ARK: Robust Knockoffs Inference via Coupling

Lan Gao

Department of Business Analytics and Statistics
Haslam College of Business
University of Tennessee Knoxville

lgao13@utk.edu

In collaboration with Yingying Fan and Jinchi Lv

BIRS-IASM Workshop, Hangzhou, China
Dec 14, 2023

Review of Model-X Knockoffs Framework

Variable Selection with False Discovery Rate Control

- Given response Y and p covariates $X_1, \cdots, X_p$
- Identify relevant covariates
- Relevant subset $\mathcal{H}_1 = \{j \in [p] : Y \not\perp\!\!\!\perp X_j | X_{-j}\}$
- Formulate as multiple hypothesis testing:

$$H_{0j} : X_j \in \mathcal{H}_0 := \mathcal{H}_1^c, \qquad j = 1, \cdots, p$$

- Aim to control FDR

$$\mathrm{FDR} = \mathbb{E}[\mathrm{FDP}], \qquad \mathrm{FDP} = \frac{|\widehat{S} \cap \mathcal{H}_0|}{|\widehat{S}| \vee 1}$$

Existing Work

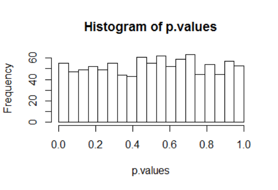
Large literature: Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Efron, 2007; Benjamini, 2010; Fan et al., 2012; ...

Most existing work relies on p-values

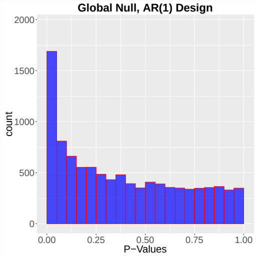
- BH procedure sorts p-values in ascending order then chooses a cutoff such that hypotheses with p-value below the cutoff are rejected
- very popularly used
- theoretically guaranteed to control FDR under p-value independence and certain forms of dependence

Potential problem: Valid p-value?

A fundamental assumption for p-value based procedures: uniform distribution of p-value under null hypothesis



However, in logistic regression with $n = 500$ $p = 200$ and under global null: non-uniform null distribution (Candès, Fan, Janson and Lv (2018))



Fan, Demirkaya and Lv (2019); Sur, Chen and Candès (2017); Sur and Candès (2018); ...

Model-X Knockoffs Framework

- Introduced in Candès, Fan, Janson and Lv (2018)
  - Bypass the use of p-values to achieve FDR control
  - Model-free: any model for the conditional dependence $Y|X_1, \cdots, X_p$
  - Dimension free: any dimension (including $p > n$)
  - Known covariate distribution: joint distribution of $X = (X_1, \cdots, X_p)$ is known
  - Guarantee finite-sample FDR control
- Intuition:
  - Generate "fake" copies of $(X_1, \ldots, X_p)$ which are irrelevant to Y but mimics the dependence structure of $(X_1, \ldots, X_p)$
  - Act as controls for assessing importance of original variables

Model-X Knockoff Variables

Definition (Candés, Fan, Janson and Lv, 2018)

Model-X knockoffs for the family of random variables $X = (X_1, \cdots, X_p)$ are a new family of random variables $\widetilde{X} = (\widetilde{X}_1, \cdots, \widetilde{X}_p)$ constructed such that

- for any subset $S \subset \{1, \cdots, p\}$,

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

- $\widetilde{X} \perp\!\!\!\perp Y | X$

Example: $(X_1, X_2, X_3, \widetilde{X}_1, \widetilde{X}_2, \widetilde{X}_3)_{\text{swap}(2,3)} = (X_1, \widetilde{X}_2, \widetilde{X}_3, \widetilde{X}_1, X_2, X_3)$

Example: Generating Knockoff Variables

Example: Consider $X \overset{d}{\sim} N(0, \Sigma)$

- construct $(X, \widetilde{X}) \overset{d}{\sim} N(0, \Sigma^{\mathsf{aug}})$, where

$$\Sigma^{\mathsf{aug}} = \begin{pmatrix} \Sigma & \Sigma - \mathrm{diag}(\mathbf{r}) \\ \Sigma - \mathrm{diag}(\mathbf{r}) & \Sigma \end{pmatrix}$$

  and $\mathrm{diag}(\mathbf{r}) \geq 0$ such that $\Sigma^{\mathsf{aug}} > 0$. Then

$$\widetilde{X} | X \overset{d}{=} N((I_p - \Sigma^{-1}\mathrm{diag}(\mathbf{s}))X, 2\mathrm{diag}(\mathbf{s}) - \mathrm{diag}(\mathbf{s})\Sigma^{-1}\mathrm{diag}(\mathbf{s}))$$

