# New perspectives on knockoffs construction

Patrizia Berti [a], Emanuela Dreassi [b], Fabrizio Leisen [c,*], Luca Pratelli [d], Pietro Rigo [e]

[a] *Dipartimento di Matematica Pura ed Applicata "G. Vitali", Università di Modena e Reggio-Emilia, via Campi 213/B, 41100 Modena, Italy*
[b] *Dipartimento di Statistica, Informatica, Applicazioni, Università di Firenze, viale Morgagni 59, 50134 Firenze, Italy*
[c] *School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK*
[d] *Accademia Navale, viale Italia 72, 57100 Livorno, Italy*
[e] *Dipartimento di Scienze Statistiche "P. Fortunati", Università di Bologna, via delle Belle Arti 41, 40126 Bologna, Italy*

## ARTICLE INFO

## ABSTRACT

Let $\Lambda$ be the collection of all probability distributions for $(X, \widetilde{X})$, where $X$ is a fixed random vector and $\widetilde{X}$ ranges over all possible knockoff copies of $X$ (in the sense of Candes et al. (2018)). Three topics are developed in this paper: (i) A new characterization of $\Lambda$ is proved; (ii) A certain subclass of $\Lambda$, defined in terms of copulas, is introduced; (iii) The (meaningful) special case where the components of $X$ are conditionally independent is treated in depth. In real problems, after observing $X = x$, each of points (i)–(ii)–(iii) may be useful to generate a value $\widetilde{x}$ for $\widetilde{X}$ conditionally on $X = x$.

© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

The availability of massive data along with new scientific problems have reshaped statistical thinking and data analysis. High-dimensionality has significantly challenged the boundaries of traditional statistical theory, in particular in the regression framework. Variable selection methods are fundamental to discover meaningful relationships between an outcome and all the measured covariates.

A new approach to regression problems, hereafter referred to as the *knockoff procedure* (KP), has been recently introduced by Barber and Candes; see Barber and Candes (2015), Barber et al. (2020), Bates et al. (2021), Candes et al. (2018), Sesia et al. (2019). KP aims to control the false discovery rate among all the variables included in the model. Indeed, KP is relevant for at least two reasons. Firstly, there are not many variable selection methods able to control the false discovery rate with finite-sample guarantees, mainly when the number $p$ of covariates far exceeds the sample size $n$. Secondly, KP makes assumptions that are substantially different from those commonly encountered in a regression set up.

---

* Corresponding author.
  *E-mail address:* fabrizio.leisen@gmail.com (F. Leisen).

Let $X_i$ and $Y$ be real random variables, where $i = 1, \ldots, p$ for some integer $p \geq 2$. Here, the $X_i$ should be regarded as covariates and $Y$ as the response variable. Letting

$$X = (X_1, \ldots, X_p),$$

one of the main features of KP is to model the probability distribution of $X$ rather than the conditional distribution of $Y$ given $X$. Quoting from Candes et al. (2018, p. 554):

*The usual set-up for inference in conditional models is to assume a strong parametric model for the response conditional on the covariates, such as a homoscedastic linear model, but to assume as little as possible about, or even to condition on, the covariates. We do the exact opposite by assuming that we know everything about the covariate distribution but nothing about the conditional distribution $Y|X_1, \ldots, X_p$. Hence, we merely shift the burden of knowledge. Our philosophy is, therefore, to model X, not Y, whereas, classically, Y (given X) is modeled and X is not.*

Real situations where to model $X$ is more appropriate than to model $Y|X$ are actually common. An effective example, in a genetic framework, is in Barber et al. (2020, Sect. 1).

As highlighted, the main target of KP is variable selection, taking the false discovery rate under control. We refer to Barber and Candes (2015), Barber et al. (2020), Bates et al. (2021), Candes et al. (2018), Sesia et al. (2019) for a description of KP and a discussion of its statistical behavior. In this paper, we deal with knockoff random variables, possibly the basic ingredient of KP.

### 1.1. Two related problems

From now on, the probability distribution of any random element $U$ is denoted by $\mathcal{L}(U)$ and the coordinates of a point $x \in \mathbb{R}^n$ are indicated by $x_1, \ldots, x_n$. Moreover, we let

$$I = \{1, \ldots, p\}.$$

For $x \in \mathbb{R}^{2p}$ and $S \subset I$, define $f_S(x) \in \mathbb{R}^{2p}$ by swapping $x_i$ with $x_{p+i}$ for each $i \in S$ and leaving all other coordinates fixed. Then, $f_S : \mathbb{R}^{2p} \to \mathbb{R}^{2p}$ is a permutation. For instance, for $p = 2$, one obtains $f_S(x) = (x_3, x_2, x_1, x_4)$ if $S = \{1\}$, $f_S(x) = (x_1, x_4, x_3, x_2)$ if $S = \{2\}$ and $f_S(x) = (x_3, x_4, x_1, x_2)$ if $S = \{1, 2\}$. Let

$$\mathcal{F} = \{f_S : S \subset I\}$$

where $f_\emptyset$ is the identity map on $\mathbb{R}^{2p}$.

A *knockoff copy of X*, or merely a *knockoff*, is a $p$-variate random variable $\widetilde{X} = (\widetilde{X}_1, \ldots, \widetilde{X}_p)$ such that

- $f(X, \widetilde{X}) \sim (X, \widetilde{X})$ for every $f \in \mathcal{F}$;
- $\widetilde{X} \perp\!\!\!\perp Y \mid X$.

The condition $\widetilde{X} \perp\!\!\!\perp Y \mid X$ is automatically true if $\widetilde{X} = g(X)$ for some measurable function $g$. More generally, such a condition is guaranteed whenever $\widetilde{X}$ is constructed "without looking" at $Y$. This is exactly the case of this paper. Hence, the condition $\widetilde{X} \perp\!\!\!\perp Y \mid X$ is neglected.

We also note that a knockoff trivially exists. It suffices to let $\widetilde{X} = X$. This trivial knockoff, however, is not useful in practice. Roughly speaking, for KP to work nicely, $\widetilde{X}$ should be "as independent of $X$ as possible".

Let $\Lambda$ denote the collection of all knockoff distributions, namely

$$\Lambda = \{\mathcal{L}(X, \widetilde{X}) : \widetilde{X} \text{ a knockoff copy of } X\}.$$

For KP to apply, a knockoff copy $\widetilde{X}$ of $X$ is required. Accordingly, the following two problems arise:

(i) How to build a knockoff $\widetilde{X}$ ?
(ii) Is it possible to characterize $\Lambda$ ?

Questions (i) and (ii) are connected. A characterization of $\Lambda$, if effective, should suggest how to obtain $\widetilde{X}$. Anyhow, both (i) and (ii) have been answered.

As to (i), a first construction of $\widetilde{X}$ is Algorithm 1 of Candes et al. (2018, p. 563). Even if nice, however, this construction is not computationally efficient except from some special cases. See Bates et al. (2021, Sect. 2.2), Candes et al. (2018, Sect. 7.2.1 ), Sesia et al. (2019, p. 6).

As to (ii), a characterization of $\Lambda$ is in Bates et al. (2021, Theo. 1). Such a characterization, based on conditional distributions, is effective. In fact, exploiting it and the Metropolis algorithm, some further (efficient) constructions of $\widetilde{X}$ have been singled out.

## 1.2. Our contribution

This paper is about problems (i)-(ii). Three distinct issues are discussed.

- In Section 2, a new characterization of $\Lambda$ is proved. Such a characterization is based on invariance arguments and provides a simple description of $\Lambda$. However, the characterization seems to have a theoretical content only. Apart from a few special cases, in fact, it does not help to build a knockoff in practice.
- In Section 3, a certain (proper) subclass $\Lambda_0 \subset \Lambda$ is introduced. The elements of $\Lambda_0$ admit a simple and explicit representation in terms of copulas. In particular, to work with $\Lambda_0$ is straightforward when $\mathcal{L}(X)$ corresponds to an Archimedean copula. Furthermore, if $\mathcal{L}(X, \widetilde{X}) \in \Lambda_0$, the conditional distribution $\mathcal{L}(\widetilde{X} \mid X)$ can be written in closed form. Therefore, after observing $X = x$, a value $\widetilde{x}$ for $\widetilde{X}$ can be drawn from $\mathcal{L}(\widetilde{X} \mid X = x)$ directly. This is quite different from the usual methods for obtaining $\widetilde{x}$; see e.g. Barber and Candes (2015), Barber et al. (2020), Bates et al. (2021), Candes et al. (2018), Sesia et al. (2019).
- In Section 4, we focus on the case where $X_1, \ldots, X_p$ are conditionally independent, in the sense that

$$P(X_1 \in A_1, \ldots, X_p \in A_p) = E\left\{\prod_{i=1}^{p} P(X_i \in A_i \mid Z)\right\} \tag{1}$$

for some random element $Z$ and all Borel sets $A_1, \ldots, A_p \subset \mathbb{R}$. This section includes our main results. Indeed, under (1), to build a reasonable knockoff $\widetilde{X}$ is straightforward. In addition, with a suitable choice of $Z$, to realize condition (1) is quite simple in practice. It suffices to regard $Z$ as a random parameter, equipped with a prior distribution, and to implement a sort of Bayesian procedure. For instance, to obtain a knockoff $\widetilde{X}$ such that $\operatorname{cov}(X_i, \widetilde{X}_i) = 0$ for each $i \in I$ is very easy; see Examples 16–18 for details. From the statistician's point of view, the advantage is twofold. Firstly, condition (1) is easy to be realized and able to describe a number of real situations. Secondly, if $X$ is modeled by (1), to build $\widetilde{X}$ is straightforward. In particular, as in Section 3, the conditional distribution $\mathcal{L}(\widetilde{X} \mid X)$ can be usually written in closed form.

