High-Dimensional Knockoffs Inference (part I)

Jinchi Lv
Data Sciences and Operations Department
Marshall School of Business
University of Southern California

http://faculty.marshall.usc.edu/jinchi-lv

USC Summer School on Uncertainty Quantification (08/09/2024)

Outline of Fan, Demirkaya, Li and L. (2020)

1) Graphical nonlinear knockoffs (Fan, Demirkaya, Li and L., 2020)

Applications

A primary goal in modern data analysis is to identify the important predictors in a sea of noise variables, e.g.,

- In economics, researchers are interested in which demographic/socioeconomic variables affect future income
- In the technology industry, people seek out specific software characteristics they can change to increase user engagement
- In political science, people want to study which demographic or socioeconomic variables determine political opinions

Problem setup

- Given response Y and p covariates $X_1, \cdots, X_p$, we aim to identify relevant covariates $\mathcal{S}_0$

- $\mathcal{S}_0$: the smallest set such that Y is independent of $X_{\mathcal{S}_0^c}$ given $X_{\mathcal{S}_0}$
  - Related to the concept of Markov blanket (Pearl, 1988, Section 3.2.1)

- Formulated as multiple hypothesis testing:

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

- More explicitly, aim to control FDR

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

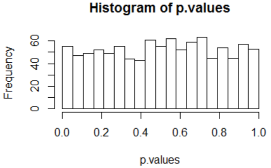
Existing work

Most existing work relies on p-value (e.g., Benjamini and Hochberg, 1995; Benjamini, 2010; Benjamini and Yekutieli, 2001; ...)

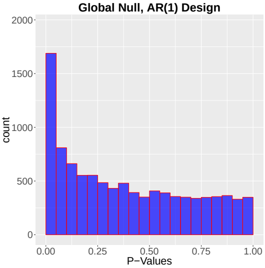
- BH procedure: sort p-values in ascending order and then choose a cutoff such that hypotheses with p-values below the cutoff are rejected
- very popularly used
- theoretically guaranteed to control FDR under p-value independence and certain forms of dependence

Potential problem with 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 L., 2018)



Theoretical characterization

Consider GLM model with regression coefficient $\boldsymbol{\beta}_0$.

Theorem (Fan, Demirkaya and L., 2019)

1. * *Under some regularity conditions, if* $p = o(n^{1/2})$, *the MLE* $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \cdots, \widehat{\beta}_p)^T$ *satisfies that*

$$\sqrt{n}(\widehat{\beta}_j - \beta_{0,j}) \to_d N(0, \sigma_j^2) \text{ for each } j = 1, \cdots, p.$$

2. *Under global null* $\boldsymbol{\beta}_0 = \mathbf{0}$ *and correlated Gaussian design* $N(\mathbf{0}, \boldsymbol{\Sigma}_0)$, *if* $p = O(n^{\alpha_0})$ *with* $\alpha_0 \in [0, 2/3)$, *then*

$$\sqrt{n}\widehat{\beta}_j \to_d N(0, \sigma_j^2) \text{ for each } j = 1, \cdots, p.$$

3. *In logistic regression model, under global null and correlated Gaussian design, the asymptotic normalities in 2) above fail to hold when* $p \sim n^{2/3}$.

**Fan and L. (2011), van de Geer et al. (2014), Javanmard and Montanari (2014), ...*

Caution when using p-values based on MLE

In GLMs, if one wants to use p-values based on MLE for testing $H_{0,j} : \beta_{0j} = 0$

- When $p = o(n^{1/2})$, p-value is asymptotically valid
- Under global null of $\boldsymbol{\beta}_0 = \mathbf{0}$, the exact breakdown point is $p \sim n^{2/3}$

Remark: For GLMs, Sur, Chen and Candès (2017) derived the asymptotic distribution of LRT when $p/n \to \gamma$ with $\gamma < 1/2$ under global null $\boldsymbol{\beta}_0 = \mathbf{0}$

The knockoff filter

Bypass the use of p-values

- Fix-X knockoffs (Barber and Candès, 2015 & 2016)
  - Originally introduced the knockoff filter
  - Geometric construction of knockoff variables
  - Gaussian linear model
- Model-X knockoffs (Candès, Fan, Janson and L., 2018): probabilistic construction of knockoffs
  - A new read of the original knockoff filter
  - 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 $\mathbf{x} = (X_1, \cdots, X_p)$ is known
- Theoretically guaranteed to achieve finite-sample FDR control

