Using Knockoffs to Find Important Variables with Statistical Guarantees

Lucas Janson

Harvard University Department of Statistics

Models, Inference, and Algorithms @ Broad, December 12, 2018

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),
- $X_1, \dots, X_p$ 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:

- Biology/genomics/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:

- Biology/genomics/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:

- Biology/genomics/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 $X_{\text{-}j}$:

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

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 $X_{\text{-}j}$:

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

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

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 $X_{\text{-}j}$:

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

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 $X_{\text{-}j}$:

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

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)**:

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

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

Group Knockoffs

"What if two variables are so correlated as to be indistinguishable?"

Group Knockoffs

"What if two variables are so correlated as to be indistinguishable?"

Insufficient info to select either variable confidently (needed for FDR control)

Group Knockoffs

"What if two variables are so correlated as to be indistinguishable?"

Insufficient info to select either variable confidently (needed for FDR control)

Single-variable resolution impossible: **wrong question**

- Group variables with their highly-correlated neighbors: $\biguplus_{k=1}^{m} G_k = \{1, \ldots, p\}$

Group Knockoffs

"What if two variables are so correlated as to be indistinguishable?"

Insufficient info to select either variable confidently (needed for FDR control)

Single-variable resolution impossible: **wrong question**

- Group variables with their highly-correlated neighbors: $\biguplus_{k=1}^{m} G_k = \{1, \ldots, p\}$
- Redefine null hypothesis on per-group basis: group G_k is unimportant if

$$Y \perp\!\!\!\perp X_{G_k} \mid X_{\text{-}G_k}$$

Group Knockoffs

"What if two variables are so correlated as to be indistinguishable?"

Insufficient info to select either variable confidently (needed for FDR control)

Single-variable resolution impossible: **wrong question**

- Group variables with their highly-correlated neighbors: $\biguplus_{k=1}^{m} G_k = \{1, \ldots, p\}$
- Redefine null hypothesis on per-group basis: group G_k is unimportant if

$$Y \perp\!\!\!\perp X_{G_k} \mid X_{\text{-}G_k}$$

- Redefine FDR: for selected set of groups $\hat{S}_G$,

$$\mathrm{FDR}_G = \mathbb{E}\left[\frac{\#\{k \text{ in } \hat{S}_G \;:\; G_k \text{ contains no important variables}\}}{\#\{j \text{ in } \hat{S}_G\}}\right] \le q \quad \text{(e.g., 10\%)}$$

Group Knockoffs

"What if two variables are so correlated as to be indistinguishable?"

Insufficient info to select either variable confidently (needed for FDR control)

Single-variable resolution impossible: **wrong question**

- Group variables with their highly-correlated neighbors: $\biguplus_{k=1}^{m} G_k = \{1, \dots, p\}$
- Redefine null hypothesis on per-group basis: group G_k is unimportant if

$$Y \perp\!\!\!\perp X_{G_k} \mid X_{\text{-}G_k}$$

- Redefine FDR: for selected set of groups $\hat{S}_G$,

$$\mathrm{FDR}_G = \mathbb{E}\left[\frac{\#\{k \text{ in } \hat{S}_G \,:\, G_k \text{ contains no important variables}\}}{\#\{j \text{ in } \hat{S}_G\}}\right] \le q \quad \text{(e.g., 10\%)}$$

Everything in this talk works for this setting! (Dai and Barber, 2016)

Sneak Peak

Model-X knockoffs uses knowledge of X's distribution to solve the controlled variable selection problem with

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

Sneak Peak

Model-X knockoffs uses knowledge of X's distribution to solve the controlled variable selection problem with

- Any model for Y and $X_1, \ldots, X_p$
- 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

Model-X knockoffs uses knowledge of X's distribution to solve the controlled variable selection problem with

- Any model for Y and $X_1, \ldots, X_p$
- 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

Model-X knockoffs uses knowledge of X's distribution to solve the controlled variable selection problem with

- Any model for Y and $X_1, \ldots, X_p$
- 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%

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

Existing Methods for Controlled Variable Selection

- Marginal p-values
  - Excellent exploratory tool
  - Answer wrong question $Y \perp\!\!\!\perp X_j$ instead of $Y \perp\!\!\!\perp X_j \mid X_{\text{-}j}$
  - Can lose power, interpretation, and FDR control when X_j are correlated

Existing Methods for Controlled Variable Selection

- Marginal p-values
  - Excellent exploratory tool
  - Answer wrong question $Y \perp\!\!\!\perp X_j$ instead of $Y \perp\!\!\!\perp X_j \mid X_{\text{-}j}$
  - Can lose power, interpretation, and FDR control when X_j are correlated
- Bayesian inference
  - Great way of incorporating prior information
  - Computation constrains to very simple priors which may not match actual prior knowledge
  - Inference (esp. in high dimensions) is sensitive to choice of prior

Existing Methods for Controlled Variable Selection

- Marginal p-values
  - Excellent exploratory tool
  - Answer wrong question $Y \perp\!\!\!\perp X_j$ instead of $Y \perp\!\!\!\perp X_j \mid X_{\text{-}j}$
  - Can lose power, interpretation, and FDR control when X_j are correlated
- Bayesian inference
  - Great way of incorporating prior information
  - Computation constrains to very simple priors which may not match actual prior knowledge
  - Inference (esp. in high dimensions) is sensitive to choice of prior
- Machine learning
  - Excellent for prediction
  - Cross-validation comes with no statistical guarantees
  - Statistical analysis exists only for simplest methods (lasso) and makes unrealistic assumptions

Knockoffs

View from 10,000 feet

You have:

- n data samples of Y and X stacked into $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$

View from 10,000 feet

You have:

- n data samples of Y and X stacked into $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$
- Algorithm to compute **variable importance measure** Z_j of each X_j for Y
  - This need not be based on any statistical model, or have any statistical properties at all
  - For instance, you could fit any machine learning method and use the drop in prediction accuracy when $\boldsymbol{X}_j$ is removed from the data

View from 10,000 feet

You have:

- n data samples of Y and X stacked into $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$
- Algorithm to compute **variable importance measure** Z_j of each X_j for Y
  - This need not be based on any statistical model, or have any statistical properties at all
  - For instance, you could fit any machine learning method and use the drop in prediction accuracy when $\boldsymbol{X}_j$ is removed from the data
- Desired FDR level q but no way to use Z_j to control it

View from 10,000 feet

You have:

- n data samples of Y and X stacked into $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$
- Algorithm to compute **variable importance measure** Z_j of each X_j for Y
  - This need not be based on any statistical model, or have any statistical properties at all
  - For instance, you could fit any machine learning method and use the drop in prediction accuracy when $\boldsymbol{X}_j$ is removed from the data
- Desired FDR level q but no way to use Z_j to control it

Knockoffs allows you to:

- Select a subset of the variables based on your variable importance measure and nothing else, while controlling the FDR exactly (no asymptotics)

View from 10,000 feet

You have:

- n data samples of Y and X stacked into $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$
- Algorithm to compute **variable importance measure** Z_j of each X_j for Y
  - This need not be based on any statistical model, or have any statistical properties at all
  - For instance, you could fit any machine learning method and use the drop in prediction accuracy when $\boldsymbol{X}_j$ is removed from the data
- Desired FDR level q but no way to use Z_j to control it

Knockoffs allows you to:

- Select a subset of the variables based on your variable importance measure and nothing else, while controlling the FDR exactly (no asymptotics)

$$\boldsymbol{y}, \quad \boldsymbol{X}_1, \dots, \boldsymbol{X}_p$$

View from 10,000 feet

You have:

- n data samples of Y and X stacked into $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$
- Algorithm to compute **variable importance measure** Z_j of each X_j for Y
  - This need not be based on any statistical model, or have any statistical properties at all
  - For instance, you could fit any machine learning method and use the drop in prediction accuracy when $\boldsymbol{X}_j$ is removed from the data
- Desired FDR level q but no way to use Z_j to control it

Knockoffs allows you to:

- Select a subset of the variables based on your variable importance measure and nothing else, while controlling the FDR exactly (no asymptotics)

$$\begin{array}{rc} \boldsymbol{y}, & \boldsymbol{X}_1, \ldots, \boldsymbol{X}_p \\ & \downarrow \\ \text{Variable importances} & Z_1, \ldots, Z_p \end{array}$$

View from 10,000 feet

You have:

- n data samples of Y and X stacked into $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$
- Algorithm to compute **variable importance measure** Z_j of each X_j for Y
  - This need not be based on any statistical model, or have any statistical properties at all
  - For instance, you could fit any machine learning method and use the drop in prediction accuracy when $\boldsymbol{X}_j$ is removed from the data
- Desired FDR level q but no way to use Z_j to control it

Knockoffs allows you to:

- Select a subset of the variables based on your variable importance measure and nothing else, while controlling the FDR exactly (no asymptotics)

$$\boxed{\begin{array}{rc} \boldsymbol{y}, & \boldsymbol{X}_1, \ldots, \boldsymbol{X}_p \\ & \downarrow \\ \text{Variable importances} & Z_1, \ldots, Z_p \end{array}}$$

View from 10,000 feet

You have:

- n data samples of Y and X stacked into $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times p}$
- Algorithm to compute **variable importance measure** Z_j of each X_j for Y
  - This need not be based on any statistical model, or have any statistical properties at all
  - For instance, you could fit any machine learning method and use the drop in prediction accuracy when $\boldsymbol{X}_j$ is removed from the data
- Desired FDR level q but no way to use Z_j to control it

