Panning for Gold:

Model-X Knockoffs for High Dimensional Controlled Variable Selection

Candès, Emmanuel J and Fan, Jianqing and Janson, Lucas and Lv, Jinchi

Presented by: Sai Zhang

April 3, 2023

Outline

1 Introduction

2 Problem and Methodology

3 Numerical Simulation

4 Discussion

Introduction

Inspiration: Variable Selection

Considering the problem of estimating

$$F_{\mathbf{Y}|\mathbf{X}}$$

where the *outcome* $\mathbf{Y}$ is determined by the set of p potential determining *covariates* $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_p)$

Inspiration: Variable Selection

Considering the problem of estimating

$$F_{\mathbf{Y}|\mathbf{X}}$$

where the *outcome* $\mathbf{Y}$ is determined by the set of p potential determining *covariates* $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_p)$

How to achieve **sparsity**?

This Paper: The Problem

$$(X_{i,1}, \cdots, X_{i,p}, Y_i) \overset{\text{i.i.d.}}{\sim} F_{\mathbf{XY}} \qquad i = 1, \cdots, n$$

Main assumptions:

- **NO** *knowledge* of the conditional distribution $\mathbf{Y} \mid \mathbf{X}_1, \cdots, \mathbf{X}_p$

This Paper: The Problem

$$(X_{i,1}, \cdots, X_{i,p}, Y_i) \overset{\text{i.i.d.}}{\sim} F_{\mathbf{XY}} \qquad i = 1, \cdots, n$$

Main assumptions:

- **NO** *knowledge* of the conditional distribution $\mathbf{Y} \mid \mathbf{X}_1, \cdots, \mathbf{X}_p$
- the joint distribution of the covariates $F_{\mathbf{X}}$ is **known**

This Paper: The Problem

$$(X_{i,1}, \cdots, X_{i,p}, Y_i) \overset{\text{i.i.d.}}{\sim} F_{\mathbf{XY}} \qquad\qquad i = 1, \cdots, n$$

Main assumptions:

- **NO** *knowledge* of the conditional distribution $\mathbf{Y} \mid \mathbf{X}_1, \cdots, \mathbf{X}_p$
- the joint distribution of the covariates $F_{\mathbf{X}}$ is **known**

Sparsity: the *smallest* subset $\mathcal{S} \subset \{1, \cdots, p\}$ s.t. conditional on $\{\mathbf{X}_j\}_{j\in\mathcal{S}}$, $\mathbf{Y}$ is independent of all *other* variables

This Paper: The Problem

$$(X_{i,1}, \cdots, X_{i,p}, Y_i) \overset{\text{i.i.d.}}{\sim} F_{\mathbf{XY}} \qquad\qquad i = 1, \cdots, n$$

Main assumptions:

- **NO** *knowledge* of the conditional distribution $\mathbf{Y} \mid \mathbf{X}_1, \cdots, \mathbf{X}_p$
- the joint distribution of the covariates $F_{\mathbf{X}}$ is **known**

Sparsity: the *smallest* subset $\mathcal{S} \subset \{1, \cdots, p\}$ s.t. conditional on $\{\mathbf{X}_j\}_{j\in\mathcal{S}}$, $\mathbf{Y}$ is independent of all *other* variables

controlling the **type-I error**: *false discovery rate (FDR)* (Benjamini and Hochberg, 1995)

$$FDR := \mathbb{E}[FDP] = \mathbb{E}\left[\frac{\#\left\{j : j \in \hat{\mathcal{S}} \setminus \mathcal{S}\right\}}{\#\left\{j : j \in \hat{\mathcal{S}}\right\}}\right]$$

This Paper: Framework

Following the **knockoff** procedure (Barber and Candes, 2015)

- construct *knockoff* variables
  - **NOT** associated with the response conditional on the original covariates
  - **MIRRORING** the structure of the original covariates

This Paper: Framework

Following the **knockoff** procedure (Barber and Candes, 2015)

- construct *knockoff* variables
  - **NOT** associated with the response conditional on the original covariates
  - **MIRRORING** the structure of the original covariates
- use the knockoff variables as *controls* for the real covariates
  - select real covariates considerably **MORE associated** with the response than their knockoff counterparts

This Paper: Framework

Following the **knockoff** procedure (Barber and Candes, 2015)

- construct *knockoff* variables
  - **NOT** associated with the response conditional on the original covariates
  - **MIRRORING** the structure of the original covariates
- use the knockoff variables as *controls* for the real covariates
  - select real covariates considerably **MORE associated** with the response than their knockoff counterparts

Extending it to $p > n$ and non-linear models

adding the *model-X* perspective: treating X_{ij} as **random**

Contributions

The first paper focused on the case of Gaussian linear model with $p \leq n$ and fixed design matrix $\mathbf{X}$

Contributions

The first paper focused on the case of Gaussian linear model with $p \leq n$ and fixed design matrix $\mathbf{X}$

- **set-up**
  assuming full knowledge about the *covariate distribution*
  instead of assuming a *pamametric model* for the response conditional on the covariates

Contributions

The first paper focused on the case of Gaussian linear model with $p \leq n$ and fixed design matrix $\mathbf{X}$

- **set-up**
  assuming full knowledge about the *covariate distribution*
  instead of assuming a *pamametric model* for the response conditional on the covariates
- **advantages**
  - extend the knockoff framework to *high-dimension* settings
  - can accomodate *any* model

Contributions

The first paper focused on the case of Gaussian linear model with $p \leq n$ and fixed design matrix $\mathbf{X}$

- **set-up**
  assuming full knowledge about the *covariate distribution*
  instead of assuming a *pamametric model* for the response conditional on the covariates
- **advantages**
  - extend the knockoff framework to *high-dimension* settings
  - can accomodate *any* model
  - *Selection with inference*: obtain *valid p-values* while rigorously controlling finite sample type-I error, instead of
    - × high-dimension $n < p$: rely on *strong sparsity* and *parametric* assumptions
    - × marginal testing: testing *unconditional* independence between $\mathbf{Y}$ and $\mathbf{X}_j$

Problem and Methodology

Problem: Variable Selection

For n IID sample

$$(X_{i,1}, \cdots, X_{i,p}, Y_i) \overset{\text{i.i.d.}}{\sim} F_{\mathbf{XY}} \qquad i = 1, \cdots, n$$

The problem is find the *smallest* subset $\mathcal{S}$, s.t., conditionally on $\{\mathbf{X}_j\}_{j\in\mathcal{S}}$, $\mathbf{Y}$ is independent of *all other* variables.

Problem: Variable Selection

For n IID sample

$$(X_{i,1}, \cdots, X_{i,p}, Y_i) \overset{\text{i.i.d.}}{\sim} F_{\mathbf{XY}} \qquad\qquad i = 1, \cdots, n$$

The problem is find the *smallest* subset $\mathcal{S}$, s.t., conditionally on $\{\mathbf{X}_j\}_{j\in\mathcal{S}}$, $\mathbf{Y}$ is independent of *all other* variables.