- Important observation: knockoff variables are not unique

The Knockoffs Procedure

1. Construct model-X knockoff variables
   - $\widetilde{\mathbf{X}} \sim F(\cdot|\mathbf{X})$
2. Compute knockoff statistics $\widetilde{W}_j$'s
   - A large positive $\widetilde{W}_j$: X_j is more important than $\widetilde{X}_j$
   - Null variables: $\widetilde{W}_j$ should be symmetric around 0
   - Eg: Regression Coefficient Difference
     $\widetilde{W}_j = |\hat{\beta}_j| - |\hat{\beta}_{j+p}|$;
     Marginal correlation difference $\widetilde{W}_j = |\mathbf{X}_j^T\mathbf{y}| - |\widetilde{\mathbf{X}}_j^T\mathbf{y}|$
3. Select relevant variables:
   - Find the knockoff threshold $\hat{\tau} > 0$
   - Select only variables with $\widetilde{W}_j \geq \hat{\tau}$

Ideal knockoffs procedure

Ideal distribution
$F(\cdot\,|\boldsymbol{X})$

Ideal knockoff variables
$\widetilde{\boldsymbol{X}} \sim F(\cdot\,|\boldsymbol{X})$

Ideal knockoff statistics $\widetilde{W}_j's$

Set of selected variables
S

Exact FDR Control

Theorem (Candés, Fan, Janson and Lv, 2018)

Letting

$$\hat{\tau}_+ = \min\left\{t > 0 : \frac{1 + \#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\}} \leq q\right\} \quad \textbf{(Knockoffs+)}$$

and setting $\hat{S} = \{j : W_j \geq \hat{\tau}_+\}$, controls the usual FDR,

$$\mathbb{E}\left[\frac{|\hat{S} \cap \mathcal{H}_0|}{|\hat{S}| \vee 1}\right] \leq q.$$

Exact FDR Control

Theorem (Candés, Fan, Janson and Lv, 2018)

Letting

$$\hat{\tau}_+ = \min\left\{t > 0 : \frac{1 + \#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\}} \leq q\right\} \quad \textbf{(Knockoffs+)}$$

and setting $\hat{S} = \{j : W_j \geq \hat{\tau}_+\}$, controls the usual FDR,

$$\mathbb{E}\left[\frac{|\hat{S} \cap \mathcal{H}_0|}{|\hat{S}| \vee 1}\right] \leq q.$$

Intuition:

$$\begin{aligned}
\mathrm{FDR} &= E\left[\frac{\#\text{selected null variables}}{\#\text{selected variables}}\right] \\
&= E\left[\frac{\#\{\text{null } W_j \geq \hat{\tau}\}}{\#\{W_j \geq \hat{\tau}\}}\right] \approx E\left[\frac{\#\{\text{null } W_j \leq -\hat{\tau}\}}{\#\{W_j \geq \hat{\tau}\}}\right] \\
&\leq E\left[\frac{\#\{W_j \leq -\hat{\tau}\}}{\#\{W_j \geq \hat{\tau}\}}\right] \leq q
\end{aligned}$$

Practical Implementation

Ideal knockoffs procedure

Ideal distribution

$F(\cdot\,|\boldsymbol{X})$

Ideal knockoff variables

$\widetilde{\boldsymbol{X}} \sim F(\cdot\,|\boldsymbol{X})$

Ideal knockoff statistics $\widetilde{W}_j's$

Set of selected variables

S

Approximate knockoffs procedure

Approximate distribution

$\hat{F}(\cdot\,|\boldsymbol{X})$

Approximate knockoff variables

$\hat{X} \sim \hat{F}(\cdot\,|\boldsymbol{X})$

Approximate knockoff statistics $\widehat{W}_j's$

Set of selected variables

$\hat{S}$

Practical Implementation

- Exact knowleadge of the joint distribution of the covariates $X = (X_1, \ldots, X_p)$ is unavailable
- What if misspecified or estimated feature distribution is applied to generate knockoffs?

Robustness of Approximate Knockoffs Procedure

Numerical Evidence

- $Y = X\boldsymbol{\beta} + \varepsilon$, where $\|\beta\|_0 = 30$, $p = 1000$ and $n = 500$
- $X \overset{d}{\sim} t_\nu(0, \Sigma)$, $\Sigma_{i,j} = \rho^{|i-j|}$
- Mis-specify the covariate distribution as $N(0, \hat{\Sigma})$ and generate $\widetilde{X}$ from the mis-specified distribution
- Knockoff statistics constructed as lasso coefficient difference

ρ	0				0.5			
ν	5	10	50	100	5	10	50	100
FDR	0.203	0.204	0.194	0.179	0.145	0.162	0.171	0.157