Knockoffs allows you to:

- Select a subset of the variables based on your variable importance measure and nothing else, while controlling the FDR exactly (no asymptotics)

$$\boxed{\begin{array}{rc} \boldsymbol{y}, & \boldsymbol{X}_1, \ldots, \boldsymbol{X}_p \\ & \downarrow \\ \text{Variable importances} & Z_1, \ldots, Z_p \end{array}} \xrightarrow{\text{knockoffs}} \hat{S} \subseteq \{1, \ldots, p\} \text{ s.t. FDR} \leq q$$

Overview of the Knockoffs Procedure

(1) **Construct knockoffs**:

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

Overview of 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**:
- Compute variable importance measures for all variables and their knockoffs
- For each variable, compute W_j as how much *more* important the original variable is than its knockoff

Overview of 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**:

- Compute variable importance measures for all variables and their knockoffs
- For each variable, compute W_j as how much *more* important the original variable is than its knockoff

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

- Order the variables by decreasing $|W_j|$ and proceed down list
- Select variables with positive W_j until an $\widehat{\text{FDR}}$ goes above q

Overview of 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**:
- Compute variable importance measures for all variables and their knockoffs
- For each variable, compute W_j as how much *more* important the original variable is than its knockoff

(3) **Find the knockoff threshold**:
- Order the variables by decreasing $|W_j|$ and proceed down list
- Select variables with positive W_j until an $\widehat{\text{FDR}}$ goes above 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 $1/2$

A Picture for Intuition

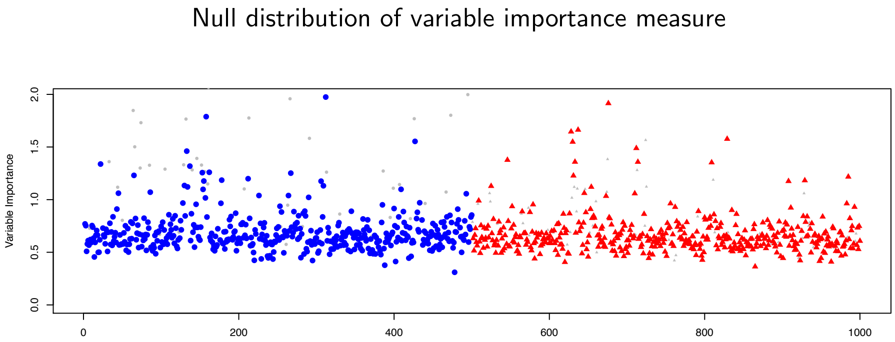


Figure: Variable importances for 500 variables and their knockoffs. Colored points are nulls, grey are non-nulls.

Step (1): Construct Knockoffs

Knockoff Construction

Valid knockoffs are defined by

(1) Swap exchangeability:

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

Knockoff Construction

Valid knockoffs are defined by

(1) Swap exchangeability:

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

(2) Nullity: $\widetilde{\boldsymbol{X}} \perp\!\!\!\perp \boldsymbol{y} \mid \boldsymbol{X}$ (don't look at $\boldsymbol{y}$ when constructing $\widetilde{\boldsymbol{X}}$)

Knockoff Construction

Valid knockoffs are defined by

(1) Swap exchangeability:

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

(2) Nullity: $\widetilde{\boldsymbol{X}} \perp\!\!\!\perp \boldsymbol{y} \mid \boldsymbol{X}$ (don't look at $\boldsymbol{y}$ when constructing $\widetilde{\boldsymbol{X}}$)

Example: $(X_1, \ldots, X_p) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, need

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

Knockoff Construction

Valid knockoffs are defined by

(1) Swap exchangeability:

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

(2) Nullity: $\widetilde{\boldsymbol{X}} \perp\!\!\!\perp \boldsymbol{y} \mid \boldsymbol{X}$ (don't look at $\boldsymbol{y}$ when constructing $\widetilde{\boldsymbol{X}}$)

Example: $(X_1, \ldots, X_p) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, need

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

Semidefinite program construction for $\boldsymbol{s}$:

$$\begin{array}{rl} \text{minimize} & \sum_j |\Sigma_{jj} - s_j| \\ \text{subject to} & s_j \geq 0 \\ & \text{diag}\{\boldsymbol{s}\} \preceq 2\boldsymbol{\Sigma}, \end{array}$$

Other Knockoff Constructions

Valid knockoff variables can always be generated:

Algorithm 1 Sequential Conditional Independent Pairs

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

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

end

Other Knockoff Constructions

Valid knockoff variables can always be generated:

Algorithm 1 Sequential Conditional Independent Pairs

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

Efficient knockoff constructions for the following X distributions:

- Multivariate Gaussian (Candès et al., 2018)
- Discrete Markov chains (Sesia et al., 2018)
- Hidden Markov models (Sesia et al., 2018)
- Gaussian mixture models (Gimenez et al., 2018)
- General graphical models (ongoing work with Wenshuo and others)

Other Knockoff Constructions

Valid knockoff variables can always be generated:

Algorithm 1 Sequential Conditional Independent Pairs

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

Efficient knockoff constructions for the following X distributions:

- Multivariate Gaussian (Candès et al., 2018)
- Discrete Markov chains (Sesia et al., 2018)
- Hidden Markov models (Sesia et al., 2018)
- Gaussian mixture models (Gimenez et al., 2018)
- General graphical models (ongoing work with Wenshuo and others)

Approximate knockoff constructions (no theoretical guarantees):

- Second-order knockoffs (tend to work well with regression-based statistics) (Candès et al., 2018)
- Deep learning, including GANs (empirically valid in low-dimensions $n > p$) (Romano et al., 2018; Liu and Zheng, 2018; Anonymous, 2018)

Why all the Fuss?

Why all the Fuss?

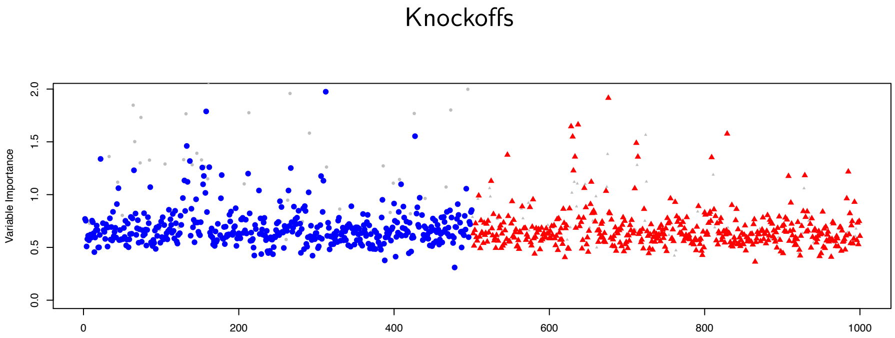


Figure: Variable importances for 500 variables and their knockoffs. Colored points are nulls, grey are non-nulls.

Why all the Fuss?

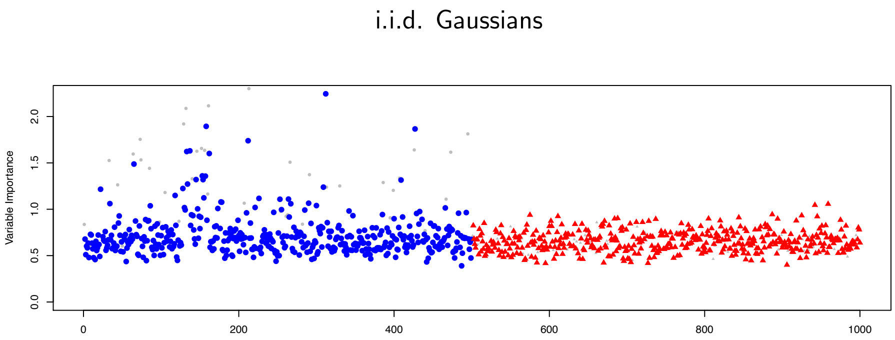


Figure: Variable importances for 500 variables and their knockoffs. Colored points are nulls, grey are non-nulls.

Why all the Fuss?

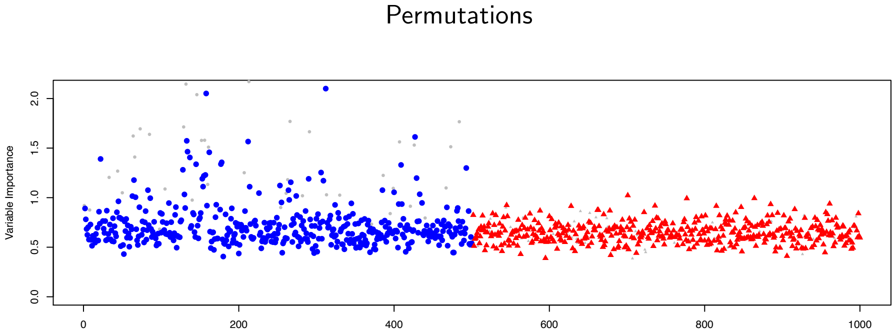


Figure: Variable importances for 500 variables and their knockoffs. Colored points are nulls, grey are non-nulls.