Definition 1

$\mathbf{X}_j$ is *null* iff $\mathbf{Y}$ is independent of $\mathbf{X}_j$ conditional on the other variables $\mathbf{X}_{-j} = \{\mathbf{X}_1, \cdots, \mathbf{X}_p\} \setminus \{\mathbf{X}_j\}$. The subset of of all null variables is denoted by $\mathcal{H}_0 \subset \{1, \cdots, p\}$. Variable $\mathbf{X}_j$ is *non-null* or *relevant*, if $j \in \mathcal{H}_0$

Problem: Controlling FDR

For a selection rule the selects a subset $\hat{\mathcal{S}}$ of covariates, we have

$$\text{FDR} := \mathbb{E}\left[\frac{|\hat{\mathcal{S}} \cap \mathcal{H}_0|}{|\hat{\mathcal{S}}|}\right]$$

Problem: Controlling FDR

For a selection rule the selects a subset $\hat{\mathcal{S}}$ of covariates, we have

$$\mathrm{FDR} := \mathbb{E}\left[\frac{|\hat{\mathcal{S}} \cap \mathcal{H}_0|}{|\hat{\mathcal{S}}|} = \frac{\#\left\{j : j \in \hat{\mathcal{S}} \setminus \mathcal{S}\right\}}{\#\left\{j : j \in \hat{\mathcal{S}}\right\}}\right]$$

Problem: Controlling FDR

For a selection rule the selects a subset $\hat{\mathcal{S}}$ of covariates, we have

$$\mathrm{FDR} := \mathbb{E}\left[\frac{|\hat{\mathcal{S}} \cap \mathcal{H}_0|}{|\hat{\mathcal{S}}|} = \frac{\#\left\{j : j \in \hat{\mathcal{S}} \setminus \mathcal{S}\right\}}{\#\left\{j : j \in \hat{\mathcal{S}}\right\}}\right]$$

Proposition 1 (GLM)

For a family of random variables $\mathbf{X}_1, \cdots, \mathbf{X}_p$ s.t. one *cannot* perfectly predict any of them from knowledge of the others. If the likelihood of $\mathbf{Y}$ follows a GLM, then

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{X}_j \mid \mathbf{X}_{-j} \Leftrightarrow \beta_j = 0$$

hence, $\mathcal{H}_0$ from Def. 1 is exactly the set $\{j : \beta_j = 0\}$. proof

Methodology: Model-X Knockoffs

Definition 2: MX Knockoffs

MX knockoffs for the family of random variables $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_p)$ are a new family of random variables $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_p)$ constructed with the following 2 properties:

(a) for any subset $S \subset \{1, \cdots, p\}$,

$$(\mathbf{X}, \tilde{\mathbf{X}})_{\mathrm{swap}(S)} \stackrel{\mathrm{d}}{=} (\mathbf{X}, \tilde{\mathbf{X}})$$

(b) $\tilde{\mathbf{X}} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}$ if there is a response $\mathbf{Y}$

Methodology: Model-X Knockoffs

Definition 2: MX Knockoffs

MX knockoffs for the family of random variables $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_p)$ are a new family of random variables $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_p)$ constructed with the following 2 properties:

(a) for any subset $S \subset \{1, \cdots, p\}$,

$$(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap}(S)} \overset{\text{d}}{=} (\mathbf{X}, \tilde{\mathbf{X}})$$

(b) $\tilde{\mathbf{X}} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}$ if there is a response $\mathbf{Y}$

$(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap}(S)}$ is obtained from $(\mathbf{X}, \tilde{\mathbf{X}})$ by swapping the entries $\mathbf{X}_j$ and $\tilde{\mathbf{X}}_j$ for each $j \in S$. Example:

$$\left(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{X}}_3\right)_{\text{swap}(\{2,3\})} \overset{\text{d}}{=} \left(\mathbf{X}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{X}}_3, \tilde{\mathbf{X}}_1, \mathbf{X}_2, \mathbf{X}_3\right)$$

Methodology: Model-X Knockoffs

Suppose $\mathbf{X} \sim \mathcal{N}(0, \mathbf{\Sigma})$, then $(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap}(S)}$ satisfies $(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap}(S)} \overset{d}{=} (\mathbf{X}, \tilde{\mathbf{X}})$ if

$$(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap}(S)} \overset{d}{=} (\mathbf{X}, \tilde{\mathbf{X}}) \sim \mathcal{N}(0, \mathbf{G}), \quad \text{where } \mathbf{G} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{\Sigma} - \text{diag}(s) \\ \mathbf{\Sigma} - \text{diag}(s) & \mathbf{\Sigma} \end{pmatrix}$$

where $\text{diag}(s)$ is any *diagonal matrix* s.t. $\mathbf{G}$ is *positive semidefinite*.

For $\mathbf{P}$, the permutation matrix encoding the swap,

$$\mathbf{PGP} = \mathbf{G}$$

Constructing Knockoff Variable

$$(\mathbf{X}, \tilde{\mathbf{X}})_{\mathrm{swap}(S)} \stackrel{d}{=} (\mathbf{X}, \tilde{\mathbf{X}}) \sim \mathcal{N}(0, \mathbf{G}), \quad \text{where } \mathbf{G} = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma} - \mathrm{diag}(s) \\ \boldsymbol{\Sigma} - \mathrm{diag}(s) & \boldsymbol{\Sigma} \end{pmatrix}$$

then we can sample the knockoff vector $\tilde{\mathbf{X}}$ from the conditional distribution

$$\tilde{\mathbf{X}} \mid \mathbf{X} \stackrel{d}{=} \mathcal{N}(\mu, \mathbf{V})$$

where

$$\mu = \mathbf{X} - \mathbf{X}\boldsymbol{\Sigma}^{-1}\mathrm{diag}(s)$$
$$\mathbf{V} = 2\mathrm{diag}(s) - \mathrm{diag}(s)\boldsymbol{\Sigma}^{-1}\mathrm{diag}(s)$$

Exchangeability of Null Covariates and Their Knockoffs

For *MX knockoffs*, swapping *null* covariates with their knockoffs would **not** change the joint distribution of the original covariate $\mathbf{X}$ and their knockoffs $\tilde{\mathbf{X}}$, conditional on the repsonse $\mathbf{Y}$:

Lemma 1: MX Knockoffs

Take any subset $S \subset \mathcal{H}_0$ of nulls, then

$$\left(\mathbf{X}, \tilde{\mathbf{X}}\right) \mid \mathbf{y} \overset{\mathrm{d}}{=} \left(\mathbf{X}, \tilde{\mathbf{X}}\right)_{\mathrm{swap}(S)} \mid \mathbf{y}$$

Feature Statistics

To provide evidence *against* the hypothesis that $\mathbf{X}_j$ is null, we compute statistics W_j for each $j \in \{1, \cdots, p\}$

$$W_j = w_j \left\{ (\mathbf{X}, \tilde{\mathbf{X}}), \mathbf{y} \right\}$$

and check whether W_j is large enough.