Table 1: FDR control based on 100 replications; target $q = 0.2$

Numerical Evidence

- $Y = X\beta + \varepsilon$, where $\|\beta\|_0 = 30$, $p = 1000$ and $n = 500$
- $X \overset{d}{\sim} t_\nu(0, \Sigma)$, $\Sigma_{i,j} = \rho^{|i-j|}$
- Mis-specify the covariate distribution as $N(0, \hat{\Sigma})$ and generate $\widetilde{X}$ from the mis-specified distribution
- Knockoff statistics constructed as lasso coefficient difference

ρ	0				0.5			
ν	5	10	50	100	5	10	50	100
FDR	0.203	0.204	0.194	0.179	0.145	0.162	0.171	0.157

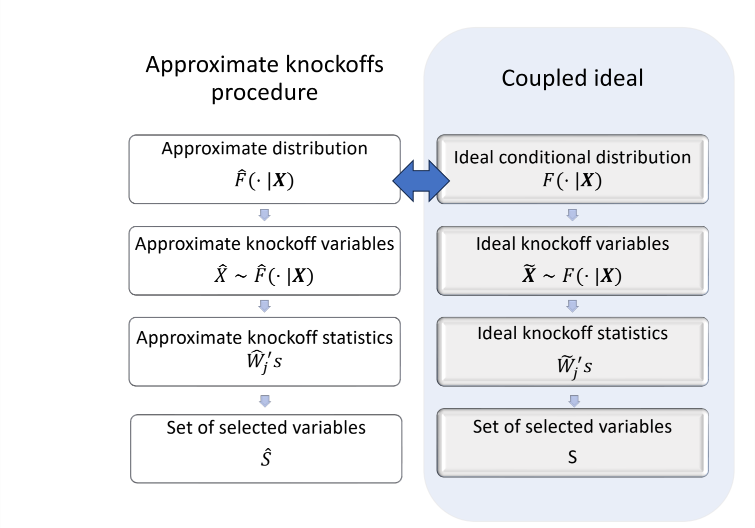
Table 1: FDR control based on 100 replications; target $q = 0.2$

- Additional numerical evidence: Candès et al.(2018); Sesia et al., (2019); Jordon et al. (2018); Lu et al. (2018); Zhu et al. (2021); Romano et al.(2020), ...
- Can we theoretically characterize to what extent it can be robust to the mis-specified distribution?

Existing results

- Fan, Lv, Sharifvaghefi, and Uematsu (2020): linear model setting where the features follow a latent factor model with parametric idiosyncratic noise
- Fan, Demirkaya, Li, and Lv (2020): theoretical guarantee on the robustness when the features have the joint Gaussian distribution, assuming Lipschitz continuity for the FDR function

Barber Candés and Samworth (2020)



Close in distribution, but not in realizations!

Continued

Barber, Candés and Samworth (2020):

- If an approximate feature distribution $\widehat{F}(\cdot)$ is used, the resulting FDR is bounded by

$$\text{FDR} \leq \min_{\epsilon \geq 0} \left\{ q \cdot e^{\epsilon} + \mathbb{P}\Big(\max_{j \in \mathcal{H}_0} \widehat{KL}_j > \epsilon \Big) \right\},$$

  where $\widehat{KL}_j$'s measure the distance between the approximate and coupled true conditional distributions of $X_j | X_{-j}$

- General results without assumption on the parametric forms of covariate distribution

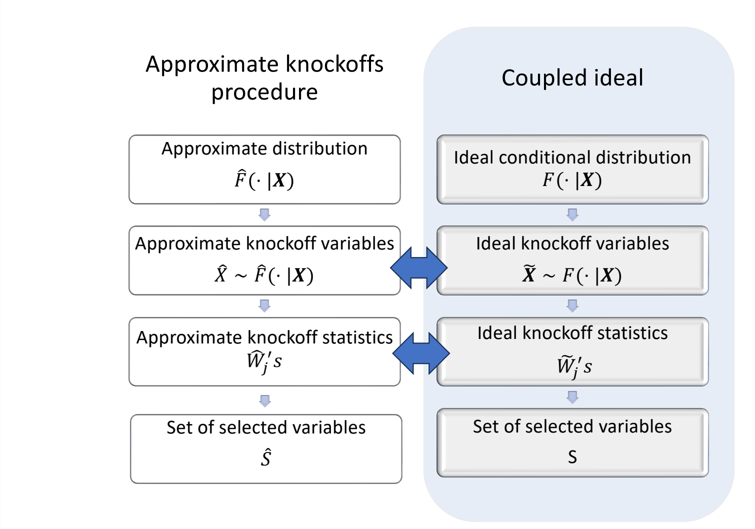
Continued

Barber, Candés and Samworth (2020):

- Two important assumptions:
  - $\widehat{F}(\cdot)$ should be learned independently from the traning data
  - $\max_{j \in \mathcal{H}_0} \widehat{KL}_j \xrightarrow{p} 0$
- These two assumptions do not always describe the practical implementation
  - Independent training data may not be available
  - In the t-distribution example, their theory requires at least $\nu^2 \gg n \min(n, p)$ for ensuring $\limsup_{n\to\infty} \text{FDR} \leq q$

Robust Knockoffs Inference via Coupling

Coupling Idea



Key coupling idea: need to be close in realization, not distribution!

