Metropolized Knockoff Sampling

Stephen Bates[†], Emmanuel Candès[†‡], Lucas Janson[*] and Wenshuo Wang[*]

[†]Department of Statistics, Stanford University [‡]Department of Mathematics, Stanford University [*]Department of Statistics, Harvard University

Overview

The knockoff filter [1] is a flexible framework which enables variable selection with rigorous finite-sample statistical guarantees. A remaining challenge is the construction of knockoff distributions and sampling mechanisms across a wide range of covariate models.

1. We provide a sequential characterization of **every** valid knockoff distribution.
2. We introduce a class of algorithms which use conditional independence information to **efficiently** generate knockoffs.
3. We develop a **concrete** and **easy-to-use** knockoff sampler for a large number of distributions with a family of MCMC tools.

Background

Figure 1: Illustration of the model-X knockoff filter [1].

- **Non-asymptotic** guarantees, valid in **high dimensions**
- **No assumption** on $\mathcal{L}(Y|X)$
- **Any** black-box feature importance measure
- **Knockoff sampling**: generate *knockoffs*, random variables $\tilde{X} \in \mathbb{R}^p$ for $X \in \mathbb{R}^p$ such that for each $j = 1, \ldots, p$,

$$(X, \tilde{X})_{\text{swap}(j)} \overset{d}{=} (X, \tilde{X}); \tag{1}$$

swap(j) means permuting X_j and $\tilde{X}_j$. For example, $(X_1, X_2, X_3, \tilde{X}_1, \tilde{X}_2, \tilde{X}_3)_{\text{swap}(2)}$ is the vector $(X_1, \tilde{X}_2, X_3, \tilde{X}_1, X_2, \tilde{X}_3)$.

SCIP, the only generic knockoff sampler prior to this work [1], has two substantial limitations.

- Only known for very special models (e.g., discrete Markov chains [3] and Gaussian distributions [1])
- Not able to cover all valid knockoff distributions

Procedure 1: Sequential Conditional Independent Pairs (SCIP)

for $j = 1$ **to** p **do**
  Sample $\tilde{X}_j$ from $\mathcal{L}(X_j \mid X_{\text{-}j}, \tilde{X}_{1:(j-1)})$, conditionally independently from X_j
end

Sequential Formulation

Theorem 1: Sequential characterization of knockoff distributions

Pairwise exchangeability (1) holds if and only if both of the following conditions hold:

Conditional exchangeability For each $j \in \{1, \ldots, p\}$,

$$(X_j, \tilde{X}_j) \mid X_{\text{-}j}, \tilde{X}_{1:(j-1)} \overset{d}{=} (\tilde{X}_j, X_j) \mid X_{\text{-}j}, \tilde{X}_{1:(j-1)}. \tag{2}$$

Knockoff symmetry For each $j \in \{1, \ldots, p\}$ and any Borel set A,

$$\mathbb{P}((X_j, \tilde{X}_j) \in A \mid X_{\text{-}j}, \tilde{X}_{1:(j-1)}) \tag{3}$$

does not change if we swap previously sampled knockoffs with the original features.

- Equation (2) resembles a time-reversible Markov chain (MC).

The Metropolized Knockoff Sampler

- SCEP: a **completely general** strategy for generating knockoffs

Procedure 2: Sequential Conditional Exchangeable Pairs (SCEP)

for $j = 1$ **to** p **do**
  Sample $\tilde{X}_j$ by taking one step of a time-reversible MC starting from X_j.
  The transition kernel must be *faithful*, i.e., it depends on the previous pairs symmetrically so knockoff symmetry (3) holds and must admit $\mathcal{L}(X_j \mid X_{\text{-}j}, \tilde{X}_{1:(j-1)})$ as a stationary distribution.
end

- Time-reversible MC. With stationary distribution π, **Metropolis–Hastings** (MH) operates as follows: generate proposal x^* from some distribution $q(\cdot \mid x)$ and set

$$y = \begin{cases} x^* & \text{with prob. } \alpha, \\ x & \text{with prob. } 1 - \alpha, \end{cases} \qquad \alpha = \min\left(1, \frac{\pi(x^*)q(x \mid x^*)}{\pi(x)q(x^* \mid x)}\right).$$