Model-X knockoffs framework (*Candès, Fan, Janson and L., 2018*)

Intuition:

- Generate "fake" copies of original covariates which are irrelevant to Y but mimic the dependence structure of original covariates

- Act as controls for assessing importance of original variables

Model-X knockoff variables

Definition (Candès, Fan, Janson and L., 2018)

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

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

$$(\mathbf{x}', \widetilde{\mathbf{x}}')_{\text{swap}(S)} \overset{d}{=} (\mathbf{x}', \widetilde{\mathbf{x}}')$$

- $\widetilde{\mathbf{x}} \perp\!\!\!\perp Y | \mathbf{x}$

The knockoffs procedure

(1) Construct model-X knockoff variables using the joint distribution of $\mathbf{x}$

(2) Compute knockoff statistics W_j's

- Positive W_j: original more important, strength measured by magnitude
- Null variables: W_j should be symmetric around 0

(3) Find the knockoff threshold:

- Order the variables by decreasing $|W_j|$ and proceed down list
- Select only variables with positive W_j exceeding some threshold $\hat{\tau}$

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

Choice of threshold

Intuition of FDR control

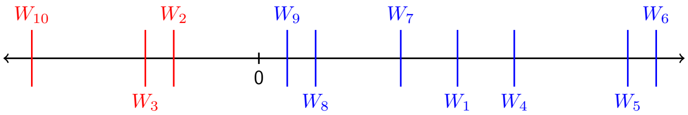
$$\begin{aligned}\text{FDR} &= E\left[\frac{\#\text{selected null variables}}{\#\text{selected variables}}\right] \\ &= E\left[\frac{\#\{\text{null } W_j \geq \hat{\tau}\}}{\#\{W_j \geq \hat{\tau}\}}\right] \\ &\approx E\left[\frac{\#\{\text{null } -W_j \geq \hat{\tau}\}}{\#\{W_j \geq \hat{\tau}\}}\right]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