Outline for Our Theoretical Results

- A specific example to illustrate the idea
- General theory

Example 1: Knockoffs for Multivariate t-Distribution

- $X \overset{d}{\sim} t_\nu(\mathbf{0}, \Omega^{-1})$ with unknown ν and Ω^{-1}
- Precision matrix $\Theta := [\mathrm{Cov}(X)]^{-1} = \frac{\nu-2}{\nu}\Omega$
- Effective estimator $\widehat{\Theta}$ constructed using data $\mathbf{X}$
- Common practice: construct approximate knockoffs data matrix $\widehat{\mathbf{X}}$ from Gaussian distribution with matched first two moments ($N(\mathbf{0}, \widehat{\Theta}^{-1})$)

$$\widehat{\mathbf{X}} = \mathbf{X}(I_p - r\widehat{\Theta}) + \mathbf{Z}(2rI_p - r^2\widehat{\Theta})^{1/2},$$

where $r > 0$ is a positive constant and $\mathbf{Z}$ independent of $(\mathbf{X}, \mathbf{y})$ and consists of i.i.d. $N(0, 1)$

- A misspecified feature distribution is used

Example 1: Knockoff Variables Coupling

Questions: Does there exist an ideal knockoff variable matrix $\widetilde{\mathbf{X}}$ that is close in realization to $\widehat{\mathbf{X}}$?

Example 1: Knockoff Variables Coupling

$$\widehat{\mathbf{X}} = \mathbf{X}(I_p - r\widehat{\Theta}) + \mathbf{Z}(2rI_p - r^2\widehat{\Theta})^{1/2}$$

- Coupled ideal knockoffs data matrix:

$$\widetilde{\mathbf{X}} = \mathbf{X}(I_p - r\Omega) + \mathrm{diag}(\frac{1}{\sqrt{\mathbf{Q}/\nu}})\mathbf{Z}(2rI_p - r^2\Omega)^{1/2},$$

  where $\mathrm{diag}(\frac{1}{\sqrt{\mathbf{Q}/\nu}}) = \mathrm{diag}(\frac{1}{\sqrt{Q_1/\nu}}, \ldots, \frac{1}{\sqrt{Q_n/\nu}})$ with $\{Q_i\}_{i=1}^n \overset{d}{\sim} \mathcal{X}_\nu^2$ i.i.d.

- Important: r and $\mathbf{Z}$ are exactly the same as those used in constructing $\widehat{\mathbf{X}}$

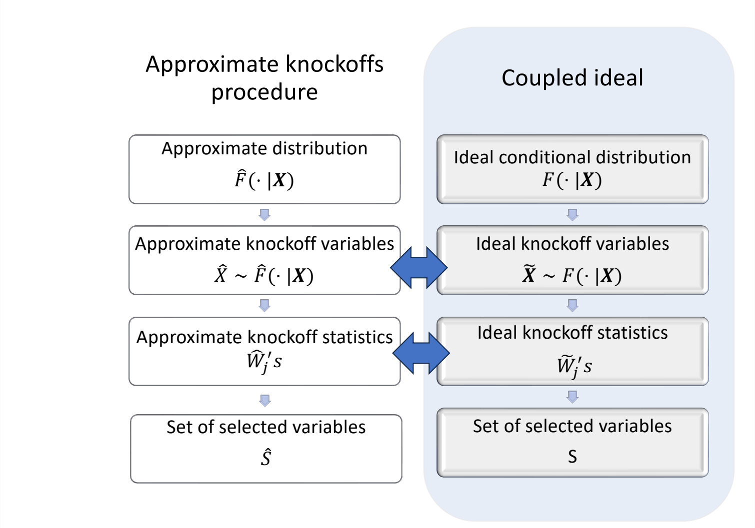
Example 1: Knockoff Variables Coupling

- **Proposition 1.** Assume $\max_{1\le j\le p}(\|\Omega_j\|_0 + \|\widehat{\Theta}_j\|_0) \le \rho_n$ almost surely, and $\rho_n\sqrt{\frac{\log p}{n}} \to 0$ and $\rho_n \nu^{-1/2} \to 0$, and that with probability $1 - o(1)$, it holds that $\|\widehat{\Omega} - \Omega\|_2 \lesssim \rho_n(n^{-1}\log p)^{1/2}$. Then as $\nu \ge 9$ and $\log p = o(n^{1-4/\nu})$, we have with probability $1 - o(1)$

$$\max_{1\le j\le p} n^{-1/2}\|\hat{\mathbf{X}}_j - \widetilde{\mathbf{X}}_j\|_2 \lesssim \rho_n(n^{-1}\log p)^{1/2} + \nu^{-1/2}.$$