- **Challenge**: $\pi(x)$, $\pi(x^*)$ hard to compute; e.g., the target $\mathcal{L}(X_2 \mid X_{\text{-}2}, \tilde{X}_1)$ has density proportional to $\mathbb{P}(X = x)\mathbb{P}(\tilde{X}_1 = \tilde{x}_1 \mid X = x)$:

$$\mathbb{P}(X = x)\Bigg[q(\tilde{x}_1 \mid x_1) \min\left(1, \frac{q(x_1 \mid \tilde{x}_1)\mathbb{P}(X_1 = \tilde{x}_1, X_{\text{-}1} = x_{\text{-}1})}{q(\tilde{x}_1 \mid x_1)\mathbb{P}(X_1 = x_1, X_{\text{-}1} = x_{\text{-}1})}\right)$$
$$+ \delta(\tilde{x}_1 - x_1)\underbrace{\int q(x^* \mid x_1)\left(1 - \min\left(1, \frac{q(x_1 \mid x^*)\mathbb{P}(X_1 = x^*, X_{\text{-}1} = x_{\text{-}1})}{q(x^* \mid x_1)\mathbb{P}(X_1 = x_1, X_{\text{-}1} = x_{\text{-}1})}\right)\right) dx^*}_{\text{intractable integral}}\Bigg].$$

- **Solution**: *condition on the proposals.* Let the target be $\mathcal{L}(X_j \mid X_{\text{-}j}, \tilde{X}_{1:j-1}, X^*_{1:j-1})$ rather than $\mathcal{L}(X_j \mid X_{\text{-}j}, \tilde{X}_{1:j-1})$; returning to the previous example, $\mathcal{L}(X_2 \mid X_{\text{-}2}, \tilde{X}_1, X^*_1)$ has density now proportional to

$$\mathbb{P}(X = x)q(x^*_1 \mid x_1)\Bigg[\delta(\tilde{x}_1 - x^*_1) \min\left(1, \frac{q(x_1 \mid \tilde{x}_1)\mathbb{P}(X_1 = \tilde{x}_1, X_{\text{-}1} = x_{\text{-}1})}{q(\tilde{x}_1 \mid x_1)\mathbb{P}(X_1 = x_1, X_{\text{-}1} = x_{\text{-}1})}\right)$$
$$+ \delta(\tilde{x}_1 - x_1)\left(1 - \min\left(1, \frac{q(x_1 \mid x^*_1)\mathbb{P}(X_1 = x^*_1, X_{\text{-}1} = x_{\text{-}1})}{q(x^*_1 \mid x_1)\mathbb{P}(X_1 = x_1, X_{\text{-}1} = x_{\text{-}1})}\right)\right)\Bigg].$$

The **intractable integral goes away**! Now we introduce the main algorithm.

Algorithm 1: **Metropolized knockoff sampling (Metro)**.

for $j = 1$ **to** p **do**
  Sample $X^*_j = x^*_j$ from a faithful proposal distribution q_j.
  Accept the proposal with probability

$$\min\left(1, \frac{q_j(x_j|x^*_j)\mathbb{P}\big(X_{\text{-}j}=x_{\text{-}j}, X_j=x^*_j, \tilde{X}_{1:(j-1)}=\tilde{x}_{1:(j-1)}, X^*_{1:(j-1)}=x^*_{1:(j-1)}\big)}{q_j(x^*_j|x_j)\mathbb{P}\big(X_{\text{-}j}=x_{\text{-}j}, X_j=x_j, \tilde{X}_{1:(j-1)}=\tilde{x}_{1:(j-1)}, X^*_{1:(j-1)}=x^*_{1:(j-1)}\big)}\right).$$

  Upon acceptance, set $\tilde{x}_j = x^*_j$; otherwise, set $\tilde{x}_j = x_j$.
end
Return $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_p)$

- Two **general** and **concrete** methods for choosing proposal distributions.
  - **Covariance-guided proposals**. Choose proposals pretending X is Gaussian (known, [1]) and let MH do the correction.
  - **Multiple-try Metropolis**. A clever way to include multiple proposals in one step for higher probability to accept [2].

Graphical Structure and Time Complexity

Theorem 2: Complexity lower bound for knockoff sampling

If a knockoff sampling procedure is given the support of X and is only allowed to make queries of the unnormalized density of X, then the total number N of queries of the unnormalized density must obey $N \geq 2^{\#\{j: X_j \neq \tilde{X}_j\}} - 1$ a.s.

- **Graphical model**. Let $X \in \mathbb{R}^p$ be a random vector whose density factors over a graph G:

$$\mathbb{P}(x) \propto \Phi(x) = \prod_{c \in C} \phi_c(x_c); \tag{4}$$

here, C is the set of maximal cliques of graph G and Φ is unnormalized version of $\mathbb{P}$.

- **Junction tree**. A junction tree of a graph provides a way of ordering the variables so they behave like high-order Markov chains. See our paper for the algorithm!

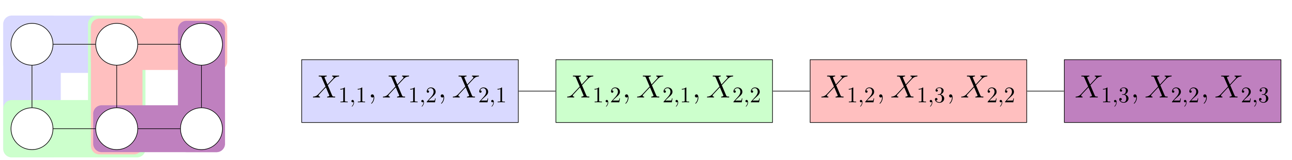

Figure 2: A junction tree of treewidth 2 for the 2×3 grid, which happens to be a chain.

