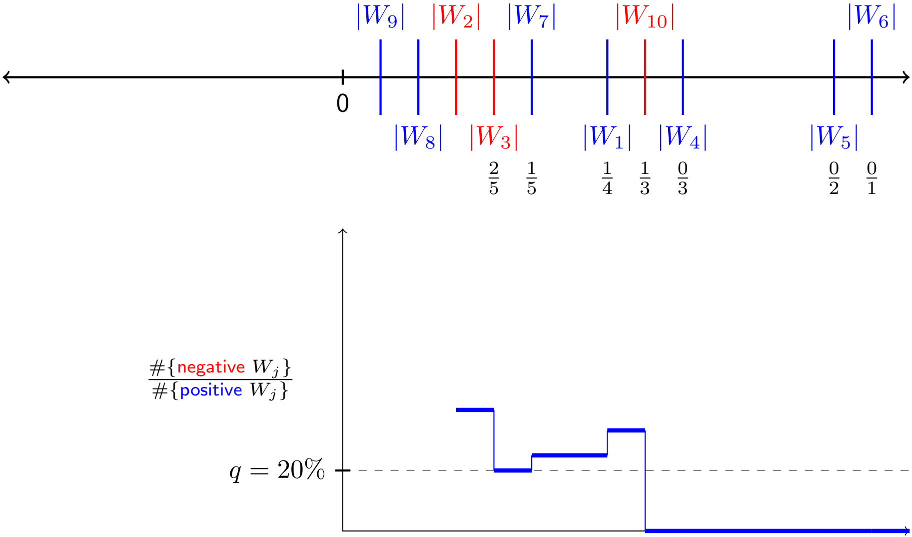
Example with $p = 10$ and $q = 20\% = 1/5$:





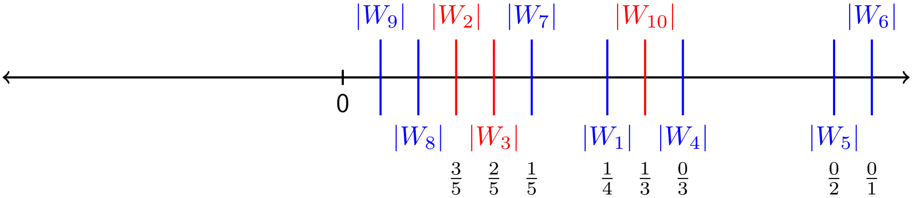
Find the Knockoff Threshold

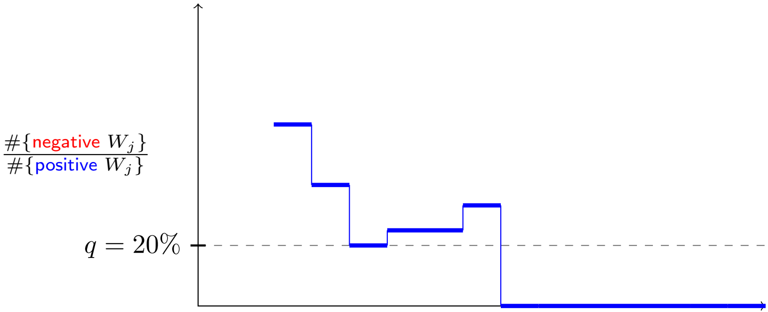
Example with $p = 10$ and $q = 20\% = 1/5$:



Find the Knockoff Threshold

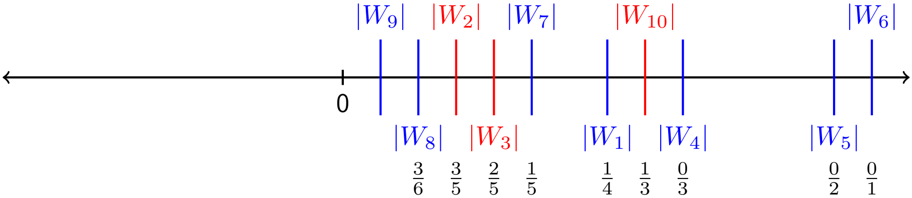
Example with $p = 10$ and $q = 20\% = 1/5$:

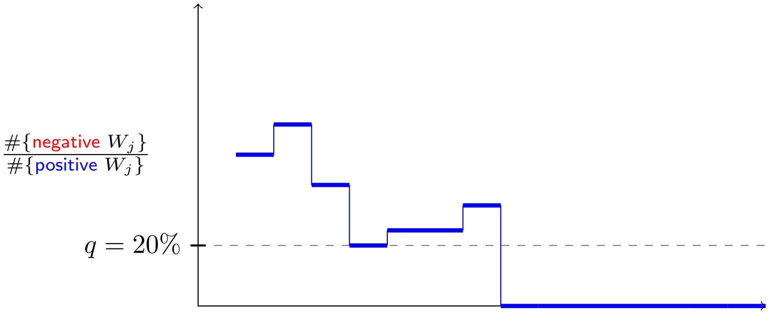




Find the Knockoff Threshold

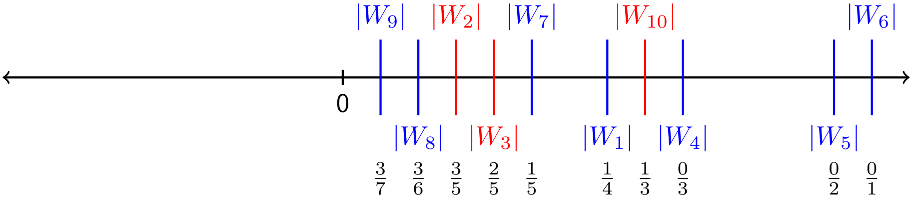
Example with $p = 10$ and $q = 20\% = 1/5$:

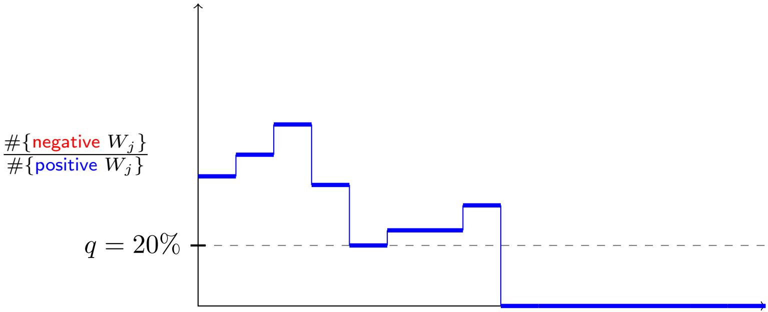




Find the Knockoff Threshold

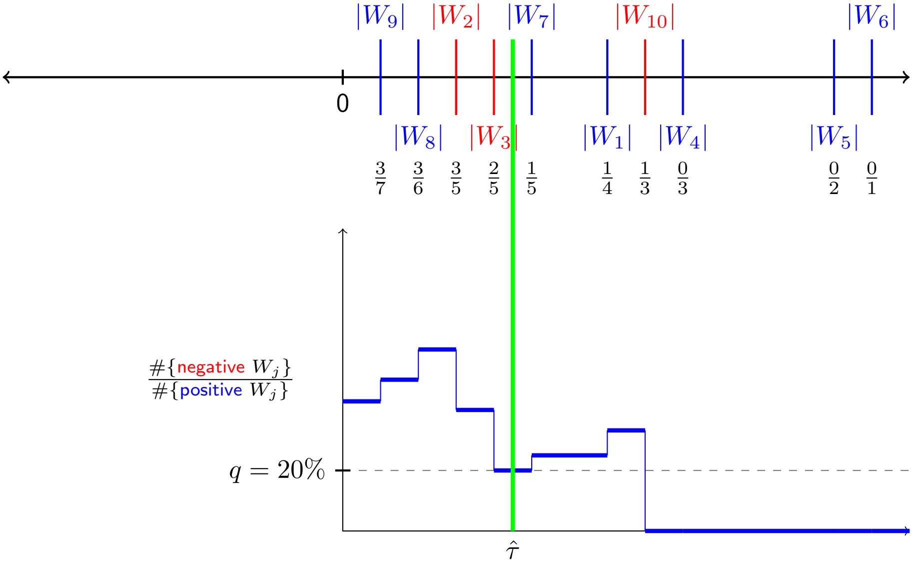
Example with $p = 10$ and $q = 20\% = 1/5$:





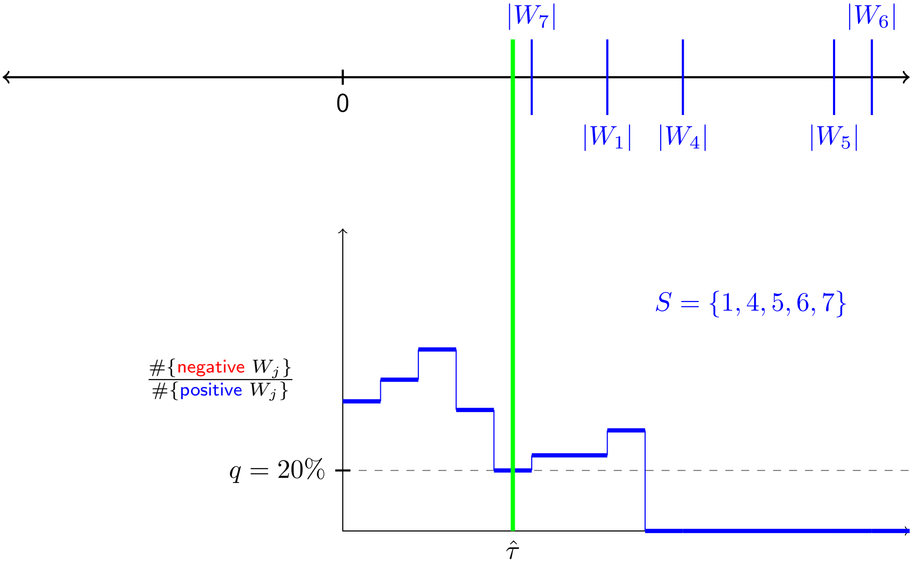
Find the Knockoff Threshold

Example with $p = 10$ and $q = 20\% = 1/5$:



Find the Knockoff Threshold

Example with $p = 10$ and $q = 20\% = 1/5$:



Intuition for FDR Control

$$\text{FDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right)$$

Intuition for FDR Control

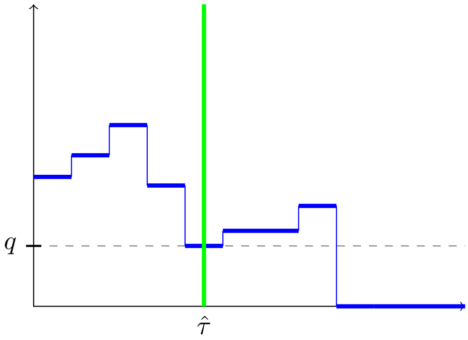
$$\text{FDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right)$$
$$= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$