- The rate $\rho_n(n^{-1}\log p)^{1/2}$ for precision matrix estimation has been verified in many existing works (Cai, Liu and Luo, 2011; Fan, Liao and Liu, 2016; Fan and Lv, 2016)

- $\nu^{-1/2}$ measures the effect of mis-specified distribution



- How $\widehat{W}_j$'s depend on $\widehat{X}_j$'s depend on the specific construction
- Consider regression coefficient difference (RCD) in linear model as an example

Example 1: RCD Knockoff Statistics Coupling

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- $\boldsymbol{\varepsilon} \overset{d}{\sim} N(0, \sigma^2 I_n)$
- $s = \|\boldsymbol{\beta}\|_0$
- Sub-Gaussian features $X \in \mathbb{R}^p$ with $\mathbb{E}(X) = \mathbf{0}$ and $\mathrm{Cov}(X) = \Sigma$
- Augmented design matrices $\widehat{X}^{\text{aug}} = (X, \widehat{X})$ and $\widetilde{X}^{\text{aug}} = (X, \widetilde{X})$

Example 1: RCD Knockoff Statistics Coupling

- Debias lasso (Zhang & Zhang, 2014) estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_j)_{1 \le j \le 2p}$ based on $(\widehat{\mathbf{X}}^{\text{aug}}, \mathbf{y})$:

$$\hat{\beta}_j = \hat{\beta}_j^{lasso} + \frac{\hat{\mathbf{z}}_j^\top \left(\mathbf{y} - \widehat{\mathbf{X}}^{\text{aug}} \hat{\boldsymbol{\beta}}^{lasso}\right)}{\hat{\mathbf{z}}_j^\top \widehat{\mathbf{X}}_j^{\text{aug}}},$$

where $\hat{\mathbf{z}}_j = \widehat{\mathbf{X}}_j^{\text{aug}} - \widehat{\mathbf{X}}_{-j}^{\text{aug}} \widehat{\gamma}_j$ and $\widehat{\gamma}_j$ is lasso coefficient by regressing $\widehat{\mathbf{X}}_j$ on $\widehat{X}_{-j}$

- Coupled debiased Lasso estimator $\widetilde{\boldsymbol{\beta}} = (\widetilde{\beta}_j)_{1 \le j \le p}$ based on $(\widetilde{\mathbf{X}}^{\text{aug}}, \mathbf{y})$:

$$\widetilde{\beta}_j = \widetilde{\beta}_j^{lasso} + \frac{\widetilde{\mathbf{z}}_j^\top \left(\mathbf{y} - \widetilde{\mathbf{X}}^{\text{aug}} \widetilde{\boldsymbol{\beta}}^{lasso}\right)}{\widetilde{\mathbf{z}}_j^\top \widetilde{\mathbf{X}}_j^{\text{aug}}} \quad \text{for } 1 \le j \le 2p,$$

where $\widetilde{\mathbf{z}}_j$ is the score vector given by

$$\widetilde{\mathbf{z}}_j = \widetilde{\mathbf{X}}_j^{\text{aug}} - \widetilde{\mathbf{X}}_{-j}^{\text{aug}} \widetilde{\gamma}_j$$

Example 1: RCD Knockoff Statistics Coupling

- Key for pairing: all the regularization parameters in the lasso algorithm used to compute $\widetilde{\boldsymbol{\beta}}^{lasso}$ and $\widetilde{\boldsymbol{\gamma}}_j$ are the same as those applied to compute $\widehat{\boldsymbol{\beta}}^{lasso}$ and $\widehat{\boldsymbol{\gamma}}_j$
- RCD knockoff statistics $\widehat{W}_j = |\widehat{\beta}_j| - |\widehat{\beta}_{j+p}|$
- Coupled perfect counterpart $\widetilde{W}_j = |\widetilde{\beta}_j| - |\widetilde{\beta}_{j+p}|$

Corollary 1. Assume conditions in Proposition 1. Under some regularity conditions on model sparsity and restricted eigenvalues, we have with probability $1 - o(1)$

$$\max_{1\le j\le p} |\widehat{W}_j - \widetilde{W}_j| \lesssim \Big(\underbrace{\rho_n(n^{-1}\log p)^{1/2} + \nu^{-1/2}}_{\text{coupling error of } \widehat{X}}\Big) \times \underbrace{(s\sqrt{n^{-1}\log p})}_{L_1 \text{ estimation error of } \widehat{\beta}}$$