To close the paper, in Section 5, the results mentioned above are translated into practical algorithms. This section, written with applications in mind, aims to show how such results can be exploited in real problems.

## 1.3. Further notation

In the sequel,

$$\widetilde{X} = (\widetilde{X}_1, \ldots, \widetilde{X}_p)$$

is *any* $p$-variate random variable (defined on the same probability space as $X$). Moreover, $\mathcal{B}_n$ is the Borel $\sigma$-field on $\mathbb{R}^n$ and $m_n$ the Lebesgue measure on $\mathcal{B}_n$.

As in Berti et al. (2022, Sect. 4), we denote by $\mathcal{S}(a, b)$ the *symmetric $\alpha$-stable law with parameters $a$ and $b$*, where $a \in \mathbb{R}$, $b > 0$ and $\alpha \in (0, 2]$. This means that $\mathcal{S}(a, b)$ is the probability distribution of $a + b^{1/\alpha} L$ where $L$ is a real random variable with characteristic function

$$E\left\{\exp(i t L)\right\} = \exp\left(-\frac{|t|^\alpha}{2}\right) \qquad \text{for all } t \in \mathbb{R}.$$

Note that $\mathcal{S}(a, b) = \mathcal{N}(a, b)$ if $\alpha = 2$ and $\mathcal{S}(a, b) = \mathcal{C}(a, b)$ if $\alpha = 1$, where $\mathcal{C}(a, b)$ is the Cauchy distribution with density $f(x) = \frac{2b}{\pi} \frac{1}{b^2 + 4(x-a)^2}$ (the standard Cauchy distribution corresponds to $a = 0$ and $b = 2$).

Finally, for any measures $\mu$ and $\nu$ (defined on the same $\sigma$-field) we write $\mu \ll \nu$ to mean that $\mu$ is absolutely continuous with respect to $\nu$, that is, $\mu(A) = 0$ whenever $A$ is measurable and $\nu(A) = 0$.

## 2. A characterization of $\Lambda$

Let $\mathcal{P}$ be the collection of $\mathcal{F}$-invariant probabilities, namely, those probability measures $\lambda$ on $\mathcal{B}_{2p}$ satisfying

$$\lambda \circ f^{-1} = \lambda \qquad \text{for all } f \in \mathcal{F}.$$

We begin by noting that $\lambda \in \mathcal{P}$ if and only if

$$\lambda = \frac{\sum_{f \in \mathcal{F}} \pi \circ f^{-1}}{2^p} \tag{2}$$

for *some* probability measure $\pi$ on $\mathcal{B}_{2p}$. In fact, if $\lambda \in \mathcal{P}$, condition (2) trivially holds with $\pi = \lambda$ (since $\operatorname{card}(\mathcal{F}) = 2^p$). Conversely, if $g \in \mathcal{F}$ and $\lambda$ meets (2) for some $\pi$, then

$$\lambda \circ g^{-1} = \frac{\sum_{f \in \mathcal{F}} \pi \circ f^{-1} \circ g^{-1}}{2^p} = \frac{\sum_{f \in \mathcal{F}} \pi \circ (g \circ f)^{-1}}{2^p} = \lambda$$

where the last equality is because $\mathcal{F}$ is a group under composition.

The above characterization of $\mathcal{P}$ has the following consequence.

**Theorem 1.** *$\lambda \in \Lambda$ if and only if condition* (2) *holds for some probability measure $\pi$ on $\mathcal{B}_{2p}$ such that*

$$\frac{1}{2^p} \sum_{f \in \mathcal{F}} \pi \left\{ x \in \mathbb{R}^{2p} : f(x) \in A \times \mathbb{R}^p \right\} = P(X \in A) \tag{3}$$

*for each $A \in \mathcal{B}_p$.*

**Proof.** If $\lambda \in \Lambda$, conditions (2)–(3) trivially hold with $\pi = \lambda$. Conversely, under (2)–(3), one obtains

$$\lambda(A \times \mathbb{R}^p) = \frac{\sum_{f \in \mathcal{F}} \pi \circ f^{-1}(A \times \mathbb{R}^p)}{2^p} = P(X \in A) \qquad \text{for all } A \in \mathcal{B}_p.$$

Hence, up to enlarging the probability space where $X$ is defined, there exists a $p$-variate random variable $\widetilde{X}$ such that $\mathcal{L}(X, \widetilde{X}) = \lambda$. Since $\lambda \in \mathcal{P}$ (because of condition (2)) $\widetilde{X}$ is a knockoff copy of $X$, namely, $\lambda \in \Lambda$. $\square$

From the theoretical point view, Theorem 1 provides a simple and clear description of $\Lambda$. Unfortunately, however, to select a probability $\pi$ satisfying condition (3) is very hard. Thus, in most cases, Theorem 1 is not practically useful. Nevertheless, it may give some indications.

**Example 2** (*X Has a Density with Respect to a Product Measure*). Let

$$\nu = \nu_1 \times \cdots \times \nu_p$$

be a product measure on $\mathcal{B}_p$, where each $\nu_i$ is a $\sigma$-finite measure on $\mathcal{B}_1$. For instance, $\nu_i = m_1$ for all $i \in I$. Or else, $\nu_i = m_1$ for some $i$ and $\nu_j =$ counting measure (on a countable subset of $\mathbb{R}$) for some $j$. And so on. In this example, we assume $\mathcal{L}(X) \ll \nu$. Hence, $X$ has a density $h$ with respect to $\nu$, namely

$$P(X \in A) = \int_A h \, d\nu \qquad \text{for all } A \in \mathcal{B}_p.$$

Fix a probability measure $\pi$ on $\mathcal{B}_{2p}$ satisfying condition (3) and define $\lambda$ through condition (2). Then, Theorem 1 implies $\lambda \in \Lambda$. In addition, since the measure $\nu \times \nu$ is $\mathcal{F}$-invariant, one obtains $\lambda \ll \nu \times \nu$ provided $\pi \ll \nu \times \nu$. Precisely, if $\pi \ll \nu \times \nu$ and $g$ is a density of $\pi$ with respect to $\nu \times \nu$, then

$$q = \frac{\sum_{f \in \mathcal{F}} g \circ f}{2^p}$$

is a density of $\lambda$ with respect to $\nu \times \nu$. This formula is practically useful. In fact, since $\lambda \in \Lambda$, there is a knockoff $\widetilde{X}$ such that $\mathcal{L}(X, \widetilde{X}) = \lambda$. Hence, after observing $X = x$, a value $\widetilde{x}$ for such $\widetilde{X}$ can be drawn from the conditional density

$$\frac{q(x, \widetilde{x})}{h(x)} = \frac{\sum_{f \in \mathcal{F}} g[f(x, \widetilde{x})]}{2^p h(x)} \qquad \text{where } x, \widetilde{x} \in \mathbb{R}^p.$$

Obviously, to make this example concrete, one needs a probability measure $\pi$ satisfying condition (3) and $\pi \ll \nu \times \nu$. As noted above, to find $\pi$ is usually hard. However, a probability $\pi$ with the required properties is in Example 9.

We close this section by determining those $\pi$ which satisfy Eq. (2) for a given $\lambda \in \mathcal{P}$.

**Theorem 3.** *Fix $\lambda \in \mathcal{P}$ and any probability measure $\pi$ on $\mathcal{B}_{2p}$. The following statements are equivalent:*

(a) *Condition* (2) *holds, namely, $\lambda = \frac{\sum_{f \in \mathcal{F}} \pi \circ f^{-1}}{2^p}$;*

(b) *$\pi$ admits a density with respect to $\lambda$, say $q$, and*

$$\sum_{f \in \mathcal{F}} q[f(x)] = 2^p \qquad \text{for $\lambda$-almost all } x \in \mathbb{R}^{2p};$$

(c) *$\pi = \lambda$ on $\mathcal{G}$, where $\mathcal{G} = \left\{ A \in \mathcal{B}_{2p} : f^{-1}(A) = A \text{ for all } f \in \mathcal{F} \right\}$.*

The proof of Theorem 3 is postponed to the final Appendix.

