Power and tuning for the knockoff filter

Brook Luers

Department of Statistics, University of Michigan

luers@umich.edu

The problem: variable selection in regression

$$Y = X\beta + \epsilon, \ \epsilon \sim N(0, I)$$

Observe X ($n \times p$ covariate matrix) and Y (outcome vector).

β : p-dimensional parameter

Which covariates x_j are relevant?

Identify nonzero entries of β from Y, X

The knockoff filter: intuition

(Barber and Candès [2015])

Step 1: construct "knockoffs" $\tilde{X} := (\tilde{x}_1, \ldots, \tilde{x}_p)$

- $\tilde{X}$ mimics correlations of X
- $\tilde{X}$ known to have no effect in the regression
- $\tilde{X}$ correlated with X

Sample correlations:

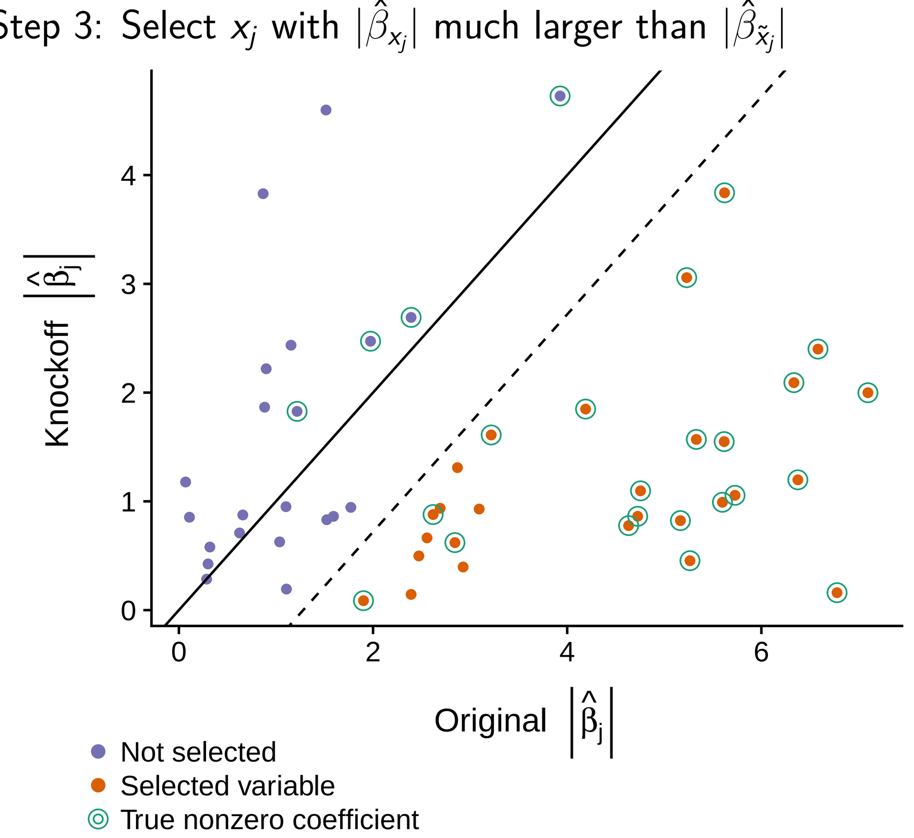
	x_2	x_3	x_4	x_5
x_1	0.589	0.347	0.212	0.138
x_2	1.000	0.579	0.341	0.197

	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$\tilde{x}_5$
$\tilde{x}_1$	0.589	0.347	0.212	0.138
$\tilde{x}_2$	1.000	0.579	0.341	0.197

	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$\tilde{x}_5$
x_1	0.589	0.347	0.212	0.138
x_2	0.635	0.579	0.341	0.197

Step 2: fit the "augmented" regression using $(X, \tilde{X})$

Step 3: Select x_j with $|\hat{\beta}_{x_j}|$ much larger than $|\hat{\beta}_{\tilde{x}_j}|$



If $|\hat{\beta}_{x_j}| \approx |\hat{\beta}_{\tilde{x}_j}|$:

- suggests x_j no better than $\tilde{x}_j$
- $\tilde{x}_j$ known to have no effect

FDR control:

- the knockoff filter limits the expected fraction of selected x_j which are, in fact, null in the model $Y = X\beta + \epsilon$

Summary

The knockoff filter (Barber and Candès [2015]) is a variable selection technique for linear regression with finite-sample control of the false discovery rate (the expected proportion of selected variables which, in fact, have no effect in the regression model). To control the FDR, the knockoff filter constructs synthetic variables which mimic the observed covariates but are known to be irrelevant to the regression. Constructing these synthetic variables involves tuning parameters which can increase collinearity and reduce power. We propose alternative tuning choices for the knockoff filter which limit collinearity and improve power in ridge and ordinary least squares regressions.