Example 1: FDR Control

Theorem 1. Under some regularity conditions on the model sparsity, restricted eigenvalues and signal strength, if $\nu \gg s^2(\log p)^{2+2/\gamma}$ with $\gamma \in (0,1)$, we have

$$\limsup_{n\to\infty} FDR(\hat{S}) \le q.$$

Example 1: Compare with Existing Result

Consider the special case of $\Omega = I_p$ and known

- Barber Candés and Samworth (2020): Theorem 1 therein requires at least $\nu \gg \sqrt{n \min(n, p)}$ for ensuring $\limsup_{n\to\infty} \mathrm{FDR} \le q$
- Under the setting of linear regression model and RCD knockoff statistics, our theory requires $\nu \gg s^2(\log p)^{2+2/\gamma}$ for $\limsup_{n\to\infty} \mathrm{FDR} \le q$, where $0 < \gamma < 1$
- In Barber Candés and Samworth (2020), large n means large ν; our condition on ν is free of n
- This improvement shows some potential advantage of our coupling technique in the robustness analyses

Additional examples in our paper

General Theory

General Theory

- Three steps:

Step 1 Construct approximate knockoffs data matrix $\widehat{\mathbf{X}}$ using a working distribution $\widehat{F}(\cdot)$

Step 2 Compute approximate knockoff statistics $\{\widehat{W}_j\}_{j=1}^p$, where W_j is a function of $((\mathbf{X}, \widehat{\mathbf{X}}), \mathbf{y})$

Step 3 Select relevant set $\hat{S} = \{1 \leq j \leq p : \widehat{W}_j > T\}$ with a data-driven threshold based on $\{\widehat{W}_j\}_{j=1}^p$

- Our theory has three layers, corresponding reversely to the different steps above

Approximate knockoffs procedure

Approximate distribution $\hat{F}(\cdot \mid \boldsymbol{X})$

↓

Approximate knockoff variables $\hat{X} \sim \hat{F}(\cdot \mid \boldsymbol{X})$

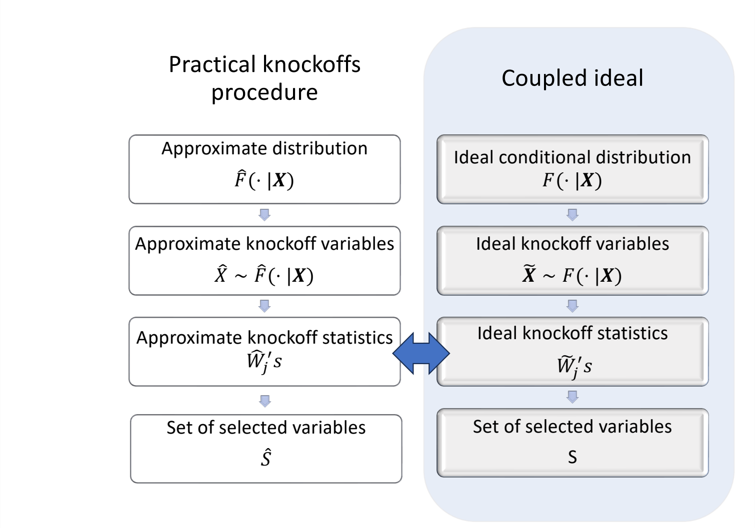
↓

Approximate knockoff statistics $\widehat{W}_j' s$

↓

Set of selected variables $\hat{S}$

Layer 1: A General Theory



Layer 1: Continued

Condition 1 (coupling accuracy on $\{\widehat{W}_j\}_{j=1}^p$). There exist perfect knockoff statistics $\{\widetilde{W}_j\}_{j=1}^p$ and a sequence $b_n \to 0$ such that

$$\mathbb{P}\Big(\max_{1 \le j \le p} |\widehat{W}_j - \widetilde{W}_j| \le b_n \Big) \to 1$$

- Assume general conditions on concentration of $\widetilde{W}_j$, signal strength, and weak dependence between $\{\widetilde{W}_j\}$; no specific model assumptions
- In addition, denote $G(t) = p_0^{-1} \sum_{j \in \mathcal{H}_0} \mathbb{P}(\widetilde{W}_j \ge t)$ and $a_n \to \infty$ as the number of strong signals. Assume

$$(\log p)^{1/\gamma} \sup_{t \in (0,\, G^{-1}(\frac{c_1 q a_n}{p})]} \frac{G(t - b_n) - G(t + b_n)}{G(t)} \to 0 \qquad (1)$$