As an application of Theorem 3, in the next example, $\lambda$ is a well known knockoff distribution and we look for a probability $\pi$ satisfying Eq. (2) with respect to $\lambda$.

**Example 4.** Suppose $X \sim \mathcal{N}(0, \Sigma)$ and take a diagonal matrix $D$ such that

$$G = \begin{pmatrix} \Sigma & \Sigma - D \\ \Sigma - D & \Sigma \end{pmatrix}$$

is semidefinite positive. If $(X, \widetilde{X}) \sim \mathcal{N}(0, G)$, then $\widetilde{X}$ is a knockoff copy of $X$; see e.g. Candes et al. (2018, p. 559). Fix $G$ as above and define $\lambda = \mathcal{N}(0, G)$. Define also

$$q(x) = 2^p \, \frac{\phi(x)}{\sum_{g \in \mathcal{F}} \phi[g(x)]} \qquad \text{for all } x \in \mathbb{R}^{2p},$$

where $\phi$ is any strictly positive Borel function on $\mathbb{R}^{2p}$. Since $\mathcal{F}$ is a group,

$$2^{-p} \sum_{f \in \mathcal{F}} q[f(x)] = \sum_{f \in \mathcal{F}} \frac{\phi[f(x)]}{\sum_{g \in \mathcal{F}} \phi[g \circ f(x)]} = \frac{\sum_{f \in \mathcal{F}} \phi[f(x)]}{\sum_{g \in \mathcal{F}} \phi[g(x)]} = 1.$$

Since $\operatorname{card}(\mathcal{F}) = 2^p$ and $\lambda \in \mathcal{P}$,

$$\int q(x) \, \lambda(dx) = \sum_{f \in \mathcal{F}} \int \frac{\phi(x)}{\sum_{g \in \mathcal{F}} \phi[g(x)]} \, \lambda(dx)$$

$$= \sum_{f \in \mathcal{F}} \int \frac{\phi[f(x)]}{\sum_{g \in \mathcal{F}} \phi[g(x)]} \, \lambda(dx) = \int \frac{\sum_{f \in \mathcal{F}} \phi[f(x)]}{\sum_{g \in \mathcal{F}} \phi[g(x)]} \, \lambda(dx) = 1.$$

Therefore, thanks to Theorem 3,

$$\pi(dx) = q(x) \, \lambda(dx)$$

is a probability measure on $\mathcal{B}_{2p}$ satisfying Eq. (2).

## 3. Constructing knockoffs via copulas

In this section, $F$ and $F_i$ are the distribution functions of $X$ and $X_i$, respectively. Moreover, for any distribution function $G$ on $\mathbb{R}^n$, we write $\lambda_G$ to denote the probability measure on $\mathcal{B}_n$ induced by $G$.

A *n-copula*, or merely a copula, is a distribution function on $\mathbb{R}^n$ with uniform (on the interval $(0, 1)$) univariate marginals. By Sklar's theorem, for any distribution function $G$ on $\mathbb{R}^n$ there is a $n$-copula $C$ such that

$$G(x) = C\big[G_1(x_1), \ldots, G_n(x_n)\big] \qquad \text{for all } x \in \mathbb{R}^n,$$

where $G_1, \ldots, G_n$ are the univariate marginals of $G$.

Let us *fix* a $p$-copula $C$ such that

$$F(x) = C\big[F_1(x_1), \ldots, F_p(x_p)\big] \qquad \text{for all } x \in \mathbb{R}^p.$$

Note that $C$ is unique whenever $F_1, \ldots, F_p$ are continuous. Note also that, since $F$ is known, $C$ can be regarded to be known as well.

In order to manufacture a knockoff, a naive idea is to let

$$H(x) = C\Big[D_1\big(F_1(x_1), F_1(x_{p+1})\big), \ldots, D_p\big(F_p(x_p), F_p(x_{2p})\big)\Big] \tag{4}$$

for all $x \in \mathbb{R}^{2p}$, where $D_1, \ldots, D_p$ are any 2-copulas. Such an $H$ is a possible candidate to be the distribution function of $(X, \widetilde{X})$ for some knockoff copy $\widetilde{X}$ of $X$.

Unfortunately, $H$ may fail to be a distribution function on $\mathbb{R}^{2p}$. However $\lambda_H \in \Lambda$, more or less by definition, whenever $H$ is a distribution function and $D_1, \ldots, D_p$ are symmetric (i.e., $D_i(u_2, u_1) = D_i(u_1, u_2)$ for all $u \in [0, 1]^2$ and $i \in I$).

**Theorem 5.** *Suppose that $H$ is a distribution function on $\mathbb{R}^{2p}$. Then,*

$$\lambda_H\big\{x \in \mathbb{R}^{2p} : f(x) \in A \times \mathbb{R}^p\big\} = P(X \in A) \tag{5}$$

*for all $f \in \mathcal{F}$ and $A \in \mathcal{B}_p$. In particular, $\lambda_H$ satisfies condition (3) (namely, (3) holds if $\pi = \lambda_H$). Moreover, $\lambda_H \in \Lambda$ whenever $D_1, \ldots, D_p$ are symmetric.*

**Proof.** To prove condition (5), just note that

$$\lim_{x_{p+1}, \ldots, x_{2p} \to \infty} H\big[f(x)\big] = F(x_1, \ldots, x_p) \qquad \text{for all } f \in \mathcal{F} \text{ and } x \in \mathbb{R}^{2p}.$$

Moreover, if $D_1, \ldots, D_p$ are symmetric, then

$$H\big[f(x)\big] = H(x) \qquad \text{for all } f \in \mathcal{F} \text{ and } x \in \mathbb{R}^{2p}.$$

Hence, $\lambda_H \in \mathcal{P}$ and Theorem 1 implies $\lambda_H \in \Lambda$. $\quad\square$

For Theorem 5 to work, the obvious drawback is how to choose $D_1, \ldots, D_p$ in such a way that $H$ is a distribution function. However, when this drawback can be overcome, an explicit expression for $\mathcal{L}(X, \widetilde{X})$ is available where $\widetilde{X}$ is a knockoff copy of $X$. Hence, the conditional distribution of $\widetilde{X}$ given $X$ can be written in closed form.

As an example, suppose that $H$ is a distribution function and $C, D_1, \ldots, D_p, F_1, \ldots, F_p$ are all absolutely continuous with respect to the Lebesgue measure of appropriate dimension. Then, $H$ is absolutely continuous with respect to the Lebesgue measure of dimension $2p$. Moreover, if $\widetilde{X}$ is such that $\mathcal{L}(X, \widetilde{X}) = \lambda_H$, the conditional density of $\widetilde{X}$ given $X = x$ can be written as

$$
p(\widetilde{x} \mid x) = \frac{1}{\varphi\big[F_1(x_1), \ldots, F_p(x_p)\big] \prod_{i=1}^{p} f_i(x_i)} \cdot \frac{\partial^{2p} H}{\partial x_p \ldots \partial x_1 \partial \widetilde{x}_p \ldots \partial \widetilde{x}_1}(x, \widetilde{x})
$$

where $x, \widetilde{x} \in \mathbb{R}^p$ and $\varphi$ and $f_i$ are the densities of $C$ and $F_i$, respectively. This formula will be used in Section 5.

We next discuss the choice of $D_1, \ldots, D_p$. As already noted, not every choice is admissible.

**Example 6** (*H May Fail to Be a Distribution Function*). Let $p = 2$ and $C(u) = (u_1 + u_2 - 1)^+$ for $u \in [0, 1]^2$. Then, with $D_1 = C$, one obtains

$$
\begin{aligned}
\lim_{x_4 \to \infty} H(x) &= C\Big[D_1\big(F_1(x_1), F_1(x_3)\big), \ D_2\big(F_2(x_2), 1\big)\Big] \\
&= C\Big[D_1\big(F_1(x_1), F_1(x_3)\big), \ F_2(x_2)\Big] \\
&= \Big(F_2(x_2) + \big(F_1(x_1) + F_1(x_3) - 1\big)^+ - 1\Big)^+ \\
&= \big(F_1(x_1) + F_1(x_3) + F_2(x_2) - 2\big)^+.
\end{aligned}
$$

Therefore, $\lim_{x_4 \to \infty} H(x)$ is not a distribution function on $\mathbb{R}^3$, so that $H$ is not a distribution function on $\mathbb{R}^4$.

Let

$$
\Lambda_0 = \big\{\lambda \in \Lambda : \text{ the distribution function of } \lambda \text{ admits representation (4)}\big\}.
$$

Despite Example 6, a possible question is whether $\Lambda_0 = \Lambda$.