False discovery rate (FDR)

p hypotheses $H_j : \beta_j = 0$

reject $H_j \implies$ select x_j as relevant in the regression

$$\text{FDR} = \text{E}\left(\frac{\#\{\text{selected } x_j \text{ with } \beta_j = 0\}}{\max\left[1, \#\{\text{selected } x_j\}\right]}\right)$$

FDR: expected fraction of selected x_j that are actually null in the regression

For a chosen $q \in (0, 1)$, the knockoff filter guarantees FDR $\leq q$.

The knockoff filter: details

Denote $\Sigma := X^\mathsf{T} X$

Step 1:
Construct knockoffs, $\tilde{X}$ ($n \times p$), so that G is positive semidefinite:

$$G := \begin{bmatrix} X & \tilde{X} \end{bmatrix}^\mathsf{T} \begin{bmatrix} X & \tilde{X} \end{bmatrix} = \begin{bmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{bmatrix}$$

$$\implies \tilde{x}_j^\mathsf{T} x_j = 1 - s_j, \ \tilde{x}_j^\mathsf{T} x_k = x_j^\mathsf{T} x_k, \ j \neq k$$

$S = \text{diag}(s_1, \ldots, s_j)$ where $0 \leq s_j \leq 1$

$\mathbf{s}$: p-dimensional **tuning parameter**

Step 2:
Compute $W_1, \ldots, W_p$, so that

- $(W_1, \ldots, W_p) = f\left(G, [X \ \tilde{X}]^\mathsf{T} Y\right)$
- swapping x_j and $\tilde{x}_j$ changes the sign of W_j
- large, positive W_j suggest that $\beta_j \neq 0$

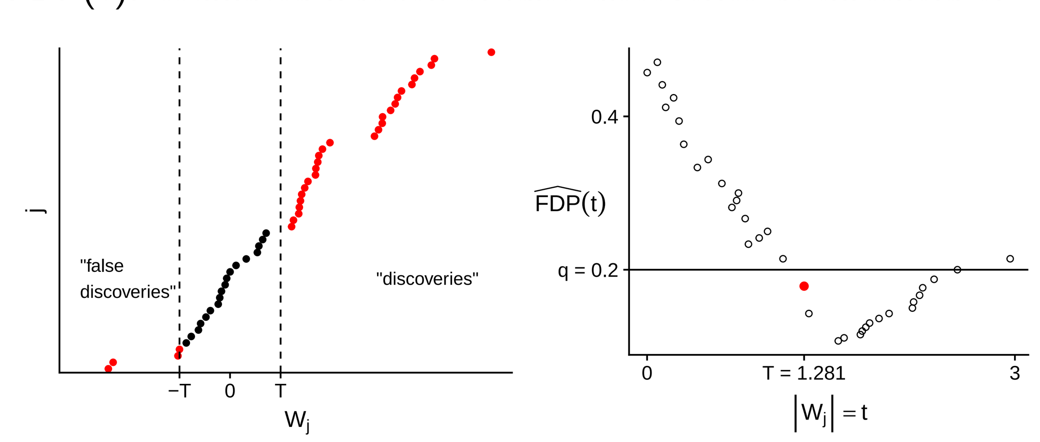
e.g. $W_j = |\hat{\beta}_j| - |\hat{\beta}_{j+p}|$ using OLS coefficients

Step 3:
Given $q \in (0, 1)$, select variables j such that $W_j \geq T$,

$$T = \min_{|W_j|}\left\{t = |W_j| : \widehat{\text{FDP}}(t) \leq q\right\}.$$

$$\widehat{\text{FDP}}(t) := \frac{1 + \#\{k : W_k \leq -t\}}{\max\{1, \#\{k : W_k \geq t\}\}}$$

$\widehat{\text{FDP}}(t)$: estimated fraction of false discoveries at threshold t.