Theorem 3: Computational efficiency of Metro

Let X be a random vector with a density which factors over a graph G as in (4). Let T be a junction tree of width w (the size of the largest vertex of T minus one) for the graph G. Under the conditions above, Metro uses $O(p2^w)$ queries of Φ.

Numerical Experiments

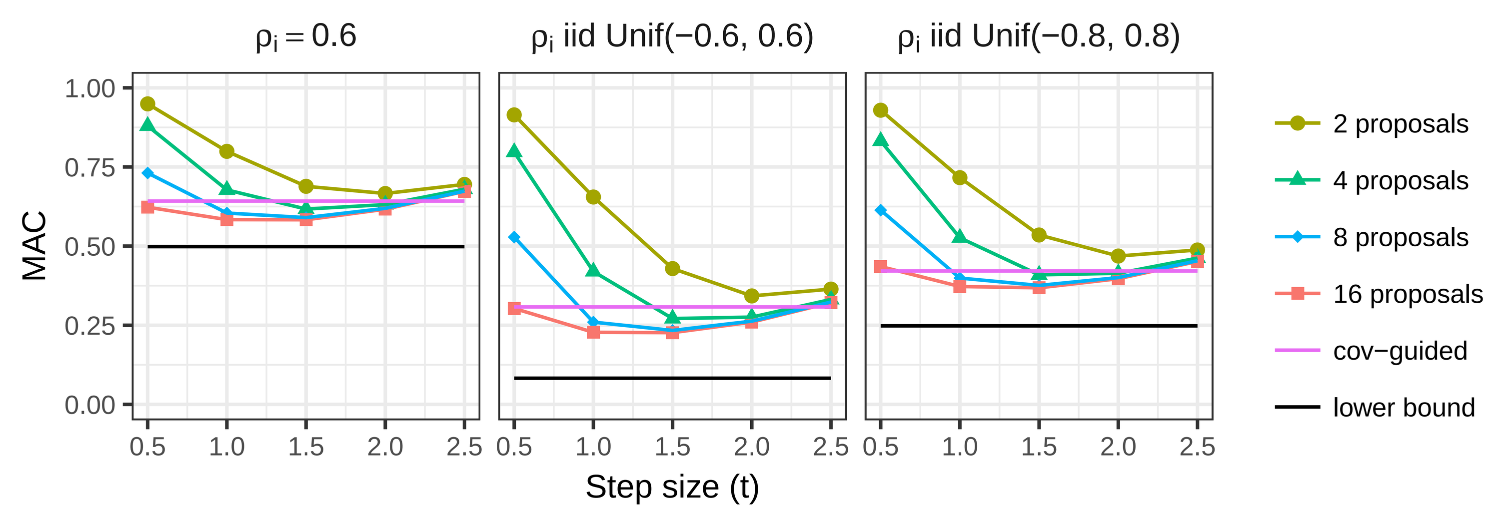

Figure 3: Simulation results for the t-distributed Markov chains. The unit of step sizes is $\sqrt{1/(\boldsymbol{\Sigma}^{-1})_{jj}}$. All standard errors are below 0.001. The mean absolute correlation (MAC) is defined as the average of $|\text{corr}(X_j, \tilde{X}_j)|$ from $j = 1$ to p. ***Many more simulations in the paper!***

References

[1] E. Candès, Y. Fan, L. Janson, and J. Lv. Panning for gold: Model-X knockoffs for high-dimensional controlled variable selection. *Journal of the Royal Statistical Society: Series B*, 80(3):551–577, 2018.

[2] J. S. Liu, F. Liang, and W. H. Wong. The multiple-try method and local optimization in Metropolis sampling. *Journal of the American Statistical Association*, 95(449):121–134, 2000.

[3] M. Sesia, C. Sabatti, and E. J. Candès. Gene hunting with hidden Markov model knockoffs. *Biometrika*, 106(1):1–18, 08 2018. doi: 10.1093/biomet/asy033.

Authors are listed in alphabetical order. E. C. was partially supported by the Office of Naval Research under grant N00014-16- 1-2712, by the National Science Foundation via DMS 1712800, and by a generous gift from Two Sigma. S. B. was supported by a Ric Weiland Graduate Fellowship. S. B. and E. C. would like to thank Yaniv Romano and Matteo Sesia for useful comments on an early version of this work. L. J. and W. W. would like to thank Jun Liu for fruitful discussions on MCMC.

wenshuo_wang@g.harvard.edu