**Example 7** ($\Lambda_0$ *Is a Proper Subset Of* $\Lambda$). Let $U = (U_1, \ldots, U_{2p})$ and $V = (V_1, \ldots, V_{2p})$ be any random variables. Then, $\mathcal{L}(U) = \mathcal{L}(V)$ provided:

$$
\mathcal{L}(U) \in \Lambda_0, \ \mathcal{L}(V) \in \Lambda_0, \quad \text{and} \quad \mathcal{L}(U_i, U_{p+i}) = \mathcal{L}(V_i, V_{p+i}) \text{ for each } i \in I.
$$

After noting this fact, take $U$ and $V$ exchangeable and such that

$$
\mathcal{L}(U) \neq \mathcal{L}(V) \quad \text{but} \quad \mathcal{L}(U_1, \ldots, U_p) = \mathcal{L}(V_1, \ldots, V_p).
$$

Suppose also that $X \sim (U_1, \ldots, U_p)$. Since $U$ is exchangeable, $\mathcal{L}(U) \in \mathcal{P}$. By Theorem 1 and $X \sim (U_1, \ldots, U_p)$, one obtains $\mathcal{L}(U) \in \Lambda$. Similarly, $\mathcal{L}(V) \in \Lambda$. Hence, at least one between $\mathcal{L}(U)$ and $\mathcal{L}(V)$ belongs to $\Lambda \setminus \Lambda_0$. In fact, $\mathcal{L}(U) \neq \mathcal{L}(V)$ but $\mathcal{L}(U_i, U_j) = \mathcal{L}(V_i, V_j)$ for all $i \neq j$ (because of exchangeability).

We next give conditions for $H$ to be a distribution function.

**Theorem 8.** *$H$ is a distribution function on $\mathbb{R}^{2p}$ whenever*

(j) *$D_i$ is of class $C^2$ for each $i \in I$;*
(jj) *$C$ has a density $\varphi$ with respect to $m_p$;*
(jjj) *$\varphi$ is of class $C^p$ and, at each point $u \in [0, 1]^{2p}$, one obtains*

$$
\frac{\partial^p}{\partial u_{2p} \ldots \partial u_{p+1}} \ \varphi\Big[D_1(u_1, u_{p+1}), \ldots, D_p(u_p, u_{2p})\Big] \prod_{i=1}^{p} \frac{\partial}{\partial u_i} D_i(u_i, u_{p+i}) \geq 0.
$$

*Under such conditions, one also obtains $\lambda_H \ll m_{2p}$ whenever $\mathcal{L}(X_i) \ll m_1$ for each $i \in I$.*

Condition (jjj) is a technical constraint, required to guarantee the existence and positivity of the partial derivatives of $H$, and has no heuristic interpretation (known to us). We also recall that $\mathcal{L}(X_i) \ll m_1$ means that the probability distribution of $X_i$ is absolutely continuous with respect to the Lebesgue measure $m_1$.

The proof of Theorem 8 is deferred to Appendix. Here, we give three final examples.

**Example 9** (*Asymmetric Copulas*). Suppose that $H$ is a distribution function on $\mathbb{R}^{2p}$. If $D_i$ is not symmetric for some $i \in I$, as we assume, then usually $\lambda_H \notin \Lambda$. However, Theorem 5 implies that $\lambda_H$ satisfies condition (3). Therefore, Theorem 1

yields

$$\lambda := \frac{\sum_{f \in \mathcal{F}} \lambda_H \circ f^{-1}}{2^p} \in \Lambda.$$

Furthermore, the distribution function of $\lambda$, say $G$, can be written explicitly as

$$G(x) = \frac{\sum_{f \in \mathcal{F}} H[f(x)]}{2^p} \qquad \text{for all } x \in \mathbb{R}^{2p}.$$

Suppose now that $C, D_1, \ldots, D_p$ satisfy conditions (j)-(jj)-(jjj) and $\mathcal{L}(X_i) \ll m_1$ for each $i \in I$. Then, not only $H$ is a distribution function, but $\lambda_H \ll m_{2p}$. Hence, one can let $\pi = \lambda_H$ in Example 2.

**Example 10** (*An Open Problem*). In principle, a knockoff $\widetilde{X}$ should be "as independent of $X$ as possible". Thus, it is tempting to let

$$D_i(u) = u_1 u_2 \qquad \text{for all } u \in [0, 1]^2 \text{ and } i \in I.$$

In this case, $D_1, \ldots, D_p$ are symmetric and, for all $i \in I$ and $x \in \mathbb{R}^{2p}$,

$$\lim_{x_j \to \infty, \, j \in J_i} H(x) = F_i(x_i) F_i(x_{p+i}) \qquad \text{where } J_i = \{1, \ldots, 2p\} \setminus \{i, p+i\}.$$

Therefore, *if $H$ is a distribution function and $(X, \widetilde{X}) \sim \lambda_H$, then*

$$\widetilde{X} \text{ is a knockoff copy of } X \text{ and } \widetilde{X}_i \text{ is independent of } X_i \text{ for each } i \in I.$$

Thus, a (natural) question is: *If each $D_i$ is the independence copula, under what conditions $H$ is a distribution function* ? Some partial answers are available. For instance, $H$ is a distribution function if $C$ admits a smooth density $\varphi$ (with respect to $m_p$) such that

$$\frac{\partial^p}{\partial u_{2p} \ldots \partial u_{p+1}} \, \varphi\Big(u_1 u_{p+1}, \ldots, u_p u_{2p}\Big) \prod_{i=1}^{p} u_{p+i} \geq 0.$$

Or else, $H$ is a distribution function if $C$ is Archimedean with a suitable generator $\psi$ (just let $\psi_i(x) = \exp(-x)$ in condition (6) of Example 11). To our knowledge, however, a general answer to the above question is still unknown.

**Example 11** (*Archimedean Copulas*). An *Archimedean generator* is a continuous and strictly decreasing function $\psi : [0, \infty) \to (0, 1]$ such that $\psi(0) = 1$ and $\lim_{x \to \infty} \psi(x) = 0$. By convention, we let $\psi(\infty) = 0$ and $\psi^{-1}(0) = \infty$.

Suppose $C$ is Archimedean with generator $\psi$, that is,

$$C(u) = \psi\Big(\sum_{i=1}^{p} \psi^{-1}(u_i)\Big) \qquad \text{for all } u \in [0, 1]^p.$$

Suppose also that $\psi$ has derivatives up to order $2p$ on $(0, \infty)$ and

$$(-1)^k \psi^{(k)} \geq 0 \qquad \text{for } k = 1, \ldots, 2p,$$

where $\psi^{(k)}$ denotes the $k$th derivative of $\psi$. In view of McNeil and Neslehova (2009, Cor. 2.1), the latter condition implies that

$$C^*(u) = \psi\Big(\sum_{i=1}^{2p} \psi^{-1}(u_i)\Big), \qquad u \in [0, 1]^{2p},$$

is a $2p$-copula. Therefore, $H$ is a distribution function on $\mathbb{R}^{2p}$ as far as $D_1, \ldots, D_p$ are Archimedean with the same generator as $C$. In this case, in fact,

$$H(x) = \psi\Big\{ \sum_{i=1}^{p} \psi^{-1}\big(D_i\big(F_i(x_i), F_i(x_{p+i})\big)\big) \Big\}$$

$$= \psi\Big\{ \sum_{i=1}^{p} \psi^{-1}(F_i(x_i)) + \sum_{i=1}^{p} \psi^{-1}(F_i(x_{p+i})) \Big\}$$

$$= C^*\Big\{ F_1(x_1), \ldots, F_p(x_p), F_1(x_{p+1}), \ldots, F_p(x_{2p}) \Big\} \qquad \text{for all } x \in \mathbb{R}^{2p}.$$

In addition, since $D_1, \ldots, D_p$ are symmetric, one also obtains $\lambda_H \in \Lambda$.

More generally, suppose that $D_i$ is Archimedean with generator $\psi_i$ for each $i \in I$. Then, $H$ is a distribution function and $\lambda_H \in \Lambda$ provided

$$(-1)^k \psi_i^{(k)} \geq 0 \quad \text{and} \quad (-1)^{k-1} \left( \psi^{-1} \circ \psi_i \right)^{(k)} \geq 0 \tag{6}$$

for all $i \in I$ and $k = 1, \ldots, 2p$; see Okhrin et al. (2013, p. 190) and Savu and Trede (2010, p. 297). If all the generators $\psi, \psi_1, \ldots, \psi_p$ belong to the same parametric family, such us the Gumbel or the Clayton, condition (6) reduces to a simple restriction on the parameters; see Embrechts et al. (2003).

A last general remark is that the idea underlying Theorems 5 and 8 could be realized, possibly in a better way, involving special types of copulas. For instance, a possibility could be using pair copulas; see e.g. Aas et al. (2009).

