Using Probabilistic Knockoffs of Binary Variables to Control the False Discovery Rate

Aaron Maurer

July 29th, 2015

Overview

1. Original Knockoffs: What They Do and Where They Fail
2. Making Knockoffs Work With GLMs
3. Random Binary Knockoffs: The Theory
4. Random Binary Knockoffs: Performance
5. Where to next?

Variable Selection in Linear Regression

Assume

$$\mathbf{y} = X\beta + \mathbf{z}$$

where $\mathbf{y} \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times p}$, $\beta \in \mathbb{R}^p$, and $\mathbf{z}$ is Gaussian noise. Also, assume sparsity:

$$\beta_i = 0 \quad \forall i \notin S$$

How do we pick estimate $\hat{S}$?

False Discover Rate

A common goal for a method that generates $\hat{S}$ is to control the false discovery rate

$$\mathrm{FDR} = \mathrm{E}\left[\frac{|\{j : \beta_j = 0 \;\&\; j \in \hat{S}\}|}{\max\{|\hat{S}|, 1\}}\right]$$

In other words, control portion of elements in $\hat{S}$ which aren't in S.

FDR is controlled at level q if $q <$FDR irrespective of true β.

Knockoffs

Knockoff variables can be used to control FDR in linear regression.

- The idea is to create a forgery of each variable; if the forgeries seem about as good predictors as the originals, the originals are lousy predictors.
- For each variable X_i, create a knockoff feature $\tilde{X}_i$. Such that.

$$\tilde{X}^T \tilde{X} = X^T X \quad \& \quad X^T \tilde{X} = X^T X - \mathrm{diag}\{\mathbf{s}\}$$

  Where $\mathrm{diag}\{X^T X\} - s$ is small but $\mathrm{diag}\{X^T X\} - s \succeq 0$
- $\tilde{X}_i$ and X_i will have same correlation with other variables, but only low correlation with each other.
- Given $\mathbf{s}$, $\tilde{X}$ can be generated via a rotation of X.

Knockoff Filter

These knockoffs can be used in the knockoff filter method.

- Fit full path of LASSO regression on $[\, X \,\tilde{X} \,]$.

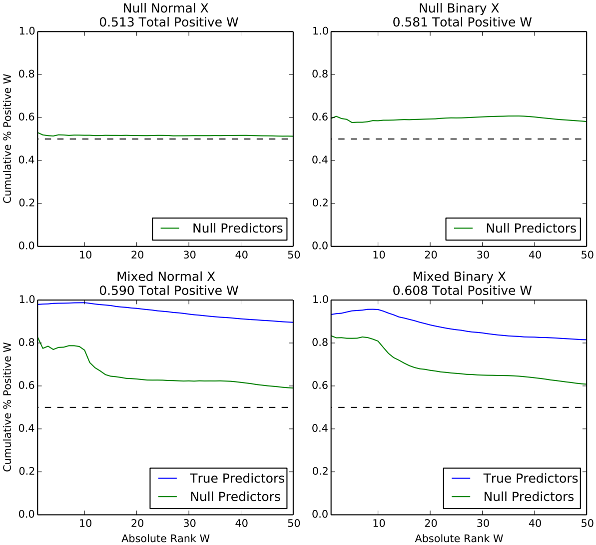
$$\beta(\lambda) = \arg\min_{\mathbf{b}} \left\{ \frac{1}{2} \|\mathbf{y} - X_L \mathbf{b}\|_2^2 + \lambda \|b\|_1 \right\}$$

- Z_i, $\tilde{Z}_i$ largest λ such that X_i, $\tilde{X}_i$ have nonzero coefficient.
- $W_i = Z_i$ if $Z_i > \tilde{Z}_i$, otherwise $W_i = -\tilde{Z}_i$.
- Since $[\, X \,\tilde{X} \,]^T [\, X \,\tilde{X} \,]$ & $[\, X \,\tilde{X} \,]^T \mathbf{y}$ sufficient statistics for $\beta(\lambda)$, W_i symmetrically distributed around 0 when X_i null predictor.
- Thus, FDR controlled when $\hat{S} = \{i : W_i > T\}$ for

$$T = \min \left\{ t > 0 \ : \ \frac{|\{j : W_j \leq -t\}|}{\max\{|\{j : W_j \geq t\}|, 1\}} \leq q \right\}$$

Where Knockoff Filter Fails

Knockoff filter don't work for other GLMs.



Can Knockoffs Be Fixed for GLMs?

- Other GLMs don't have the same sufficient statistics as linear regression.
- Original Knockoffs don't remotely have same distribution as X, so "look" different than real variables.
- Knockoffs will likely work better if they have the same marginal distribution as originals.
- For X_i with arbitrary distribution, unclear how this might be accomplished.

Random Binary Notation

- Binary data is common in data analysis and a much more manageable family of distributions.
- We can think of observations in X as observations of random binary vector $\mathbf{x} \in \{0, 1\}^p$.
- The full family for $\mathbf{x}$ is multinomial on 2^p outcomes.
- Still useful to consider first two moments:

$$\mathrm{E}(\mathbf{x}) = \mathbf{m} \in [0, 1]^p \quad \& \quad \mathrm{E}(\mathbf{x}\mathbf{x}^T) = M \in [0, 1]^{p \times p}$$

- For arbitrary M to correspond to a random binary vector, must be case that $M - \mathbf{m}\mathbf{m}^T = \Sigma \succeq 0$

$$\max\{0, m_i + m_j - 1\} \leq M_{ij} \leq \min\{m_i, m_j\}$$

Random Binary Knockoffs

- Integer programing is np-hard, making finding finding $\tilde{X} \in \{0,1\}^{n \times p}$ to fit correlations exactly difficult.
- Instead, introduce a relaxed problem where $\tilde{X} \mid X$ is drawn randomly such that, where $\Sigma = \mathrm{Cov}(\mathbf{x})$

$$\mathrm{Cov}(\tilde{\mathbf{x}}, \mathbf{x}) = \Sigma - \mathrm{diag}\{\mathbf{s}\} \quad \& \quad \mathrm{Cov}(\tilde{\mathbf{x}}) = \Sigma$$

- Almost same correlation condition as before, just only holds in expectation.
- Switch from Gramian matrix to correlation matrix makes moment condition less likely to be violated.

Quadratic Programing

- Simplest approach to Random Binary Knockoffs is to draw the entries of $\tilde{X}$ independently based on $P \in [0,1]^{n \times p}$.
- The best possible P for the task would satisfy

$$\begin{array}{rl} \text{minimize} & \|X^T P - (M - \mathrm{diag}\{s\})\|_{fro}^2 + \sum_{i \neq j} (P_i^T P_j - M_{ij})^2 \\ \text{subject to} & \mathbf{1}^T P = \mathbf{m} \\ & 0 \leq P \leq 1 \end{array}$$

- Can be formulated as a quadratic program with slack variables

$$\begin{array}{rl} \text{minimize} & \|W\|_{fro}^2 + \|V\|_{fro}^2 \\ \text{subject to} & -W \leq X^T P - (M - \mathrm{diag}\{s\}) \leq W \\ & -V_{ij} \leq P_i^T P_j - M_{ij} \leq V_{ij} \quad \forall i \neq j \\ & \mathbf{1}^T P = \mathbf{m} \\ & 0 \leq P \leq 1 \end{array}$$

- Huge optimization problem, likely computationally impractical.