Robustness

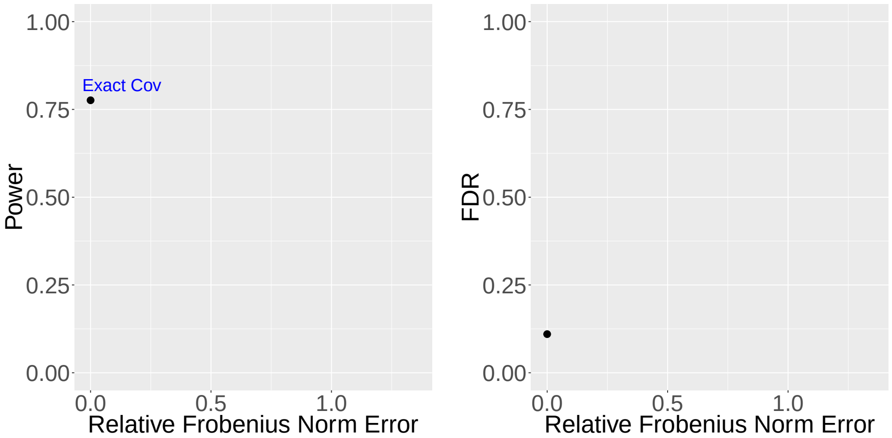


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

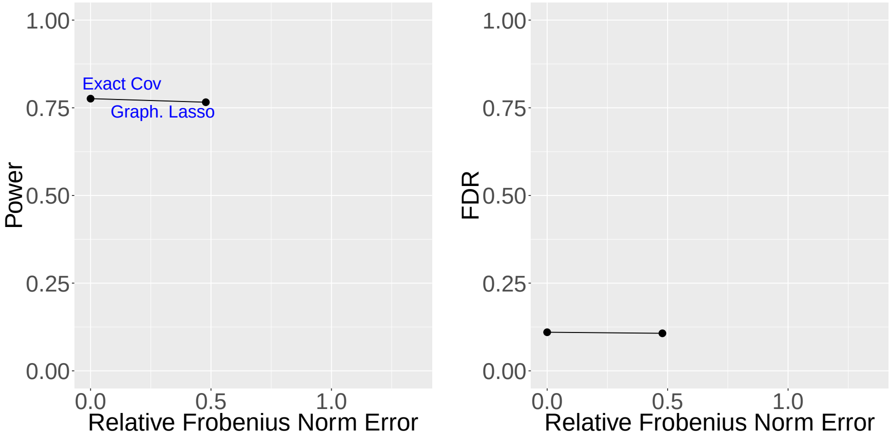


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

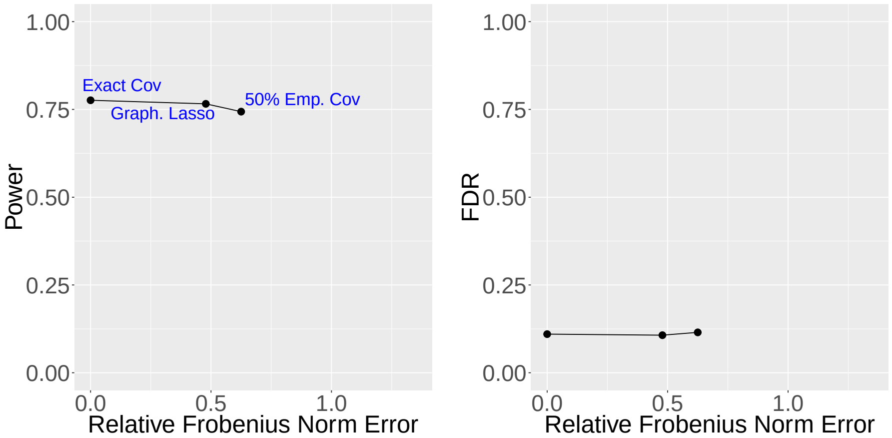


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

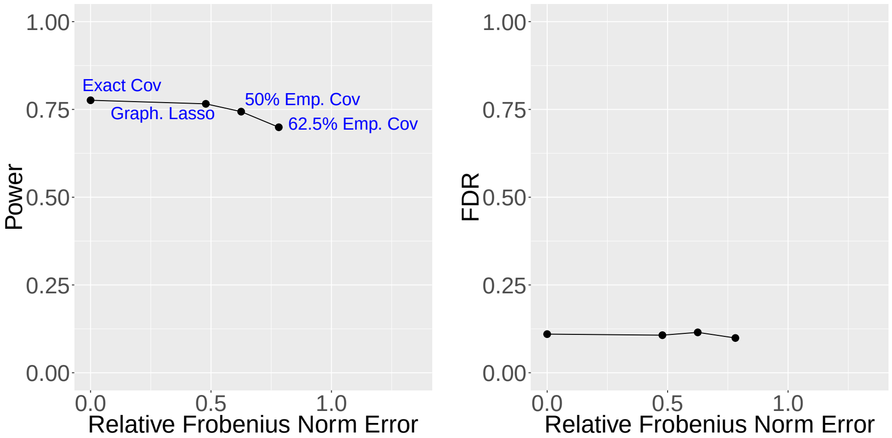


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

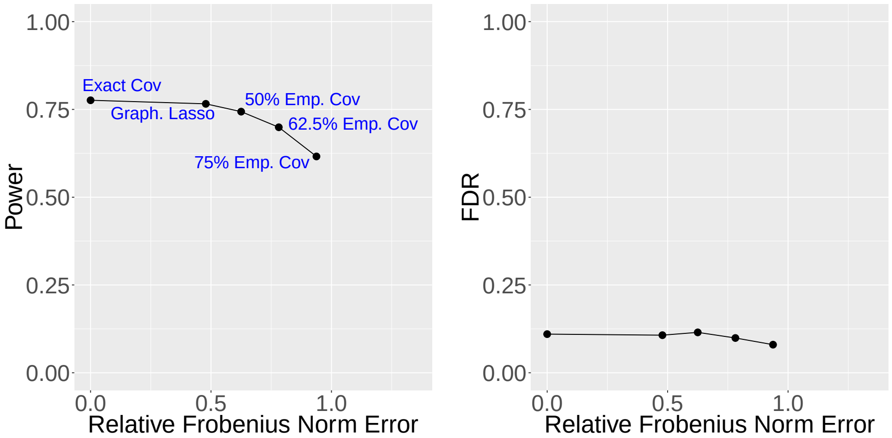


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

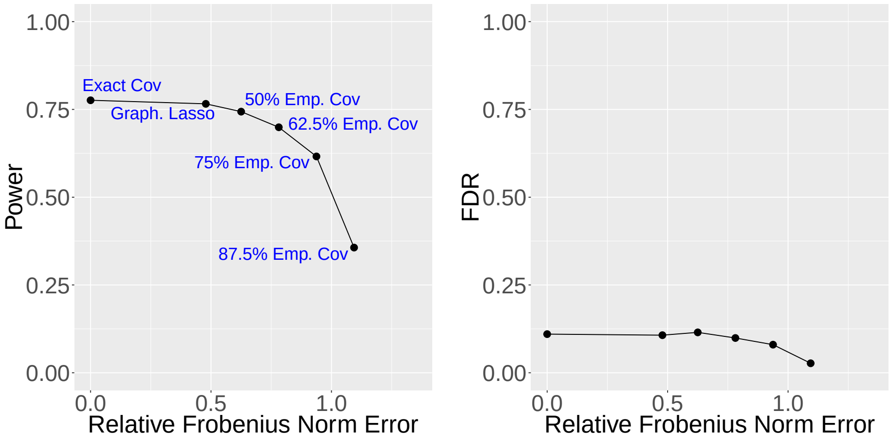


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Robustness

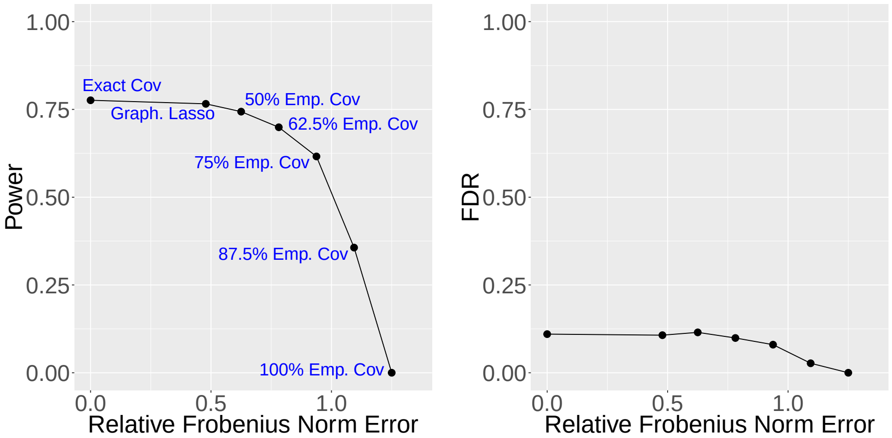


Figure: Covariates are **AR(1) with autocorrelation coefficient 0.3**. $n = 800$, $p = 1500$, and target FDR is 10%. Y comes from a binomial linear model with logit link function with 50 nonzero entries.