## 4. Conditional independence

To build a (reasonable) knockoff is not hard if $X$ is conditionally independent given some random element $Z$. We begin by making this claim precise.

**Theorem 12.** *Suppose that, for some random element Z, one obtains*

$$P\left(X_1 \in A_1, \ldots, X_p \in A_p\right) = E\left\{ \prod_{i=1}^{p} P(X_i \in A_i \mid Z) \right\} \tag{7}$$

*for all $A_1, \ldots, A_p \in \mathcal{B}_1$. Let $\lambda$ be the (only) probability measure on $\mathcal{B}_{2p}$ such that*

$$\lambda(A_1 \times \cdots \times A_{2p}) = E\left\{ \prod_{i=1}^{p} P(X_i \in A_i \mid Z) \prod_{i=1}^{p} P(X_i \in A_{p+i} \mid Z) \right\}$$

*whenever $A_i \in \mathcal{B}_1$ for all $i = 1, \ldots, 2p$. Then, $\lambda \in \Lambda$.*

**Proof.** For all $A_1, \ldots, A_{2p} \in \mathcal{B}_1$, define

$$\lambda_0(A_1 \times \cdots \times A_{2p}) = E\left\{ \prod_{i=1}^{p} P(X_i \in A_i \mid Z) \prod_{i=1}^{p} P(X_i \in A_{p+i} \mid Z) \right\}.$$

Such a $\lambda_0$, defined on

$$\mathcal{R} = \left\{ A_1 \times \cdots \times A_{2p} : A_i \in \mathcal{B}_1, \ i = 1, \ldots, 2p \right\},$$

uniquely extends to a probability measure $\lambda$ on $\mathcal{B}_{2p}$. By definition,

$$\lambda \circ f^{-1}(A) = \lambda_0 \circ f^{-1}(A) = \lambda_0(A) = \lambda(A)$$

whenever $f \in \mathcal{F}$ and $A \in \mathcal{R}$. Hence, $\lambda \in \mathcal{P}$. Finally, if $A_i = \mathbb{R}$ for $i > p$, condition (7) yields

$$\lambda(A_1 \times \cdots \times A_p \times \mathbb{R}^p) = E\left\{ \prod_{i=1}^{p} P(X_i \in A_i \mid Z) \right\} = P\left(X_1 \in A_1, \ldots, X_p \in A_p\right).$$

Therefore, $\lambda \in \Lambda$. $\square$

In real problems, to take advantage of Theorem 12, one needs to select a random element $Z$ satisfying condition (7). As an extreme example, suppose $Z = X$. Then, condition (7) holds and $P(X_i \in A_i \mid X) = 1_{A_i}(X_i)$ a.s. Therefore,

$$\lambda(A_1 \times \cdots \times A_{2p}) = E\left\{ \prod_{i=1}^{p} 1_{A_i}(X_i) \prod_{i=1}^{p} 1_{A_{p+i}}(X_i) \right\}$$

$$= P\left(X_1 \in A_1 \cap A_{p+1}, \ldots, X_p \in A_p \cap A_{2p}\right).$$

Such a $\lambda$ is precisely the probability distribution of the trivial knockoff $(X, X)$ (namely, $\widetilde{X} = X$). Thus, as it could be guessed, $Z = X$ is not a good choice. We now consider some better choices.

**Example 13** (*Stable Laws*)**.** Let $U = (U_1, \ldots, U_p)$ and $Z = (Z_1, \ldots, Z_p)$ be $p$-variate random variables, with $U$ independent of $Z$ and $U_1, \ldots, U_p$ independent among them. Then, condition (7) holds whenever

$$X = U + Z.$$

As an example, fix $\alpha \in (0, 2]$ and suppose $U_i \sim \mathcal{S}(a_i, b_i)$ for all $i$. According to Section 1.3, this means that $U_i$ has a symmetric $\alpha$-stable distribution with parameters $a_i \in \mathbb{R}$ and $b_i > 0$. For $A \in \mathcal{B}_1$, write $\mathcal{S}(a, b)(A)$ to denote the value attached to $A$ by the probability measure $\mathcal{S}(a, b)$. In this notation, since $U_i + c \sim \mathcal{S}(a_i + c, \ b_i)$ for all $c \in \mathbb{R}$, one obtains

$$P(X_1 \in A_1, \ldots, X_p \in A_p \mid Z) = \prod_{i=1}^{p} \mathcal{S}(a_i + Z_i, b_i)(A_i) \qquad a.s.$$

Hence, Theorem 12 implies $\lambda \in \Lambda$ where

$$\lambda(A_1 \times \cdots \times A_{2p}) = E \left\{ \prod_{i=1}^{p} \mathcal{S}(a_i + Z_i, b_i)(A_i) \prod_{i=1}^{p} \mathcal{S}(a_i + Z_i, b_i)(A_{p+i}) \right\}.$$

**Example 14** (*Normal Distributions*). As a special case of Example 13 (with $\alpha = 2$) suppose $X \sim \mathcal{N}(\mu, \Sigma)$. Let $D$ be a diagonal matrix such that $\Sigma - D$ is semidefinite positive and $d_{ii} \geq 0$ for all $i$, where $d_{ii}$ is the $i$th diagonal element of $D$. Then, one can take $U \sim \mathcal{N}(0, D)$ and $Z \sim \mathcal{N}(\mu, \Sigma - D)$. The conditional distribution of $X$ given $Z$ is $\mathcal{N}(Z, D)$. Since $D$ is diagonal, $X_1, \ldots, X_p$ are conditionally independent, given $Z$, with $X_i \sim \mathcal{N}(Z_i, d_{ii})$. Define

$$\lambda(A_1 \times \cdots \times A_{2p}) = E \left\{ \prod_{i=1}^{p} \mathcal{N}(Z_i, d_{ii})(A_i) \prod_{i=1}^{p} \mathcal{N}(Z_i, d_{ii})(A_{p+i}) \right\}.$$

Then, by Theorem 12, there is a knockoff copy $\widetilde{X}$ of $X$ such that $(X, \widetilde{X}) \sim \lambda$. Finally, it is easily seen that

$$\lambda = \mathcal{N}(\mu^*, G) \quad \text{where} \quad \mu^* = \begin{pmatrix} \mu \\ \mu \end{pmatrix} \text{ and } G = \begin{pmatrix} \Sigma & \Sigma - D \\ \Sigma - D & \Sigma \end{pmatrix}.$$

A concrete example (suggested by an anonymous referee) is the so called "equicorrelated" Gaussian distribution, namely, $\sigma_{ii} = b$ and $\sigma_{ij} = a$ for all $i$ and all $j \neq i$, where $0 < a < b$ are fixed constants. In this case, it suffices to take $d_{ii} \in (0, b - a)$ for all $i$.

The probability $\lambda$ obtained in Example 14 is already known to be an element of $\Lambda$; see e.g. Candes et al. (2018, p. 559). Instead, in the next example, Theorem 12 yields a new knockoff distribution.

**Example 15** (*Mixtures of Normal Distributions*). Let

$$X = ZU$$

where $Z$ is a random $p \times p$ diagonal matrix and $U$ a $p$-dimensional column vector. Suppose $U \sim \mathcal{N}(0, I)$ and $Z$ independent of $U$. Then, the probability distribution of $X$ can be written as

$$P(X \in A) = E \left\{ \mathcal{N}(0, ZZ)(A) \right\} \qquad \text{for all } A \in \mathcal{B}_p.$$

Probability distributions of this type play a role in various frameworks. For instance, they arise as the limit laws in the CLT for exchangeable random variables; see e.g. Berti et al. (2004, Sect. 3). In any case, since $ZZ$ is diagonal, $X_1, \ldots, X_p$ are conditionally independent given $Z$ with $X_i \sim \mathcal{N}(0, Z_{ii}^2)$. Hence, Theorem 12 implies $\lambda \in \Lambda$ where

$$\lambda(A_1 \times \cdots \times A_{2p}) = E \left\{ \prod_{i=1}^{p} \mathcal{N}(0, Z_{ii}^2)(A_i) \prod_{i=1}^{p} \mathcal{N}(0, Z_{ii}^2)(A_{p+i}) \right\}.$$

A further example, where conditional independence is exploited to obtain a knockoff, is in Bates et al. (2020).

In applications, to assign $\mathcal{L}(X)$ is one of the main statistician's tasks. Hence, a reasonable strategy is to model $X$ so as to realize conditional independence, with respect to some latent variable $Z$, and then to obtain a knockoff $\widetilde{X}$ via Theorem 12. As already noted, the advantage is twofold. On one hand, conditional independence is easy to be realized and able to describe various real situations. On the other hand, to build $\widetilde{X}$ is straightforward whenever $X$ is conditionally independent. In the rest of this section, the statistician is assumed to adopt this strategy. Thus, he/she decides to model $X$ as conditionally independent with respect to some $Z$. Note that, in this framework, $\mathcal{L}(X)$ *is regarded as a statistician's choice (and not as an external constraint to be satisfied)*. The next example is fundamental.

