Simultaneous false discovery proportion bounds via knockoffs and closed testing

Jinzhou Li, Marloes H. Maathuis, Jelle J. Goeman

September 17, 2022

Abstract

We propose new methods to obtain simultaneous false discovery proportion bounds for knockoff-based approaches. We first investigate an approach based on Janson and Su's k-familywise error rate control method and interpolation. We then generalize it by considering a collection of k values. We show that the bound of Katsevich and Ramdas is a special case of this method and can be uniformly improved. We further generalize the method by using closed testing with a multi-weighted-sum local test statistic. This closed testing based approach allows us to derive uniform improvement and other generalizations over previous methods, and we develop an efficient shortcut for its implementation. We compare the performance of our proposed methods in simulations and apply them to the UK Biobank data set.

1 Introduction

Many modern data analysis tasks are of an exploratory nature. In such cases, the simultaneous inference framework (Genovese and Wasserman, 2006; Goeman and Solari, 2011) is more flexible and can be more suitable than the commonly used false discovery rate (FDR) control framework (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001). Specifically, for a preset nominal FDR level, the only output of a FDR control framework is a rejection set R. In order for the FDR control guarantee (1) holds, users are not allowed to decide the nominal FDR level in a post-hoc manner or do anything to R. For example, they can not change to a smaller (or larger) FDR level even if the currently used one returns too many (or few) rejections than expected, or they can not remove any rejection from R even though some rejections might violate their scientific expertise. The simultaneous inference framework, however, does not suffer from such issues. Instead, it allows users to check any of their interested sets freely, and it will return high probability false discovery proportion (FDP) bounds (see (2)) for all these sets. As such, the simultaneous inference framework brings more flexibility and can be better suited for exploratory research.

Formally, let $\mathcal{N} \subseteq [p] = \{1, \dots, p\}$ be the index set of true null hypotheses among p null hypotheses. The FDP of a rejection set $R \subseteq [p]$ is defined as

$$\mathrm{FDP}(R) = \frac{|R \cap \mathcal{N}|}{\max\{1, |R|\}},$$

where $|\cdot|$ is the cardinality of a set. For a given data set and a fixed $\alpha \in (0, 1)$, the FDR control framework aims to produce a rejection set R such that

$$\mathrm{FDR}(R) = \mathbb{E}[\mathrm{FDP}(R)] \leq \alpha, \tag{1}$$

while the simultaneous inference framework aims to obtain a function $\overline{\mathrm{FDP}} : 2^{[p]} \to [0, 1]$, where $2^{[p]}$ denotes the power set of $[p]$, such that

$$\mathbb{P}(\mathrm{FDP}(R) \leq \overline{\mathrm{FDP}}(R), \forall R \subseteq [p]) \geq 1 - \alpha. \tag{2}$$

When (2) is satisfied, we say that $\overline{\mathrm{FDP}}(R)$ is a simultaneous FDP bound of R. The FDP upper bound is equivalent to the false discovery upper bound and the true discovery lower bound, in the sense that obtaining one, the other two can be calculated directly (see Appendix D.1 for more details). In this paper we focus on FDP upper bound.

Almost all existing methods of obtaining simultaneous bounds are based on p-values. For example, Genovese and Wasserman (2004); Meinshausen (2006); Hemerik et al. (2019) propose methods to obtain simultaneous FDP bounds for sets of the form $R_t = \{i \in [p] : p_i \leq t\}$, where p_i is the p-value for the i-th hypothesis, and Genovese and Wasserman (2006); Goeman and Solari (2011); Goeman et al. (2019); Blanchard et al. (2020); Vesely et al. (2021) propose simultaneous FDP bounds for all sets $R \subseteq [p]$.

For error-controlled variable selection problems, a novel framework called knockoffs (Barber and Candès, 2015; Candès et al., 2018) was developed without resorting to p-values. The fundamental idea of the knockoff method is to create synthetic variables (i.e., knockoffs) that serve as "negative controls". By using the knockoffs together with the original variables as inputs of an algorithm (e.g., lasso, random forests, or neural networks), one obtains knockoff statistics $W_1, \dots, W_p$, which possess the so-called coin-flip property (see Definition 2.1). Based on this property, a variable selection method was developed to control the FDR (Barber and Candès, 2015) or the k-familywise error rate (k-FWER, Janson and Su (2016)).

The knockoff method has been extended to many different settings, including group variable selection (Dai and Barber, 2016), multilayer variable selection (Katsevich and Sabatti, 2019), high-dimensional linear model (Barber and Candès, 2019), Gaussian graphical model (Li and Maathuis, 2021), multi-environment (Li et al., 2021), and time series settings (Chi et al., 2021). It has also been successfully applied to many real applications (see, e.g., Bates et al., 2020; Sesia et al., 2020; He et al., 2021; Sesia et al., 2021).

Almost all existing knockoff-based methods aim to control the FDR, but as we mentioned before, simultaneous FDP bounds could be a better alternative for exploratory research. This motivates us to consider the problem of obtaining simultaneous FDP bounds for knockoff-based approaches. To the best of our knowledge, there is only one existing paper by Katsevich and Ramdas (2020) that investigated this problem. Katsevich and Ramdas (2020) actually studied a more general problem about the relationship between FDR control and simultaneous FDP bounds. They proposed a general approach to obtain simultaneous FDP bounds for a broad range of FDR control procedures, and one of the results is the simultaneous FDP bounds for knockoff-based methods.

In this paper, we focus on the variable selection setting for convenience, but the same proposed methods can be applied as long as the coin-flip property (see Definition 2.1) holds. We start by investigating a simple approach that obtains simultaneous FDP bounds based on the k-FWER controlled set (Janson and Su, 2016) and interpolation (Blanchard et al., 2020; Goeman et al., 2021). We found that there are two issues with such an approach: it is unclear how to choose the tuning parameter k, which heavily affects the results, and the FDP bounds for the sets that are much larger or smaller than the k-FWER controlled set can be very conservative due to the nature of interpolation.

Addressing these two issues motivates us to consider several values of k at the same time. Specifically, the idea is to first get a collection of sets $S_1, \ldots, S_m$, such that the k_i-FWER of set S_i is controlled jointly for all i, then obtain the simultaneous FDP bounds by using interpolation based on these sets. This idea is a special case of a more general procedure called joint familywise error rate control proposed by Blanchard et al. (2020), where the set S_i is called the reference set in their paper, and related ideas can be traced back to Genovese and Wasserman (2006) and Van der Laan et al. (2004). In particular, Blanchard et al. (2020) considers p-value settings, and they propose methods to obtain reference set S_i, $i \in [m]$, such that the i-FWER of S_i is jointly controlled.

We propose a new method to obtain reference sets $S_1, \ldots, S_m$ with joint k_i-FWER control based on the knockoff statistics, and the simultaneous FDP bounds are then obtained by interpolation. This new method is uniformly better than the previous method based on only one k-FWER controlled set from Janson and Su (2016), and it largely addresses the second issue we mentioned above. The first issue remains as there are still tuning parameters to set, but it is less severe than before. We suggest some choices and propose a two-step approach to obtain the tuning parameters. In addition, we prove that the simultaneous FDP bounds from Katsevich and Ramdas (2020) is a special case of our method, and uniformly better bounds can be obtained.

Next, we turn to the closed testing framework and propose a general method built on it. In particular, we propose a multi-weighted-sum local test statistic based on the knockoff statistics. We prove that all previously mentioned approaches are special cases (or shortcuts) of this method, and we can derive uniform improvements and generalizations over them. The closed testing procedure is computationally intractable in its standard form, so we also develop an efficient and general shortcut for the implementation of this closed testing based method.

The remaining part of the paper is structured as follows. Section 2 gives a brief review of the knockoff statistics, the k-FWER control method using knockoffs, and the closed testing framework. In Section 3, we propose a method that obtains simultaneous FDP bounds based on joint k-FWER control and interpolation. In Section 4, we propose a closed testing based method to obtain simultaneous FDP bounds. We evaluate the numerical performance of the proposed methods in simulation studies and apply them to the UK Biobank dataset in Section 5. We close the paper with a discussion and some potential future research directions in Section 6.

2 Preliminaries

2.1 Knockoff statistics and the coin-flip property

The knockoff statistics $W_1, \ldots, W_p$ are obtained based on data and some specific knockoff method. In this paper, we do not specify the data generating model or how to obtain the knockoff statistics. Instead, we directly work on the knockoff statistic vector $W = (W_1, \ldots, W_p)$, and the only assumption we have is that W possesses the following coin-flip property.

Definition 2.1. *For knockoff statistic vector $W = (W_1, \ldots, W_p)$, let $D_i = sign(W_i)$. Let $\mathcal{N} \subseteq [p] = \{1, \ldots, p\}$ be the index set of null variables. We say that W possesses the* ***coin-flip property*** *if conditional on $(|W_1|, |W_2|, \ldots, |W_p|)$, D_i's are i.i.d. with distribution $\mathbb{P}(D_i = 1) = \mathbb{P}(D_i = -1) = 1/2$ for $i \in \mathcal{N}$.*

The coin-flip property holds in the two seminal papers Barber and Candès (2015) and Candès et al. (2018). Analogous to the p-value setting, where multiple testing procedures are designed based on the property that the null p-value is stochastically larger than the uniform random variable, the error control of knockoff methods is achieved by making use of the coin-flip property.