Feature Statistics

To provide evidence *against* the hypothesis that $\mathbf{X}_j$ is null, we compute statistics W_j for each $j \in \{1, \cdots, p\}$

$$W_j = w_j \left\{ (\mathbf{X}, \tilde{\mathbf{X}}), \mathbf{y} \right\}$$

and check whether W_j is large enough.

Following Barber and Candes (2015), impose a *flip sign property*: swapping the jth variable with its knockoff has the effect of changing the sign of W_j

$$w_j \left\{ (\mathbf{X}, \tilde{\mathbf{X}})_{\mathrm{swap}(S)}, \mathbf{y} \right\} = \begin{cases} w_j \left\{ (\mathbf{X}, \tilde{\mathbf{X}}), \mathbf{y} \right\}, & j \notin S \\ -w_j \left\{ (\mathbf{X}, \tilde{\mathbf{X}}), \mathbf{y} \right\}, & j \in S \end{cases}$$

Feature Statistics

Consider a statistic $\mathbf{T}$ for each original and knockoff variable

$$\mathbf{T} \triangleq (\mathbf{Z}, \tilde{\mathbf{Z}}) = (Z_1, \cdots, Z_p, \tilde{Z}_1, \cdots, \tilde{Z}_p) = t\left\{(\mathbf{X}, \tilde{\mathbf{X}}), \mathbf{y}\right\}$$

if the components of $\mathbf{T}$ are switched in the same way:

$$(\mathbf{Z}, \tilde{\mathbf{Z}})_{\text{swap}(S)} = t\left\{(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap}(S)}, \mathbf{y}\right\}$$

set

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

where f_j is any antisymmetric function $(\underline{f(v,u) = -f(u,v)})$ to achieve the flip sign condition.

Feature Statistics: *Lasso coefficient difference (LCD)*

Consider the Lasso *augmented with knockoffs*

$$\min_{b\in\mathbb{R}^{2p}} \frac{1}{2}\|y-(\mathbf{X},\tilde{\mathbf{X}})b\|_2^2+\lambda\|b\|_1$$

which has solution $\hat{b}(\lambda)=(\hat{b}_1(\lambda),\cdots,\hat{b}_p(\lambda),\hat{b}_{p+1}(\lambda),\cdots,\hat{b}_{2p}(\lambda))$, then

$$W_j=Z_j-\tilde{Z}_j=|\hat{b}_j(\lambda)|-|\hat{b}_{j+p}(\lambda)|$$

Feature Statistics: *Lasso coefficient difference (LCD)*

Consider the Lasso *augmented with knockoffs*

$$\min_{b \in \mathbb{R}^{2p}} \frac{1}{2}\|y - (\mathbf{X}, \tilde{\mathbf{X}})b\|_2^2 + \lambda\|b\|_1$$

which has solution $\hat{b}(\lambda) = (\hat{b}_1(\lambda), \cdots, \hat{b}_p(\lambda), \hat{b}_{p+1}(\lambda), \cdots, \hat{b}_{2p}(\lambda))$, then

$$W_j = Z_j - \tilde{Z}_j = |\hat{b}_j(\lambda)| - |\hat{b}_{j+p}(\lambda)|$$

- a large positive value of W_j provides some evidence that the distribution of $\mathbf{Y}$ depends on $\mathbf{X}_j$

Feature Statistics: *Lasso coefficient difference (LCD)*

Consider the Lasso *augmented with knockoffs*

$$\min_{b\in\mathbb{R}^{2p}} \frac{1}{2}\|y-(\mathbf{X},\tilde{\mathbf{X}})b\|_2^2+\lambda\|b\|_1$$

which has solution $\hat{b}(\lambda)=(\hat{b}_1(\lambda),\cdots,\hat{b}_p(\lambda),\hat{b}_{p+1}(\lambda),\cdots,\hat{b}_{2p}(\lambda))$, then

$$W_j=Z_j-\tilde{Z}_j=|\hat{b}_j(\lambda)|-|\hat{b}_{j+p}(\lambda)|$$

- a large positive value of W_j provides some evidence that the distribution of $\mathbf{Y}$ depends on $\mathbf{X}_j$
- the value of λ can be chosen in any data-dependent fashion for a pair of $\mathbf{y}$ and $(\mathbf{X},\tilde{\mathbf{X}})$

Feature Statistics: *Lasso coefficient difference (LCD)*

$$W_j = Z_j - \tilde{Z}_j = |\hat{b}_j(\lambda)| - |\hat{b}_{j+p}(\lambda)|$$

Lemma 2

Conditional on $(|W_1|, \cdots, |W_p|)$, the sign of the null W_js ($j \in \mathcal{H}_0$) are i.i.d. coin flips

Feature Statistics: *Lasso coefficient difference (LCD)*

$$W_j = Z_j - \tilde{Z}_j = |\hat{b}_j(\lambda)| - |\hat{b}_{j+p}(\lambda)|$$

Lemma 2

Conditional on $(|W_1|, \cdots, |W_p|)$, the sign of the null W_js ($j \in \mathcal{H}_0$) are i.i.d. coin flips

Proof: for a sequence independent random variables $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_p)$ s.t. $\epsilon_j = \pm 1$ with probability $\frac{1}{2}$ if $j \in \mathcal{H}_0$, and $\epsilon_j = 1$ otherwise, put $S = \{j : \epsilon_j = -1\} \subset \mathcal{H}_0$

Feature Statistics: *Lasso coefficient difference (LCD)*

$$W_j = Z_j - \tilde{Z}_j = |\hat{b}_j(\lambda)| - |\hat{b}_{j+p}(\lambda)|$$

Lemma 2

Conditional on $(|W_1|, \cdots, |W_p|)$, the sign of the null W_js $(j \in \mathcal{H}_0)$ are i.i.d. coin flips

Proof: for a sequence independent random variables $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_p)$ s.t. $\epsilon_j = \pm 1$ with probability $\frac{1}{2}$ if $j \in \mathcal{H}_0$, and $\epsilon_j = 1$ otherwise, put $S = \{j : \epsilon_j = -1\} \subset \mathcal{H}_0$

- flip sign property: $W_{\mathrm{swap}(S)} \triangleq w\left\{(\mathbf{X}, \tilde{\mathbf{X}})_{\mathrm{swap}(S)}, \mathbf{y}\right\} \boxed{\overset{\mathrm{d}}{=}} \epsilon \odot W = (\epsilon_1 W1, \cdots, \epsilon_p W_p)$

Feature Statistics: *Lasso coefficient difference (LCD)*

$$W_j = Z_j - \tilde{Z}_j = |\hat{b}_j(\lambda)| - |\hat{b}_{j+p}(\lambda)|$$

Lemma 2

Conditional on $(|W_1|, \cdots, |W_p|)$, the sign of the null W_js $(j \in \mathcal{H}_0)$ are i.i.d. coin flips

Proof: for a sequence independent random variables $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_p)$ s.t. $\epsilon_j = \pm 1$ with probability $\frac{1}{2}$ if $j \in \mathcal{H}_0$, and $\epsilon_j = 1$ otherwise, put $S = \{j : \epsilon_j = -1\} \subset \mathcal{H}_0$