Step (2): Compute Knockoff Statistics

Ingredients for Knockoff Statistics

- Variable importance measures for all original and knockoff variables

$$Z_1, \ldots, Z_p, \widetilde{Z}_1, \ldots, \widetilde{Z}_p$$

Ingredients for Knockoff Statistics

- Variable importance measures for all original and knockoff variables

$$Z_1, \ldots, Z_p, \widetilde{Z}_1, \ldots, \widetilde{Z}_p$$

- Antisymmetric function $f_j : \mathbb{R}^2 \to \mathbb{R}$, i.e., $\quad f_j(a, b) = -f_j(b, a)$

Ingredients for Knockoff Statistics

- Variable importance measures for all original and knockoff variables

$$Z_1, \ldots, Z_p, \widetilde{Z}_1, \ldots, \widetilde{Z}_p$$

- Antisymmetric function $f_j : \mathbb{R}^2 \to \mathbb{R}$, i.e., $\quad f_j(a, b) = -f_j(b, a)$
- $W_j = f_j(Z_j, \widetilde{Z}_j)$

Ingredients for Knockoff Statistics

- Variable importance measures for all original and knockoff variables

$$Z_1, \ldots, Z_p, \widetilde{Z}_1, \ldots, \widetilde{Z}_p$$

- Antisymmetric function $f_j : \mathbb{R}^2 \to \mathbb{R}$, i.e., $\quad f_j(a, b) = -f_j(b, a)$
- $W_j = f_j(Z_j, \widetilde{Z}_j)$

Example 1:

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

Ingredients for Knockoff Statistics

- Variable importance measures for all original and knockoff variables

$$Z_1, \dots, Z_p, \widetilde{Z}_1, \dots, \widetilde{Z}_p$$

- Antisymmetric function $f_j : \mathbb{R}^2 \to \mathbb{R}$, i.e., $f_j(a, b) = -f_j(b, a)$
- $W_j = f_j(Z_j, \widetilde{Z}_j)$

Example 1:

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

$$W_j = |\beta_j| - |\tilde{\beta}_j| \qquad \text{(Lasso Coefficient Difference statistic)}$$

Ingredients for Knockoff Statistics

- Variable importance measures for all original and knockoff variables

$$Z_1, \ldots, Z_p, \widetilde{Z}_1, \ldots, \widetilde{Z}_p$$

- Antisymmetric function $f_j : \mathbb{R}^2 \to \mathbb{R}$, i.e., $\quad f_j(a,b) = -f_j(b,a)$
- $W_j = f_j(Z_j, \widetilde{Z}_j)$

Example 1:

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

$$W_j = |\beta_j| - |\tilde{\beta}_j| \qquad \text{(Lasso Coefficient Difference statistic)}$$

Example 2:

- Fit machine learning method (e.g., deep learning) to $\boldsymbol{y}$ with features $[\boldsymbol{X}\, \widetilde{\boldsymbol{X}}]$
- Z is increase in cross-validation error when variable is dropped; same f_j

$$(\mathsf{CV}_{\text{-}j} - \mathsf{CV}) - \left(\widetilde{\mathsf{CV}}_{\text{-}j} - \mathsf{CV}\right) = \mathsf{CV}_{\text{-}j} - \widetilde{\mathsf{CV}}_{\text{-}j}$$

Adaptivity and Prior Information in Z

Adaptivity

- Z can be **any** variable importance measure

Adaptivity and Prior Information in Z

Adaptivity

- Z can be **any** variable importance measure
- Higher-level adaptivity: CV to choose best-fitting model for inference

Adaptivity and Prior Information in Z

Adaptivity

- Z can be **any** variable importance measure
- 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 Z

Adaptivity

- Z can be **any** variable importance measure
- 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 Z

Adaptivity

- Z can be **any** variable importance measure
- 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 Z

Adaptivity

- Z can be **any** variable importance measure
- 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**

Why Does it Work?

Recall swap exchangeability property: for any j,

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

Why Does it Work?

Recall swap exchangeability property: for any j,

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

Coin-flipping property for W_j:

Why Does it Work?

Recall swap exchangeability property: for any j,

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

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

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

Why Does it Work?

Recall swap exchangeability property: for any j,

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

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

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

Why Does it Work?

Recall swap exchangeability property: for any j,

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

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

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

Why Does it Work?

Recall swap exchangeability property: for any j,

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

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

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

Why Does it Work?

Recall swap exchangeability property: for any j,

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

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

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

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

Why Does it Work?

Recall swap exchangeability property: for any j,

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

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

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

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

Why Does it Work?

Recall swap exchangeability property: for any j,

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

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

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

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

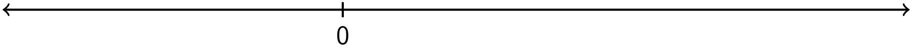
Step (3): Find the Knockoff Threshold

Simple Example

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

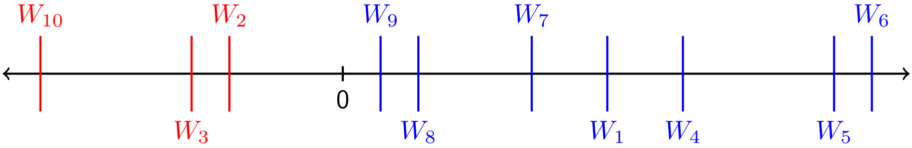
Simple Example

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



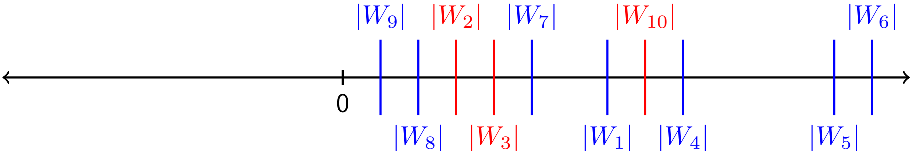
Simple Example

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



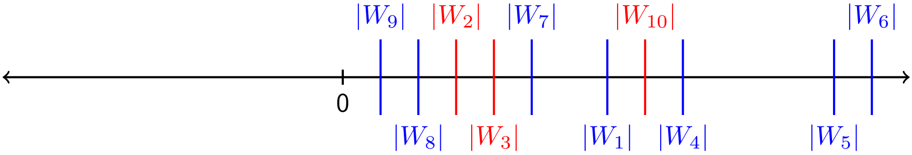
Simple Example

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



Simple Example

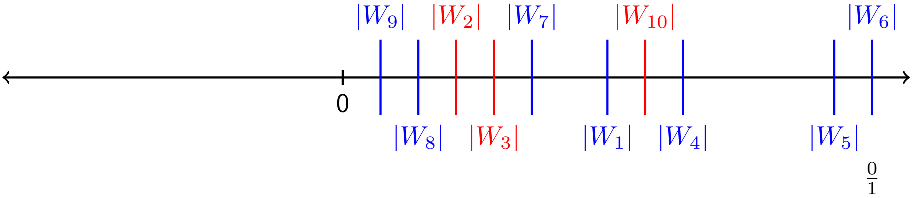
Example with $p = 10$ and $q = 20\% = 1/5$:



$$\widehat{\mathsf{FDR}} = \frac{\#\{\text{negative } W_j\}}{\#\{\text{positive } W_j\}}$$

Simple Example

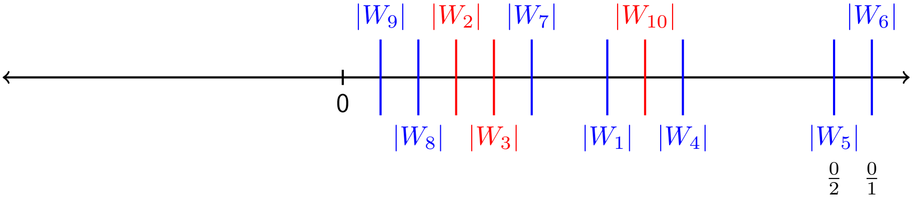
Example with $p = 10$ and $q = 20\% = 1/5$:



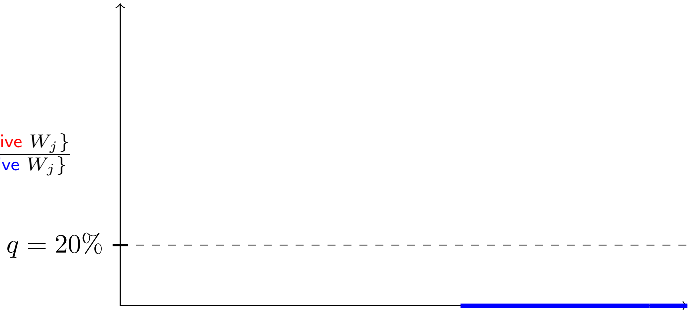
$$\widehat{\mathsf{FDR}} = \frac{\#\{\text{negative } W_j\}}{\#\{\text{positive } W_j\}}$$

Simple Example

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

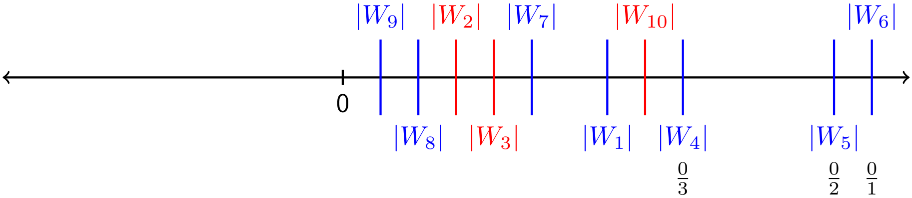


$$\widehat{\mathsf{FDR}} = \frac{\#\{\text{negative } W_j\}}{\#\{\text{positive } W_j\}}$$