Intuition for FDR Control

$$\begin{aligned}
\text{FDR} &= \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right) \\
&= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right) \\
&\approx \mathbb{E}\left(\frac{\#\{\text{null negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)
\end{aligned}$$

Intuition for FDR Control

$$
\begin{aligned}
\text{FDR} &= \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right) \\
&= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right) \\
&\approx \mathbb{E}\left(\frac{\#\{\text{null negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right) \\
&\leq \mathbb{E}\left(\frac{\#\{\text{negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)
\end{aligned}
$$



GWAS Application

Genetic Analysis of Crohn's Disease

2007 case-control study by WTCCC

- $n \approx 5,000$, $p \approx 375,000$; preprocessing mirrored original analysis

Genetic Analysis of Crohn's Disease

2007 case-control study by WTCCC

- $n \approx 5,000$, $p \approx 375,000$; preprocessing mirrored original analysis
- **Strong spatial structure**: second-order knockoffs generated using genetic covariance estimate (Wen and Stephens, 2010)

Genetic Analysis of Crohn's Disease

2007 case-control study by WTCCC

- $n \approx 5,000$, $p \approx 375,000$; preprocessing mirrored original analysis
- **Strong spatial structure**: second-order knockoffs generated using genetic covariance estimate (Wen and Stephens, 2010)
- Entire analysis took 6 hours of serial computation time; **1 hour** in parallel

Genetic Analysis of Crohn's Disease

2007 case-control study by WTCCC

- $n \approx 5,000$, $p \approx 375,000$; preprocessing mirrored original analysis
- **Strong spatial structure**: second-order knockoffs generated using genetic covariance estimate (Wen and Stephens, 2010)
- Entire analysis took 6 hours of serial computation time; **1 hour** in parallel
- Knockoffs made **twice as many discoveries** as original analysis

Genetic Analysis of Crohn's Disease

2007 case-control study by WTCCC

- $n \approx 5,000$, $p \approx 375,000$; preprocessing mirrored original analysis
- **Strong spatial structure**: second-order knockoffs generated using genetic covariance estimate (Wen and Stephens, 2010)
- Entire analysis took 6 hours of serial computation time; **1 hour** in parallel
- Knockoffs made **twice as many discoveries** as original analysis
  - Some new discoveries confirmed in larger study

Genetic Analysis of Crohn's Disease

2007 case-control study by WTCCC

- $n \approx 5,000$, $p \approx 375,000$; preprocessing mirrored original analysis
- **Strong spatial structure**: second-order knockoffs generated using genetic covariance estimate (Wen and Stephens, 2010)
- Entire analysis took 6 hours of serial computation time; **1 hour** in parallel
- Knockoffs made **twice as many discoveries** as original analysis
  - Some new discoveries confirmed in larger study
  - Some corroborated by work on nearby genes: promising candidates

Genetic Analysis of Crohn's Disease

2007 case-control study by WTCCC

- $n \approx 5,000$, $p \approx 375,000$; preprocessing mirrored original analysis
- **Strong spatial structure**: second-order knockoffs generated using genetic covariance estimate (Wen and Stephens, 2010)
- Entire analysis took 6 hours of serial computation time; **1 hour** in parallel
- Knockoffs made **twice as many discoveries** as original analysis
  - Some new discoveries confirmed in larger study
  - Some corroborated by work on nearby genes: promising candidates
  - Similar result when HMM knockoffs applied to same data (Sesia et al., 2017)

Discussion

Summary and Next Steps

By conditioning on Y and modeling X, knockoffs can be applied to high-dimensional and nonlinear problems, where it is powerful, flexible, and appears robust

Summary and Next Steps

By conditioning on Y and modeling X, knockoffs can be applied to high-dimensional and nonlinear problems, where it is powerful, flexible, and appears robust

Some future directions for research:

- *Theoretical*: rigorous guarantees on robustness

Summary and Next Steps

By conditioning on Y and modeling X, knockoffs can be applied to high-dimensional and nonlinear problems, where it is powerful, flexible, and appears robust

Some future directions for research:

- *Theoretical*: rigorous guarantees on robustness
- *Methodological*: develop knockoff constructions for new X distributions

Summary and Next Steps

By conditioning on Y and modeling X, knockoffs can be applied to high-dimensional and nonlinear problems, where it is powerful, flexible, and appears robust

Some future directions for research:

- *Theoretical*: rigorous guarantees on robustness
- *Methodological*: develop knockoff constructions for new X distributions
- *Applied*: team up with domain experts who know/control their X, e.g., gene knockout/knockdown, climate change modeling

Summary and Next Steps

By conditioning on Y and modeling X, knockoffs can be applied to high-dimensional and nonlinear problems, where it is powerful, flexible, and appears robust

Some future directions for research:

- *Theoretical*: rigorous guarantees on robustness
- *Methodological*: develop knockoff constructions for new X distributions
- *Applied*: team up with domain experts who know/control their X, e.g., gene knockout/knockdown, climate change modeling

Thank you!

Appendix

References

Barber, R. F. and Candès, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.*, 43(5):2055–2085.

Candès, E., Fan, Y., Janson, L., and Lv, J. (2016). Panning for gold: Model-free knockoffs for high-dimensional controlled variable selection. *arXiv preprint arXiv:1610.02351*.

Dai, R. and Barber, R. F. (2016). The knockoff filter for fdr control in group-sparse and multitask regression. *arXiv preprint arXiv:1602.03589*.

Sesia, M., Sabatti, C., and Candès, E. (2017). Gene hunting with knockoffs for hidden markov models. *arXiv preprint arXiv:1706.04677*.