Throughout the paper, we assume without loss of generality that all W_i's are non-zero, there is no tie in $|W|$, and the indices are ordered, that is, $|W_1| > |W_2| > \cdots > |W_p| > 0$. Otherwise one can preprocess W by relabeling the indices, discarding the zero ones, and breaking ties while keeping the order unchanged. This preprocessing step has no influence on the remaining part of the paper because the preprocessed W still possesses the crucial coin-flip property.

2.2 K-FWER control via knockoffs

We review the k-FWER control method using knockoffs by Janson and Su (2016). This method will be a crucial component of our starting point for the interpolation-based method proposed in Section 3.

Specifically, for given $k \geq 1$ and $\alpha \in (0, 1)$, Janson and Su (2016) proposed to use the threshold

$$T(v) = \max\{|W_i| : \#\{j \in [p] : |W_j| \geq |W_i| \text{ and } W_j < 0\} = v\} \tag{3}$$

with $T(v) = \min_{i \in [p]}\{|W_i|\}$ if $\#\{j \in [p] : W_j < 0\} < v$, where

$$v = \operatorname{argmax}\left\{v' \in [p] : \sum_{i=k}^{\infty} 2^{-i-v'} \binom{i + v' - 1}{i} \leq \alpha\right\}. \tag{4}$$

The discovery set is then obtained by

$$S(v) = \{i \in [p] : W_i \geq T(v)\}. \tag{5}$$

Formula (4) for obtaining v comes from a key observation of Janson and Su (2016) that $|S(v) \cap \mathcal{N}|$ is stochastically dominated by a negative binomial random variable $NB(v)$ with distribution $\mathrm{NB}(v, 1/2)$ (note that the summation in (4) relates to the cumulative distribution function of $\mathrm{NB}(v', 1/2)$). So the k-FWER of $S(v)$ is controlled because

$$\mathbb{P}(|S(v) \cap \mathcal{N}| \leq k-1) \geq \mathbb{P}(NB(v) \leq k-1) \geq 1-\alpha. \tag{6}$$

Based on formula (4), one can calculate v for given k and α. As pointed by Janson and Su (2016), in the case of $|S(v)| < k-1$, one can always make the size of $S(v)$ to be $k-1$ without violating the k-FWER guarantee (6). In addition, due to the discrete nature of the negative binomial distribution, it can be the case that the largest v obtained by formula (4) doesn't exhaust the α level, in the sense that the summation inside (4) might not be close to α. To exhaust the α level, Janson and Su (2016) suggested a randomization scheme, which is more powerful than the non-randomized version, see Janson and Su (2016) for details.

2.3 Closed testing

The closed testing procedure was first proposed by Marcus et al. (1976) for familywise error rate control (FWER). Goeman and Solari (2011) observed that it could also be used to obtain simultaneous FDP bounds. This procedure will serve as the basis of our proposed method in Section 4.

Specifically, for hypotheses $H_1, \dots, H_p$, the closed testing procedure proceeds in three steps:

Step 1. Locally test all intersection hypotheses $H_I = \cap_{i \in I} H_i$ for $I \subseteq [p]$ at level $\alpha \in (0,1)$. Here we use the convention that $H_\emptyset$ is a true hypothesis and we will always accept it. We use notation $\phi_I \in \{0,1\}$ to denote the local test result, with 1 indicating rejection of H_I. For convenience, we sometimes say that a set I is rejected to mean that H_I is rejected.

Step 2. Obtain the closed testing results based on local test results: for all $I \subseteq [p]$, H_I is rejected by closed testing if all supersets J of I are locally rejected. We use

$$\phi_I^{[p]} = \min\{\phi_J : I \subseteq J \subseteq [p]\}$$

to denote the closed testing result for hypothesis H_I.

Step 3. For any set $R \subseteq [p]$, let

$$t_\alpha(R) = \max_{I \in 2^R} \{|I| : \phi_I^{[p]} = 0\} \tag{7}$$

be the size of the largest subset of R that is not rejected by closed testing. Then the simultaneous FDP bound for R is

$$\overline{\text{FDP}}^{ct}(R) = \frac{t_\alpha(R)}{\max\{1, |R|\}}. \tag{8}$$

The first two steps guarantee that the FWER of the closed testing results is controlled. To see this, note that the event that $H_{\mathcal{N}}$ is not locally rejected (this event happens with a probability larger than $1-\alpha$ for a valid local test) implies that any true null hypothesis H_I with $I \subseteq \mathcal{N}$ will not be rejected by closed testing, so there is no false rejection.

Step 3 is the new step from Goeman and Solari (2011) which gives us simultaneous FDP bounds. To see this, note that the event that $H_{\mathcal{N}}$ is not locally rejected implies that $|R \cap \mathcal{N}| \leq t_\alpha(R)$ by the definition of $t_\alpha(R)$ and the fact that $R \cap \mathcal{N}$ will not be rejected by closed testing for any set $R \subseteq [p]$. Therefore, given valid local tests, we have $\mathbb{P}(|R \cap \mathcal{N}| \leq t_\alpha(R), \forall R \subseteq [p]) \geq \mathbb{P}(\phi_{\mathcal{N}} = 0) \geq 1 - \alpha$, which leads to simultaneous FDP bounds by dividing $\max\{1, |R|\}$ on both sides inside the probability.

The closed testing procedure is computationally intractable in its standard form as there are $2^p - 1$ hypotheses to test in step 1. In some cases, however, there exist shortcuts to derive the closed testing results efficiently (see, e.g., Goeman and Solari, 2011; Dobriban, 2020; Goeman et al., 2021; Tian et al., 2021; Vesely et al., 2021).

3 Simultaneous FDP bounds via knockoffs and joint k-FWER control with interpolation

3.1 K-FWER control with interpolation

Let $R \subseteq [p]$ be the interested set for which we want to obtain simultaneous FDP bound. The first method we will investigate is based on the k-FWER controlled set $S(v)$ from Janson and Su (2016) (see (5)) and interpolation. In our case, the idea of using interpolation to get the FDP bound is simple: elements in R but not $S(v)$ (i.e. $R \backslash S(v)$) will be treated as false discoveries because we have no information about them, and elements in both R and $S(v)$ (i.e. $R \cap S(v)$) has less than $\min\{|R \cap S(v)|, |S(v) \cap \mathcal{N}|\}$ false discoveries, for which we have a high probability upper bound due to the k-FWER control of $S(v)$.

Formally, with probability larger than $1 - \alpha$, we have

$$\begin{aligned} |R \cap \mathcal{N}| &= |R \cap S(v) \cap \mathcal{N}| + |(R \backslash S(v)) \cap \mathcal{N}| \leq \min\{|R \cap S(v)|, |S(v) \cap \mathcal{N}|\} + |R \backslash S(v)| \\ &\leq \min\{|R \cap S(v)|, k-1\} + |R \backslash S(v)| = \min\{|R|, k - 1 + |R \backslash S(v)|\}. \end{aligned} \tag{9}$$

Hence

$$\overline{\text{FDP}}_{k,v}(R) = \frac{\min\{|R|, k - 1 + |R \backslash S(v)|\}}{\max\{1, |R|\}} \tag{10}$$

is a simultaneous FDP bound because

$$\mathbb{P}(\text{FDP}(R) \leq \overline{\text{FDP}}_{k,v}(R), \forall R \subseteq [p]) \geq \mathbb{P}(|S(v) \cap \mathcal{N}| \leq k - 1) \geq 1 - \alpha.$$

We mention that interpolation is actually a more general technique, which reduces to the simplified form (9) in our special case. See Blanchard et al. (2020) and Goeman et al. (2021) for more details.

As mentioned in Section 2.2, when $|S(v)| < k - 1$, one can always obtain a larger discovery set S' with $|S'| = k - 1$, and the k-FWER control guarantee still holds. Hence one may think that the simultaneous FDP bounds (10) based on S' might be better (i.e., smaller). However, as we show in Appendix A, the FDP bounds (10) based on S' and $S(v)$ are actually equivalent, so we will use $S(v)$ for simplicity.

Based on the k-FWER controlled set $S(v)$, the obtained simultaneous FDP bound (10) is free and conservative, and there are two issues: (a) it is not clear a priori how to choose the tuning parameter k, which can heavily affect the performance of the simultaneous FDP bounds. (b) the obtained simultaneous FDP bounds can be very conservative for a set R that is much smaller or larger than $S(v)$. To better illustrate these two issues, we present Figure 1 showing the simultaneous FDP bounds for sets $R_i = \{j \le i : W_j > 0\}$, $i \in [p]$, in four simulation settings (see Appendix F.1 for more details).