- flip sign property: $W_{\text{swap}(S)} \triangleq w\left\{(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap}(S)}, \mathbf{y}\right\} \boxed{\overset{\text{d}}{=}} \epsilon \odot W = (\epsilon_1 W1, \cdots, \epsilon_p W_p)$
- Lemma 1: $W_{\text{swap}(S)} \overset{\text{d}}{=} W$

which establishes $W \overset{\text{d}}{=} \epsilon \odot W$

Methodology: False Discovery Rate Control

Since the null W_js ($j \in \mathcal{H}_0$) are i.i.d. coin flips conditional on $(|W_1|, \cdots, |W_p|)$ (Lemma 2), they are *symmetric*

$$\#\{j : W_j \leq -t, j \in \mathcal{H}_0\} \stackrel{d}{=} \#\{j : W_j \geq t, j \in \mathcal{H}_0\}$$

and for any fixed threshold $t > 0$

$$\#\{j : W_j \leq -t\} \geq \#\{j : W_j \leq -t, j \in \mathcal{H}_0\}$$

so for the false discovery proportion FDP:

$$\mathrm{FDP}(t) = \frac{\#\{j : W_j \geq t, j \in \mathcal{H}_0\}}{\#\{j : W_j \geq t\}}$$

Methodology: False Discovery Rate Control

Since the null W_js ($j \in \mathcal{H}_0$) are i.i.d. coin flips conditional on $(|W_1|, \cdots, |W_p|)$ (Lemma 2), they are *symmetric*

$$\#\{j : W_j \leq -t, j \in \mathcal{H}_0\} \stackrel{d}{=} \#\{j : W_j \geq t, j \in \mathcal{H}_0\}$$

and for any fixed threshold $t > 0$

$$\#\{j : W_j \leq -t\} \geq \#\{j : W_j \leq -t, j \in \mathcal{H}_0\}$$

so for the false discovery proportion FDP, an *upward-biased* estimate is:

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

FDR Process (Barber and Candes, 2015)

Theorem 1: Modified FDR

Choose a threshold $\tau > 0$ by setting

$$\tau = \min\left\{t > 0 : \frac{\#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\}} \leq q\right\}$$

where q is the target FDR level. Then the procedure selecting the variables:

$$\hat{S} = \{j : W_j \geq \tau\}$$

controlling the *modified FDR* defined as

$$\mathrm{mFDR} = \mathbb{E}\left[\frac{|\left\{j \in \hat{S} \cap \mathcal{H}_0\right\}|}{|\hat{S}| + 1/q}\right] \leq q$$

These results are *non-asymptotic* and hold *conditionally* on the response $\mathbf{y}$

FDR Process (Barber and Candes, 2015)

Theorem 1: Usual FDR

Choose a threshold $\tau > 0$ by setting

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

where q is the target FDR level. Then the procedure selecting the variables:

$$\hat{S} = \{j : W_j \geq \tau\}$$

controlling the *modified FDR* defined as

$$\mathrm{mFDR} = \mathbb{E}\left[\frac{|\left\{ j \in \hat{S} \cap \mathcal{H}_0 \right\}|}{|\hat{S}| \vee 1} \right] \leq q$$

These results are *non-asymptotic* and hold *conditionally* on the response $\mathbf{y}$

Constructing Model-X Knockoffs

Definition 2: MX Knockoffs

MX knockoffs for the family of random variables $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_p)$ are a new family of random variables $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_p)$ constructed with the following 2 properties:

(a) for any subset $S \subset \{1, \cdots, p\}$,

$$(\mathbf{X}, \tilde{\mathbf{X}})_{\mathrm{swap}(S)} \stackrel{\mathrm{d}}{=} (\mathbf{X}, \tilde{\mathbf{X}})$$

(b) $\tilde{\mathbf{X}} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}$ if there is a response $\mathbf{Y}$

Constructing Model-X Knockoffs

Definition 2: MX Knockoffs

MX knockoffs for the family of random variables $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_p)$ are a new family of random variables $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_p)$ constructed with the following 2 properties:

(a) for any subset $S \subset \{1, \cdots, p\}$,

$$(\mathbf{X}, \tilde{\mathbf{X}})_{\mathrm{swap}(S)} \stackrel{\mathrm{d}}{=} (\mathbf{X}, \tilde{\mathbf{X}})$$

(b) $\tilde{\mathbf{X}} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}$ if there is a response $\mathbf{Y}$

which gives:

Proposition 2: MX Knockoffs

The random variables $(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_p)$ are *MX knockoffs* for $(\mathbf{X}_1, \cdots, \mathbf{X}_p)$ if and only if for any $j \in \{1, \cdots, p\}$, the pair $(\mathbf{X}_j, \tilde{\mathbf{X}}_j)$ is *exchangeable* conditional on all the other variables and their knockoffs.

Constructing Model-X Knockoffs

Goal: constructing pairs that are *conditionally exchangeable*

Constructing Model-X Knockoffs

Goal: constructing pairs that are *conditionally exchangeable*

Algorithm 1: Sequential Conditional Independent Pairs

$j = 1$
while $j \leq p$ **do**
  sample $\tilde{\mathbf{X}}_j$ from $\mathcal{L}(\mathbf{X}_j \mid \mathbf{X}_{-j}, \tilde{\mathbf{X}}_{1:j-1})$
  $j = j + 1$
end

Constructing Model-X Knockoffs

Goal: constructing pairs that are *conditionally exchangeable*

Algorithm 1: Sequential Conditional Independent Pairs

$j = 1$
while $j \le p$ **do**
  sample $\tilde{\mathbf{X}}_j$ from $\mathcal{L}(\mathbf{X}_j \mid \mathbf{X}_{-j}, \tilde{\mathbf{X}}_{1:j-1})$
  $j = j + 1$
end

Example: $p = 3$

Constructing Model-X Knockoffs

Goal: constructing pairs that are *conditionally exchangeable*

Algorithm 1: Sequential Conditional Independent Pairs

$j = 1$
while $j \leq p$ **do**
  sample $\tilde{\mathbf{X}}_j$ from $\mathcal{L}(\mathbf{X}_j \mid \mathbf{X}_{-j}, \tilde{\mathbf{X}}_{1:j-1})$
  $j = j + 1$
end

Example: $p = 3$

- $j = 1$: sample $\tilde{\mathbf{X}}_1$ from $\mathcal{L}(\mathbf{X}_1 \mid \mathbf{X}_{2:3})$

Constructing Model-X Knockoffs

Goal: constructing pairs that are *conditionally exchangeable*

Algorithm 1: Sequential Conditional Independent Pairs

$j = 1$
while $j \leq p$ **do**
    sample $\tilde{\mathbf{X}}_j$ from $\mathcal{L}(\mathbf{X}_j \mid \mathbf{X}_{-j}, \tilde{\mathbf{X}}_{1:j-1})$
    $j = j + 1$
end