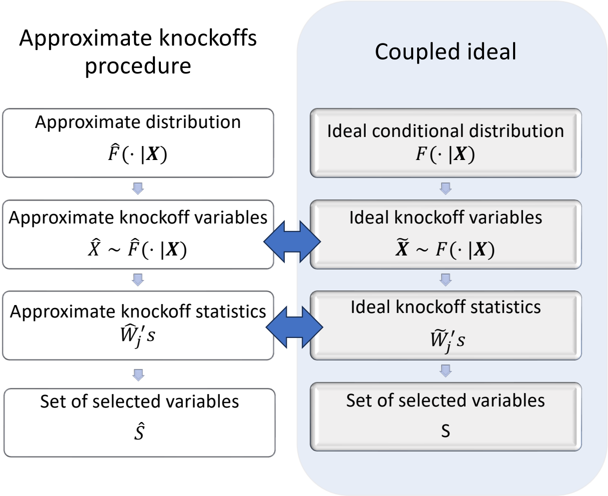
Layer 1: FDR Control

Theorem 2. Under the above conditions, we have

$$\limsup_{n\to\infty} FDR(\hat{S}) \leq q.$$

- A general theory of on the asymptotic FDR control for the approximate knockoffs inference
- Layer 1 is related to the third step in ARK procedure (selecting relevant features based on $\{\widehat{W_j}\}_{j=1}^{p}$)

Layer 2: Knockoff Variables Coupling



Layer 2: Continued

- $\widehat{\mathbf{X}} \in \mathbb{R}^{n \times p}$: approximate knockoff variable matrix
- $\widetilde{\mathbf{X}}$: perfect counterpart
- their realizations need to be close

Condition 2. (Coupling accuracy on $\widehat{\mathbf{X}}$) There exists a perfect knockoff data matrix $\widetilde{\mathbf{X}}$ and a sequence $\Delta_n \to 0$ such that

$$\mathbb{P}\Big(\max_{1 \le j \le p} n^{-1/2} \|\widehat{\mathbf{X}}_j - \widetilde{\mathbf{X}}_j\|_2 \le \Delta_n \Big) \to 1$$

In the t distribution example: $\Delta_n \sim \rho_n \sqrt{(\log p)/n} + \nu^{-1/2}$

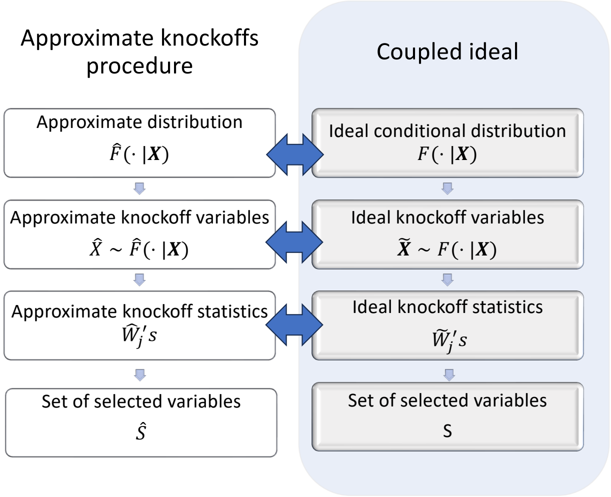
Layer 2: Knockoff Variables Coupling

- Layer 2 is related to the second step in ARK procedure (computine approximate knockoff statistics $\{\widehat{W}_j\}_{j=1}^p$ based on $(\widehat{\mathbf{X}}, \mathbf{X}, \mathbf{y})$
- Different knockoff statistics depends on the feature matrix differently
- We have verified that for RCD, it holds that

$$\text{Condition 2} \Rightarrow \max_j |\widehat{W}_j - \widetilde{W}_j| \leq o_p(\Delta_n \cdot s\sqrt{n^{-1}\log p})$$

- Our paper also verified a similar result for marginal correlation difference statistics

Layer 3: Existence of Coupled Knockoff Variables



Layer 3: Existence of Coupled Knockoff Variables

- Layer 3: We present specific constructions for the coupled knockoff variables and verify Condition 2

- Layer 3 is corresponding to the first step in the ARK procedure (Constructing knockoffs variables using approximate feature distribution $\widehat{F}(\cdot)$)

- Intuition: $\widehat{\mathbf{X}} \to \widetilde{\mathbf{X}}$ and $\widehat{W_j} \to \widetilde{W_j}$ as $\widehat{F} \to F$

- We have shown three examples of misspecified or estimated feature distributions that Condition 2 holds:

$$\mathbb{P}\left(\max_{1 \leq j \leq p} n^{-1/2} \|\widehat{\mathbf{X}}_j - \widetilde{\mathbf{X}}_j\|_2 \leq \Delta_n \right) \to 1$$

Summary

- We have investigated the robustness of the model-X knockoffs framework
- The approximate knockoffs procedure can achieve asymptotic FDR control as the sample size diverges in the high-dimensional setting
- Our theoretical analysis relies on the key idea of coupling – to pair statistics in the approximate knockoffs procedure with those in a perfect knockoffs procedure so that they are close in realizations
- We also showcase specific constructions of such coupled approximate and perfect knockoff variables
- Our paper also contains results for k-FWER control

Thank you!