Wen, X. and Stephens, M. (2010). Using linear predictors to impute allele frequencies from summary or pooled genotype data. *Ann. Appl. Stat.*, 4(3):1158–1182.

WTCCC (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661–678.

Simulations in Low-Dimensional Linear Model

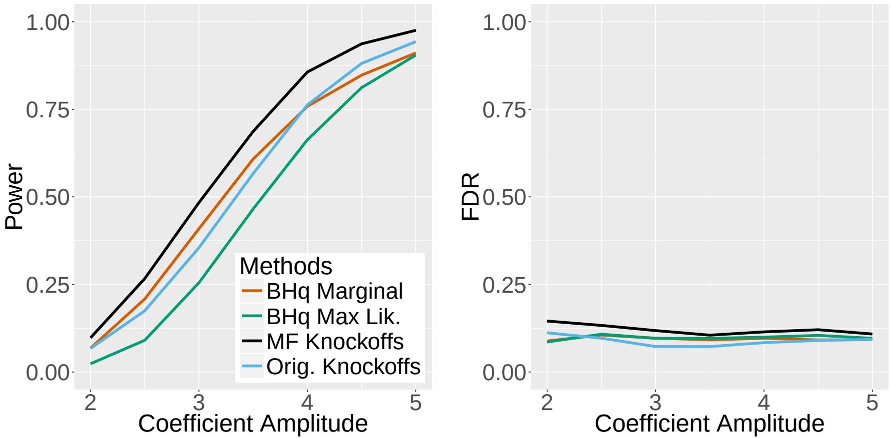


Figure: Power and FDR (target is 10%) for MF knockoffs and alternative procedures. The design matrix is i.i.d. $\mathcal{N}(0, 1/n)$, $n = 3000$, $p = 1000$, and y comes from a Gaussian linear model with 60 nonzero regression coefficients having equal magnitudes and random signs. The noise variance is 1.

Simulations in Low-Dimensional Nonlinear Model

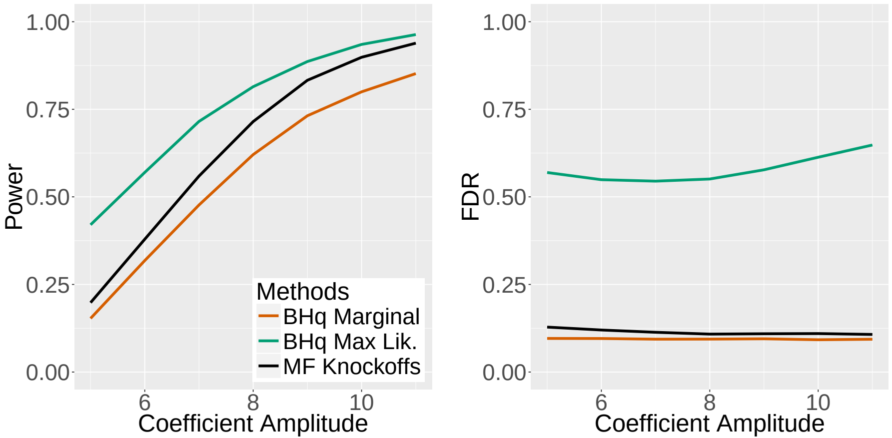


Figure: Power and FDR (target is 10%) for MF knockoffs and alternative procedures. The design matrix is i.i.d. $\mathcal{N}(0, 1/n)$, $n = 3000$, $p = 1000$, and y comes from a binomial linear model with logit link function, and 60 nonzero regression coefficients having equal magnitudes and random signs.

Simulations in High Dimensions

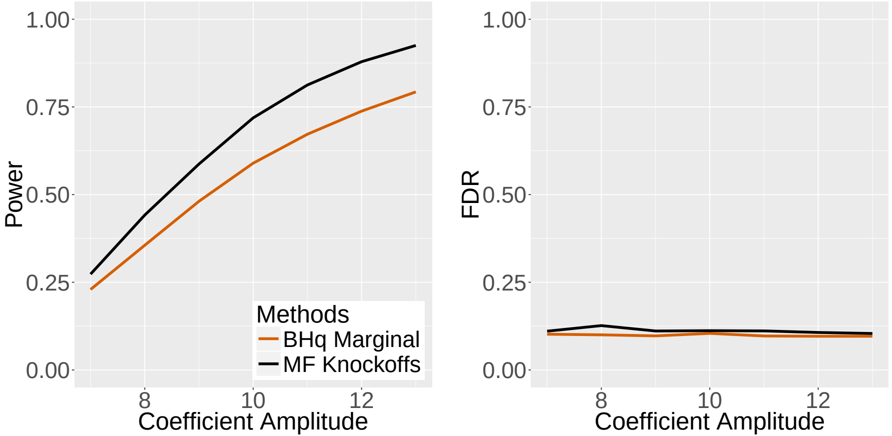


Figure: Power and FDR (target is 10%) for MF knockoffs and alternative procedures. The design matrix is i.i.d. $\mathcal{N}(0, 1/n)$, $n = 3000$, $p = 6000$, and y comes from a binomial linear model with logit link function, and 60 nonzero regression coefficients having equal magnitudes and random signs.