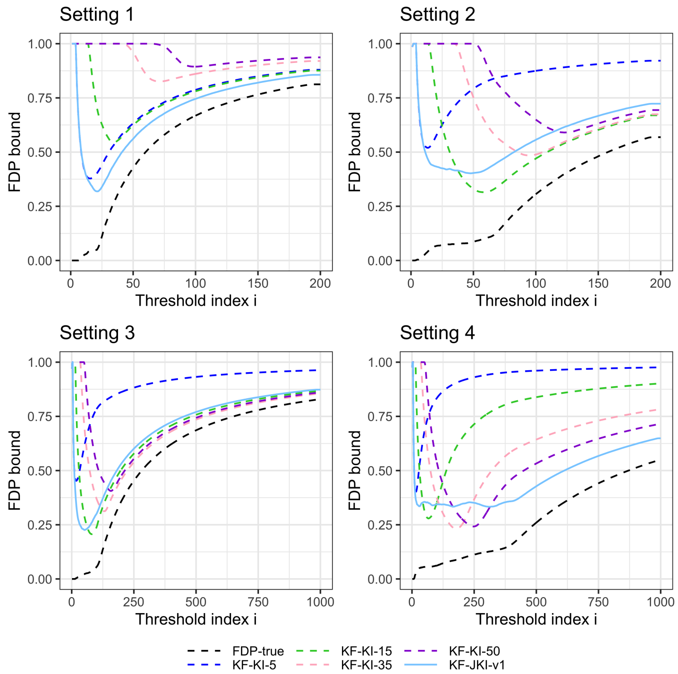

Figure 1: Simultaneous FDP bound (10) with $k \in \{5, 15, 35, 50\}$ (denoted by KF-KI-5, KF-KI-15, KF-KI-35, and KF-KI-50, respectively) and our later proposed Algorithm 1 with inputs $v = (v_i = i)_{i=1}^p$ and k obtained by using Algorithm 2 (denoted by KF-JKI-v1). The black solid line indicates the true FDP. All results are average values over 200 replications.

3.2 Joint k-FWER control with interpolation

Addressing the above two issues motivates our new method. The idea is that since the FDP bounds based on different k can be better in different settings, we use many k-FWER controlled sets to construct them. As we mentioned in Section 1, this idea is a special case of a more general method proposed by Blanchard et al. (2020).

Formally, for $\alpha \in (0,1)$, $v = (v_1, \dots, v_m)$ and $k = (k_1, \dots, k_m)$ (here we slightly abuse notation as k and v were scalars before), if the following joint k-FWER control holds (we will discuss how to obtain it later)

$$\mathbb{P}(|S(v_1) \cap \mathcal{N}| \leq k_1 - 1, \dots, |S(v_m) \cap \mathcal{N}| \leq k_m - 1) \geq 1 - \alpha, \tag{11}$$

where $S(v_i) = \{j \in [p] : W_j \geq T(v_i)\}$ and $T(v_i)$ is defined by formula (3). Then for any $R \subseteq [p]$, we can obtain the FDP bound by

$$\overline{\mathrm{FDP}}^m_{(k,v)}(R) = \min_{i \in [m]} \overline{\mathrm{FDP}}_{k_i, v_i}(R), \tag{12}$$

where $\overline{\mathrm{FDP}}_{k_i,v_i}(R)$ is defiend by (10). $\overline{\mathrm{FDP}}^m_{(k,v)}(R)$ is a valid simultaneous FDP bound because

$$\begin{aligned} &\mathbb{P}(\mathrm{FDP}(R) \leq \overline{\mathrm{FDP}}^m_{(k,v)}(R), \forall R \subseteq [p]) \\ \geq &\mathbb{P}(|S(v_1) \cap \mathcal{N}| \leq k_1 - 1, \dots, |S(v_m) \cap \mathcal{N}| \leq k_m - 1) \geq 1 - \alpha. \end{aligned}$$

The previous FDP bound (10) is a special case of bound (12) with $m = 1$, $S(v_1) = S(v)$ and $k_1 = k$, and it is clear that the FDP bound (12) with $m > 1$, $S(v_1) = S(v)$ and $k_1 = k$ is uniformly better (i.e., smaller) than bound (10).

As mentioned in Section 2.2, the randomization scheme is one way to exhaust the α level, but it suffers from the previously mentioned two issues. The joint k-FWER control (11) can be seen as another way to exhaust the α level. As a result, it outputs many error-controlled sets on which we construct simultaneous FDP bounds. As we will see later, this method largely alleviates the two issues.

To obtain vectors $v = (v_1, \dots, v_m)$ and $k = (k_1, \dots, k_m)$ satisfying the joint k-FWER control guarantee (11), we follow similar ideas from Janson and Su (2016) and generalize their result. In particular, we have the following lemma:

Lemma 3.1. *Let $\mathcal{N} \subseteq [p]$ be the set of null variables. For knockoff statistic vector $W = (W_1, \dots, W_p)$ satisfying the coin-flip property and $|W_1| > \cdots > |W_p| > 0$, let $S(v) = \{i \in [p] : W_i \geq T(v)\}$, where $T(v)$ is defined by formula* (3)*. Then, for positive integers $v_1 < \cdots < v_m$ and $k_1 < \cdots < k_m$, we have*

$$\begin{aligned} &\mathbb{P}(|S(v_1) \cap \mathcal{N}| \leq k_1 - 1, \dots, |S(v_m) \cap \mathcal{N}| \leq k_m - 1) \\ \geq &\mathbb{P}(NB^p(v_1) \leq k_1 - 1, \dots, NB^p(v_m) \leq k_m - 1), \end{aligned} \tag{13}$$

where $NB^p(v_1), \dots, NB^p(v_m)$ are early-stopped negative binomial random variables defined on the same Bernoulli trials of length p.

Based on Lemma 3.1, for $\alpha \in (0,1)$, if vectors $v = (v_1, \ldots, v_m)$ and $k = (k_1, \ldots, k_m)$ satisfy

$$\mathbb{P}(NB^p(v_1) \le k_1 - 1, \ldots, NB^p(v_m) \le k_m - 1) \ge 1 - \alpha, \tag{14}$$

then the joint k-FWER control guarantee (11) holds, and $\overline{\mathrm{FDP}}^m_{(k,v)}(R)$ (see (12)) is a simultaneous FDP bound.

In the case of Janson and Su (2016) of getting the k-FWER controlled set, $p = \infty$, $m = 1$ and k is user-chosen in the inequality (14), so the optimal v can be naturally obtained via optimization (4) because it is clear that the largest v leads to the most discoveries. In our case of getting simultaneous FDP bounds with $m > 1$, things are more complicated because the vectors v and k satisfying inequality (14) are not unique, and the best v and k depend on the unknown distribution of W. Hence there are no uniformly best v and k, and they can be viewed as tuning parameters and need to be chosen. We will offer some suggestions and discuss this in more detail in the following sections.

We summarize the method of obtaining simultaneous FDP bounds based on joint k-FWER and interpolation in Algorithm 1. For a quick illustration, we present the performance of this algorithm in Figure 1, where the tuning parameters $v = (v_i = i)_{i=1}^p$ and k obtained by our later proposed Algorithm 2. One can see that our proposed method can help to alleviate the two issues, and it generally performs well in all four settings.

Algorithm 1 : Simultaneous FDP bounds via knockoffs and joint k-FWER control with interpolation

Input: (R, W, α, v, k), where $R \subseteq [p]$ is any non-empty set, $W \in \mathbb{R}^p$ is the vector of the knockoff statistics, $\alpha \in (0,1)$ is the nominal level for the simultaneous FDP control, $v = (v_1, \ldots, v_m)$ and $k = (k_1, \ldots, k_m)$ are vectors satisfying inequality (14).

Output: the FDP upper bound for set R.

1: For $i \in [m]$, obtain $S(v_i) = \{j \in [p] : W_j \ge T(v_i)\}$, where $T(v_i)$ is defined by (3).

2: Obtain the FDP upper bound for set R by formula (12).

3.3 Uniform improvement of Katsevich and Ramdas (2020)

We first present a valid choice of tuning parameters v and k implied by the method of Katsevich and Ramdas (2020) in Proposition 3.1. In addition, we show that the interpolation version (17) of the FDP bound from Katsevich and Ramdas (2020) is a special case of our proposed FDP bound (12). The relationship between (17) and the work of Katsevich and Ramdas (2020) is explained in Appendix D.3.

Proposition 3.1. *For $\alpha \in (0,1)$, $p \ge 1$, $0 < v_1 < \cdots < v_m$ and $i \in [m]$, let*

$$k^{raw}_{v_i} = \min_{j \ge 1}\{c_j : j - c_j + 1 = v_i\} \quad \textit{with} \quad c_j = \left\lfloor \frac{c(\alpha)\cdot(1+j)}{1+c(\alpha)} \right\rfloor + 1, \tag{15}$$

where $\lfloor a \rfloor$ denotes the largest integer smaller than or equal to $a \in \mathbb{R}$ and $c(\alpha) = \frac{\log(\alpha^{-1})}{\log(2-\alpha)}$. We have the following two results:

(i) Let $NB^p(v_1), \ldots, NB^p(v_m)$ be early-stopped negative binomial random variables defined on the same Bernoulli trials of length p, then

$$\mathbb{P}(NB^p(v_1) \leq k_{v_1}^{raw} - 1, \ldots, NB^p(v_m) \leq k_{v_m}^{raw} - 1) \geq 1 - \alpha. \tag{16}$$

(ii) For knockoff statistic vector $W = (W_1, \ldots, W_p)$, $v = (1, \ldots, p)$, $k^{raw} = (k_1^{raw}, \ldots, k_p^{raw})$ and $R \subseteq [p]$, let $\overline{FDP}^p_{(k^{raw},v)}(R)$ be the output of Algorithm 1 with input $(R, W, \alpha, v, k^{raw})$, and let

$$\overline{FDP}^{KR}(R) = \frac{\min_{i\in[p]}\{|R\backslash S_i| + \bar{V}(|S_i^-|), |R|\}}{\max\{1, |R|\}}, \tag{17}$$

where $S_i = \{j \leq i : W_j > 0\}$, $S_i^- = \{j \leq i : W_j < 0\}$, and $\bar{V}(|S_i^-|) = \lfloor c(\alpha) \cdot (1 + |S_i^-|)\rfloor$. Then

$$\overline{FDP}^{KR}(R) = \overline{FDP}^p_{(k^{raw},v)}(R).$$

The first statement of Proposition 3.1 shows that for any given vector $v = (v_1, \ldots, v_m)$, the vector $k^{raw} = (k_{v_1}^{raw}, \ldots, k_{v_m}^{raw})$ defined by (15) satisfies inequality (14), thus they are valid choices for Algorithm 1. The second statement shows that the method of Katsevich and Ramdas (2020) is a special case of our proposed method. Specifically, it is equivalent to the FDP bound obtained by using Algorithm 1 with tuning parameters $v = (v_i = i)_{i=1}^p$ and $k^{raw} = (k_1^{raw}, \ldots, k_p^{raw})$.