**Example 16** (*Parametric Constructions of Knockoffs*). Suppose $X$ is modeled as

$$P(X_1 \in A_1, \ldots, X_p \in A_p) = \int_{\Theta} \prod_{i=1}^{p} Q_i(A_i, \theta) \gamma(d\theta),$$

where $Q_1(\cdot, \theta), \ldots, Q_p(\cdot, \theta)$ are probabilities on $\mathcal{B}_1$, indexed by some parameter $\theta \in \Theta$, and $\gamma$ is a mixing probability on $\Theta$. As an example, one could take

$$Q_i(\cdot, \theta) = \mathcal{N}(\mu_i, \sigma_i^2) \quad \text{and} \quad \theta = (\mu_1, \ldots, \mu_p, \sigma_1^2, \ldots, \sigma_p^2).$$

In this case, $\gamma$ would be a probability measure on $\Theta = \mathbb{R}^p \times (0, \infty)^p$.

More generally, fix a $\sigma$-finite measure $\nu_i$ on $\mathcal{B}_1$ and suppose $Q_i(\cdot, \theta)$ has a density $f_i(\cdot, \theta)$ with respect to $\nu_i$, namely

$$Q_i(A, \theta) = \int_A f_i(t, \theta)\, \nu_i(dt) \qquad \text{for all } i \in I,\ A \in \mathcal{B}_1 \text{ and } \theta \in \Theta.$$

Define $\lambda$ to be the probability measure on $\mathcal{B}_{2p}$ with density $q$ with respect to $\nu \times \nu$, where $\nu = \nu_1 \times \cdots \times \nu_p$ and

$$q(y) = q(y_1, \ldots, y_{2p}) = \int_\Theta \prod_{i=1}^p f_i(y_i, \theta) \prod_{i=1}^p f_i(y_{p+i}, \theta)\, \gamma(d\theta) \qquad \text{for all } y \in \mathbb{R}^{2p}.$$

Then, $\lambda \in \Lambda$ because of Theorem 12. Therefore, after observing $X = x$, a value $\widetilde{x}$ for the knockoff $\widetilde{X}$ can be drawn from the conditional density

$$\frac{q(x, \widetilde{x})}{h(x)},$$

where $x, \widetilde{x} \in \mathbb{R}^p$ and $h(x) = \int_\Theta \prod_{i=1}^p f_i(x_i, \theta)\, \gamma(d\theta)$ is the marginal density of $X$.

Example 16 is general enough to cover a wide range of real situations.

We now briefly discuss the choice of $\gamma$. It may be helpful to recall that, once $Q_1(\cdot, \theta), \ldots, Q_p(\cdot, \theta)$ have been selected, to choose $\gamma$ is equivalent to choose the probability distribution of $X$.

**Example 17** (*Choice of $\gamma$*). It is tempting to regard the mixing measure $\gamma$ as a prior distribution. Even if not mandatory, this interpretation is helpful. Hence, in the sequel, $\gamma$ is referred to as *the prior*. Let $Q_i(\cdot, \theta)$, $f_i(\cdot, \theta)$ and $\lambda \in \Lambda$ be as in Example 16. Two (distinct) criterions to select $\gamma$ are as follows.

Roughly speaking, $\gamma$ tunes the dependence between $X$ and $\widetilde{X}$, where $\widetilde{X}$ is such that $\mathcal{L}(X, \widetilde{X}) = \lambda$. Define in fact

$$Q(\cdot, \theta) = Q_1(\cdot, \theta) \times \cdots \times Q_p(\cdot, \theta).$$

Then, $Q(\cdot, \theta)$ is a probability measure on $\mathcal{B}_p$ and

$$P(X \in A,\ \widetilde{X} \in B) - P(X \in A)\, P(\widetilde{X} \in B) = \tag{8}$$
$$= \int_\Theta Q(A, \theta)\, Q(B, \theta)\, \gamma(d\theta) - \int_\Theta Q(A, \theta)\, \gamma(d\theta) \int_\Theta Q(B, \theta)\, \gamma(d\theta)$$

for all $A, B \in \mathcal{B}_p$. Thus, a first criterion is to choose $\gamma$ so as to make (8) small for some $A$ and $B$. This is just a rough and naive indication, difficult to realize in practice, but it may be potentially useful.

To state the second criterion, denote by $h_\gamma$ the marginal density of $X$ when the prior is $\gamma$, namely

$$h_\gamma(x) = \int_\Theta \prod_{i=1}^p f_i(x_i, \theta)\, \gamma(d\theta) \qquad \text{for all } x \in \mathbb{R}^p.$$

Suppose now that $X = x$ is observed. Then, $h_\gamma$ can be seen as the integrated likelihood of $x$ with respect to the prior $\gamma$. From a Bayesian point of view, it is desirable that $h_\gamma(x)$ is high. Therefore, a second criterion is to choose $\gamma$ so as to maximize the map $\gamma \mapsto h_\gamma(x)$. For instance, the choice between two conflicting priors $\gamma_1$ and $\gamma_2$ could be seen as a model selection problem. Accordingly, we could choose between $\gamma_1$ and $\gamma_2$ based on the *Bayes factor* $h_{\gamma_1}(x)/h_{\gamma_2}(x)$. A practical advantage is that we can profit on the broad literature on Bayes factors and related topics.

Another useful feature of Example 16 is highlighted in the next example.

**Example 18** (*Uncorrelated Knockoffs*). Under some assumptions on $Q_i(\cdot, \theta)$, one obtains

$$\text{cov}(X_i, \widetilde{X}_i) = 0 \qquad \text{for all } i \in I \text{ and all priors } \gamma.$$

Fix in fact $i \in I$ and suppose the mean of $Q_i(\cdot, \theta)$ exists and does not depend on $\theta$, say

$$\int_\mathbb{R} t\, Q_i(dt, \theta) = a_i \qquad \text{for some } a_i \in \mathbb{R} \text{ and all } \theta \in \Theta.$$

Then, independently of $\gamma$, Fubini's theorem yields

$$\text{cov}(X_i, \widetilde{X}_i) = \int_\Theta a_i^2 \, d\gamma - \left( \int_\Theta a_i \, d\gamma \right)^2 = a_i^2 - a_i^2 = 0.$$

For instance, $\text{cov}(X_i, \widetilde{X}_i) = 0$ provided $Q_i(\cdot, \theta) = \mathcal{N}(0, \sigma_i^2(\theta))$ for all $\theta$.

We conclude our discussion of Example 16 with a practical example.

**Example 19** (*Conditionally Independent Poisson Data*). Let $\theta = (\theta_1, \ldots, \theta_p)$ and $Q_i(\cdot, \theta)$ a Poisson distribution with parameter $\theta_i$. We consider two different choices of the prior $\gamma$.

First, let $\gamma = \gamma_1 \times \cdots \times \gamma_p$ where each $\gamma_i$ is a Gamma distribution with parameters $a_i$ and $b_i$. In this case, since $\theta_1, \ldots, \theta_p$ are independent under $\gamma$, the calculations are straightforward:

$$q(y_1, \ldots, y_{2p}) = \prod_{i=1}^p \int_0^\infty \left( \frac{\theta_i^{y_i}}{y_i!} e^{-\theta_i} \frac{\theta_i^{y_{i+p}}}{y_{i+p}!} e^{-\theta_i} \right) \frac{b_i^{a_i}}{\Gamma(a_i)} \theta_i^{a_i-1} e^{-b_i\theta_i} \, d\theta_i$$

$$= \prod_{i=1}^p \frac{1}{y_i! y_{i+p}!} \frac{b_i^{a_i}}{\Gamma(a_i)} \frac{\Gamma(a_i + y_i + y_{i+p})}{(b_i + 2)^{a_i+y_i+y_{i+p}}}.$$

Similarly,

$$h(y_1, \ldots, y_p) = \prod_{i=1}^p \frac{1}{y_i!} \frac{b_i^{a_i}}{\Gamma(a_i)} \frac{\Gamma(a_i + y_i)}{(b_i + 1)^{a_i+y_i}}.$$

Therefore, after observing $X = x$, a value $\tilde{x}$ for the knockoff $\widetilde{X}$ can be drawn from the conditional density

$$\frac{q(x, \tilde{x})}{h(x)} = \prod_{i=1}^p \frac{1}{\tilde{x}_i!} \frac{\Gamma(a_i + x_i + \tilde{x}_i)}{\Gamma(a_i + x_i)} \frac{(b_i + 1)^{a_i+x_i}}{(b_i + 2)^{a_i+x_i+\tilde{x}_i}}.$$