Simulations in High Dimensions with Dependence

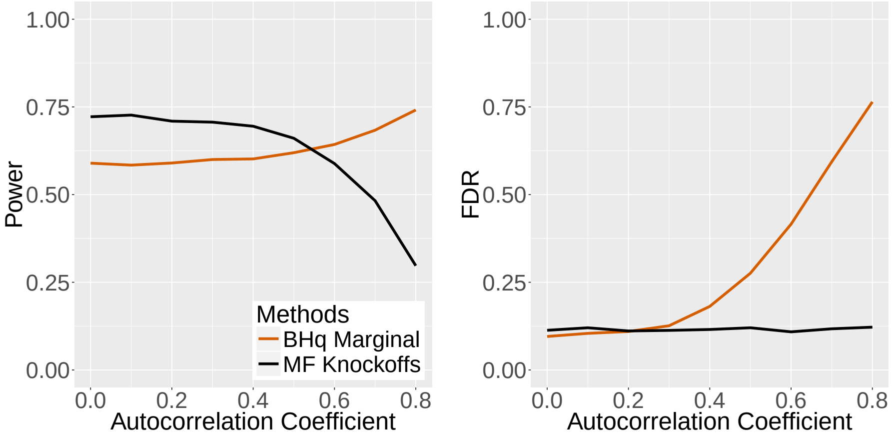


Figure: Power and FDR (target is 10%) for MF knockoffs and alternative procedures. The design matrix has AR(1) columns, and marginally each $X_j \sim \mathcal{N}(0, 1/n)$. $n = 3000$, $p = 6000$, and y follows a binomial linear model with logit link function, and 60 nonzero coefficients with random signs and randomly selected locations.

Checking Sensitivity to Misspecification Error

	Concern about misspecification	
	$Y \mid X$	X
Canonical (model Y, not X)	Yes	No
model X, not Y	No	Yes

Checking Sensitivity to Misspecification Error

	Concern about misspecification	
	$Y \mid X$	X
Canonical (model Y, not X)	Yes	No
model X, not Y	No	Yes
Misspecification replicated in simulation?	No	Yes

Checking Sensitivity to Misspecification Error

	Concern about misspecification: $Y \mid X$	Concern about misspecification: X
Canonical (model Y, not X)	Yes	No
model X, not Y	No	Yes
Misspecification replicated in simulation?	No	Yes

Can actually **check sensitivity** to misspecification error!

Robustness on Real Data

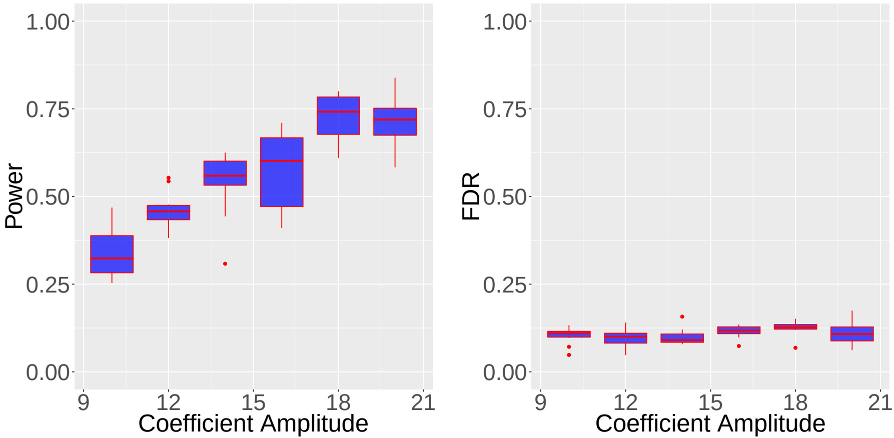


Figure: Power and FDR (target is 10%) for model-free knockoffs applied to subsamples of a chromosome 1 of real genetic design matrix; $n \approx 1,400$.

Computation of Second-Order Knockoffs

$\mathrm{Cov}(X_1, \ldots, X_p) = \boldsymbol{\Sigma}$, need:

$$\mathrm{Cov}(X_1, \ldots, X_p, \tilde{X}_1, \ldots, \tilde{X}_p) = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} \\ \boldsymbol{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} & \boldsymbol{\Sigma} \end{bmatrix}$$

Computation of Second-Order Knockoffs

$\mathrm{Cov}(X_1, \ldots, X_p) = \mathbf{\Sigma}$, need:

$$\mathrm{Cov}(X_1, \ldots, X_p, \tilde{X}_1, \ldots, \tilde{X}_p) = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} \\ \mathbf{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} & \mathbf{\Sigma} \end{bmatrix}$$

- **Equicorrelated (EQ)** (fast, less powerful): $s_j^{\mathsf{EQ}} = 2\lambda_{\mathsf{min}}(\mathbf{\Sigma}) \wedge 1$ for all j

Computation of Second-Order Knockoffs

$\mathrm{Cov}(X_1, \ldots, X_p) = \boldsymbol{\Sigma}$, need:

$$\mathrm{Cov}(X_1, \ldots, X_p, \tilde{X}_1, \ldots, \tilde{X}_p) = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} \\ \boldsymbol{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} & \boldsymbol{\Sigma} \end{bmatrix}$$

- **Equicorrelated (EQ)** (fast, less powerful): $s_j^{\mathsf{EQ}} = 2\lambda_{\mathsf{min}}(\boldsymbol{\Sigma}) \wedge 1$ for all j
- **Semidefinite program (SDP)** (slower, more powerful):

$$\begin{array}{rl} \text{minimize} & \sum_j |1 - s_j^{\mathsf{SDP}}| \\ \text{subject to} & s_j^{\mathsf{SDP}} \geq 0 \\ & \mathrm{diag}\{s^{\mathsf{SDP}}\} \preceq 2\boldsymbol{\Sigma}, \end{array}$$