This suggests to choose the threshold $\hat{\tau}$ by examining the ratio

$$\frac{\#\{-W_j \geq \hat{\tau}\}}{\#\{W_j \geq \hat{\tau}\}}$$

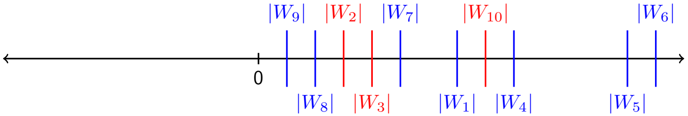
Example

10 variables, target FDR $q = 0.2$



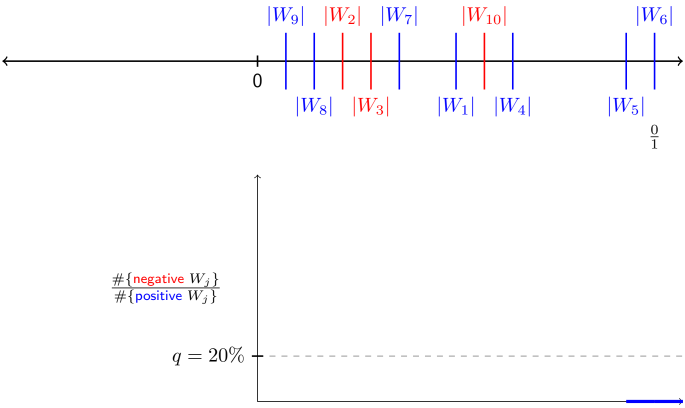
Example

10 variables, target FDR $q = 0.2$



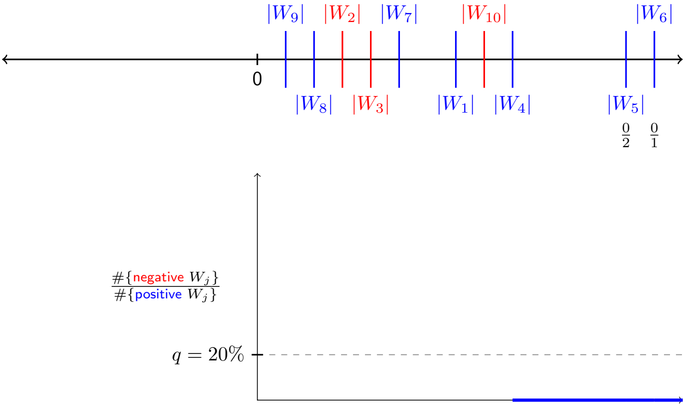
Example

10 variables, target FDR $q = 0.2$



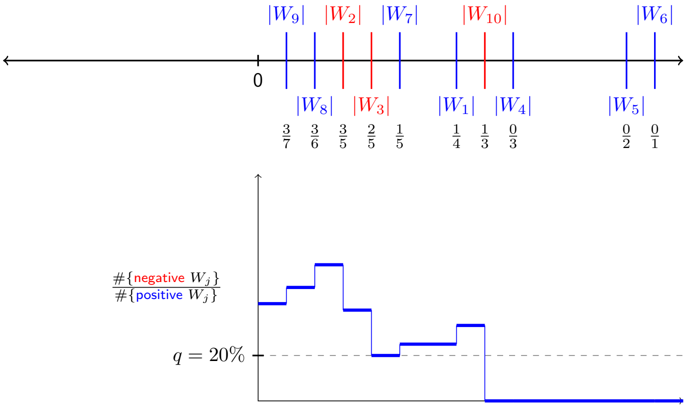
Example

10 variables, target FDR $q = 0.2$



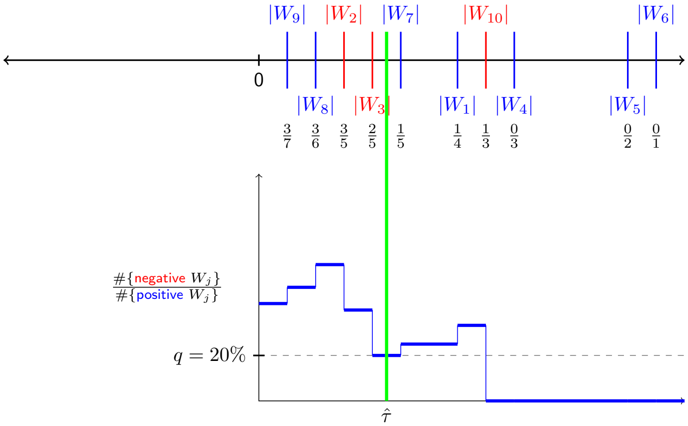
Example

10 variables, target FDR $q = 0.2$



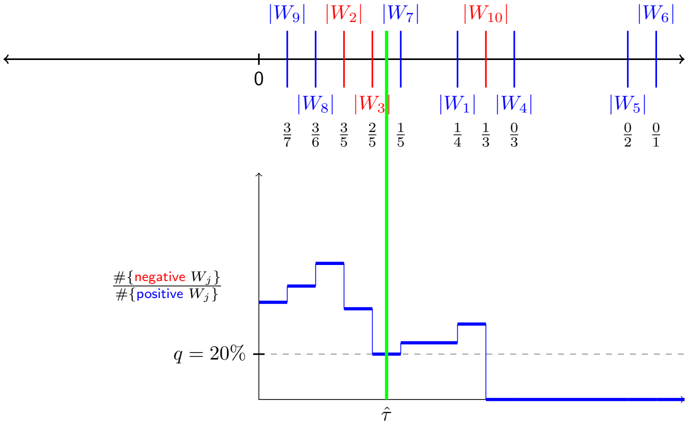
Example

10 variables, target FDR $q = 0.2$



Example

10 variables, target FDR $q = 0.2$



Set of selected variables = $\{1,4,5,6,7\}$

Summary

- Review:
  - Arbitrary dependence structure of y on $\mathbf{x}$
  - Arbitrary dimensionality
  - Exact finite sample FDR control (has been proved)
  - Need to know the joint distribution of $\mathbf{x}$ in order to construct valid knockoff variables
  - Can be regarded as a wrapper

- What are missing?
  - Power justification*
  - Implementable knockoff variable construction
  - Robustness analysis to unknown covariate distribution

---

*Weinstein, Barber and Candès (2017): Gaussian linear model with i.i.d. Gaussian design