For vector $v = (v_1, \ldots, v_m)$, using $k = (k_{v_1}^{raw}, \ldots, k_{v_m}^{raw})$ in Algorithm 1 is not optimal and can be further improved. Specifically, the constant $c(\alpha)$ in (15), based on which we obtain k^{raw} from v, tends to be too large. That is, $\mathbb{P}(NB^p(v_1) \leq k_{v_1}^{raw} - 1, \ldots, NB^p(v_m) \leq k_{v_m}^{raw} - 1)$ and $1 - \alpha$ are not close, so that the α level is not exhausted. To this end, we propose a two-step approach to improve k^{raw} as follows. Let j_i^* be the index such that $k_{v_i}^{raw} = c_{j_i^*}$. In the first step, we refine $k_{v_i}^{raw}$ by replacing the constant $c(\alpha)$ in $c_{j_i^*}$ (see (15)) with the smallest constant such that inequality (14) still holds. In most cases, however, the first step may still not exhaust the α level due to the discrete nature of the probability. Hence in the second step, we greedily update k_i under the constraint (14). In practical implementation, we update $c(\alpha)$ in a discrete manner in the first step, and we use Monte-Carlo simulation to approximate the target probability $\mathbb{P}(NB^p(v_1) \leq k_1 - 1, \ldots, NB^p(v_m) \leq k_m - 1)$, since exact calculation of this probability is computationally infeasible for large m. We summarize this two-step approach in Algorithm 2.

To illustrate the improvement of using Algorithm 2 over k^{raw}, we present the corresponding vectors k^{raw}, k^{step1} and k^{step2} based on $v = (v_i = i)_{i=p}^p$ in plot (A) of Figure 2 (see Appendix F.2 for the plots of other types of v proposed in the next section). One can see that k_i^{step2} is smaller than k_i^{step1} and k_i^{raw} for the same v_i, hence it will lead to better FDP bounds. We also present the corresponding target probabilities for the four types of v (see (18) in next section) in Table 1. One can see that k^{step2} exhausted the α level while k^{step1} and k^{raw} didn't.

Algorithm 2 : Two-step approach to obtain $k_1, \ldots, k_m$

Input: $(v_1, \ldots, v_m, \alpha, \delta, p)$, where $v_1 < \cdots < v_m$, $\alpha \in (0,1)$ is the nominal level for the simultaneous FDP control, δ is the step size used in the first step, and p is the parameter in inequality (14).

Output: $k_1, \ldots, k_m$.

1: For $i \in [m]$, obtain $j_i^* = \operatorname{argmin}_{j \geq 1}\{c_j : j - c_j + 1 = v_i\}$, where c_j is defined by (15). Refine $c(\alpha)$ by using step size δ: find the largest $N \geq 0$ such that inequality (14) holds with $(v_1, \ldots, v_m)$ and $(k_1^{step1}, \ldots, k_m^{step1})$, where $k_i^{step1} = \left\lfloor \frac{c^{step1} \cdot (1+j_i^*)}{1+c^{step1}} \right\rfloor + 1$ and $c^{step1} = c(\alpha) - N\delta$.

2: Update $k_1^{step1}, \ldots, k_m^{step1}$ in a greedy way: find the smallest $k_1^{step2} \leq k_1^{step1}$ such that inequality (14) holds with $(v_1, \ldots, v_m)$ and $(k_1^{step2}, k_2^{step1}, \ldots, k_m^{step1})$, then keep k_1^{step2} and find the smallest $k_2^{step2} \leq k_2^{step1}$ such that inequality (14) holds with $(v_1, \ldots, v_m)$ and $(k_1^{step2}, k_2^{step2}, \ldots, k_m^{step1})$. Iterate this process until we get all $k_1^{step2}, \ldots, k_m^{step2}$, and return them as output.

	$v_i = i$	$v_1 = 1, v_2 = 2,$ $v_i = v_{i-1} + v_{i-2}$	$v_i = \lfloor i^2/2 \rfloor$	$v_i = 2^{i-1}$
k^{raw}	0.9635 (0.0006)	0.9633 (0.0006)	0.9637 (0.0006)	0.9631 (0.0007)
k^{step1}	0.9575 (0.0007)	0.9568 (0.0006)	0.9583 (0.0006)	0.9560 (0.0008)
k^{step2}	0.9500 (0)	0.9500 (0)	0.9500 (0)	0.9500 (0)

Table 1: The empirical target probabilities $\mathbb{P}(NB^p(v_1) \leq k_1 - 1, \ldots, NB^p(v_m) \leq k_m - 1)$ for k^{raw}, k^{step1} and k^{step2} based on four types of v, with standard deviation in the brackets. Here we use $p = 1000$ and let $v_i < 150$. The results are based on 100 replications.

For vector $v = (1, 2, \ldots, p)$, let $k = (k_1, \ldots, k_p)$ be the output of Algorithm 2. It is clear that $k_i \leq k_i^{raw}$ for all $i \in [p]$, hence $\overline{\text{FDP}}^p_{(k,v)}(R) \leq \overline{\text{FDP}}^p_{(k_{raw},v)}(R) = \overline{\text{FDP}}^{\text{KR}}(R)$ for any $R \subseteq [p]$ by the second result of Proposition 3.1. In other words, the FDP bound $\overline{\text{FDP}}^p_{(k,v)}(R)$ is at least as good as the one from Katsevich and Ramdas (2020). As we will see in Section 5, it is actually strictly better in many simulations.

It might have been expected that the FDP bound of Katsevich and Ramdas (2020) could be uniformly improved, because that bound was obtained using a general martingale approach which can also be applied to non-knockoff settings, while our method makes more targeted use of the Bernoulli trials, which are fundamental for the knockoff-based approach. This enables us to obtain better FDP bounds in that specific context. A natural follow-up question is: Can we further improve $\overline{\text{FDP}}^p_{(k,v)}(R)$? As we will show in Section 4, the answer is yes. The FDP bound $\overline{\text{FDP}}^p_{(k,v)}(R)$ is still not admissible (Goeman et al., 2021), and a further uniform improvement can be obtained by connecting it to the closed testing framework.

3.4 Other choices of tuning parameters

Instead of using $v = (v_i = i)_{i=1}^{p}$, one may use other types of v because this one might not be the best in some settings. In this paper, we consider the following four types of vector v (ordered in a way that v_i grows ever faster):

$$\begin{aligned} &\text{Type-1:}\quad v_i = i, && \text{Type-2:}\quad v_i = \lfloor i^2/2 \rfloor, \\ &\text{Type-3:}\quad v_1 = 1, v_2 = 2, v_i = v_{i-1} + v_{i-2}, && \text{Type-4:}\quad v_i = 2^{i-1}. \end{aligned} \tag{18}$$

We use $v_1 = 1$ for all of them because it gives the best FDP bounds when the obtained knockoff statistic vector is optimal. Namely, when all true W_i's are positive and have larger absolute values than any null W_i. Different growth rates of v lead to different k_i for the same v_i and different reference sets, so the resulting FDP bounds can be better for different sets in different settings.

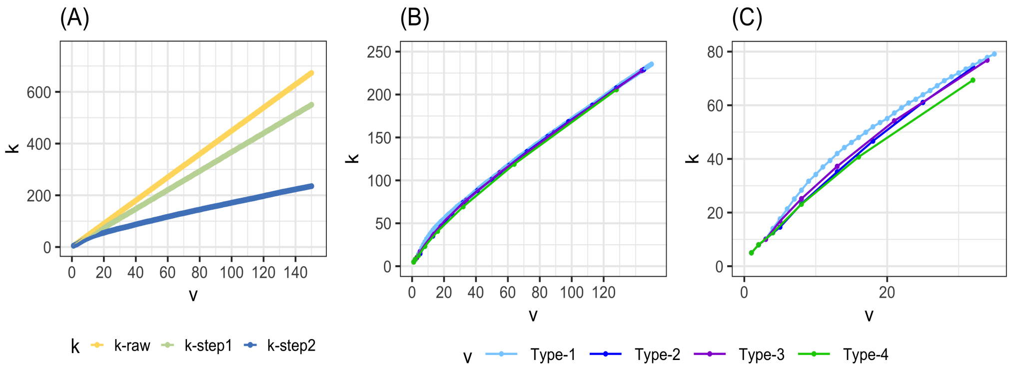

Figure 2: Plot (A): The vectors k^{raw}, k^{step1} and k^{step2} based on the Type-1 v in (18). Plot (B): The output vector k by using Algorithm 2 with four types of v in (18). Plot (C): The zoom-in version of Plot (B). For all three plots, we use $p = 1000$ and let $v_i < 150$. All results are average values over 100 replications.