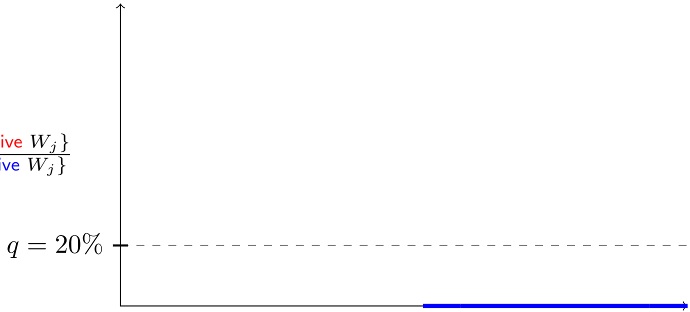


Simple Example

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

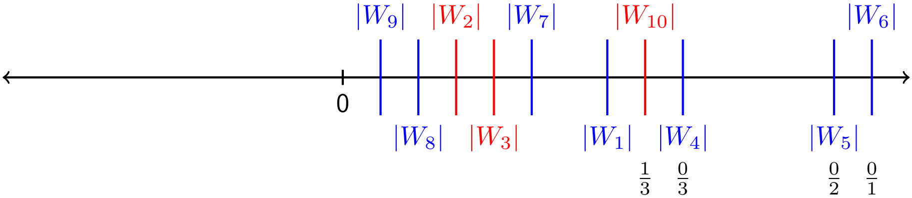


$$\widehat{\text{FDR}} = \frac{\#\{\text{negative } W_j\}}{\#\{\text{positive } W_j\}}$$

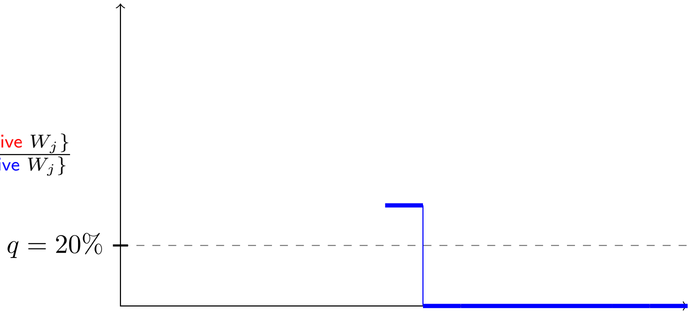


Simple Example

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

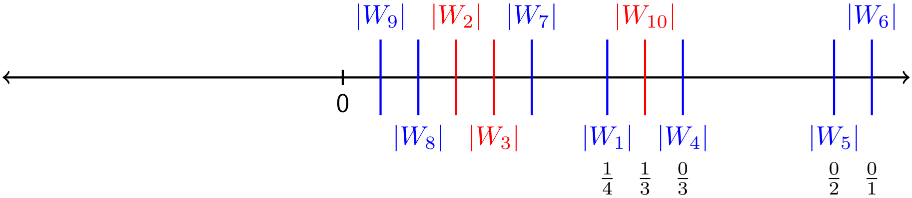


$$\widehat{\mathrm{FDR}} = \frac{\#\{\text{negative } W_j\}}{\#\{\text{positive } W_j\}}$$

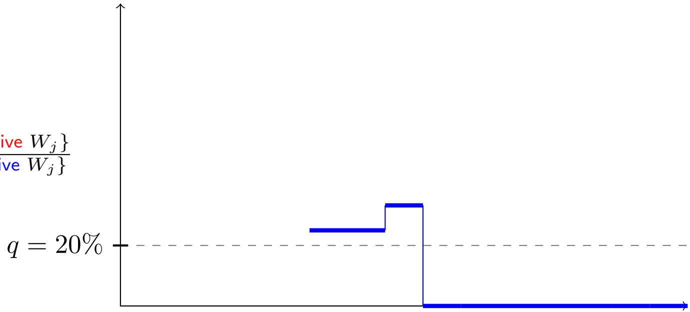


Simple Example

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

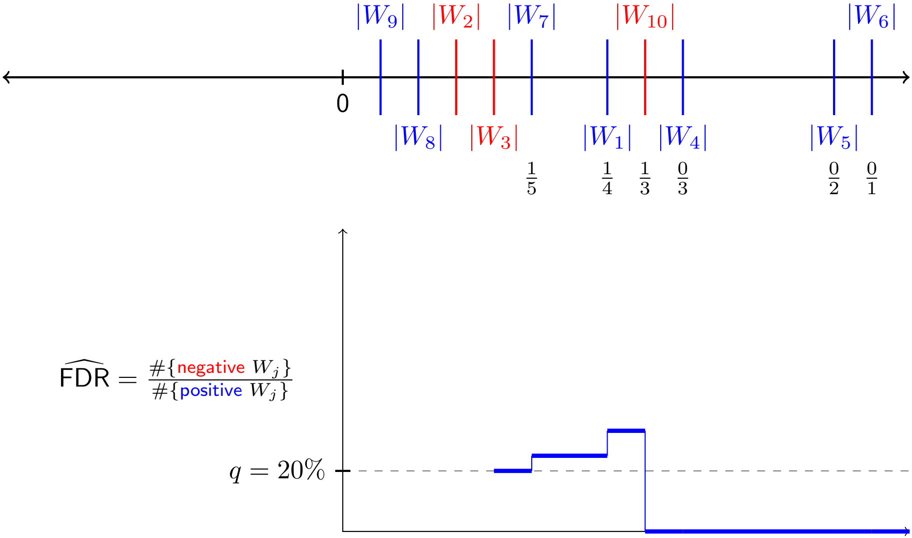


$$\widehat{\text{FDR}} = \frac{\#\{\text{negative } W_j\}}{\#\{\text{positive } W_j\}}$$



Simple Example

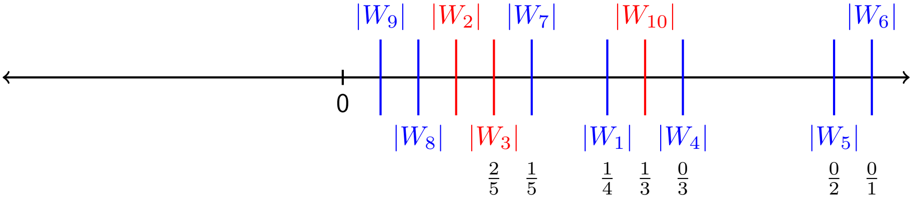
Example with $p = 10$ and $q = 20\% = 1/5$:



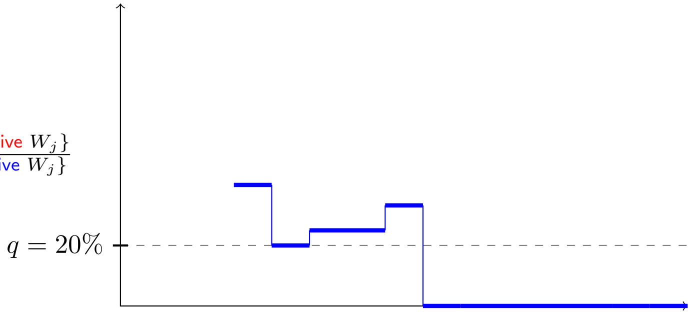
$$\widehat{\mathsf{FDR}} = \frac{\#\{\text{negative } W_j\}}{\#\{\text{positive } W_j\}}$$

Simple Example

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

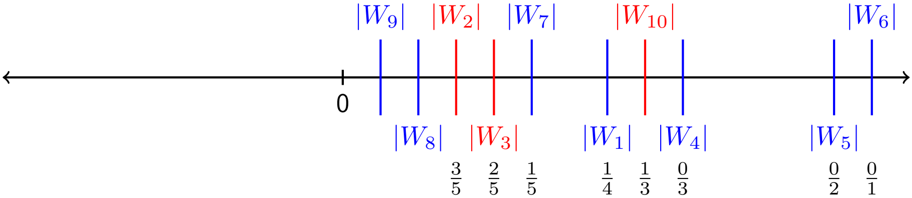


$$\widehat{\text{FDR}} = \frac{\#\{\text{negative } W_j\}}{\#\{\text{positive } W_j\}}$$

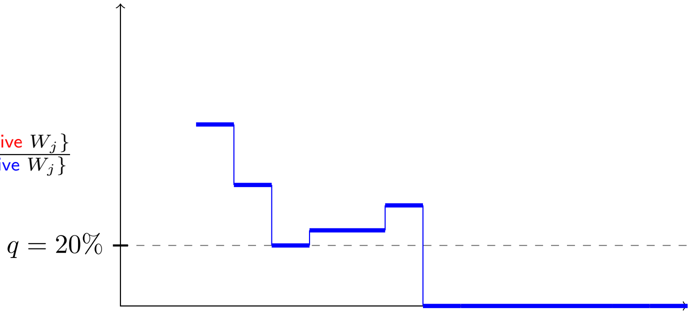


Simple Example

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

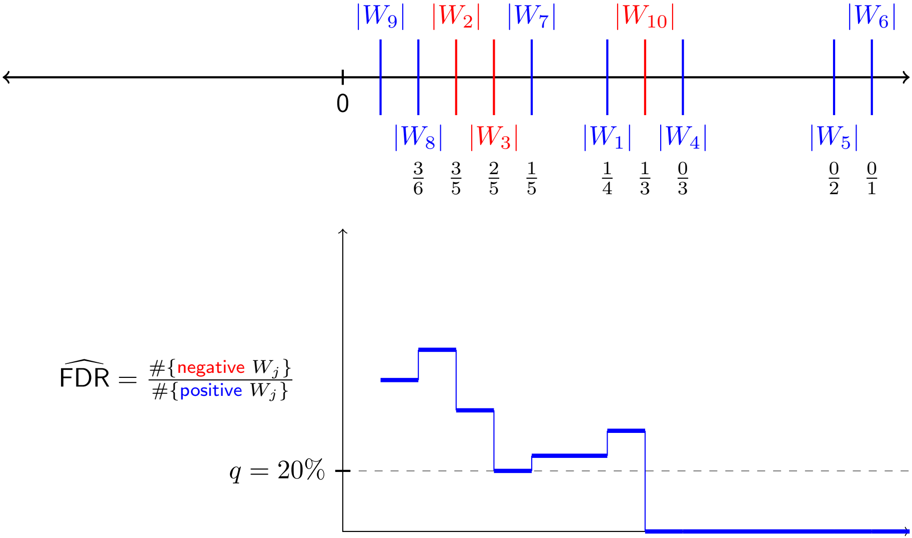


$$\widehat{\mathrm{FDR}} = \frac{\#\{\text{negative } W_j\}}{\#\{\text{positive } W_j\}}$$