Overview

- Outline
  - Asymptotic power analysis for model-X knockoffs
  - RANK: a graphical nonlinear knockoff filter
  - Robustness analysis of RANK to estimation error

Gaussian graphic model

- If $\mathbf{x} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0)$, then $\widetilde{\mathbf{x}}$ can be generated according to

$$\begin{pmatrix} \mathbf{x} \\ \widetilde{\mathbf{x}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_0 - \mathrm{diag}\{\mathbf{s}\} \\ \boldsymbol{\Sigma}_0 - \mathrm{diag}\{\mathbf{s}\} & \boldsymbol{\Sigma}_0 \end{pmatrix}\right),$$

  or equivalently,

$$\widetilde{\mathbf{x}}|\mathbf{x} \sim N\Big(\mathbf{x} - \mathrm{diag}\{\mathbf{s}\}\boldsymbol{\Sigma}_0^{-1}\mathbf{x}, 2\mathrm{diag}\{\mathbf{s}\} - \mathrm{diag}\{\mathbf{s}\}\boldsymbol{\Sigma}_0^{-1}\mathrm{diag}\{\mathbf{s}\}\Big), \quad (1)$$

  where $\mathrm{diag}\{\mathbf{s}\}$ is a diagonal matrix controlling the power; nuisance parameters.

  Assume implicitly that $2\mathrm{diag}\{\mathbf{s}\} - \mathrm{diag}\{\mathbf{s}\}\boldsymbol{\Sigma}_0^{-1}\mathrm{diag}\{\mathbf{s}\}$ has smallest eigenvalue bounded below from 0.

Sesia, Sabatti and Candès (2017) extended model-X knockoffs to the setting when covariate distribution is HMM

Asymptotic power analysis – I

- Power depends on signal strength

- Focusing on linear model for easy characterization of signal strength

$$\mathbf{y} = \mathbf{X}\beta_0 + \varepsilon,$$

  - $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n\times p}$, and $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ with i.i.d. rows.
  - $s := |\mathrm{supp}(\boldsymbol{\beta}_0)| = o(n)$

- Construction of knockoff statistics
  - Regress $\mathbf{y}$ on augmented design matrix $[\mathbf{X}, \widetilde{\mathbf{X}}] \in \mathbb{R}^{n\times 2p}$ using Lasso

$$\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \cdots, \widehat{\beta}_{2p})^T = \arg\min_{\mathbf{b}\in\mathbb{R}^{2p}} \left\{ (2n)^{-1} \|\mathbf{y} - [\mathbf{X}, \widetilde{\mathbf{X}}]\mathbf{b}\|_2^2 + \lambda \|\mathbf{b}\|_1 \right\}.$$

  - LCD: $W_j = |\widehat{\beta}_j| - |\widehat{\beta}_{p+j}|$, $j = 1, \cdots, p$

  - The coin flipping property is satisfied

Asymptotic power analysis – II

Let $\widehat{\mathcal{S}}$ be the set of variables selected by knockoff filter. Then

$$\text{Power}(\widehat{\mathcal{S}}) = \mathbb{E}\Big[\frac{|\widehat{\mathcal{S}} \cap \text{supp}(\boldsymbol{\beta}_0)|}{|\text{supp}(\boldsymbol{\beta}_0)|}\Big]$$

Technical conditions:

- *Condition 1*. $\boldsymbol{\varepsilon}$ has i.i.d. sub-Gaussian components
- *Condition 2*. $\{n/(\log p)\}^{1/2} \min_{j \in \mathcal{S}_0} |\beta_{0,j}| \to \infty$ as n increases
- *Condition 3*. With asymptotic probability one, $|\widehat{\mathcal{S}}| \geq cs$ with some constant $c \in (2(qs)^{-1}, 1)$

Remark: Condition 2 is to ensure Lasso has asymptotic power 1

Asymptotic power analysis – III

Lemma (Fan, Demirkaya, Li and L., 2020)

Assume that Condition 1 holds and there exists some constant $c \in (2(qs)^{-1}, 1)$ such that $|\{j : |\beta_{0,j}| \gg [sn^{-1}(\log p)]^{1/2}\}| \geq cs$. Then Condition 3 holds.