Standard tuning

Heuristic: increase s_j to reduce $x_j^\mathsf{T} \tilde{x}_j = 1 - s_j$

"SDP" knockoffs:

$$\min \sum_j (1 - s_j)$$

subject to $0 \leq s_j \leq 1, \ 2\Sigma - S \succeq 0$
(constraints guarantee G p.s.d)

"Equicorrelated" knockoffs: $s_j = \min\{1, 2\lambda_{\min}(\Sigma)\}$ for all j

If (smallest eigenvalue) $\lambda_{\min}(\Sigma) \leq \frac{1}{2}$, then

- equicorrelated knockoffs produce singular G
- SDP knockoffs produce singular G when
  - any $s_j = 0$ or
  - $2\Sigma - S$ has a null eigenvalue.

Standard tuning leads to linearly dependent $(X \ \tilde{X})$ unless $\Sigma = X^\mathsf{T} X$ nearly orthogonal

Proposal: improved tuning

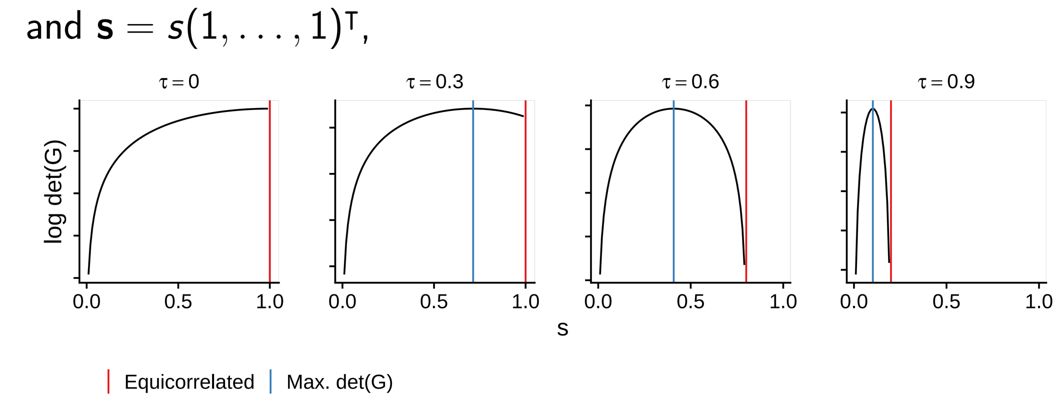
Choose $\mathbf{s}$ to reduce collinearity: maximize $\det(G)$

- $\det(G) = \det(S)\det(2\Sigma - S)$
- $\frac{\partial \log \det(G)}{\partial s_j} = \frac{1}{s_j} - \left[(2\Sigma - S)^{-1}\right]_{j,j}$
- $\log \det(G)$ is convex in the vector $\mathbf{s} \in \mathbb{R}^p$

e.g. if Σ has an exchangeable structure

$$\Sigma = \begin{pmatrix} 1 & \tau & \tau & \cdots \\ \tau & 1 & \tau & \cdots \\ \vdots & & & \ddots \end{pmatrix},$$

and $\mathbf{s} = s(1, \ldots, 1)^\mathsf{T}$,



Smaller $s \implies$ larger correlation $x_j^\mathsf{T} \tilde{x}_j = 1 - s_j$

Proposed tuning: improve conditioning of G, allow larger $x_j^\mathsf{T} \tilde{x}_j$