Second, let $\gamma$ be a Dirichlet distribution with parameters $a_1, \ldots, a_p$. Denote by

$$S = \left\{ \theta \in \mathbb{R}^p : \theta_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^p \theta_i = 1 \right\}$$

the $p$-dimensional simplex, and by

$$m(n_1, \ldots, n_p) = \int_S \theta_1^{n_1} \ldots \theta_p^{n_p} \, \gamma(d\theta)$$

the mixed moment of $\gamma$ of order $(n_1, \ldots, n_p)$. Explicit formulae for $m(n_1, \ldots, n_p)$ are available; see e.g. Kotz et al. (2000), page 488, equation (49.7). Since $\gamma(S) = 1$, one obtains

$$q(y_1, \ldots, y_{2p}) = \int_S \exp\left( -2 \sum_{i=1}^p \theta_i \right) \prod_{i=1}^p \theta_i^{y_i+y_{i+p}} \prod_{i=1}^p \frac{1}{y_i! y_{i+p}!} \gamma(d\theta)$$

$$= e^{-2} m(y_1 + y_{p+1}, \ldots, y_p + y_{2p}) \prod_{i=1}^p \frac{1}{y_i! y_{i+p}!}.$$

Similarly,

$$h(y_1, \ldots, y_p) = e^{-1} m(y_1, \ldots, y_p) \prod_{i=1}^p \frac{1}{y_i!}.$$

Hence, the conditional density of $\widetilde{X}$ given $X = x$ can be written as

$$\frac{q(x, \tilde{x})}{h(x)} = e^{-1} \frac{m(x_1 + \tilde{x}_1, \ldots, x_p + \tilde{x}_p)}{m(x_1, \ldots, x_p)} \prod_{i=1}^p \frac{1}{\tilde{x}_i!}.$$

## 5. Sampling strategies

In Sections 3 and 4, exploiting copulas and conditional independence, two general methods for constructing knockoffs have been introduced. In this section, having applications in mind, such methods are translated into practical algorithms.

Two classical MCMC algorithms, the Metropolis–Hastings sampler and the Gibbs sampler via data augmentation, are proposed. Obviously, our proposals are not the only possible ones. The literature on MCMC is huge (see e.g. Brooks et al., 2011) and some better sampling strategies could be available. The only goal of this section is to point out that the material of Sections 3–4 can be easily used in applied settings.

We denote by $x = (x_1, \ldots, x_p)$ and $\widetilde{x} = (\widetilde{x}_1, \ldots, \widetilde{x}_p)$ two points of $\mathbb{R}^p$. Here, $x$ should be regarded as the observed value of $X$ and $\widetilde{x}$ as the value to be sampled of the knockoff $\widetilde{X}$.

### 5.1. A Metropolis–Hastings approach to copula knockoffs

In the notation of Section 3, we assume that $C, D_1, \ldots, D_p, F_1, \ldots, F_p$ are all absolutely continuous with respect to the Lebesgue measure of appropriate dimension. Algorithm 1 provides a strategy to sample $\widetilde{x}$ via the copula construction of Section 3.

---

**Algorithm 1** Copula knockoffs: general algorithm

---

1. Choose the distribution functions $F_1, \ldots, F_p$ on $\mathbb{R}$, a $p$-copula C and a family of 2-copulas $D_1, \ldots, D_p$ in such a way that

$$H(x, \widetilde{x}) = C\Big[D_1\big(F_1(x_1), F_1(\widetilde{x}_1)\big), \ldots, D_p\big(F_p(x_p), F_p(\widetilde{x}_p)\big)\Big]$$

is a distribution function on $\mathbb{R}^{2p}$
2. Sample $\widetilde{x}$ from the conditional density

$$p(\widetilde{x} \mid x) = \frac{1}{\varphi\big[F_1(x_1), \ldots, F_p(x_p)\big] \prod_{i=1}^{p} f_i(x_i)} \cdot \frac{\partial^{2p} H}{\partial x_p \ldots \partial x_1 \partial \widetilde{x}_p \ldots \partial \widetilde{x}_1}(x, \widetilde{x})$$

where $\varphi$ and $f_i$ are the densities of $C$ and $F_i$, respectively

---

Sampling from $p(\widetilde{x} \mid x)$ may be not straightforward. However, since

$$p(\widetilde{x} \mid x) \propto \frac{\partial^{2p} H}{\partial x_p \ldots \partial x_1 \partial \widetilde{x}_p \ldots \partial \widetilde{x}_1}(x, \widetilde{x}),$$

a Metropolis–Hastings sampler is available. One such sampler is provided by Algorithm 2.

---

**Algorithm 2** Metropolis–Hastings sampler

---

1. Choose the initial value $\widetilde{x}^{(0)}$
2. Choose the proposal distribution $K(\cdot|x)$ (usually a Markov Kernel)
**for** $j \leftarrow 1$ to $M$ **do**
   3. Sample $y$ from the proposal $K(\cdot|\widetilde{x}^{(j-1)})$
   4. Compute the acceptance probability

$$\alpha(\widetilde{x}^{(j-1)}, y) = \min\left\{1, \frac{p(y \mid x)}{p(\widetilde{x}^{(j-1)} \mid x)}\frac{K(\widetilde{x}^{(j-1)}|y)}{K(y|\widetilde{x}^{(j-1)})}\right\} = \min\left\{1, \frac{\frac{\partial^{2p} H}{\partial x_p \ldots \partial x_1 \partial \widetilde{x}_p \ldots \partial \widetilde{x}_1}(x, y)}{\frac{\partial^{2p} H}{\partial x_p \ldots \partial x_1 \partial \widetilde{x}_p \ldots \partial \widetilde{x}_1}(x, \widetilde{x}^{(j-1)})}\frac{K(\widetilde{x}^{(j-1)}|y)}{K(y|\widetilde{x}^{(j-1)})}\right\}$$

   5. Set $\widetilde{x}^{(j)} = y$ with probability $\alpha(\widetilde{x}^{(j-1)}, y)$ and $\widetilde{x}^{(j)} = \widetilde{x}^{(j-1)}$ with probability $1 - \alpha(\widetilde{x}^{(j-1)}, y)$
**end for**
6. Return the sample $\widetilde{x}^{(j)}, j = 1, \ldots, M$

---

### 5.2. A data augmentation approach to conditional independence knockoffs

We now outline how to sample knockoffs via the conditional independence strategy proposed in Example 16. The main steps of the procedure are summarized by Algorithm 3.

Step 4 of Algorithm 3 could be difficult since the numerator and denominator of the conditional density $q(x, \widetilde{x})/h(x)$ are often not in closed form. Hence, $q(x, \widetilde{x})/h(x)$ may be not explicit and computational methods come to the fore. The Metropolis–Hastings could be problematic since we have to evaluate integrals in the acceptance rate. This can be time consuming. An alternative approach is a Data Augmentation strategy where both $\widetilde{x}$ and $\theta$ are sampled at the same time.

---

**Algorithm 3** Conditional independence knockoffs: general algorithm

---

1. Choose a density $f_i(\cdot, \theta)$, with respect to some reference measure $\nu_i$, for $X_i$, $i = 1, \ldots, p$
2. Choose a mixing probability $\gamma(d\theta)$
3. Compute

$$q(x, \widetilde{x}) = \int_{\Theta} \prod_{i=1}^{p} f_i(x_i, \theta) \prod_{i=1}^{p} f_i(\widetilde{x}_i, \theta) \, \gamma(d\theta) \quad \text{and} \quad h(x) = \int_{\Theta} \prod_{i=1}^{p} f_i(x_i, \theta) \, \gamma(d\theta)$$

4. Sample $\widetilde{x}$ from the conditional density $q(x, \widetilde{x})/h(x)$

---

Suppose $\Theta$ is an open subset of $\mathbb{R}^k$ for some $k$, and $\gamma$ has a density with respect to Lebesgue measure on $\Theta$, say $\gamma(d\theta) = p(\theta) \, d\theta$. Then,

$$\frac{q(x, \widetilde{x})}{h(x)} = \int_{\Theta} \frac{\bar{q}(x, \widetilde{x}, \theta)}{h(x)} \, d\theta \quad \text{where} \quad \bar{q}(x, \widetilde{x}, \theta) = p(\theta) \prod_{i=1}^{p} f_i(x_i, \theta) \prod_{i=1}^{p} f_i(\widetilde{x}_i, \theta).$$

This is quite convenient since it makes easier to implement a Gibbs sampler on the augmented space with $\theta$. The full conditional distributions are straightforward by noting that

$$\frac{\bar{q}(x, \widetilde{x}, \theta)}{h(x)} \propto p(\theta) \prod_{i=1}^{p} f_i(x_i, \theta) \prod_{i=1}^{p} f_i(\widetilde{x}_i, \theta).$$

Algorithm 4 provides a Gibbs sampler for $\bar{q}(x, \widetilde{x}, \theta)$.