Theorem (Fan, Demirkaya, Li and L., 2020)

Under Conditions 1–3 and some other regularity conditions, if $\log p = o(n)$, with asymptotic probability one, we have $\mathrm{Power}(\widehat{\mathcal{S}}) \to 1$ as $n \to \infty$

Remark: The results can be easily generalized to non-Gaussian design case

Graphical Nonlinear Knockoffs

- From now on, focus on $\mathbf{x} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0)$ with unknown $\boldsymbol{\Sigma}_0$.
  - Still allows for arbitrary dependence of y on $\mathbf{x}$
  - Still allows for high dimensionality of $p \gg n$
- Challenges:
  - ✓ Estimation of $\boldsymbol{\Omega}_0 = \boldsymbol{\Sigma}_0^{-1}$ when $p \gg n$
  - ? Knockoff variables are only approximate (no coin flipping property)
  - ? Is FDR still under control with approximate knockoff variables?
  - ? How does it affect power?

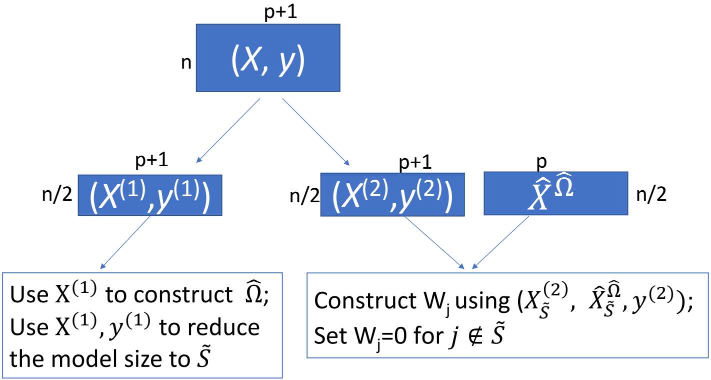
Estimation of precision matrix $\mathbf{\Omega}_0$

- Large literature on this; Glasso (Friedman et. al, 2008), CLIME (Cai et. al, 2011), ISEE (Fan and L., 2016), ...

- In our numerical analysis, we use ISEE (Fan and L., 2016)
  - Main idea: convert the problem of precision matrix estimation into that of covariance matrix estimation by the *innovated* transformation

- For our theory, consider the following class of estimators
  - *Condition 4*. Assume that $\widehat{\mathbf{\Omega}}$ satisfies $\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2 \leq C_2 a_n$ with probability $1 - O(p^{-c_1})$ for some $C_2, c_1 > 0$ and $a_n \to 0$.

The RANK procedure for graphical nonlinear knockoffs



Why data splitting?

- Conjecture: only a technical assumption

- Main challenges in proofs:
  - Coin flipping property is violated; original proof does not apply
  - $\widetilde{\mathbf{X}}^{\mathbf{\Omega}_1} \in \mathbb{R}^{n\times p}$ and $\widetilde{\mathbf{X}}^{\mathbf{\Omega}_2} \in \mathbb{R}^{n\times p}$ are not close *in distribution* even if $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ are close
- Solution:
  - Reduce the dimensionality to $\widetilde{\mathcal{S}}$ using half of the data
  - Use $\widetilde{\mathbf{X}}_{\widetilde{\mathcal{S}}}^{\mathbf{\Omega}_0}$ as a bridge and show

$$\mathrm{FDR}(\mathbf{\Omega}; \widetilde{\mathcal{S}}) \approx \mathrm{FDR}(\mathbf{\Omega}_0; \widetilde{\mathcal{S}}) \text{ for } \mathbf{\Omega} \approx \mathbf{\Omega}_0$$

  - Prove $\mathrm{FDR}(\mathbf{\Omega}_0; \widetilde{\mathcal{S}}) \leq q$ with q some target FDR level (Independence of $\widetilde{\mathcal{S}}$ and $\mathbf{X}^{(2)}$ is crucial for ensuring coin flipping property in this step!)

Connection with Barber and Candès (2016)

- Data splitting was used in Barber and Candès (2016) for fixed-X knockoffs in Gaussian linear model when $p > n$

- Main differences:

BC16

- Gaussian linear model
- Need sure screening property for dimension reduction step for FDR control

RANK

- Arbitrary dependence structure of y on $\mathbf{x}$
- No screening property needed for FDR control
- Might be just a technical assumption

Robustness of FDR

Technical conditions: $\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_2 \le C_2 a_n$ with probability $1 - O(p^{-c_1})$; the reduced model size $|\widetilde{S}| \le K_n$.