Computation of Second-Order Knockoffs

$\mathrm{Cov}(X_1, \ldots, X_p) = \mathbf{\Sigma}$, need:

$$\mathrm{Cov}(X_1, \ldots, X_p, \tilde{X}_1, \ldots, \tilde{X}_p) = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} \\ \mathbf{\Sigma} - \mathrm{diag}\{\boldsymbol{s}\} & \mathbf{\Sigma} \end{bmatrix}$$

- **Equicorrelated (EQ)** (fast, less powerful): $s_j^{\mathsf{EQ}} = 2\lambda_{\mathsf{min}}(\mathbf{\Sigma}) \wedge 1$ for all j
- **Semidefinite program (SDP)** (slower, more powerful):

$$\begin{array}{rl} \text{minimize} & \sum_j |1 - s_j^{\mathsf{SDP}}| \\ \text{subject to} & s_j^{\mathsf{SDP}} \geq 0 \\ & \mathrm{diag}\{s^{\mathsf{SDP}}\} \preceq 2\mathbf{\Sigma}, \end{array}$$

- **(New) Approximate SDP**:
  - Approximate $\mathbf{\Sigma}$ as block diagonal so that SDP separates
  - Bisection search scalar multiplier of solution to account for approximation
  - faster than SDP, more powerful than EQ, and easily parallelizable

Sequential Independent Pairs Generates Valid Knockoffs

Algorithm 1 Sequential Conditional Independent Pairs

for $j = \{1, \ldots, p\}$ **do**
  Sample $\tilde{X}_j$ from $\mathcal{L}(X_j \,|\, X_{\text{-}j}, \tilde{X}_{1:j-1})$ conditionally independently of X_j
end

Sequential Independent Pairs Generates Valid Knockoffs

Algorithm 1 Sequential Conditional Independent Pairs

for $j = \{1, \ldots, p\}$ **do**
  Sample $\tilde{X}_j$ from $\mathcal{L}(X_j \,|\, X_{\text{-}j},\, \tilde{X}_{1:j-1})$ conditionally independently of X_j
end

Proof sketch (discrete case):

- Denote PMF of $(X_{1:p}, \tilde{X}_{1:j-1})$ by $\mathcal{L}(X_{\text{-}j}, X_j, \tilde{X}_{1:j-1})$

Sequential Independent Pairs Generates Valid Knockoffs

Algorithm 1 Sequential Conditional Independent Pairs

for $j = \{1, \dots, p\}$ **do**

  Sample $\tilde{X}_j$ from $\mathcal{L}(X_j \mid X_{\text{-}j}, \tilde{X}_{1:j-1})$ conditionally independently of X_j

end

Proof sketch (discrete case):

- Denote PMF of $(X_{1:p}, \tilde{X}_{1:j-1})$ by $\mathcal{L}(X_{\text{-}j}, X_j, \tilde{X}_{1:j-1})$
- Conditional PMF of $\tilde{X}_j \mid X_{1:p}, \tilde{X}_{1:j-1}$ is

$$\frac{\mathcal{L}(X_{\text{-}j}, \tilde{X}_j, \tilde{X}_{1:j-1})}{\sum_u \mathcal{L}(X_{\text{-}j}, u, \tilde{X}_{1:j-1})}.$$

Sequential Independent Pairs Generates Valid Knockoffs

Algorithm 1 Sequential Conditional Independent Pairs

for $j = \{1, \ldots, p\}$ **do**
  Sample $\tilde{X}_j$ from $\mathcal{L}(X_j \mid X_{\text{-}j}, \tilde{X}_{1:j-1})$ conditionally independently of X_j
end

Proof sketch (discrete case):

- Denote PMF of $(X_{1:p}, \tilde{X}_{1:j-1})$ by $\mathcal{L}(X_{\text{-}j}, X_j, \tilde{X}_{1:j-1})$
- Conditional PMF of $\tilde{X}_j \mid X_{1:p}, \tilde{X}_{1:j-1}$ is

$$\frac{\mathcal{L}(X_{\text{-}j}, \tilde{X}_j, \tilde{X}_{1:j-1})}{\sum_u \mathcal{L}(X_{\text{-}j}, u, \tilde{X}_{1:j-1})}.$$

- Joint PMF of $(X_{1:p}, \tilde{X}_{1:j})$ is

$$\frac{\mathcal{L}(X_{\text{-}j}, X_j, \tilde{X}_{1:j-1})\mathcal{L}(X_{\text{-}j}, \tilde{X}_j, \tilde{X}_{1:j-1})}{\sum_u \mathcal{L}(X_{\text{-}j}, u, \tilde{X}_{1:j-1})}$$

Sequential Independent Pairs Generates Valid Knockoffs

Algorithm 1 Sequential Conditional Independent Pairs

for $j = \{1, \ldots, p\}$ **do**
  Sample $\tilde{X}_j$ from $\mathcal{L}(X_j \,|\, X_{\text{-}j}, \tilde{X}_{1:j-1})$ conditionally independently of X_j
end

Proof sketch (discrete case):

- Denote PMF of $(X_{1:p}, \tilde{X}_{1:j-1})$ by $\mathcal{L}(X_{\text{-}j}, X_j, \tilde{X}_{1:j-1})$
- Conditional PMF of $\tilde{X}_j \,|\, X_{1:p}, \tilde{X}_{1:j-1}$ is

$$\frac{\mathcal{L}(X_{\text{-}j}, \tilde{X}_j, \tilde{X}_{1:j-1})}{\sum_u \mathcal{L}(X_{\text{-}j}, u, \tilde{X}_{1:j-1})}.$$

