Power Analysis of Knockoff Filters for Correlated Designs

Jingbo Liu and Philippe Rigollet

Neurips 2019

Dec. 2019

Massachusetts Institute of Technology
Feature Selection Problem

Problem: Given samples from the observation model

\[ Y = \sum_{j=1}^{p} \theta_j X_j + N, \]

find \( j \) for which \( |\theta_j| > 0 \).

Example: finding a few interesting genes among a haystack of nulls.

False discovery rate: \( FDR = \mathbb{E} \left[ \frac{|\mathcal{H}_0 \cap \hat{\mathcal{H}}_1|}{|\hat{\mathcal{H}}_1|} \right] \)

- \( \mathcal{H}_0 \) null.
- \( \hat{\mathcal{H}}_1 \) declared true hypothesis.
Classical FDR Control: Benjamini-Hochberg

Idea: if $X$ is identity matrix and the null distribution is known, simply reject null hypothesis when $p$-value is small.

How to do more general cases?
Using LASSO?

Passing a threshold test for $\hat{\theta}$ should work.

But what threshold to use?

Figure from [Barber, Candès 2015]
Knockoff Filter of Barber Candès

**Idea:** generate knockoff (fake) variables $\tilde{X}^p$ known to be uncorrelated with $Y$, so as to estimate the right threshold to apply to the $\hat{\theta}$ computed from Lasso.
Knockoff Filter of Barber Candès

**Idea:** generate knockoff (fake) variables $\tilde{X}^p$ known to be uncorrelated with $Y$, so as to estimate the right threshold to apply to the $\hat{\theta}$ computed from Lasso.

But... simply generating $\tilde{X}^p$ independent of $X^p$ won’t work...
Knockoff Filter of Barber Candès

Idea: generate knockoff (fake) variables $\tilde{X}^p$ known to be uncorrelated with $Y$, so as to estimate the right threshold to apply to the $\hat{\theta}$ computed from Lasso.

But... simply generating $\tilde{X}^p$ independent of $X^p$ won’t work...

Surprisingly, the following always bounds FDR [Barber, Candès 2015]:

- Design **knockoff mechanism** $P_{\tilde{X}^p | X^p}$ satisfying exchangeability

\[(X^p, \tilde{X}^p)_{\text{swap}(S)} = (X^p, \tilde{X}^p), \quad \forall S \subseteq [p].\]

- Generate knockoff samples $(\tilde{X}_{i,j})_{i=1\ldots n, j=1\ldots p}$.
- Solve LASSO for $Y$ and $[X, \tilde{X}]$.
- Use the lower tail of the symmetric statistic $W_j := |\beta_j| - |\beta_{j+p}|$ to determine the threshold.
Figure from [Barber, Candès 2015]
Big problem: How to Design the Knockoff Mechanism $P_{\tilde{X}^p|X^p}$?

Existing choices: [Barber, Candès 2015][Candès, Fan, Janson, Lv 2018]:
- Gaussian: exchangeability imposes that the covariance of $(X^p, \tilde{X}^p)$ has the form
  \[
  \begin{bmatrix}
  \Sigma & \Sigma - \text{diag}(s) \\
  \Sigma - \text{diag}(s) & \Sigma
  \end{bmatrix}
  \]
- **SDP-Knockoff**: Minimize $\|s\|_1$ by solving SDP
- **Equi-Knockoff**: Maximize $s_1 = \cdots = s_p$
- **ASDP-Knockoff**
- ...
- General: requires strong knowledge of $P_{X^p}$; not today

No prior theory to guide the choice
Power (Type II error) Analysis: Previous results

All knockoff mechanisms bound FDR; we want to find one to maximize the Power $\mathbb{E} \left[ \frac{|\mathcal{H}_1 \cap \hat{\mathcal{H}}_1|}{|\mathcal{H}_1|} \right]$.

Consistency results:
- Fixed $\Sigma$ and $n \to \infty$ [Fan et al. 2019].
- Fixed $n/p$, $X_{i,j} \sim \mathcal{N}(0, 1)$ i.i.d., and $\text{SNR} \to \infty$ [Weinstein, Barber, Candes 2017].

Bounds on power are complicated!
**Our Contribution 1**

**Definition**: For a given feature selection algorithm, Effective Signal Deficiency $ESD^{(p)}$ is a function of $\Sigma^{(p)}$ such that: for the class of sequences of $(\theta^{(p)}, \Sigma^{(p)})_{p\geq 1}$ admitting suitable distributional limits,

\[
\limsup_{p \to \infty} \max \{ FDR^{(p)}, 1 - POWER^{(p)} \} \to 0
\]

if and only if

\[
ESD := \limsup_{p \to \infty} ESD^{(p)} \to 0
\]
Our Contribution 1 (Continued)

Let \( \mathbf{P} = \mathbf{\Sigma}^{-1} \in \mathbb{R}^{2p \times 2p} \) be the extended precision matrix (including the predictors and the knockoff variables).

**Main Result:** if \( \delta > 1 \),

\[
ESD(p) = d_{\text{LP}} \left( \frac{1}{p} \sum_{j=1}^{p} \delta \mathbf{P}(p)_{jj}, \delta_0 \right)
\]

where \( d_{\text{LP}} \) is the Lévy-Prokhorov metric.

Remarks:
- ESD is an equivalence class, hence \( d_{\text{LP}} \) can be replaced by any metric compatible with the topology of weak convergence.
- Bounds available for \( \delta \leq 1 \).
Our Contribution 2

We introduce the **conditional independence knockoff** in which \( X_j \) and \( \tilde{X}_j \) are independent conditioned on \( X_{\setminus j}, j = 1, \ldots, p \).

- Well defined iff \( X^p \) is a walk-summable model (including all trees)
- Competitive statistical performance against previous computational-involving mechanisms:

![Graph 1](image1.png)

![Graph 2](image2.png)