---

**Algorithm 4** Data Augmentation sampler

---

1. Choose the initial value $\theta^{(0)}$
  **for** $j \leftarrow 1$ to $M$ **do**
    **for** $i \leftarrow 1$ to $p$ **do**
      2. Sample $\widetilde{x}_i^{(j)} \mid \theta^{(j-1)} \sim f_i(\cdot, \theta^{(j-1)})$
    **end for**
    3. Sample $\theta^{(j)}$ from the posterior of $\theta$ given $x, \widetilde{x}^{(j)}$
  **end for**
4. Return the sample $(\widetilde{x}^{(j)}, \theta^{(j)})$, $j = 1, \ldots, M$

---

It should be noted that Algorithms 2 and 4 provide a sample $\widetilde{x}^{(1)}, \ldots, \widetilde{x}^{(M)}$ of knockoffs rather than a single realization. This could be helpful when taking into account the uncertainty intrinsic in the simulation procedure. Note also that, at each step $j$, Algorithm 4 requires to sample from the posterior of $\theta$ given $x, \widetilde{x}^{(j)}$. In some cases, this could not be an easy step. However, it is straightforward in several scenarios, such as conjugate models.

**Acknowledgment**

**Appendix**

**Proof of Theorem 3.** First note that, since $\mathcal{F}$ is a group under composition,

$$\sum_{f \in \mathcal{F}} g \circ f^{-1} = \sum_{f \in \mathcal{F}} g \circ f \quad \text{for any real function } g \text{ on } \mathbb{R}^{2p}.$$

**"(a) $\Rightarrow$ (b)".** Since $\mathcal{F}$ contains the identity map, condition (a) implies $\pi \leq 2^p \lambda$. Hence, $\pi$ has a density $q$ with respect to $\lambda$. Since $\lambda \in \mathcal{P}$, condition (a) also implies

$$\int_A 2^p \, d\lambda = 2^p \, \lambda(A) = \sum_{f \in \mathcal{F}} \pi \circ f^{-1}(A) = \sum_{f \in \mathcal{F}} \int_{f^{-1}(A)} q \, d\lambda$$

$$= \sum_{f \in \mathcal{F}} \int_A q \circ f^{-1} \, d\lambda = \int_A \left( \sum_{f \in \mathcal{F}} q \circ f \right) d\lambda \quad \text{for each } A \in \mathcal{B}_{2p}.$$

**"(b) $\Rightarrow$ (c)".** If $A \in \mathcal{G}$, then $A = f^{-1}(A)$ for all $f \in \mathcal{F}$, so that

$$\pi(A) = \pi\left(f^{-1}(A)\right) = \int_{f^{-1}(A)} q \, d\lambda = \int_A q \circ f^{-1} \, d\lambda \qquad \text{for all } f \in \mathcal{F}.$$

Hence, condition (b) implies

$$2^p \, \pi(A) = \sum_{f \in \mathcal{F}} \int_A q \circ f^{-1} \, d\lambda = \int_A \left( \sum_{f \in \mathcal{F}} q \circ f \right) d\lambda = 2^p \, \lambda(A).$$

**"(c) $\Rightarrow$ (a)".** For each $x \in \mathbb{R}^{2p}$, define

$$\mu_x = \frac{\sum_{f \in \mathcal{F}} \delta_{f(x)}}{2^p}$$

where $\delta_{f(x)}$ denotes the unit mass at the point $f(x)$. Then,

$$\lambda(A) = \frac{\sum_{f \in \mathcal{F}} \lambda \circ f^{-1}(A)}{2^p} = \int \mu_x(A) \, \lambda(dx)$$

$$= \int \mu_x(A) \, \pi(dx) = \frac{\sum_{f \in \mathcal{F}} \pi \circ f^{-1}(A)}{2^p} \qquad \text{for all } A \in \mathcal{B}_{2p}$$

where the first equality follows from $\lambda \in \mathcal{P}$ and the third is because $\pi = \lambda$ on $\mathcal{G}$ and the map $x \mapsto \mu_x(A)$ is $\mathcal{G}$-measurable. $\square$

**Proof of Theorem 8.** We first recall a known fact. Let $\Phi : [0, 1]^n \to [0, 1]$ be a function such that $\Phi(u) = 0$, if $u_i = 0$ for some $i$, and $\Phi(u) = u_i$ if $u_j = 1$ for all $j \neq i$. Then, $\Phi$ is an $n$-copula and $\lambda_\Phi \ll m_n$ provided $\frac{\partial^n \Phi}{\partial u_n \ldots \partial u_1} \geq 0$ on $[0, 1]^n$.

After noting this fact, define

$$C^*(u) = C\left[ D_1(u_1, u_{p+1}), \ldots, D_p(u_p, u_{2p}) \right]$$

$$= \int_0^{D_1(u_1, u_{p+1})} \cdots \int_0^{D_p(u_p, u_{2p})} \varphi(t_1, \ldots, t_p) \, dt_1 \ldots dt_p \qquad \text{for each } u \in [0, 1]^{2p}.$$

Let $n = 2p$ and $\Phi = C^*$. By the result mentioned above, $C^*$ is a $2p$-copula and $\lambda_{C^*} \ll m_{2p}$ provided

$$\frac{\partial^{2p} C^*}{\partial u_{2p} \ldots \partial u_1} \geq 0 \qquad \text{everywhere on } [0, 1]^{2p}. \tag{9}$$

In this case, since $C^*$ is a copula, $H$ is a distribution function. Since $\lambda_{C^*} \ll m_{2p}$, one also obtains $\lambda_H \ll m_{2p}$ whenever $\mathcal{L}(X_i) \ll m_1$ for each $i \in I$. Therefore, it suffices to prove condition (9). In turn, (9) follows from condition (jjj) after noting that

$$\frac{\partial^{2p} C^*}{\partial u_{2p} \ldots \partial u_1} = \frac{\partial^p}{\partial u_{2p} \ldots \partial u_{p+1}} \, \frac{\partial^p C^*}{\partial u_p \ldots \partial u_1}$$

$$= \frac{\partial^p}{\partial u_{2p} \ldots \partial u_{p+1}} \, \varphi\left[ D_1(u_1, u_{p+1}), \ldots, D_p(u_p, u_{2p}) \right] \prod_{i=1}^p \frac{\partial}{\partial u_i} D_i(u_i, u_{p+i}). \quad \square$$

## References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance: Math. Econ. 44, 182–198.

Barber, R.F., Candes, E.J., 2015. Controlling the false discovery rate via knockoffs. Ann. Statist. 43, 2055–2085.

Barber, R.F., Candes, E.J., Samworth, R.J., 2020. Robust inference with knockoffs. Ann. Statist. 48, 1409–1431.

Bates, S., Candes, E.J., Janson, L., Wang, W., 2021. Metropolized knockoff sampling. J.A.S.A. 116, 1413–1427.

Bates, S., Sesia, M., Sabatti, C., Candes, E.J., 2020. Causal inference in genetic trio studies. Proc. Nat. Acad. Sci. USA 117 (39), 24117–24126.

Berti, P., Dreassi, E., Leisen, F., Pratelli, L., Rigo, P., 2022. Bayesian predictive inference without a prior. Statist. Sinica http://dx.doi.org/10.5705/ss.202021.0238, published online.

Berti, P., Pratelli, L., Rigo, P., 2004. Limit theorems for a class of identically distributed random variables. Ann. Probab. 32, 2029–2052.

Brooks, S., Gelman, A., Jones, G., Meng, X.L. (Eds.), 2011. Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC, New York.

Candes, E.J., Fan, Y., Janson, L., Lv, J., 2018. Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. J. R. Stat. Soc. Ser. B Stat. Methodol. 80, 551–577.

Embrechts, P., Lindskog, F., McNeil, A.J., 2003. Modelling dependence with copulas and applications to risk management. In: Rachev, S.T. (Ed.), In: Handbook of Heavy Tailed Distributions in Finance, Elsevier/North-Holland, Amsterdam.

Kotz, S., Balakrishnan, N., Johnson, N.L., 2000. Continuous Multivariate Distributions, second ed. Wiley, New York.

McNeil, A.J., Neslehova, J., 2009. Multivariate archimedean copulas, $d$-monotone functions and $l_1$-norm symmetric distributions. Ann. Statist. 37, 3059–3097.

Okhrin, O., Okhrin, Y., Schmid, W., 2013. On the structure and estimation of hierarchical Archimedean copulas. J. Econometrics 173, 189–204.

Savu, C., Trede, M., 2010. Hierarchies of Archimedean copulas. Quant. Finance 10, 295–304.

Sesia, M., Sabatti, C., Candes, E.J., 2019. Gene hunting with hidden Markov model knockoffs. Biometrika 106, 1–18.