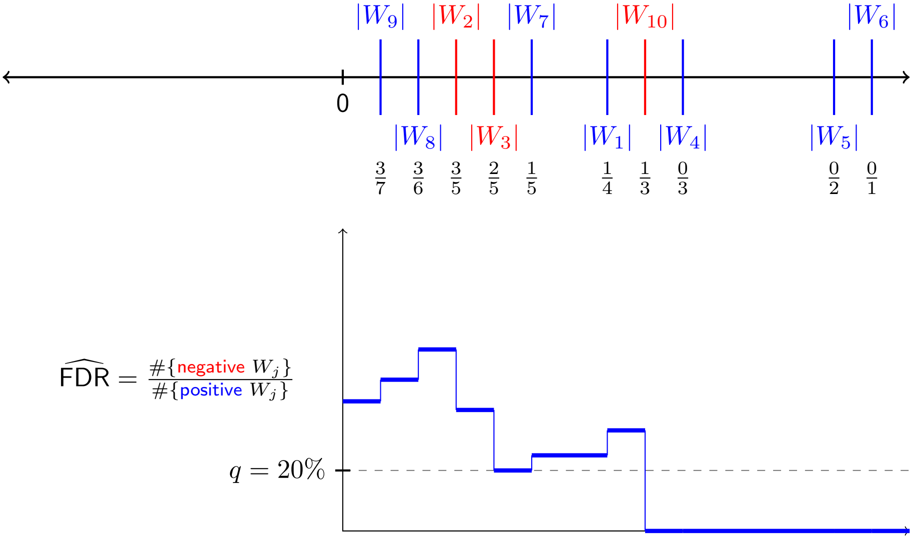
Simple Example

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



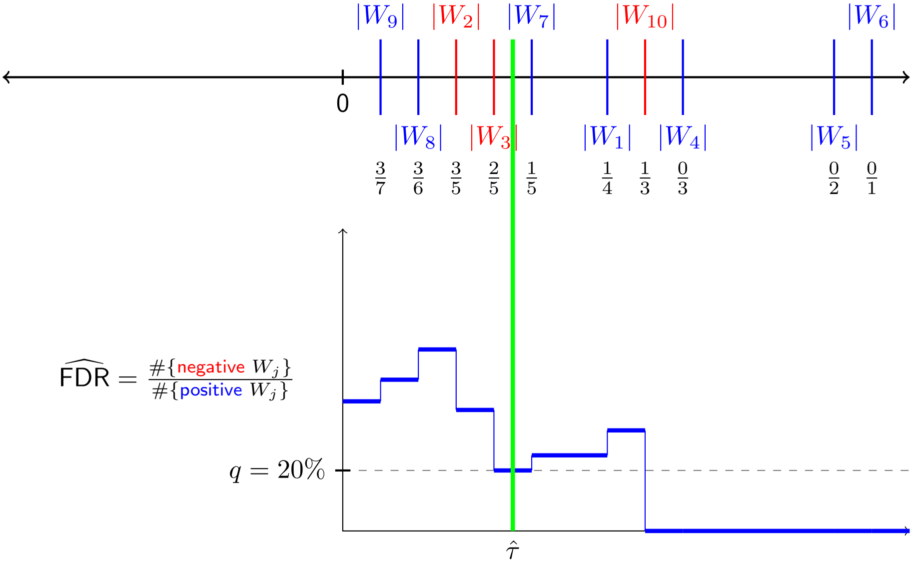
Simple Example

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



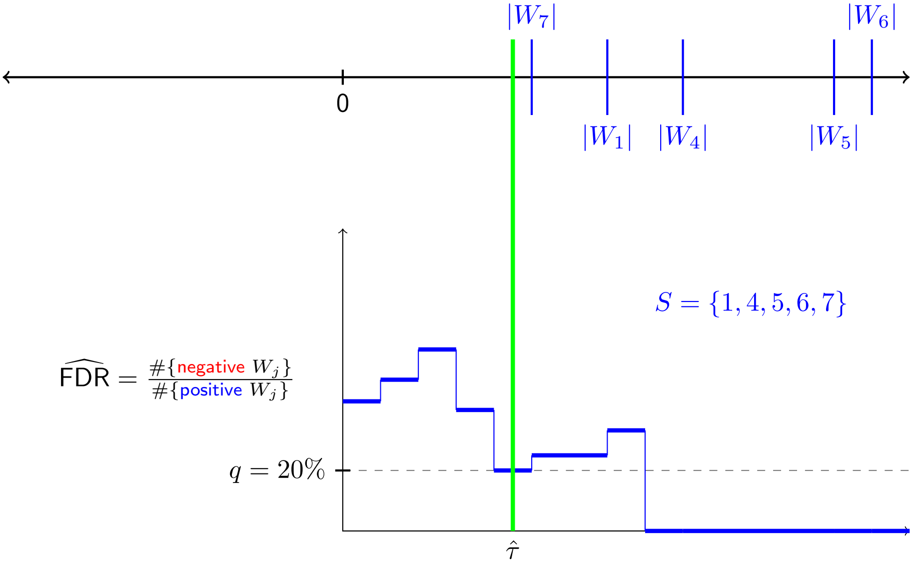
Simple Example

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



Simple Example

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



Tracking the FDR

$$\text{FDR} \;= \mathbb{E}\left[\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right]$$

Tracking the FDR

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

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

Tracking the FDR

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

Tracking the FDR

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

Tracking the FDR

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

Numerical Results

Simulations in Low-Dimensional Linear Model

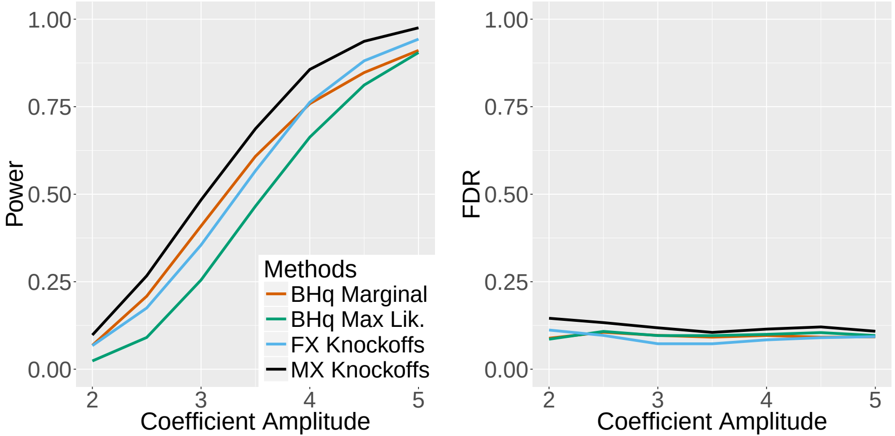


Figure: Power and FDR (target is 10%) for knockoffs and alternative procedures. The design matrix is i.i.d. $\mathcal{N}(0, 1/n)$, $n = 3000$, $p = 1000$, and y comes from a Gaussian linear model with 60 nonzero regression coefficients having equal magnitudes and random signs. The noise variance is 1.

Simulations in Low-Dimensional Nonlinear Model

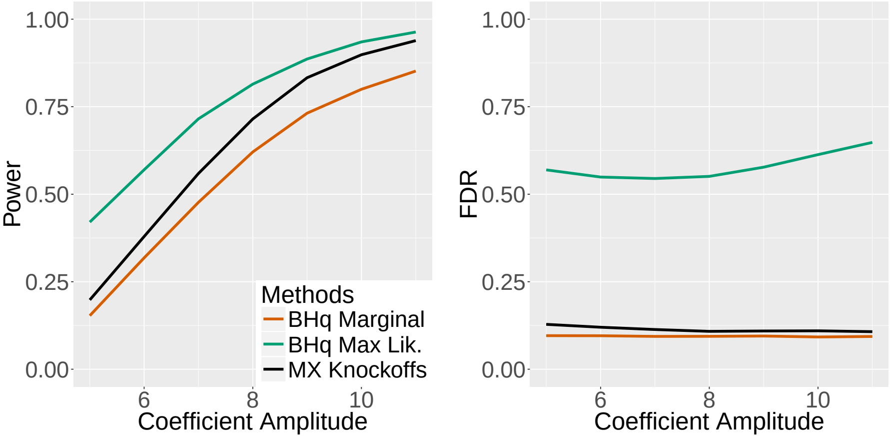


Figure: Power and FDR (target is 10%) for knockoffs and alternative procedures. The design matrix is i.i.d. $\mathcal{N}(0, 1/n)$, $n = 3000$, $p = 1000$, and y comes from a binomial linear model with logit link function, and 60 nonzero regression coefficients having equal magnitudes and random signs.