- Joint PMF of $(X_{1:p}, \tilde{X}_{1:j})$ is

$$\frac{\mathcal{L}(X_{\text{-}j}, {\color{red}X_j}, \tilde{X}_{1:j-1})\mathcal{L}(X_{\text{-}j}, {\color{red}\tilde{X}_j}, \tilde{X}_{1:j-1})}{\sum_u \mathcal{L}(X_{\text{-}j}, u, \tilde{X}_{1:j-1})}$$

Sequential Independent Pairs Generates Valid Knockoffs

Algorithm 1 Sequential Conditional Independent Pairs

for $j = \{1, \dots, p\}$ **do**
  Sample $\tilde{X}_j$ from $\mathcal{L}(X_j \mid X_{\text{-}j}, \tilde{X}_{1:j-1})$ conditionally independently of X_j
end

Proof sketch (discrete case):

- Denote PMF of $(X_{1:p}, \tilde{X}_{1:j-1})$ by $\mathcal{L}(X_{\text{-}j}, X_j, \tilde{X}_{1:j-1})$
- Conditional PMF of $\tilde{X}_j \mid X_{1:p}, \tilde{X}_{1:j-1}$ is

$$\frac{\mathcal{L}(X_{\text{-}j}, \tilde{X}_j, \tilde{X}_{1:j-1})}{\sum_u \mathcal{L}(X_{\text{-}j}, u, \tilde{X}_{1:j-1})}.$$

- Joint PMF of $(X_{1:p}, \tilde{X}_{1:j})$ is

$$\frac{\mathcal{L}(X_{\text{-}j}, {\color{red}\tilde{X}_j}, \tilde{X}_{1:j-1})\mathcal{L}(X_{\text{-}j}, {\color{red}X_j}, \tilde{X}_{1:j-1})}{\sum_u \mathcal{L}(X_{\text{-}j}, u, \tilde{X}_{1:j-1})}$$

Sequential Independent Pairs Generates Valid Knockoffs

Algorithm 1 Sequential Conditional Independent Pairs

for $j = \{1, \ldots, p\}$ **do**

  Sample $\tilde{X}_j$ from $\mathcal{L}(X_j \mid X_{\text{-}j}, \tilde{X}_{1:j-1})$ conditionally independently of X_j

end

Proof sketch (discrete case):

- Denote PMF of $(X_{1:p}, \tilde{X}_{1:j-1})$ by $\mathcal{L}(X_{\text{-}j}, X_j, \tilde{X}_{1:j-1})$
- Conditional PMF of $\tilde{X}_j \mid X_{1:p}, \tilde{X}_{1:j-1}$ is

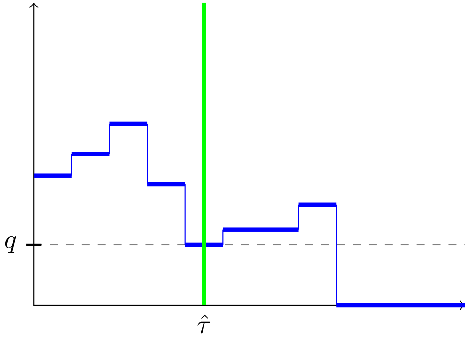
$$\frac{\mathcal{L}(X_{\text{-}j}, \tilde{X}_j, \tilde{X}_{1:j-1})}{\sum_u \mathcal{L}(X_{\text{-}j}, u, \tilde{X}_{1:j-1})}.$$

- Joint PMF of $(X_{1:p}, \tilde{X}_{1:j})$ is

$$\frac{\color{red}{\mathcal{L}}(X_{\text{-}j}, X_j, \tilde{X}_{1:j-1})\color{red}{\mathcal{L}}(X_{\text{-}j}, \tilde{X}_j, \tilde{X}_{1:j-1})}{\sum_u \color{red}{\mathcal{L}}(X_{\text{-}j}, u, \tilde{X}_{1:j-1})}$$

Proof of Control

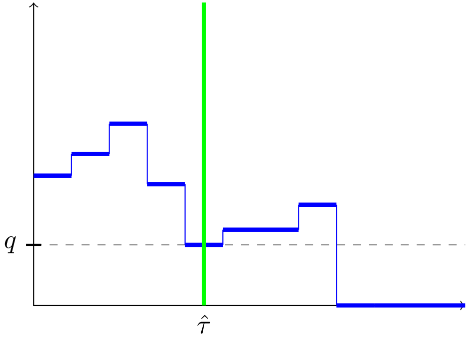
$$\text{FDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right)$$



Proof of Control

$$\text{FDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right)$$

$$= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$

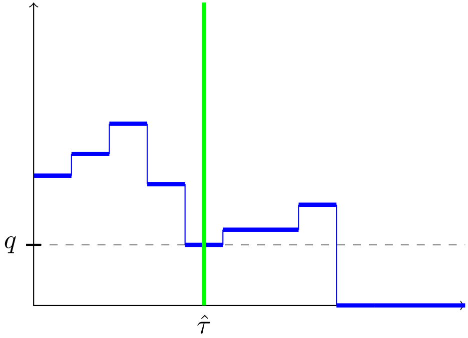


Proof of Control

$$\text{FDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right)$$

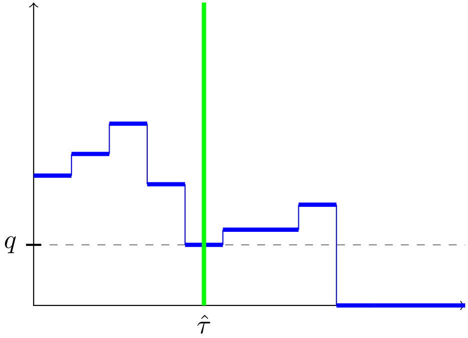
$$= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$

$$\approx \mathbb{E}\left(\frac{\#\{\text{null negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$