Theorem (Fan, Demirkaya, Li and L., 2020)

Under some regularity conditions, it holds that

$$\sup_{|\mathcal{S}| \le K_n,\ \|\boldsymbol{\Omega} - \boldsymbol{\Omega}_0\|_2 \le C_2 a_n} |\mathrm{FDR}_n(\boldsymbol{\Omega}, \mathcal{S}) - \mathrm{FDR}(\boldsymbol{\Omega}_0, \mathcal{S})| \le O(K_n^{1/2} a_n).$$

Moreover, if $K_n^{1/2} a_n \to 0$,

$$\mathrm{FDR}_n(\widehat{\boldsymbol{\Omega}}, \widetilde{\mathcal{S}}) \le q + O(K_n^{1/2} a_n) + O(p^{-c_1}),$$

with $q \in (0, 1)$ *target FDR level.*

Remark: FDR control is with respect to the original model instead of the reduced model.

Robustness of Power

- Back to linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}$ for easy characterization of signal strength
- Lasso is used as the underlying variable selection method
- Focus on relative power loss because
  - Power of knockoffs $\leq$ Power of Lasso
- WLOG, assume the sure screening property $P(\widetilde{S} \supset \mathrm{supp}(\boldsymbol{\beta}_0)) \to 1$ as $n \to \infty$ to simplify technical proof
- Remark: without sure screening property, similar conclusion is still true because the model can be regarded as projection

Robustness of Power – Continued

Some additional conditions

- *Condition 5*: $\mathbf{\Omega}_0$ is L_p-sparse; all the eigenvalues of $\mathbf{\Omega}_0$ are bounded away from 0 and ∞
- *Condition 6*: With probability $1 - O(p^{-c_2})$, $\widehat{\mathbf{\Omega}}$ is L_p'-sparse and $\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2 \leq C_2 a_n$
- *Condition 7*: $|\{j : |\beta_{0,j}| \gg [sn^{-1}(\log p)]^{1/2}\}| \geq cs$

Theorem (Fan, Demirkaya, Li and L., 2020)

Under Conditions 1–2 and 5–7 and some growth conditions on (a_n, K_n, L_p, L_p'), if $\log p = o(n^a)$, then it holds that RANK with estimated precision matrix $\widehat{\mathbf{\Omega}}$ and reduced model $\widetilde{S}$ has asymptotic power one.

Model settings

- Focus on Gaussian design $\mathbf{x} \sim N(\mathbf{0}, \mathbf{\Sigma}_0)$ for easy generation of knockoff variables
- Linear model: $\mathbf{y} = \mathbf{X}\beta_0 + \varepsilon$
- Partially linear model: $\mathbf{y} = \mathbf{X}\beta_0 + \mathbf{g}(\mathbf{U}) + \varepsilon$
- Single-index model: $\mathbf{y} = \mathbf{g}(\mathbf{X}\beta_0) + \varepsilon$
- Additive model: $\mathbf{y} = \sum_{j=1}^{p} \mathbf{g}_j(\mathbf{X}_j) + \varepsilon$
- $n = 400$ in all settings

How to choose $\operatorname{diag}\{\mathbf{s}\}$ *in generating knockoff variables?*

- Barber and Candès (2015) suggested
  - *equicorrelated* construction $s_j^{\mathrm{EQ}} = 2\lambda_{\min}(\boldsymbol{\Sigma}) \wedge 1$ for all j
  - *semidefinite program (SDP)* construction

$$\text{minimize} \sum_j |1 - s_j^{\mathrm{SDP}}| \qquad \text{subject to} \quad s_j^{\mathrm{SDP}} \geq 0, \quad \operatorname{diag}\{s^{\mathrm{SDP}}\} \preceq 2\boldsymbol{\Sigma}$$

- We (Candès, Fan, Janson and L., 2018) suggested scalable construction – approximate semidefinite program (ASDP) construction:

  Step 1. Choose an approximation $\boldsymbol{\Sigma}_{\text{approx}}$ of $\boldsymbol{\Sigma}$ and solve:

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

  Step 2. maximize γ subject to $\operatorname{diag}\{\gamma\hat{s}\} \preceq 2\boldsymbol{\Sigma}$, and set $s^{\mathrm{ASDP}} = \gamma\hat{s}$. Can be solved quickly by bisection search over $\gamma \in [0, 1]$.