Simulations in High Dimensions

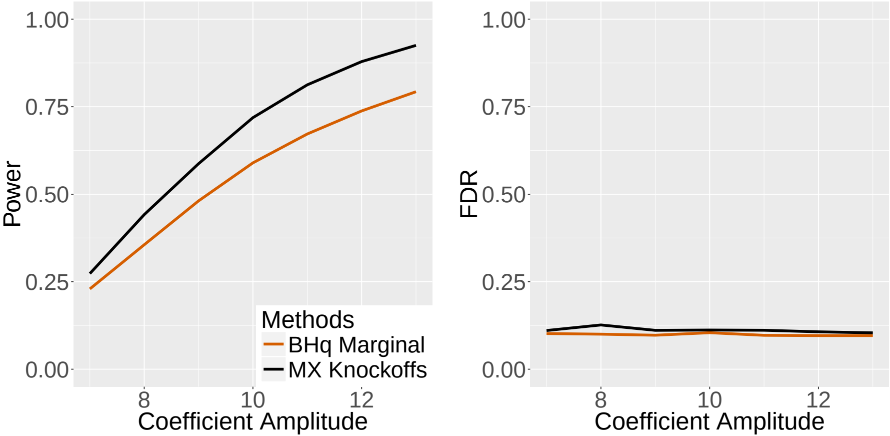


Figure: Power and FDR (target is 10%) for knockoffs and alternative procedures. The design matrix is i.i.d. $\mathcal{N}(0, 1/n)$, $n = 3000$, $p = 6000$, and y comes from a binomial linear model with logit link function, and 60 nonzero regression coefficients having equal magnitudes and random signs.

Simulations in High Dimensions with Dependence

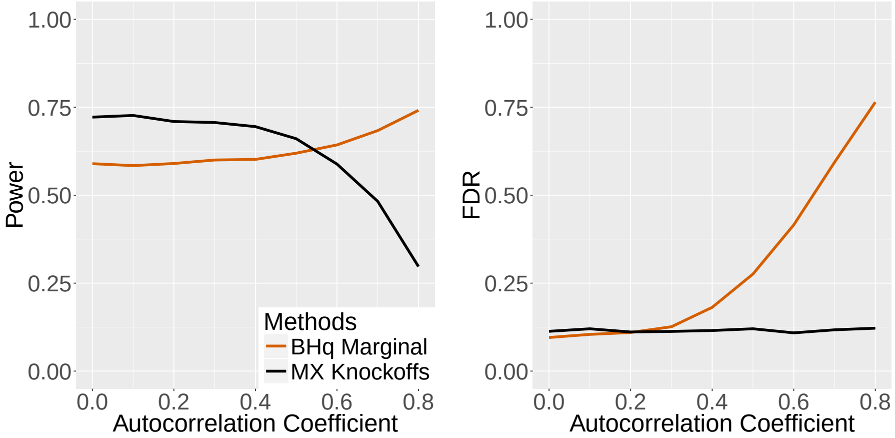


Figure: Power and FDR (target is 10%) for knockoffs and alternative procedures. The design matrix has AR(1) columns, and marginally each $X_j \sim \mathcal{N}(0, 1/n)$. $n = 3000$, $p = 6000$, and y follows a binomial linear model with logit link function, and 60 nonzero coefficients with random signs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

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 on 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 on 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 on 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 on 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 on 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 on 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 obtained with X model taken from **existing genomic imputation software** (Sesia et al., 2018)

Computation and Software

- R, Python, and Matlab packages available depending on knockoff construction; link on my website

Computation and Software

- R, Python, and Matlab packages available depending on knockoff construction; link on my website
- Knockoff construction algorithms generally scale linearly in p and n

Computation and Software

- R, Python, and Matlab packages available depending on knockoff construction; link on my website
- Knockoff construction algorithms generally scale linearly in p and n
- Given variable importances $Z_1, \ldots, Z_p, \widetilde{Z}_1, \ldots, \widetilde{Z}_p$, computation trivial

Computation and Software

- R, Python, and Matlab packages available depending on knockoff construction; link on my website
- Knockoff construction algorithms generally scale linearly in p and n
- Given variable importances $Z_1, \ldots, Z_p, \widetilde{Z}_1, \ldots, \widetilde{Z}_p$, computation trivial
- Need to compute $Z_1, \ldots, Z_p, \widetilde{Z}_1, \ldots, \widetilde{Z}_p$
    - Just compute variable importances for twice as many variables
    - Generally only constant slower than computing variable importances without knockoffs

Checking Sensitivity to Misspecification Error

	Concern about misspecification	
	$Y \mid X$	X
Canonical (fixed-X)	Yes	No
Model-X	No	Yes

Checking Sensitivity to Misspecification Error

	Concern about misspecification	
	$Y \mid X$	X
Canonical (fixed-X)	Yes	No
Model-X	No	Yes
Misspecification replicated in simulation?	No	Yes

Checking Sensitivity to Misspecification Error

	Concern about misspecification	
	$Y \mid X$	X
Canonical (fixed-X)	Yes	No
Model-X	No	Yes
Misspecification replicated in simulation?	No	Yes

Model-X: can actually **check sensitivity** to misspecification error!

Robustness on Real Data

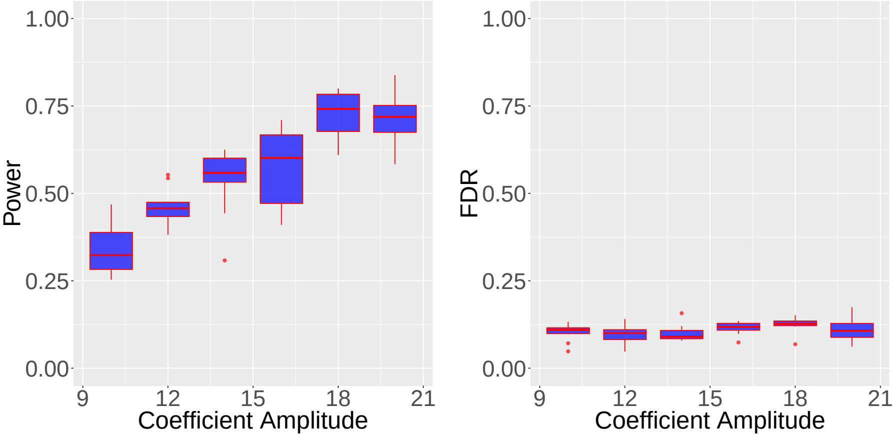


Figure: Power and FDR (target is 10%) for knockoffs applied to subsamples of a chromosome 1 of real genetic design matrix; $n \approx 1,400$.

Summary

The controlled variable selection problem is ubiquitous in modern science, and knockoffs is a powerful, flexible, and robust solution

Summary

The controlled variable selection problem is ubiquitous in modern science, and knockoffs is a powerful, flexible, and robust solution

My group is extremely interested in **making knockoffs work for you**. Please reach out if you think it could help with your research.

Summary

The controlled variable selection problem is ubiquitous in modern science, and knockoffs is a powerful, flexible, and robust solution

My group is extremely interested in **making knockoffs work for you**. Please reach out if you think it could help with your research.

Thank you!

`http://lucasjanson.fas.harvard.edu`
ljanson@fas.harvard.edu

Appendix

References

Anonymous (2018). KnockoffGAN: generating knockoffs for feature selection using generative adversarial networks. *Submitted to International Conference on Learning Representations*.

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. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):551–577.

Dai, R. and Barber, R. F. (2016). The knockoff filter for FDR control in group-sparse and multitask regression. In *Proceedings of the 33nd International Conference on Machine Learning (ICML 2016)*.

Gimenez, J. R., Ghorbani, A., and Zou, J. (2018). Knockoffs for the mass: new feature importance statistics with false discovery guarantees. *arXiv preprint arXiv:1807.06214*.

Liu, Y. and Zheng, C. (2018). Auto-encoding knockoff generator for fdr controlled variable selection. *arXiv preprint arXiv:1809.10765*.

Romano, Y., Sesia, M., and Candès, E. J. (2018). Deep knockoffs. *arXiv preprint arXiv:1811.06687*.

Sesia, M., Sabatti, C., and Candès, E. J. (2018). Gene hunting with hidden

Existing Methods: Low-Dimensional Linear Model

Suppose we assume that our data:

- follows a **linear model**:

$$Y = X_1\beta_1 + \cdots + X_p\beta_p + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2),$$

- has more observations that variables: $n \geq p$ (**low-dimensional**).

Existing Methods: Low-Dimensional Linear Model

Suppose we assume that our data:

- follows a **linear model**:

$$Y = X_1\beta_1 + \cdots + X_p\beta_p + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2),$$

- has more observations that variables: $n \geq p$ (**low-dimensional**).

Classical problem:

- Ordinary least squares (OLS) theory gives exact p-values for testing whether each $\beta_j = 0$ or not (under very mild assumptions, $\beta_j = 0 \;\Leftrightarrow\; Y \perp\!\!\!\perp X_j \,|\, X_{\text{-}j}$)
- The Benjamini-Hochberg procedure (BHq) applied to the p-values will essentially control the FDR

Existing Methods: Low-Dimensional Linear Model

Suppose we assume that our data:

- follows a **linear model**:

$$Y = X_1\beta_1 + \cdots + X_p\beta_p + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2),$$

- has more observations that variables: $n \geq p$ (**low-dimensional**).