In plots (B) and (C) of Figure 2, we compare the corresponding k based on four types of v using Algorithm 2. One can see that all of them give similar values of k_i for a large v_i, but for a small v_i, the k_i will be smaller for v with a faster growth rate. For example, the Type-1 k_i is larger than the Type-4 k_i for $v_i = 16$. This is because that there are more elements between $v_1 = 1$ and $v_i = 16$ in inequality (14) when using the Type-1 v. But note that it is still unclear which one will give a better FDP bound for a set R, because using Type-1 v will generate more reference sets than using Type-4 v, which largely influences the final FDP bound.

In practice, people may choose v based on their background knowledge. For example, a vector v with a faster growth rate is desired if it is expected that there will be blocks of negative W_i's. We found that the Type-2 v performs quite well in our simulations, so we recommend it as a default choice in practice. Lastly, for the choice of the largest value v_m, based on both its role in Algorithm 1 and 2 and our experience from simulations, we found that this value does not have a significant influence as long as it is relatively large (e.g., $p/3$), so one may set it according to the affordable computational expense.

4 Simultaneous FDP bounds via knockoffs and closed testing

In this section, we turn to the closed testing framework and propose a method to obtain simultaneous FDP bounds by combining it with the knockoff-based approach. In particular, we propose a local test based on the knockoff statistics for closed testing. We show that Algorithm 1 is a special case (or an exact shortcut) of this closed testing based approach, and it can be uniformly improved. We also present other generalizations based on closed testing and develop a shortcut for their implementation.

4.1 The multi-weighted-sum local test statistic and connections to other methods

For variable selection problem, the null hypothesis with respect to set $I \subseteq [p]$ is

$$H_I : \text{the } i\text{-th variable is null}, \ \forall i \in I.$$

For knockoff statistic vector $W = (W_1, \ldots, W_p)$ possesses the coin-flip property, $D_i = \text{sign}(W_i)$'s are i.i.d. with distribution $\mathbb{P}(D_i = 1) = \mathbb{P}(D_i = -1) = 1/2$ for $i \in I$ under the null hypothesis H_I. Based on this, we propose the following multi-weighted-sum test statistic to locally test H_I in the closed testing framework:

$$L_i^I = \sum_{j \in I} w_{i,j}^I \mathbb{1}_{D_j = 1}, \quad i \in [m], \tag{19}$$

where $m \geq 1$ and $w_{i,j}^I$ are tuning parameters. Let $z_1^I, \ldots, z_m^I$ be the corresponding critical values such that under H_I,

$$\mathbb{P}(L_1^I \leq z_1^I - 1, \ldots, L_m^I \leq z_m^I - 1) \geq 1 - \alpha. \tag{20}$$

We locally reject H_I if there exists some $i \in [m]$ such that $L_i^I \geq z_i^I$. Note that these critical values can be approximated using Monte-Carlo simulation because the distribution of D_i's under H_I is known.

Given a concrete local test, the remaining key issue of obtaining FDP bounds using closed testing is the computation, which is infeasible in the standard form. Hence, a shortcut is needed for the implementation. In the following, we show that Algorithm 1 is actually an exact shortcut for closed testing using local test statistic (19) with $w_{i,j}^I = \mathbb{1}_{r_j^I > |I| - b_i^I}$, that is,

$$L^I(b_i^I) = \sum_{j \in I} \mathbb{1}_{r_j^I > |I| - b_i^I} \mathbb{1}_{D_j = 1}, \quad i = 1, \ldots, m, \tag{21}$$

where b_i^I's, $i \in I$, are tuning parameters and r_i^I is the local rank of $|W_i|$ in $\{|W_j|, j \in I\}$. For example, $r_1^I = 3, r_3^I = 2, r_4^I = 1$ for $I = \{1, 3, 4\}$.

In particular, the following Theorem 4.1 shows that the FDP bounds obtained by closed testing using local test statistic (21) with $b_i^I = k_i - v_i + 1$ and critical value $z_i^I = k_i$

is equivalent to the FDP bounds obtain by using Algorithm 1 with tuning parameters $v = (v_1, \ldots, v_m)$ and $k = (k_1, \ldots, k_m)$.

Theorem 4.1. *For $m \geq 1$, $\alpha \in (0,1)$, $v = (v_1, \ldots, v_m)$, $k = (k_1, \ldots, k_m)$, $W = (W_1, \ldots, W_p)$ and $R \subseteq [p]$, let $\overline{FDP}^m_{(k,v)}(R)$ (see (12)) be the output of Algorithm 1 with input (R, W, v, k, α), and let $\overline{FDP}^{ct}(R)$ (see (8)) be the FDP bound based on closed testing using test statistic (21) with $b_i^I = k_i + v_i - 1$ and critical values $z_i^I = k_i$ to locally test H_I. Then $\overline{FDP}^{ct}(R) = \overline{FDP}^m_{(k,v)}(R)$.*

We point out that one can also apply a general method from Goeman et al. (2021) to connect the simultaneous FDP bound (12) to closed testing (see Appendix D.4 for a concrete example). This approach, however, only results in $\overline{\mathrm{FDP}}^{ct}(R) \leq \overline{\mathrm{FDP}}^m_{(k,v)}(R)$, which might be expected due to the generality of the method. Our proof of Theorem 4.1 is more targeted to the special case we consider here and does not rely on the method from Goeman et al. (2021). As a result, it leads to a more precise characterization $\overline{\mathrm{FDP}}^{ct}(R) = \overline{\mathrm{FDP}}^m_{(k,v)}(R)$, which means that Algorithm 1 is an exact shortcut of the corresponding closed testing procedure.

Proposition 3.1 and Theorem 4.1 directly implies that the interpolation version of the FDP bounds from Katsevich and Ramdas (2020) is also a special case of closed testing:

Corollary 4.1. *For $\alpha \in (0,1)$, $W = (W_1, \ldots, W_p)$ and $R \subseteq [p]$, let $\overline{FDP}^{KR}(R)$ be defined by (17) and $\overline{FDP}^{ct}(R)$ (see (8)) be the FDP bound based on closed testing using test statistic (21) with $m = |I|$, $b_i^I = k_i^{raw} + i - 1$ and critical values $z_i^I = k_i^{raw}$ to locally test H_I, where k_i^{raw} and c_j are defined by (15). Then $\overline{FDP}^{ct}(R) = \overline{FDP}^{KR}(R)$.*

As a side result, we show a connection between the k-FWER control method of Janson and Su (2016) and closed testing. Specifically, for a given k, the method of Janson and Su (2016) output a set $S(v)$ whose k-FWER is controlled. The following Proposition 4.1 shows that in the case of $|S(v)| \geq k-1$, the same k-FWER control guarantee for set $S(v)$ is obtained by closed testing using local test statistic (21) with $m = 1$, $b_1^I = k - v + 1$ and $z_1^I = k$. In other words, if we search over all sets whose false discovery upper bound is k based on the closed testing results, then $S(v)$ is one of them.

Proposition 4.1. *For $\alpha \in (0,1)$, $k \geq 1$ and $W = (W_1, \ldots, W_p)$. Let v be calculated by formula (4) and let $S(v)$ be defined by (5). For the closed testing procedure using test statistic (21) with $m = 1$, $b_1^I = k + v - 1$ and critical value $z_1^I = k$ to locally test H_I, let $t_\alpha(S(v))$ be the corresponding false discovery upper bound defined by (7). Then $t_\alpha(S(v)) = |S(v)|$ when $|S(v)| < k-1$ and $t_\alpha(S(v)) = k-1$ when $|S(v)| \geq k-1$.*

4.2 Uniform improvement, generalization and shortcut

Based on Theorem 4.1, uniform improvement of $\overline{\mathrm{FDP}}^m_{(k,v)}(R)$ (see (12)) can be achieved. In particular, when locally testing the null hypothesis H_I using test statistic (21), instead of using the same tuning parameter $b_i^I = k_i + v_i - 1$ and critical value $z_i^I = k_i$ for all $I \subseteq [p]$, we use the adapted $b_i^I = k_i^{|I|} + v_i - 1$ and $z_i^I = k_i^{|I|}$, where $k_i^{|I|}$ is obtained by using Algorithm 2 with $|I|$, instead of p, as the last input value. Note that the new local test is valid due to the fact (24) in Appendix C, and because it is clear that $k_i^{|I|} \leq k_i$, the new local test is uniformly better than the previous one, in the sense that it must locally reject H_I if the previous local test rejects. This then brings a uniform improvement over the previous FDP bounds.

Aside from uniform improvement, closed testing framework also enables generalizations of $\overline{\mathrm{FDP}}^m_{(k,v)}(R)$ by using other local tests. As a first attempt, we incorporate the local rank r_j^I in the weights of test statistic (19) to use $w_{i,j}^I = r_j^I \mathbb{1}_{r_j^I > |I| - b_i^I}$ instead of $w_{i,j}^I = \mathbb{1}_{r_j^I > |I| - b_i^I}$, and we implemented simulations to test the performance of this method, see Appendix F.4 for more details. In addition, the closed testing framework also provides a direct way for users to make use of their background knowledge. For instance, one may use a large weight $w_{i,j}^I$ in test statistic (19) if they believe that the i-th variable is non-null. We will not investigate more in this direction and leave it for future research.