Example 2: Gaussian Knockoffs

- $X \stackrel{d}{\sim} N(0, \Omega^{-1})$ with unknown precision matrix Ω

- Effective estimator $\widehat{\Omega}$ for Ω

- Approximate knockoff data matrix

$$\widehat{\mathbf{X}} = \mathbf{X}(I_p - r\widehat{\Omega}) + \mathbf{Z}(2rI_p - r^2\widehat{\Omega})^{1/2},$$

where $r > 0$ is a constant and $\mathbf{Z} = (\mathbf{Z}_{i,j}) \in \mathbb{R}^{n \times p}$ is independent of $(\mathbf{X}, \mathbf{y})$ with i.i.d. entries $Z_{i,j} \stackrel{d}{\sim} N(0, 1)$

- Coupled perfect knockoff data matrix

$$\widetilde{\mathbf{X}} = \mathbf{X}(I_p - r\Omega) + \mathbf{Z}(2rI_p - r^2\Omega)^{1/2},$$

- $\mathbf{Z}$ and r are identical as those used in constructing $\hat{\mathbf{X}}$.

Example 2: Knockoff Variables Coupling

Proposition 2. Assume $\max_{1\le j\le p} \|(\Omega_j\|_0 + \|\widehat{\Omega}_j\|_0) \le \rho_n$ almost surely with $\rho_n\sqrt{\frac{\log p}{n}} \to 0$, and with probability $1-o(1)$, it holds that $\|\widehat{\Omega}-\Omega\|_2 \lesssim \rho_n\sqrt{\frac{\log p}{n}}$. Then with probability $1-o(1)$

$$\max_{1\le j\le p} n^{-1/2}\|\widehat{\mathbf{X}}_j - \widetilde{\mathbf{X}}_j\|_2 \lesssim \rho_n\sqrt{\frac{\log p}{n}}.$$

Example 2: RCD Knockoff Statistics Coupling

Consider RCD based on debiased Lasso: $\widehat{W}_j = |\widehat{\beta}_j| - |\widehat{\beta}_{j+p}|$

Corollary 2. Assume conditions in Proposition 2. Under some regularity conditions on model sparsity and restricted eigenvalues, we have with probability $1 - o(1)$

$$\max_{1 \le j \le p} |\widehat{W}_j - \widetilde{W}_j| \lesssim \underbrace{\left(\rho_n \sqrt{\frac{\log p}{n}}\right)}_{\text{coupling error of } \widehat{X}} \times \underbrace{\left(s\sqrt{n^{-1}\log p}\right)}_{L_1 \text{ estimation error of } \widehat{\beta}} .$$

Example 2: RCD Knockoff Statistics Coupling

Conditions

- It holds that $|\mathcal{H}_1|^{-1} \sum_{j \in \mathcal{H}_1} \mathbb{P}(\widetilde{W}_j < -t) \leq G(t)$ for all $t \in (0, C_3\sqrt{n^{-1} \log p})$ with $C_3 > 0$ some large constant.

- $a_n := \left|\{j \in \mathcal{H}_1 : |\beta_j| \gg \sqrt{n^{-1} \log p}\}\right| \to \infty.$

Theorem 2. Under the above two conditions and some regularity conditions on sparsity and restricted eigenvalues, if $s\rho_n(\log p)^{3/2+1/\gamma} = o(\sqrt{n})$ with $\gamma \in (0, 1)$, we have

$$\limsup_{n \to \infty} FDR(\hat{S}) \leq q.$$

Example 2: Compare with Existing Result

- Barber Candés and Samworth (2020): Theorem 1 therein requires an independent unlabeled training data set with sample size N satisfying $N \gg n\rho_n(\log p)^2$ for ensuring $\limsup_{n\to\infty} \mathrm{FDR} \leq q$
- Under the setting of linear regression model and RCD knockoff statistics, our theory requires $s\rho_n(\log p)^{3/2+1/\gamma}/\sqrt{n} \to 0$ for $\limsup_{n\to\infty} \mathrm{FDR} \leq q$ for some $0 < \gamma < 1$
- Our technical analysis do not require data splitting or an independent training sample