Note: Each column of $\mathbf{X}$ *is standardized to have norm 1*

Simulation Results – linear model

Table 1: Linear model with $A = 1.5$ and $s = 30$; "+" stands for knockoff$_+$ filter; "s" stands for with data splitting

		RANK		RANK$_+$		RANKs		RANKs$_+$	
ρ	p	FDR	Power	FDR	Power	FDR	Power	FDR	Power
0	200	0.2054	1.00	0.1749	1.00	0.1909	1.00	0.1730	1.00
	400	0.2062	1.00	0.1824	1.00	0.2010	1.00	0.1801	1.00
	600	0.2263	1.00	0.1940	1.00	0.2206	1.00	0.1935	1.00
	800	0.2385	1.00	0.1911	1.00	0.2247	1.00	0.1874	1.00
	1000	0.2413	1.00	0.2083	1.00	0.2235	1.00	0.1970	1.00
0.5	200	0.2087	1.00	0.1844	1.00	0.1875	1.00	0.1692	1.00
	400	0.2144	1.00	0.1879	1.00	0.1954	1.00	0.1703	1.00
	600	0.2292	1.00	0.1868	1.00	0.2062	1.00	0.1798	1.00
	800	0.2398	1.00	0.1933	1.00	0.2052	0.9997	0.1805	0.9997
	1000	0.2412	1.00	0.2019	1.00	0.2221	0.9984	0.2034	0.9984

Continued

Table 2: Linear model with $A = 3.5$ and $s = 30$; "+" stands for knockoff$_+$ filter; "s" stands for with data splitting; "HKF" stands for Barber and Candès (2016) approach

		RANKs		RANKs$_+$		HKF		HKF$_+$	
ρ	p	FDR	Power	FDR	Power	FDR	Power	FDR	Power
0	200	0.1858	1.00	0.1785	1.00	0.1977	0.9849	0.1749	0.9837
	400	0.1895	1.00	0.1815	1.00	0.2064	0.9046	0.1876	0.8477
	600	0.2050	1.00	0.1702	1.00	0.1964	0.8424	0.1593	0.7668
	800	0.2149	1.00	0.1921	1.00	0.1703	0.7513	0.1218	0.6241
	1000	0.2180	1.00	0.1934	1.00	0.1422	0.7138	0.1010	0.5550
0.5	200	0.1986	1.00	0.1618	1.00	0.1992	0.9336	0.1801	0.9300
	400	0.1971	1.00	0.1805	1.00	0.1657	0.8398	0.1363	0.7825
	600	0.2021	1.00	0.1757	1.00	0.1253	0.7098	0.0910	0.6068
	800	0.2018	1.00	0.1860	1.00	0.1374	0.6978	0.0917	0.5792
	1000	0.2097	0.9993	0.1920	0.9993	0.1552	0.6486	0.1076	0.5524

Simulation results – partially linear model

Table 3: Partially linear model with $s = 30$; "+" stands for knockoff$_+$ filter; "s" stands for with data splitting

		RANK		RANK$_+$		RANKs		RANKs$_+$	
ρ	p	FDR	Power	FDR	Power	FDR	Power	FDR	Power
0	200	0.2117	1.00	0.1923	1.00	0.1846	0.9976	0.1699	0.9970
	400	0.2234	1.00	0.1977	1.00	0.1944	0.9970	0.1747	0.9966
	600	0.2041	1.00	0.1776	1.00	0.2014	0.9968	0.1802	0.9960
	800	0.2298	1.00	0.1810	1.00	0.2085	0.9933	0.1902	0.9930
	1000	0.2322	1.00	0.1979	1.00	0.2113	0.9860	0.1851	0.9840
0.5	200	0.2180	1.00	0.1929	1.00	0.1825	0.9952	0.1660	0.9949
	400	0.2254	1.00	0.1966	1.00	0.1809	0.9950	0.1628	0.9948
	600	0.2062	1.00	0.1814	1.00	0.2038	0.9945	0.1898	0.9945
	800	0.2264	1.00	0.1948	1.00	0.2019	0.9916	0.1703	0.9906
	1000	0.2316	1.00	0.2033	1.00	0.2127	0.9830	0.1857	0.9790