Classical problem:

- Ordinary least squares (OLS) theory gives exact p-values for testing whether each $\beta_j = 0$ or not (under very mild assumptions, $\beta_j = 0 \;\Leftrightarrow\; Y \perp\!\!\!\perp X_j \,|\, X_{\text{-}j}$)
- The Benjamini-Hochberg procedure (BHq) applied to the p-values will essentially control the FDR

Minor caveats:

- FDR control not exact (but good enough in practice)
- Sparsity not used (reduces power to find important variables)

Nonlinearity and High Dimensions

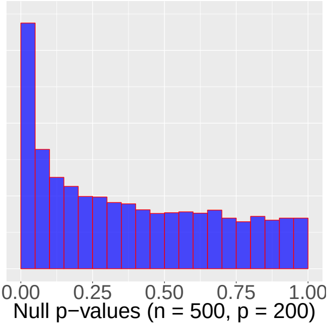
Low-dimensional ($n \geq p$) generalized linear model

- Apply BHq to asymptotic p-values

Nonlinearity and High Dimensions

Low-dimensional ($n \geq p$) generalized linear model

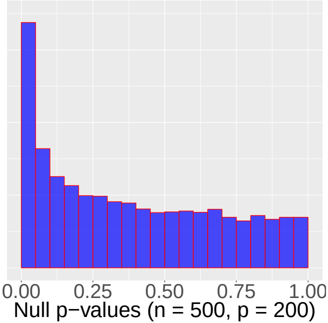
- Apply BHq to asymptotic p-values
- Can be far from valid in practice



Nonlinearity and High Dimensions

Low-dimensional ($n \geq p$) generalized linear model

- Apply BHq to asymptotic p-values
- Can be far from valid in practice



High-dimensional ($n < p$) generalized linear models

- Apply BHq to p-values from
  - Debiased lasso, e.g., Zhang and Zhang (2014), Javanmard and Montanari (2014), van de Geer et al. (2014), Cai and Guo (2015)
  - Causal inference, e.g., Belloni et al. (2014), Athey et al. (2016), Farrell (2015)
  - Inference after selection, e.g., Berk et al. (2013), Lee et al. (2016), Fithian et al. (2014)
- Asymptotic, require sparsity and random design assumptions

Sequential Independent Pairs Generates Valid Knockoffs

Algorithm 2 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 2 Sequential Conditional Independent Pairs

for $j = \{1, \dots, p\}$ **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 2 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

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

Sequential Independent Pairs Generates Valid Knockoffs

Algorithm 2 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 2 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 2 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

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

Sequential Independent Pairs Generates Valid Knockoffs

Algorithm 2 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

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

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

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

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

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

Computation of Second-Order Knockoffs

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

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

- **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):

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

Computation of Second-Order Knockoffs

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

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

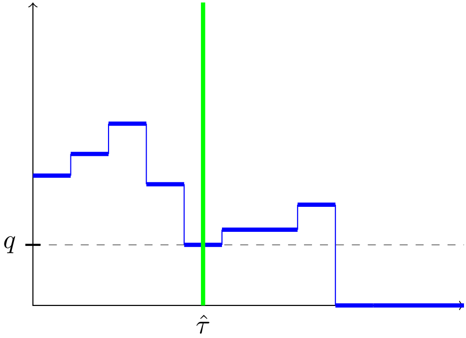
- **Equicorrelated (EQ)** (fast, less powerful): $s_j^{\mathsf{EQ}} = 2\lambda_{\mathsf{min}}(\mathbf{\Sigma}) \wedge 1$ for all j
- **Semidefinite program (SDP)** (slower, more powerful):

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

- **(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

Proof of Control

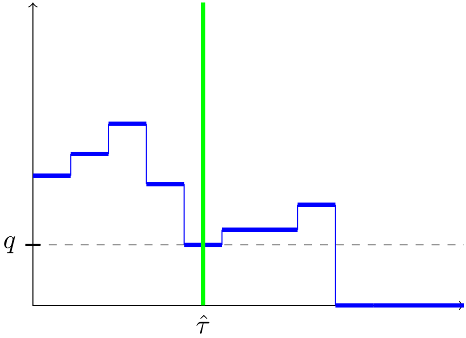
$$\text{FDR} = \mathbb{E}\left[\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right]$$



Proof of Control

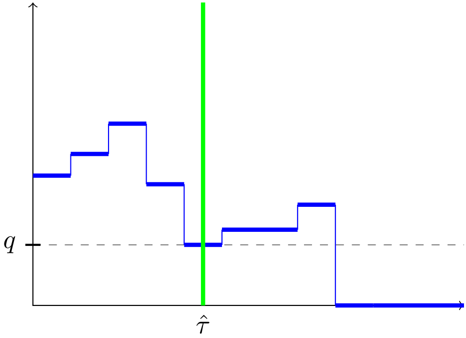
$$\text{FDR} = \mathbb{E}\left[\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{\#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right]$$

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



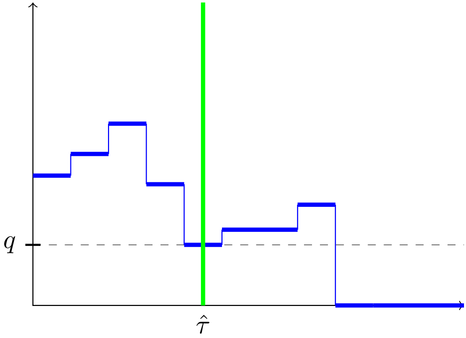
Proof of Control

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



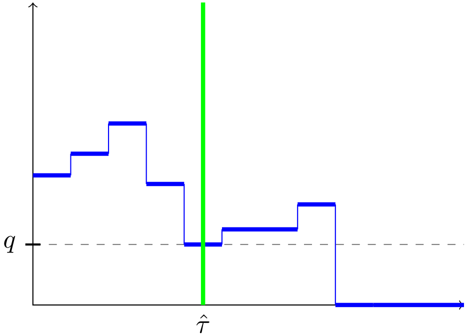
Proof of Control

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



Proof of Control

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

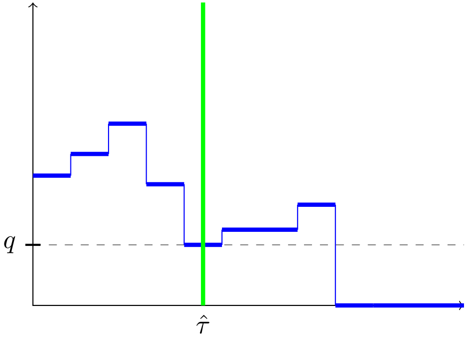


More precisely:

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

Proof of Control

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

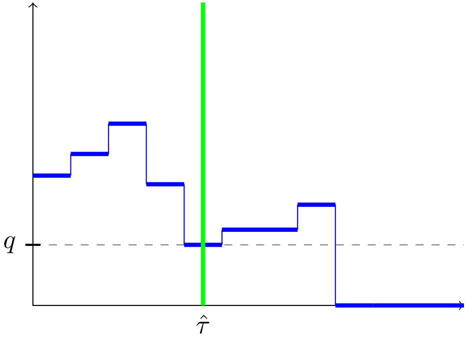


More precisely:

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

Proof of Control

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

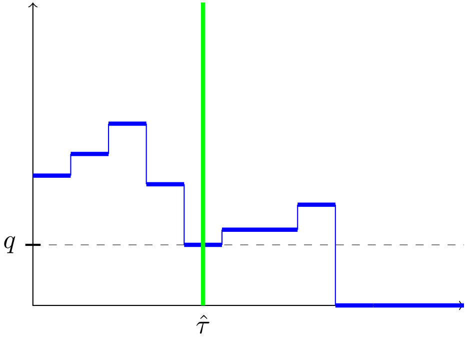


More precisely:

$$\begin{aligned}
\text{mFDR} &= \mathbb{E}\left[\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{q^{-1} + \#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right] = \mathbb{E}\left[\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{q^{-1} + \#\{\text{positive } |W_j| > \hat{\tau}\}}\right] \\
&= \mathbb{E}\left(\frac{\#\{\text{null positive } |W_j| > \hat{\tau}\}}{1 + \#\{\text{null negative } |W_j| > \hat{\tau}\}} \cdot \underbrace{\frac{1 + \#\{\text{null negative } |W_j| > \hat{\tau}\}}{q^{-1} + \#\{\text{positive} |W_j| > \hat{\tau}\}}}_{\leq q \text{ by definition of } \hat{\tau}}\right)
\end{aligned}$$

Proof of Control

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



More precisely:

$$\begin{aligned}
\text{mFDR} &= \mathbb{E}\left[\frac{\#\{\text{null } \boldsymbol{X}_j \text{ selected}\}}{q^{-1} + \#\{\text{total } \boldsymbol{X}_j \text{ selected}\}}\right] = \mathbb{E}\left[\frac{\#\{\text{null } {\color{blue}\text{positive}}\ |W_j| > \hat{\tau}\}}{q^{-1} + \#\{{\color{blue}\text{positive}}\ |W_j| > \hat{\tau}\}}\right] \\
&= \mathbb{E}\left(\underbrace{\frac{\#\{\text{null } {\color{blue}\text{positive}}\ |W_j| > \hat{\tau}\}}{1 + \#\{\text{null } {\color{red}\text{negative}}\ |W_j| > \hat{\tau}\}}}_{\substack{\text{Supermartingale } \leq 1 \\ \text{with } \hat{\tau} \text{ a stopping time}}} \cdot \underbrace{\frac{1 + \#\{\text{null } {\color{red}\text{negative}}\ |W_j| > \hat{\tau}\}}{q^{-1} + \#\{{\color{blue}\text{positive}}|W_j| > \hat{\tau}\}}}_{\leq q \text{ by definition of } \hat{\tau}}\right)
\end{aligned}$$