For the above uniformly improved (i.e., using local test statistic (21) with $b_i^I = k_i^{|I|} + v_i - 1$ and critical value $z_i^I = k_i^{|I|}$) and rank-generalization versions (i.e., using local test statistic (19) with $w_{i,j}^I = r_j^I \mathbb{1}_{r_j^I > |I| - b_i^I}$, $b_i^I = k_i^{|I|} + v_i - 1$ and critical values calculated based on inequality (20)), the analytic shortcut $\overline{\mathrm{FDP}}^m_{(k,v)}(R)$ (see Theorem 4.1) does not apply anymore, hence a new shortcut is needed for the practical implementation of closed testing. To this end, we propose a general shortcut which is valid when the following two conditions hold:

(C1) The critical values $z_1^I, \ldots, z_m^I$ depend on I only through $|I|$.

(C2) For any two sets $I_1 = \{u_1, \ldots, u_s\}$ and $I_2 = \{v_1, \ldots, v_s\}$ of the same size, there exists a permutation $\pi : [p] \to [p]$ such that if $\pi_{(j)}^{I_1} \leq \pi_{(j)}^{I_2}$ for all $j = 1, \ldots, s$, then $L_i^{I_1} \leq L_i^{I_2}$ for all $i \in [m]$, where $\pi_{(1)}^I < \cdots < \pi_{(s)}^I$ are ordered values of $\{\pi(i), i \in I\}$.

These two conditions ensure that for null hypothesis H_I with $|I| = s$, to check whether all subsets of I of size $t \in \{s, \ldots, 1\}$ are rejected by closed testing, we only need to check one set B of size t (see Algorithm 3 for the definition of set B). Then, to check whether all supersets of B of size $r \in \{0, \ldots, p - t\}$ are locally rejected, we again only need to check one set of this size. This brings us the necessary computational reduction in the implementation of closed testing. These two conditions hold for the above mentioned uniformly improved and rank-generalization versions. In particular, one such permutation that satisfies condition (C2) can be obtained by letting $\pi(i) = (l)$, where (l) is the index of W_i in the sorted $W_{(1)} < \cdots < W_{(p)}$.

We summarize the shortcut of getting FDP bounds using closed testing in Algorithm 3, and show its validity in Appendix B. The computational complexity of Algorithm 3 is $O(p^2)$ in the worst case, which is larger than the previous analytic shortcut Algorithm 1. We think that this is a fair price to pay for a more broader shortcut, in the sense that it can be applied to more types of local tests. In practice, users may first use Algorithm 1, and continuous with the uniformly improved version and using Algorithm 3 for its calculation if the result is not already satisfactory.

Algorithm 3 : A shortcut for closed testing

Input: (R, π), where $R \subseteq [p]$ is any non-empty set of size s, and π is the permutation satisfying condition (C2).

Output: the FDP upper bound for set R.

1: **for** $t = s, \ldots, 1$ **do**

Let $B = \{\pi^{-1}(\pi^R_{(1)}), \ldots, \pi^{-1}(\pi^R_{(t)})\} \subseteq R$ (i.e., the index set of the t smallest $\pi(i)$ with $i \in R$).

2: **for** $r = 0, \ldots, p - t$ **do**

if $r = 0$, **then** let $U_{t+r} = B$, **else** Let $U_{t+r} = B \cup B^c_r$, where $B^c_r = \{\pi^{-1}(\pi^{B^c}_{(1)}), \ldots, \pi^{-1}(\pi^{B^c}_{(r)})\}$ (i.e., the index set of the r smallest $\pi(i)$ with $i \in B^c$).

For $j \in [m]$, let $L^{U_{t+r}}_j$ be the test statistic related to $H_{U_{t+r}}$ and $z^{U_{t+r}}_j$ be the corresponding critical value.

if $L^{U_{t+r}}_j < z^{U_{t+r}}_j$ for all $j \in [m]$, **then return** t/s.

3: **Return** 0.

5 Simulations and a real data application

In this section, we examine the finite sample performance of our proposed methods in a range of settings and compare them to the method of Katsevich and Ramdas (2020). Since comparing FDP bounds for all sets $R \subseteq [p]$ is computationally infeasible, we follow Katsevich and Ramdas (2020) and focus on the nested sets $R_i = \{j \leq i : W_j > 0\}$, $i = 1, \ldots, p$. All simulations were carried out in R, and the code is available at xxx.

5.1 Comparison of Algorithm 1 and the method of Katsevich and Ramdas (2020)

We compare the FDP bounds of our proposed Algorithm 1 to Katsevich and Ramdas (2020). Specifically, we consider the following methods:

- KR: The method from Katsevich and Ramdas (2020). We use the R code from `https://github.com/ekatsevi/simultaneous-fdp` for its implementation.

- KF-JKI: Our proposed Algorithm 1. We implement four types of tuning parameters (v, k), where v is as described in (18) and k is obtained using Algorithm 2, and we denote them as KF-JKI-v1, KF-JKI-v2, KF-JKI-v3, and KF-JKI-v4, respectively.

All methods use the same knockoff statistic vector W, which is obtained according to the following settings:

(a) Linear regression model setting: We first generate $X \in \mathbb{R}^{n\times p}$ by drawing n i.i.d. samples from $N_p(0, \Sigma)$ with $(\Sigma)_{i,j} = 0.6^{|i-j|}$. Then we generate $\beta \in \mathbb{R}^{p\times 1}$ by randomly sampling $s \times p$ non-zero entries from $\{1, \ldots, p\}$, where the sparsity level $s \in \{0.1, 0.2, 0.3\}$, and we set the signal amplitude of the non-zero β_j to be $a/\sqrt{n}$,

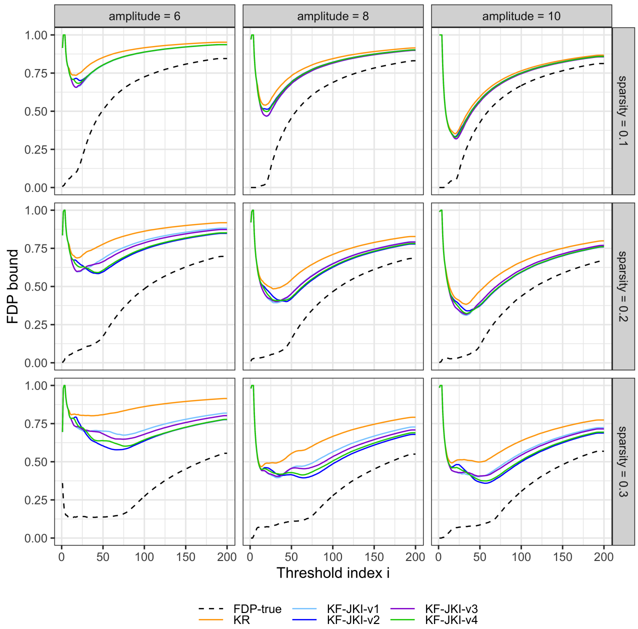

Figure 3: The FDP bounds of KR and KF-JKI in the linear regression model setting with $p = 200$ and $n = 500$. The dashed black line indicate the true FDP. All FDP bounds are the average values over 200 replications.

where n is the sample size and the amplitude $a \in \{6, 8, 10\}$. Finally, we sample $Y \sim N_n(X\beta, I)$. The design matrix X is fixed over different replications in the simulations. Based on the generated data (X, Y), we obtain W by using the fixed-X "sdp" knockoffs and the signed maximum lambda knockoff statistic. Similar settings and the same knockoff statistics are considered in Barber and Candès (2015) and Janson and Su (2016).

(b) Logistic regression model setting: For $i = 1, \dots, n$, we first sample $x_i \sim N_p(0, \Sigma)$ with $(\Sigma)_{i,j} = 0.6^{|i-j|}$, and then sample $y_i \sim \text{Bernoulli}\left(\frac{1}{1+e^{-x_i^T\beta}}\right)$, where β is generated as in (a), except now we consider amplitude $a \in \{8, 10, 12\}$. Based on the generated data (x_i, y_i), $i = 1, \dots, n$, we obtain W by using the "asdp" model-X knockoffs and the Lasso coefficient-difference statistic with 10-fold cross-validation. Similar settings and the same knockoff statistics are considered in Candès et al. (2018).

We take $p = 200$ and $n = 500$ and use the tuning parameter vector v with $v_m < 65$. The reported FDP bounds are average values over 200 replications. We also tried the case of $p = 1000$ and $n = 2500$, and the simulation results convey similar information, see Appendix F.3.