Example: $p = 3$

- $j = 1$: sample $\tilde{\mathbf{X}}_1$ from $\mathcal{L}(\mathbf{X}_1 \mid \mathbf{X}_{2:3})$
- $j = 2$: sample $\tilde{\mathbf{X}}_2$ from $\mathcal{L}(\mathbf{X}_2 \mid \mathbf{X}_1, \mathbf{X}_3, \tilde{\mathbf{X}}_1)$

Constructing Model-X Knockoffs

Goal: constructing pairs that are *conditionally exchangeable*

Algorithm 1: Sequential Conditional Independent Pairs

$j = 1$
while $j \leq p$ **do**
  sample $\tilde{\mathbf{X}}_j$ from $\mathcal{L}(\mathbf{X}_j \mid \mathbf{X}_{-j}, \tilde{\mathbf{X}}_{1:j-1})$
  $j = j + 1$
end

Example: $p = 3$

- $j = 1$: sample $\tilde{\mathbf{X}}_1$ from $\mathcal{L}(\mathbf{X}_1 \mid \mathbf{X}_{2:3})$
- $j = 2$: sample $\tilde{\mathbf{X}}_2$ from $\mathcal{L}(\mathbf{X}_2 \mid \mathbf{X}_1, \mathbf{X}_3, \tilde{\mathbf{X}}_1)$
- $j = 3$: sample $\tilde{\mathbf{X}}_3$ from $\mathcal{L}(\mathbf{X}_3 \mid \mathbf{X}_{1:2}, \tilde{\mathbf{X}}_{1:2})$

Constructing Model-X Knockoffs *Approximately*

Definition 2: MX Knockoffs

MX knockoffs $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_p)$ constructed with the following 2 properties:

(a) for any subset $S \subset \{1, \cdots, p\}$,

$$\boxed{(\mathbf{X}, \tilde{\mathbf{X}})_{\mathrm{swap}(S)} \overset{\mathrm{d}}{=} (\mathbf{X}, \tilde{\mathbf{X}})}$$

(b) $\tilde{\mathbf{X}} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}$ if there is a response $\mathbf{Y}$

Constructing Model-X Knockoffs *Approximately*

Definition 2: MX Knockoffs

MX knockoffs $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_p)$ constructed with the following 2 properties:

(a) for any subset $S \subset \{1, \cdots, p\}$,

$$\boxed{(\mathbf{X}, \tilde{\mathbf{X}})_{\mathrm{swap}(S)} \overset{\mathrm{d}}{=} (\mathbf{X}, \tilde{\mathbf{X}})}$$

(b) $\tilde{\mathbf{X}} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}$ if there is a response $\mathbf{Y}$

An approximate constructions: $(\mathbf{X}, \tilde{\mathbf{X}})_{\mathrm{swap(S)}}$ and $(\mathbf{X}, \tilde{\mathbf{X}})$ have the same first 2 moments

Constructing Model-X Knockoffs *Approximately*

Definition 2: MX Knockoffs

MX knockoffs $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_p)$ constructed with the following 2 properties:

(a) for any subset $S \subset \{1, \cdots, p\}$,

$$\boxed{(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap}(S)} \overset{\text{d}}{=} (\mathbf{X}, \tilde{\mathbf{X}})}$$

(b) $\tilde{\mathbf{X}} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}$ if there is a response $\mathbf{Y}$

An approximate constructions: $(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap(S)}}$ and $(\mathbf{X}, \tilde{\mathbf{X}})$ have the same first 2 moments

$$\text{cov}(\mathbf{X}, \tilde{\mathbf{X}}) = \mathbf{G} \qquad \mathbf{G} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{\Sigma} - \text{diag}(s) \\ \mathbf{\Sigma} - \text{diag}(s) & \mathbf{\Sigma} \end{pmatrix}$$

Constructing Model-X Knockoffs *Approximately*

Definition 2: MX Knockoffs

MX knockoffs $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_p)$ constructed with the following 2 properties:

(a) for any subset $S \subset \{1, \cdots, p\}$,

$$\boxed{(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap}(S)} \overset{\text{d}}{=} (\mathbf{X}, \tilde{\mathbf{X}})}$$

(b) $\tilde{\mathbf{X}} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}$ if there is a response $\mathbf{Y}$

An approximate constructions: $(\mathbf{X}, \tilde{\mathbf{X}})_{\text{swap(S)}}$ and $(\mathbf{X}, \tilde{\mathbf{X}})$ have the same first 2 moments

$$\text{cov}(\mathbf{X}, \tilde{\mathbf{X}}) = \mathbf{G} \qquad \mathbf{G} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{\Sigma} - \text{diag}(s) \\ \mathbf{\Sigma} - \text{diag}(s) & \mathbf{\Sigma} \end{pmatrix}$$

For Gaussian $(\mathbf{X}, \tilde{\mathbf{X}})$, s is chosen s.t. $\mathbf{G}$ is positive semidefinite

Constructing Model-X Knockoffs *Approximately*

Constructing Model-X Knockoffs *Approximately*

- *equicorrelated* construction:

$$s_j^{\mathrm{EQ}} = 2\lambda_{\min}(\mathbf{\Sigma}) \wedge 1, \ \forall j$$

minimizing the *correlation between variable knockoff pairs* subject to the constraint that all such pairs *must have the same correlation*

Constructing Model-X Knockoffs *Approximately*

- *equicorrelated* construction:

$$s_j^{\text{EQ}} = 2\lambda_{\min}(\mathbf{\Sigma}) \wedge 1, \ \forall j$$

minimizing the *correlation between variable knockoff pairs* subject to the constraint that all such pairs *must have the same correlation*

- *semidefinite programme* construction:

$$\begin{aligned} \text{minimize} \quad & \sum_j \left|1 - s_j^{\text{SDP}}\right| \\ \text{subject to} \quad & s_j^{\text{SDP}} \geq 0, \ \operatorname{diag}\left(s^{\text{SDP}}\right) \preceq 2\mathbf{\Sigma} \end{aligned}$$

minimizing the *sum of the absolute values* of variable knockoff correlations between *all* suitable *s*

Construct Model-X Knockoffs *Approximately*

- *equicorrelated* construction:

$$s_j^{\mathrm{EQ}} = 2\lambda_{\min}(\boldsymbol{\Sigma}) \wedge 1, \ \forall j$$

- *semidefinite programme* construction:

$$\begin{aligned} &\text{minimize} && \sum_j \left|1 - s_j^{\mathrm{SDP}}\right| \\ &\text{subject to} && s_j^{\mathrm{SDP}} \geq 0, \ \operatorname{diag}\left(s^{\mathrm{SDP}}\right) \preceq 2\boldsymbol{\Sigma} \end{aligned}$$

challenges with large p

Construct Model-X Knockoffs *Approximately*

- *equicorrelated* construction:

$$s_j^{\mathrm{EQ}} = 2\lambda_{\min}(\boldsymbol{\Sigma}) \wedge 1, \ \forall j$$