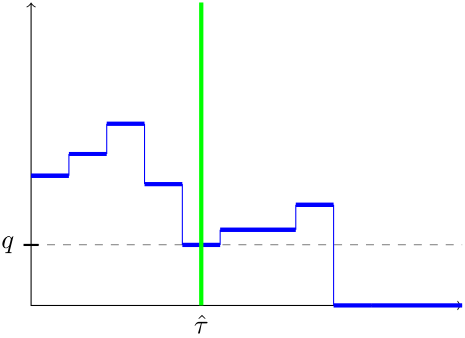
Proof of Control

$$
\begin{aligned}
\text{FDR} &= \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right)\\
&= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)\\
&\approx \mathbb{E}\left(\frac{\#\{\text{null negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)\\
&\leq \mathbb{E}\left(\frac{\#\{\text{negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)
\end{aligned}
$$



Proof of Control

$$\begin{aligned}
\text{FDR} &= \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right) \\
&= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right) \\
&\approx \mathbb{E}\left(\frac{\#\{\text{null negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right) \\
&\leq \mathbb{E}\left(\frac{\#\{\text{negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)
\end{aligned}$$

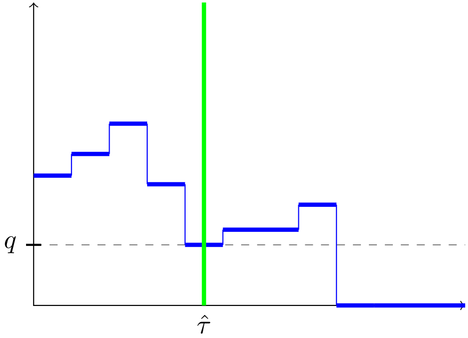


More precisely:

$$\text{mFDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{q^{-1} + \#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right) = \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{q^{-1} + \#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$

Proof of Control

$$\text{FDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right)$$
$$= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$
$$\approx \mathbb{E}\left(\frac{\#\{\text{null negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$
$$\leq \mathbb{E}\left(\frac{\#\{\text{negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$



More precisely:

$$\text{mFDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{q^{-1} + \#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right) = \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{q^{-1} + \#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$
$$= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{1 + \#\{\text{null negative } |W_j| > \hat{\tau}\}} \cdot \frac{1 + \#\{\text{null negative } |W_j| > \hat{\tau}\}}{q^{-1} + \#\{\text{positive} |W_j| > \hat{\tau}\}}\right)$$

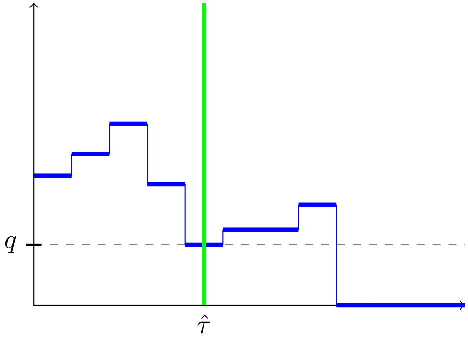
Proof of Control

$$\text{FDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right)$$

$$= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$

$$\approx \mathbb{E}\left(\frac{\#\{\text{null negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$

$$\leq \mathbb{E}\left(\frac{\#\{\text{negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$



More precisely:

$$\text{mFDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{q^{-1} + \#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right) = \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{q^{-1} + \#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$

$$= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{1 + \#\{\text{null negative } |W_j| > \hat{\tau}\}} \cdot \underbrace{\frac{1 + \#\{\text{null negative } |W_j| > \hat{\tau}\}}{q^{-1} + \#\{\text{positive} |W_j| > \hat{\tau}\}}}_{\leq q \text{ by definition of } \hat{\tau}}\right)$$

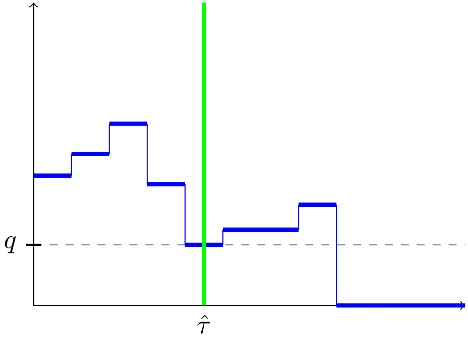
Proof of Control

$$\text{FDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right)$$

$$= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$

$$\approx \mathbb{E}\left(\frac{\#\{\text{null negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$

$$\leq \mathbb{E}\left(\frac{\#\{\text{negative } |W_j| > \hat{\tau}\}}{\#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$



More precisely:

$$\text{mFDR} = \mathbb{E}\left(\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{q^{-1} + \#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right) = \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{q^{-1} + \#\{\text{positive } |W_j| > \hat{\tau}\}}\right)$$

$$= \mathbb{E}\left(\underbrace{\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{1 + \#\{\text{null negative } |W_j| > \hat{\tau}\}}}_{\substack{\text{Supermartingale } \leq 1 \\ \text{with } \hat{\tau} \text{ a stopping time}}} \cdot \underbrace{\frac{1 + \#\{\text{null negative } |W_j| > \hat{\tau}\}}{q^{-1} + \#\{\text{positive} |W_j| > \hat{\tau}\}}}_{\leq q \text{ by definition of } \hat{\tau}}\right)$$