For the simulation results, first, as expected, the simultaneous FDP guarantee (2) holds for all methods, see Figure 9 in Appendix F.3. Now we compare the FDP bounds. Figures 3 and 4 show the FDP bounds of KR and KF-JKI with four types of (v, k) in the linear and logistic model settings, respectively. One can see that KF-JKI-v1 gives better

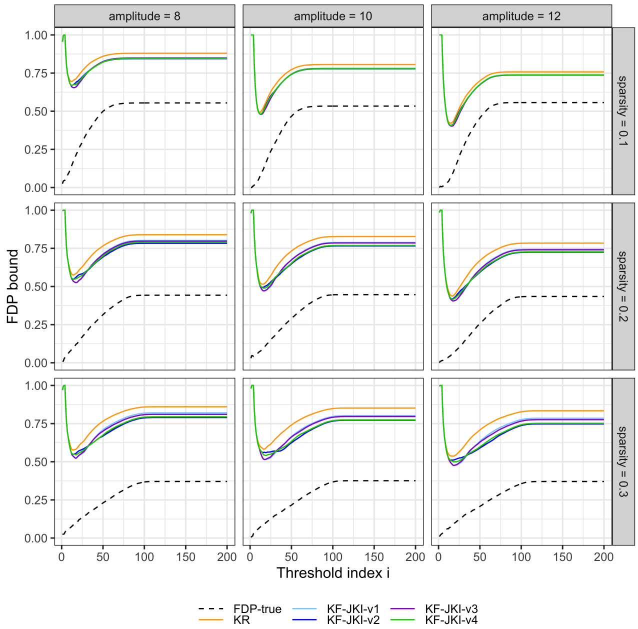

Figure 4: The FDP bounds of KR and KF-JKI in the logistic regression model setting with $p = 200$ and $n = 500$. The dashed black line indicate the true FDP. All FDP bounds are the average values over 200 replications.

FDP bounds for all R_i than KR over all settings, which verifies our claim about the uniform improvement of KF-JKI-v1 over KR in Section 3.3. In particular, the improvement is larger in denser settings with a smaller signal. All methods give similar FDP bounds in the settings with sparsity level 0.1 and large amplitude. The reason is that in such settings, almost all true W_i's are positive and have larger absolute values than any null W_i, so the FDP bounds based on the reference set $S(v_1 = 1)$ and interpolation are the best possible one can obtain. We also observe that KF-JKI with different types of (v, k) can give better FDP bounds for different R_i under different settings, and there is no uniformly best one.

5.2 Comparison of Algorithm 1 and the closed testing based approach

We now compare the FDP bounds of Algorithm 1 (i.e. KF-JKI described in Section 5.1) to its uniformly improved version based on closed testing. Specifically, we consider the following method:

- KF-CT: Our proposed method to obtain simultaneous FDP bounds based on closed testing using local test statistic (21) with $b_i^I = k_i^{|I|} + v_i - 1$ and the critical value $z_i^I = k_i^{|I|}$. We use the same four types of v as for KF-JKI and obtain $k^{|I|}$ by Algorithm 2 with $|I|$ as the last input, and we denote them as KF-CT-v1, KF-CT-v2, KF-CT-v3, and KF-CT-v4, respectively. We use the shortcut Algorithm 3 for the implementation of closed testing.

We implemented KF-JKI and KF-CT in the same linear and logistic model settings as in Section 5.1 with $p = 200$ and $n = 500$. As expected, the FDP bounds of KF-CT are indeed smaller than or equal to that of KF-JKI with the same tuning parameter vector v. In most cases, however, the corresponding FDP bounds are identical, and the improvement is typically very small in the cases where they are not identical. Hence we do not present the related plots in the paper.

Such phenomenon is due to the fact that compared to the closed testing equivalent form of KF-JKI (see Theorem 4.1), even though KF-CT uses a uniformly more powerful local test and can make more local rejections, such local rejections do not necessarily lead to more rejections by closed testing, especially in the sparse settings we consider here. Therefore the final FDP bounds can be the same.

To show that the uniform improvement by closed testing over Algorithm 1 can be non-trivial in certain cases, we consider a simulation setting that directly generates knockoff statistic vector W. Specifically, we consider a dense setting with $p = 50$ and null variable set $\mathcal{N} = \{10, \dots, 20\}$. For the knockoff statistics, we set $|W_i| = 50 - i + 1$, $\text{sign}(W_i) = 1$ for $i \notin \mathcal{N}$ and $\text{sign}(W_i) \overset{i.i.d.}{\sim} \{-1, 1\}$ with same probability $1/2$ for $i \in \mathcal{N}$. It is clear that the generated W satisfies the coin-flip property, so it is a valid knockoff statistic vector.

Figure 5 shows the simulation results in this setting. We can see that the improvements of KF-CT over KF-JKI can be large for some R_i. In addition, the improvements are different for different tuning parameters v.

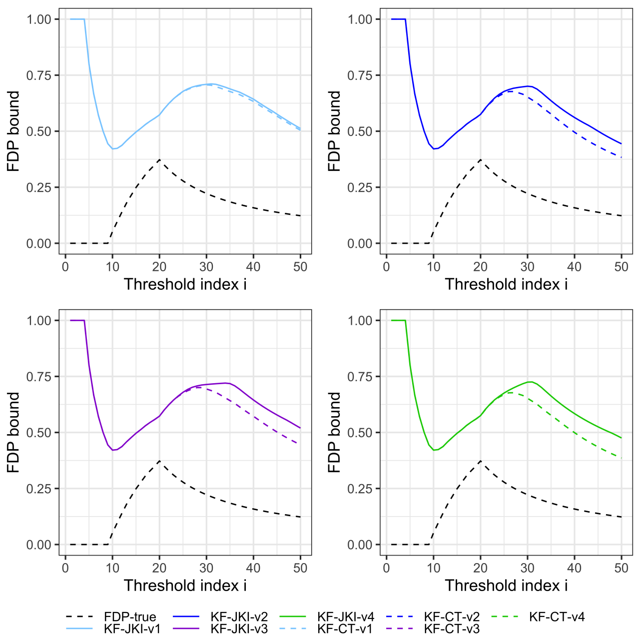

Figure 5: The FDP bounds of KF-JKI and KF-CT in the simulation setting where the knockoff statistic vector $W \in \mathbb{R}^{50}$ is directly generated. The dashed black line indicate the true FDP. All FDP bounds are the average values over 200 replications.

5.3 Real data application

In this section, we apply KR and KF-JKI (see Section 5.1) to the genome-wide association studies (GWAS) data sets from the UK Biobank resource (Bycroft et al., 2018). A GWAS data set typically consists of genotypes (covariants) and traits (response variables) of different individuals, and the goal is to identify the genotypes that are conditionally dependent with the trait (i.e., variable selection). In particular, the traits considered in our analysis are height, body mass index, platelet count, systolic blood pressure, cardiovascular disease, hypothyroidism, respiratory disease, and diabetes.

Following Katsevich and Ramdas (2020), we directly use the knockoff statistics $W = (W_1, \ldots, W_p)$ constructed by Sesia et al. (2020) based on the UK Biobank data (see `https://msesia.github.io/knockoffzoom/ukbiobank.html`). We also present the estimated FDP for set $R_i = \{j \leq i : W_j > 0\}$: $\widehat{\text{FDP}}(R_i) = \frac{1+\{j \leq i, W_j < 0\}}{\max\{|R_i|, 1\}}$. Note that the FDR controlled set returned by Barber and Candès (2015) at level $q \in (0, 1)$ is then the largest set R_i such that $\widehat{\text{FDP}}(R_i) \leq q$. We use the tuning parameter vector v with $v_m < 1200$ for KF-JKI.

Figure 6 shows the results based on the UK Biobank data with platelet count as the trait. In particular, we follow Katsevich and Ramdas (2020) to present the FDP bounds based on the number of rejections. One can see that KF-JKI with all four types of v gives smaller FDP bounds than KR. The simultaneous FDP bounds based on different v are quite different for sets with rejections less than about 3000 but are quite similar for sets with more than about 3000 rejections. The latter is because all four types of v use a similar reference set to obtain the FDP bounds, and the corresponding k values are also similar, as shown in Figure 2.

Then, we apply the methods to the data sets with other traits. We calculate the number of discoveries (i.e., rejections) by controlling the FDR at level 0.1 and the simultaneous FDP at levels 0.05, 0.1, and 0.2. Figure 7 shows the corresponding results. One can see that KF-JKI with different types of v give similar numbers of discoveries in most cases, and they are generally larger than KR. In some cases, controlling the simultaneous FDP of KF-JKI at level 0.2 gives more discoveries than controlling the FDR at level 0.1. For example, for the data set with height as the trait, KF-JKI-v1 returns 4298 discoveries, and the FDR control method returns 3284 discoveries. From the interpretation perspective, KF-JKI-v1 guarantees that there are more than 3438 true discoveries in 4298 discoveries with a probability larger than 0.95, whereas the FDR control method guarantees that there are more than 2955 true discoveries in 3284 discoveries in the expectation sense. By looking at the simultaneous FDP results for level 0.05 of the same data set, one then knows that there are more than 1542 true discoveries in 1624 discoveries with a probability larger than 0.95. One can also look at other levels or sets, and the FDP bounds are still valid. In practice, such flexibility can be very desired, and the obtained information can be very useful to guide the following experiments. Note that the FDR control method can not provide such flexibility because then the FDR control guarantee does not hold anymore.