- *semidefinite programme* construction:

$$\begin{aligned} \text{minimize} \quad & \sum_j \left| 1 - s_j^{\mathrm{SDP}} \right| \\ \text{subject to} \quad & s_j^{\mathrm{SDP}} \geq 0, \ \operatorname{diag}\left(s^{\mathrm{SDP}} \right) \preceq 2\boldsymbol{\Sigma} \end{aligned}$$

challenges with large p

- $\lambda_{\min}(\boldsymbol{\Sigma})$ tends to be extremely small: computationally easy, but **low power** of s_j^{EQ}

Construct Model-X Knockoffs *Approximately*

- *equicorrelated* construction:

$$s_j^{\mathrm{EQ}} = 2\lambda_{\min}(\boldsymbol{\Sigma}) \wedge 1, \ \ \forall j$$

- *semidefinite programme* construction:

$$\begin{aligned} \text{minimize} \quad & \sum_j \left|1 - s_j^{\mathrm{SDP}}\right| \\ \text{subject to} \quad & s_j^{\mathrm{SDP}} \geq 0, \ \operatorname{diag}\left(s^{\mathrm{SDP}}\right) \preceq 2\boldsymbol{\Sigma} \end{aligned}$$

challenges with large p

- $\lambda_{\min}(\boldsymbol{\Sigma})$ tends to be extremely small: computationally easy, but **low power** of s_j^{EQ}
- SDP (a convex problem) is computationally expensive

Construct Model-X Knockoffs *Approximately*

Following a two-step procedure:

- Step 1: choose an *approximation* $\mathbf{\Sigma}_{\text{approx}}$ of $\mathbf{\Sigma}$ and solve

$$\begin{aligned} \text{minimize} \quad & \sum_j |1 - \hat{s}_j| \\ \text{subject to} \quad & \hat{s}_j \geq 0, \ \operatorname{diag}(\hat{s}_j) \preceq 2\mathbf{\Sigma}_{\text{approx}} \end{aligned}$$

Construct Model-X Knockoffs *Approximately*

Following a two-step procedure:

- Step 1: choose an *approximation* $\boldsymbol{\Sigma}_{\text{approx}}$ of $\boldsymbol{\Sigma}$ and solve

$$\begin{aligned} &\text{minimize} && \sum_j |1 - \hat{s}_j| \\ &\text{subject to} && \hat{s}_j \geq 0, \ \operatorname{diag}(\hat{s}_j) \preceq 2\boldsymbol{\Sigma}_{\text{approx}} \end{aligned}$$

- Step 2: solve

$$\begin{aligned} &\text{maximize} && \gamma \\ &\text{subject to} && \operatorname{diag}(\gamma \hat{s}) \preceq 2\boldsymbol{\Sigma} \end{aligned}$$

  and set $s^{\text{ASDP}} = \gamma \hat{s}$

Construct Model-X Knockoffs *Approximately*

Approximate semidefinite programming

- Step 1: choose an *approximation* $\boldsymbol{\Sigma}_{\text{approx}}$ of $\boldsymbol{\Sigma}$ and solve

$$\text{minimize} \quad \sum_j |1 - \hat{s}_j| \qquad\qquad \text{subject to} \quad \hat{s}_j \geq 0, \ \operatorname{diag}(\hat{s}_j) \preceq 2\boldsymbol{\Sigma}_{\text{approx}}$$

- Step 2: solve

$$\text{maximize} \quad \gamma \qquad\qquad \text{subject to} \quad \operatorname{diag}(\gamma\hat{s}) \preceq 2\boldsymbol{\Sigma}$$

and set $s^{\text{ASDP}} = \gamma\hat{s}$

Construct Model-X Knockoffs *Approximately*

Approximate semidefinite programming

- Step 1: choose an *approximation* $\boldsymbol{\Sigma}_{\text{approx}}$ of $\boldsymbol{\Sigma}$ and solve

$$\text{minimize} \quad \sum_j |1 - \hat{s}_j| \qquad\qquad \text{subject to} \quad \hat{s}_j \geq 0, \ \text{diag}\left(\hat{s}_j\right) \preceq 2\boldsymbol{\Sigma}_{\text{approx}}$$

- Step 2: solve

$$\text{maximize} \quad \gamma \qquad\qquad \text{subject to} \quad \text{diag}\left(\gamma\hat{s}\right) \preceq 2\boldsymbol{\Sigma}$$

and set $s^{\text{ASDP}} = \gamma\hat{s}$

- *equicorrelated*: $\boldsymbol{\Sigma} = \mathbf{I} \Rightarrow \hat{s}_j = 1, \ \gamma = 2 \times \lambda_{\min}\boldsymbol{\Sigma} \wedge 1$

Construct Model-X Knockoffs *Approximately*

Approximate semidefinite programming

- Step 1: choose an *approximation* $\mathbf{\Sigma}_{\text{approx}}$ of $\mathbf{\Sigma}$ and solve

$$\text{minimize} \quad \sum_j |1 - \hat{s}_j| \qquad \text{subject to} \quad \hat{s}_j \geq 0, \ \operatorname{diag}(\hat{s}_j) \preceq 2\mathbf{\Sigma}_{\text{approx}}$$

- Step 2: solve

$$\text{maximize} \quad \gamma \qquad \text{subject to} \quad \operatorname{diag}(\gamma \hat{s}) \preceq 2\mathbf{\Sigma}$$

and set $s^{\text{ASDP}} = \gamma \hat{s}$

- *equicorrelated*: $\mathbf{\Sigma} = \mathbf{I} \Rightarrow \hat{s}_j = 1, \ \gamma = 2 \times \lambda_{\min}\mathbf{\Sigma} \wedge 1$
- *semidefinite programme*: $\mathbf{\Sigma} = \mathbf{\Sigma}, \ \hat{s}_j = s^{\text{SDP}}, \ \gamma = 1$

Numerical Simulation

Alternative Knockoff Statistics

For a statistic $\mathbf{T}$ for each original and knockoff variable

$$\mathbf{T} \triangleq (\mathbf{Z}, \tilde{\mathbf{Z}}) = (Z_1, \cdots, Z_p, \tilde{Z}_1, \cdots, \tilde{Z}_p) = t\left\{(\mathbf{X}, \tilde{\mathbf{X}}), \mathbf{y}\right\}$$

set

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

where f_j is any **antisymmetric** function $(\underline{f(v, u) = -f(u, v)})$ to achieve the flip sign condition.

Alternative Knockoff Statistics

For a statistic **T** for each original and knockoff variable

$$\mathbf{T} \triangleq (\mathbf{Z}, \tilde{\mathbf{Z}}) = (Z_1, \cdots, Z_p, \tilde{Z}_1, \cdots, \tilde{Z}_p) = t\left\{(\mathbf{X}, \tilde{\mathbf{X}}), \mathbf{y}\right\}$$

set

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

where f_j is any **antisymmetric** function $(\underline{f(v, u) = -f(u, v)})$ to achieve the flip sign condition.