Simulation study

Investigate power tradeoffs among tuning choices for $\mathbf{s}$

$n = 3000, p = 100$

$k = 50$ nonzero β_j with $|\beta_j| = 3.5$

Three population structures, $X \sim N(0, \Gamma)$,

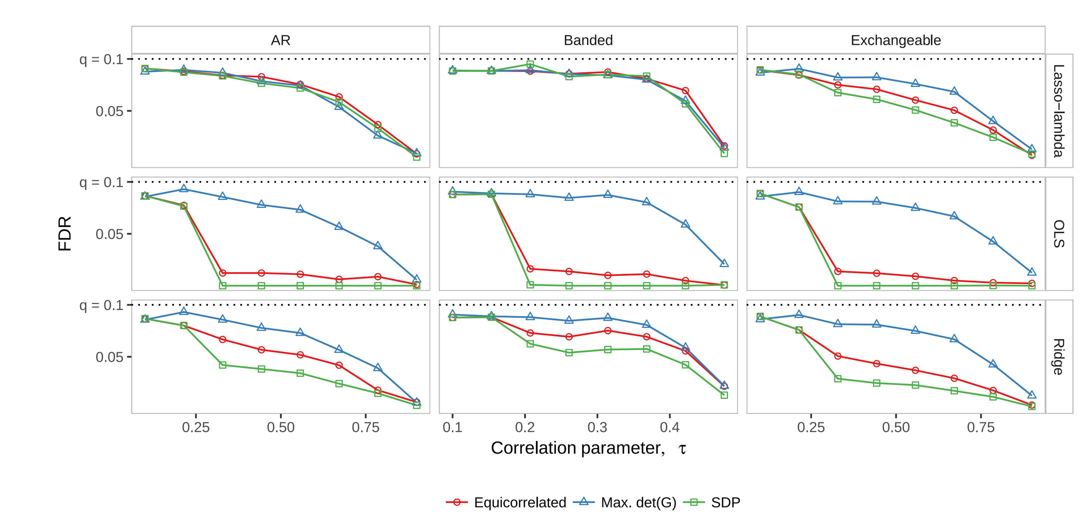
$$\Gamma = \text{Exchangeable:} \begin{pmatrix} 1 & \tau & \tau & \cdots \\ \tau & 1 & \tau & \cdots \\ \vdots & & \ddots & \end{pmatrix}, \text{Banded:} \begin{pmatrix} 1 & \tau & 0 & 0 & \cdots \\ \tau & 1 & \tau & 0 & \cdots \\ 0 & \tau & 1 & \tau & 0 \end{pmatrix}$$

$$\text{Autoregressive (AR):} \begin{pmatrix} 1 & \tau & \tau^2 & \tau^3 & \cdots \\ \tau & 1 & \tau & \tau^2 & \cdots \\ \tau^2 & \tau & 1 & \tau & \tau^2 \end{pmatrix}$$

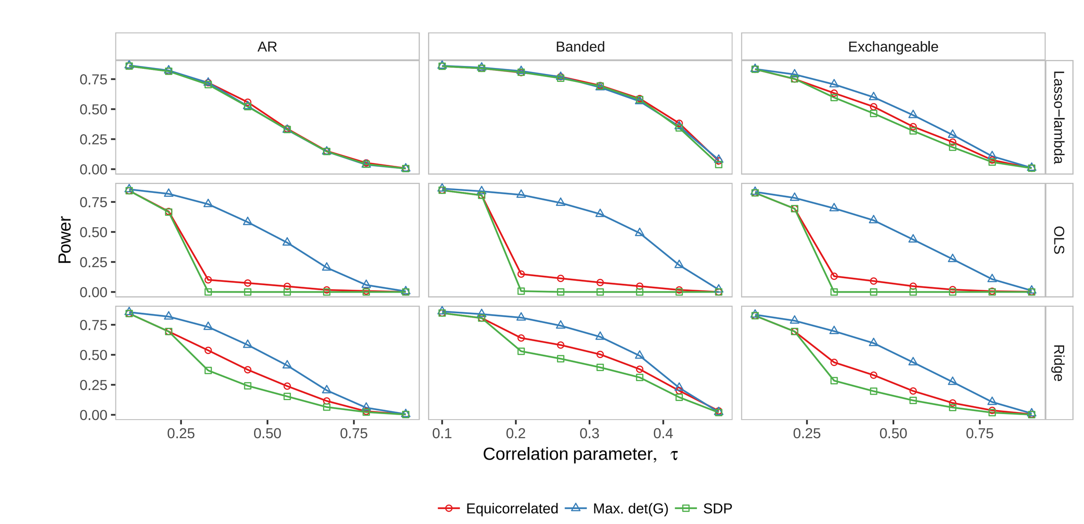
Three different W_j:

- the difference in lasso tuning parameters at which a coefficient enters the lasso model
- OLS: $|\hat{\beta}_j| - |\hat{\beta}_{j+p}|$ (use generalized inverse when G singular)
- ridge, GCV: $|\hat{\beta}_j(\lambda)| - |\hat{\beta}_{j+p}(\lambda)|$

Nominal FDR: $q = 0.1$



$$\text{Power} = \text{E}\left(\frac{\#\{\text{selected } j : \beta_j \neq 0\}}{\#\{j : \beta_j \neq 0\}}\right)$$



Maximum-determinant $\mathbf{s}$:

- no loss of power in any scenario
- improved power using ridge or OLS W_j

References

Rina Foygel Barber and Emmanuel J. Candès. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43(5):2055–2085, 2015.