Simulation results – single-index model

Table 4: Single-index model with $s = 10$; "+" stands for knockoff$_+$ filter; "s" stands for with data splitting

		RANK		RANK$_+$		RANKs		RANKs$_+$	
ρ	p	FDR	Power	FDR	Power	FDR	Power	FDR	Power
0	200	0.1893	1	0.1413	1	0.1899	1	0.1383	1
	400	0.2163	1	0.1598	1	0.245	0.998	0.1676	0.997
	600	0.2166	1	0.1358	1	0.2314	0.999	0.1673	0.998
	800	0.1964	1	0.1406	1	0.2443	0.992	0.1817	0.992
	1000	0.2051	1	0.134	1	0.2431	0.969	0.1611	0.962
0.5	200	0.2189	1	0.1591	1	0.2322	1	0.1626	1
	400	0.2005	1	0.1314	1	0.2099	0.996	0.1615	0.995
	600	0.2064	1	0.1426	1	0.2331	0.998	0.1726	0.998
	800	0.2049	1	0.1518	1	0.2288	0.994	0.1701	0.994
	1000	0.2259	1	0.1423	1	0.2392	0.985	0.185	0.983

Simulation results – additive model

Table 5: Additive model with $s = 10$; "+" stands for knockoff$_+$ filter; "s" stands for with data splitting

		RANK		RANK$_+$		RANKs		RANKs$_+$	
ρ	p	FDR	Power	FDR	Power	FDR	Power	FDR	Power
0	200	0.1926	0.9780	0.1719	0.9690	0.2207	0.9490	0.1668	0.9410
	400	0.2094	0.9750	0.1773	0.9670	0.2236	0.9430	0.1639	0.9340
	600	0.2155	0.9670	0.1729	0.9500	0.2051	0.9310	0.1620	0.9220
	800	0.2273	0.9590	0.1825	0.9410	0.2341	0.9280	0.1905	0.9200
	1000	0.2390	0.9570	0.1751	0.9350	0.2350	0.9140	0.1833	0.9070
0.5	200	0.1904	0.9680	0.1733	0.9590	0.2078	0.9370	0.1531	0.9330
	400	0.2173	0.9650	0.1701	0.9540	0.2224	0.9360	0.1591	0.9280
	600	0.2267	0.9600	0.1656	0.9360	0.2366	0.9340	0.1981	0.9270
	800	0.2306	0.9540	0.1798	0.9320	0.2332	0.9150	0.1740	0.9110
	1000	0.2378	0.9330	0.1793	0.9270	0.2422	0.8970	0.1813	0.8880

Simulation Results

- RANK and RANK_+ mimic closely RANKs and RANKs_+, suggesting that data splitting is more of a technical assumption
- FDR approximately controlled at target level of $q = 0.2$ with high power, which is in line with our theory
- Despite that both RANKs and HKF (for linear model) are based on data splitting, their practical performance is very different
- These results demonstrate model-free feature of our procedure for large-scale inference in nonlinear models

Conclusions

- Model-X knockoffs is a powerful, flexible, and robust solution for FDR control in high-dimensional variable selection
- Can be regarded as a “wrapper” that may be combined with any variable selection methods
- Extensions
  - Investigate how to generate model-X knockoff variables in more general settings (*IPAD* Fan, L., Sharifvaghefi and Uematsu, 2020; ...)
  - Integrate the idea of knockoffs inference with deep learning (*DeepPINK* Lu, Fan, L. and Noble, 2018; ...)

References

- Barber, R. F. and Candès, E. J. (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics* **43**, 2055–2085.
- Candès, E. J., 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* **80**, 551–577.
- Fan, Y., Demirkaya, E., Li, G. and Lv, J. (2020). RANK: large-scale inference with graphical nonlinear knockoffs. *Journal of the American Statistical Association* **115**, 362–379.
- Fan, Y., Demirkaya, E. and Lv, J. (2019). Nonuniformity of p-values can occur early in diverging dimensions. *Journal of Machine Learning Research* **20**, 1–33.