- *Alternative* knockoff statistics
- *Bayesian* knockoff statistics

Alternative Knockoff Statistics: LCD vs LSM

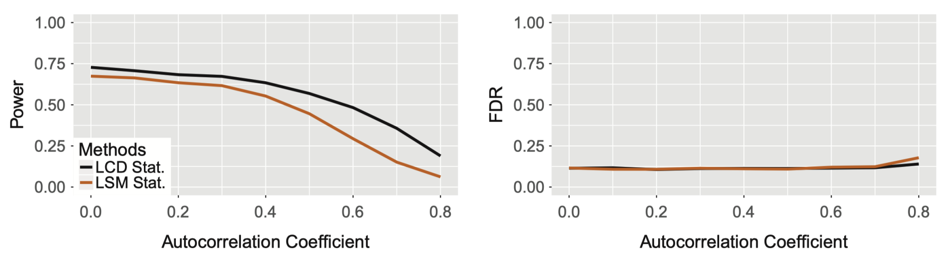
$$W_j = Z_j - \tilde{Z}_j = |\hat{b}_j(\lambda)| - |\hat{b}_{j+p}(\lambda)| \quad \text{where } \hat{b} = \arg\min_{b \in \mathbb{R}^{2p}} \frac{1}{2}\|y - (\mathbf{X}, \tilde{\mathbf{X}})b\|_2^2 + \lambda\|b\|_1 \quad \text{(LCD)}$$

$$W_j = \operatorname{sgn}(|Z_j| - |\tilde{Z}_j|)\max\left\{|Z_j|, |\tilde{Z}_j|\right\} \quad \text{where } Z_j = \sup\left\{\lambda : \hat{b}_j(\lambda) \neq 0\right\} \quad \text{(LSM)}$$

Alternative Knockoff Statistics: LCD vs LSM

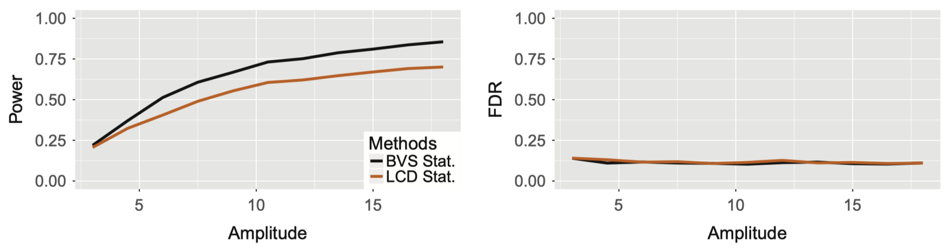
$$W_j = Z_j - \tilde{Z}_j = |\hat{b}_j(\lambda)| - |\hat{b}_{j+p}(\lambda)| \qquad \text{where } \hat{b} = \arg\min_{b \in \mathbb{R}^{2p}} \frac{1}{2}\|y - (\mathbf{X}, \tilde{\mathbf{X}})b\|_2^2 + \lambda\|b\|_1 \qquad \text{(LCD)}$$

$$W_j = \operatorname{sgn}(|Z_j| - |\tilde{Z}_j|) \max\left\{|Z_j|, |\tilde{Z}_j|\right\} \qquad \text{where } Z_j = \sup\left\{\lambda : \hat{b}_j(\lambda) \neq 0\right\} \qquad \text{(LSM)}$$



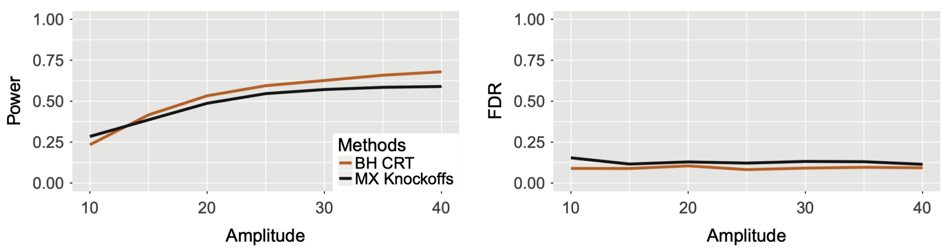
Alternative Knockoff Statistics: LCD vs BVS

BVS: $Z_j - \tilde{Z}_j$ where Z_j and $\tilde{Z}_j$ are the posterior probabilities that the jth original and knockoff coefficients are *non-zero* respectively.



Alternative Procedures: *BH* conditional randomization test

- for $(\mathbf{X}, \mathbf{y})$ and $k = 1, \cdots, K$, simulate $\mathbf{X}^{(k)}$ by simulating the jth column of $\mathbf{X}$ from $\mathcal{L}(\mathbf{X}_j \mid \mathbf{X}_{-j})$
- calculate $p-$value as $P_j = \frac{1}{K+1}\left[1 + \sum_{k=1}^{K} \mathbf{1}_{T_j(\mathbf{X}^{(k)},\mathbf{y}) \geq T_j(\mathbf{X},\mathbf{y})}\right]$



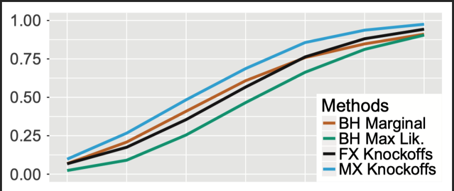
BH CRT is *very* computationally costly (by 5000 times)!

Alternative Procedures: Independent Covariates + Gaussian

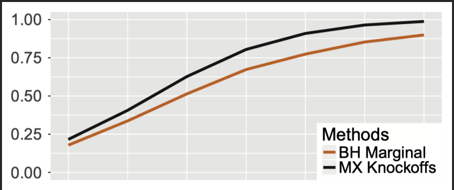
- *FX* knockoff: only applicable when $n \gg p$
- *BH* applied to asymptotic GLM p-values: only applicable when $n \gg p$
- *BH* applied to marginal test $p-$values: value for testing hypothesis of *marginal* distribution of X_j

Alternative Procedures: Independent Covariates + Gaussian

- *FX* knockoff: only applicable when $n \gg p$
- *BH* applied to asymptotic GLM p-values: only applicable when $n \gg p$
- *BH* applied to marginal test $p-$values: value for testing hypothesis of *marginal* distribution of X_j

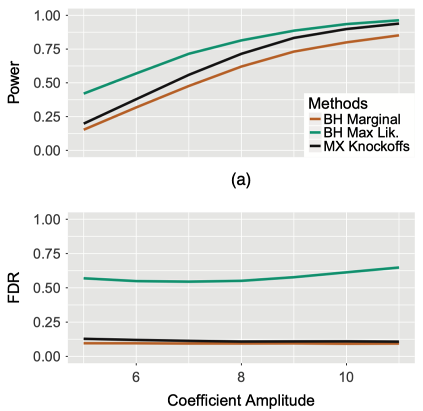


$n = 3000, p = 1000$

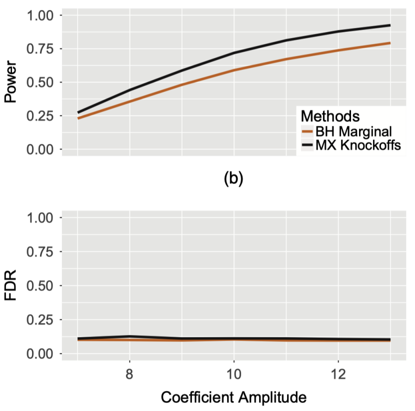


$n = 3000, p = 6000$

Alternative Procedures: Independent Covariates + Binomial

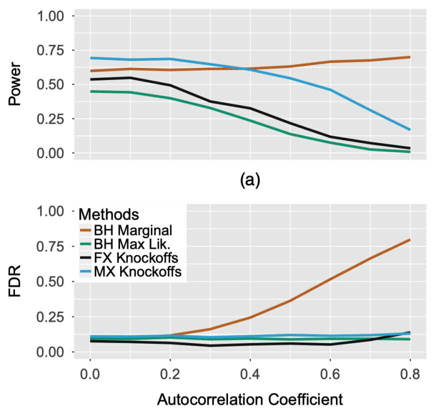


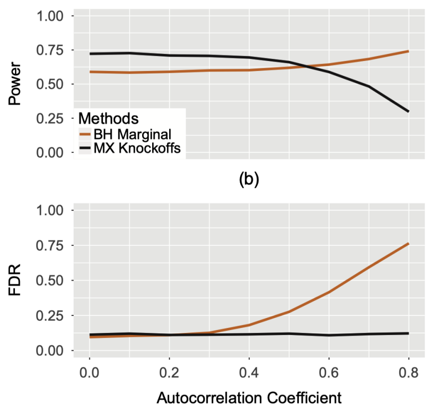
$n = 3000, p = 1000$



$n = 3000, p = 6000$

Alternative Procedures: AR(1) Covariates

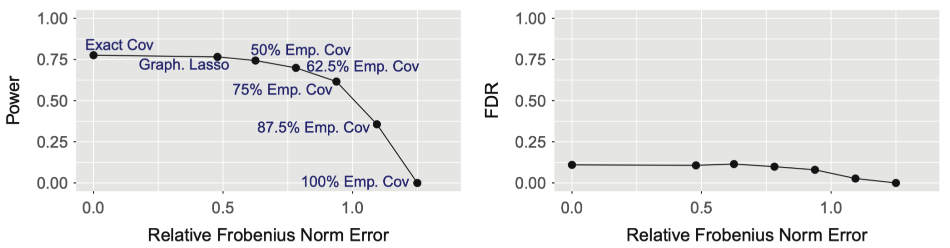




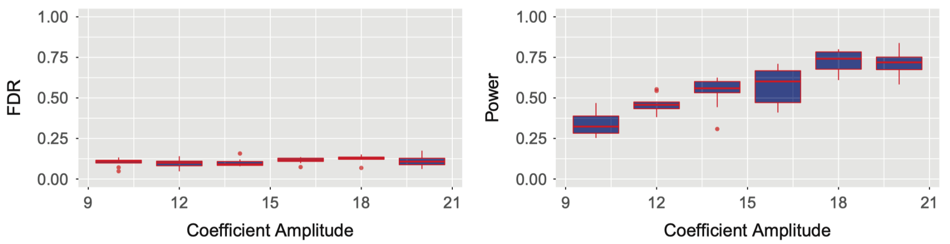
$n = 3000, p = 1000$; Gaussian

$n = 3000, p = 6000$; Binomial

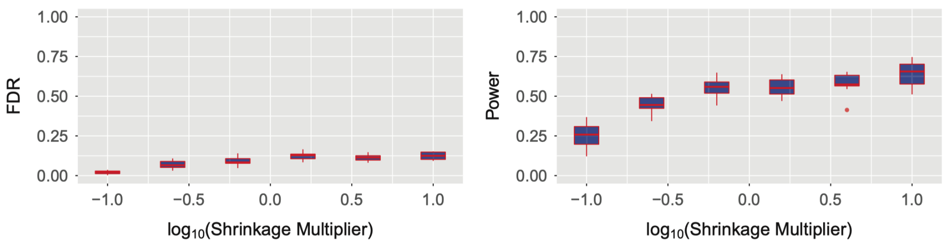
Robustness: Overfitting Error



Robustness: coefficient Amplitude



Robustness: Shrinkage Multiplier



Discussion

Discussion

Discussion

- Deployment in *highly correlated* settings

Discussion

- Deployment in *highly correlated* settings
- Open questions

Discussion

- Deployment in *highly correlated* settings
- Open questions
  - How to *construct* model-X knockoffs? How to *choose* model-X knockoffs?

Discussion

- Deployment in *highly correlated* settings
- Open questions
  - How to *construct* model-X knockoffs? How to *choose* model-X knockoffs?
  - *multiple* model-X knockoffs

Discussion

- Deployment in *highly correlated* settings
- Open questions
  - How to *construct* model-X knockoffs? How to *choose* model-X knockoffs?
  - *multiple* model-X knockoffs
  - robustness and computational cost

References I

Barber, R. F., & Candes, E. J. (2015). Controlling the false discovery rate via knockoffs. *Annals of Statistics*, *43*(5), 2055–2085.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Candès, E. J., Fan, J., Janson, L., & 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.

Thank you!

Appendix

Proof of Proposition 1

In the case of a logistic regression model:

$$\mathbb{E}[\mathbf{Y} \mid \mathbf{X}] = \mathbb{P}(\mathbf{Y} = \mathbf{1} \mid \mathbf{X}) = \frac{\exp(\eta)}{1 + \exp(\eta)} = g^{-1}(\eta), \qquad \eta = \beta_1 \mathbf{X}_1 + \cdots \beta_p \mathbf{X}_p$$

- $\beta_j = 0 \Rightarrow \mathbf{Y} \perp\!\!\!\perp \mathbf{X}_j \mid \mathbf{X}_{-j}$: if $\beta_j = 0$, then

$$P_{\mathbf{Y},\mathbf{x}_j|\mathbf{x}_{-j}}(\mathbf{y}, \mathbf{x}_j \mid \mathbf{x}_{-j}) = \underbrace{P_{\mathbf{Y}|\mathbf{x}_j,\mathbf{x}_{-j}}(\mathbf{y} \mid \mathbf{x}_j, \mathbf{x}_{-j})}_{\text{independent of } \mathbf{X}_j} P_{\mathbf{x}_j|\mathbf{x}_{-j}}(\mathbf{x}_j \mid \mathbf{x}_{-j})$$

- $\beta_j = 0 \Leftarrow \mathbf{Y} \perp\!\!\!\perp \mathbf{X}_j \mid \mathbf{X}_{-j}$: if $\mathbf{Y}$ and $\mathbf{X}_j$ are conditionally independent, the the likelihood function $\mathbb{E}[\mathbf{Y} \mid \mathbf{X}]$ must, conditional on $\mathbf{X}_{-j}$, factorize into a function of $\mathbf{Y}$ times a function of $\mathbf{X}_j$. That is, conditional on $\mathbf{X}_{-j}$, the odds ratio must not depend on $\mathbf{X}_j$, i.e., $\exp(\beta_j \mathbf{X}_j)$ must be constant, hence $\beta_j = 0$

back