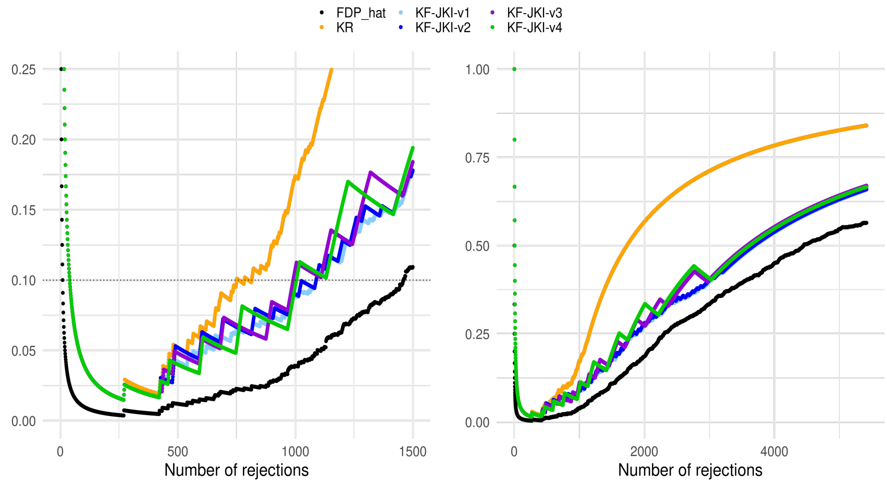

Figure 6: The FDP bounds of KR and KF-JKI based on the UK Biobank data set with the platelet count as the trait. The left plot is the zoom-in version of the right plot. The black dotted line is the estimated FDP.

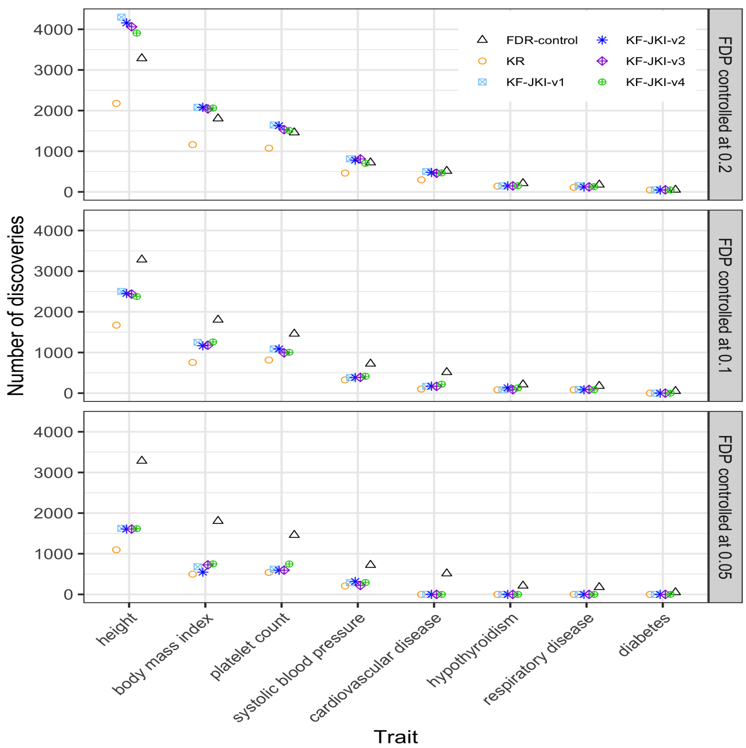

Figure 7: The number of discoveries obtained by controlling the FDR at level 0.1 and controlling the simultaneous FDP bounds at different levels for the UK Biobank data with different traits. The simultaneous FDP bounds are obtained by using KR and KF-JKI.

6 Discussion

We studied the problem of obtaining simultaneous FDP bounds for knockoff-based approaches. We first proposed a method based on joint k-FWER control and interpolation (Algorithm 1), and we showed that the only existing approach for this problem (Katsevich and Ramdas, 2020) is equivalent to Algorithm 1 with certain tuning parameters. In other words, it is a special case of Algorithm 1. We suggested other choices for the tuning parameters, and one of them guarantees to be uniformly better than the approach from (Katsevich and Ramdas, 2020). Next, we proposed a method based on closed testing and showed that the previously proposed Algorithm 1 is a special case (or exact shortcut) of this method. Using closed testing, we showed that Algorithm 1 can be further uniformly improved, and other generalizations can be derived. We also developed a new shortcut for the implementation of this closed testing based method.

Closed testing is a very general and powerful methodology in which one has the flexibility to choose different local tests. However, as we showed in simulations, different local tests can be better in different settings. One natural problem is to obtain the optimal data-dependent local tests while guaranteeing the validity of the simultaneous FDP bounds. Sample-splitting is a possible solution, but it can be sub-optimal, especially in cases where the sample size is not large. We leave this problem for future research.

The key assumption underlying our methods is that the knockoff statistics satisfy the coin-slip property. In some cases, however, such an assumption may not hold, for example, in Gaussian graphical model settings (Li and Maathuis, 2021) or in the model-X knockoff settings where the distribution of the covariates can only be estimated (Barber et al., 2020). It would be interesting to study such cases and develop methods to obtain valid simultaneous FDP bounds.

References

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

Barber, R. F. and Candès, E. J. (2019). A knockoff filter for high-dimensional selective inference. *The Annals of Statistics*, 47(5):2504–2537.

Barber, R. F., Candès, E. J., and Samworth, R. J. (2020). Robust inference with knockoffs. *The Annals of Statistics*, 48(3):1409–1431.

Bates, S., Sesia, M., Sabatti, C., and Candès, E. (2020). Causal inference in genetic trio studies. *Proceedings of the National Academy of Sciences*, 117(39):24117–24126.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188.

Blanchard, G., Neuvial, P., and Roquain, E. (2020). Post hoc confidence bounds on false positives using reference families. *The Annals of Statistics*, 48(3):1281–1303.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209.

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

Chi, C.-M., Fan, Y., Lv, J., et al. (2021). High-dimensional knockoffs inference for time series data. *ArXiv:2112.09851*.

Dai, R. and Barber, R. (2016). The knockoff filter for FDR control in group-sparse and multitask regression. In *International conference on machine learning*, volume 48, pages 1851–1859. PMLR.

Dobriban, E. (2020). Fast closed testing for exchangeable local tests. *Biometrika*, 107(3):761–768.

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *The Annals of Statistics*, 32(3):1035–1061.

Genovese, C. R. and Wasserman, L. (2006). Exceedance control of the false discovery proportion. *Journal of the American Statistical Association*, 101(476):1408–1417.

Goeman, J. J., Hemerik, J., and Solari, A. (2021). Only closed testing procedures are admissible for controlling false discovery proportions. *The Annals of Statistics*, 49(2):1218–1238.

Goeman, J. J., Meijer, R. J., Krebs, T. J., and Solari, A. (2019). Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *Biometrika*, 106(4):841–856.

Goeman, J. J. and Solari, A. (2011). Multiple testing for exploratory research. *Statistical Science*, 26(4):584–597.

He, Z., Liu, L., Belloy, M. E., Le Guen, Y., Sossin, A., Liu, X., Qi, X., Ma, S., Wyss-Coray, T., Tang, H., Sabatti, C., Candès, E., Greicius, M. D., and Ionita-Laza, I. (2021). Summary statistics knockoff inference empowers identification of putative causal variants in genome-wide association studies. *bioRxiv 2021.12.06.471440*.

Hemerik, J., Solari, A., and Goeman, J. J. (2019). Permutation-based simultaneous confidence bounds for the false discovery proportion. *Biometrika*, 106(3):635–649.

Janson, L. and Su, W. (2016). Familywise error rate control via knockoffs. *Electronic Journal of Statistics*, 10(1):960–975.

Katsevich, E. and Ramdas, A. (2020). Simultaneous high-probability bounds on the false discovery proportion in structured, regression and online settings. *The Annals of Statistics*, 48(6):3465–3487.

Katsevich, E. and Sabatti, C. (2019). Multilayer knockoff filter: Controlled variable selection at multiple resolutions. *The Annals of Applied Statistics*, 13(1):1–33.

Li, J. and Maathuis, M. H. (2021). GGM knockoff filter: False discovery rate control for Gaussian graphical models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(3):534–558.

Li, S., Sesia, M., Romano, Y., Candès, E., and Sabatti, C. (2021). Searching for consistent associations with a multi-environment knockoff filter. *ArXiv:2106.04118*.

Marcus, R., Eric, P., and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660.

Meinshausen, N. (2006). False discovery control for multiple tests of association under general dependence. *Scandinavian Journal of Statistics*, 33(2):227–237.

Sesia, M., Bates, S., Candès, E., Marchini, J., and Sabatti, C. (2021). False discovery rate control in genome-wide association studies with population structure. *Proceedings of the National Academy of Sciences*, 118(40).

Sesia, M., Katsevich, E., Bates, S., Candès, E., and Sabatti, C. (2020). Multi-resolution localization of causal variants across the genome. *Nature communications*, 11(1):1–10.

Tian, J., Chen, X., Katsevich, E., Goeman, J., and Ramdas, A. (2021). Large-scale simultaneous inference under dependence. *ArXiv:2102.11253*.

Van der Laan, M. J., Dudoit, S., and Pollard, K. S. (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Statistical Applications in Genetics and Molecular Biology*, 3(1).

Vesely, A., Finos, L., and Goeman, J. J. (2021). Permutation-based true discovery guarantee by sum tests. *ArXiv:2102.11759*.

APPENDIX

The appendix consists of the following six sections.

A The equivalence of the simultaneous FDP bounds (10) based on two specific k-FWER controlled sets

B Validity of Algorithm 3

C A fact about early-stopped negative binomial and truncated binomial statistics

D Supplementary materials relate to Goeman et al. (2021)

E Main proofs

F Supplementary materials relate to simulations

A The equivalence of the simultaneous FDP bounds (10) based on two specific k-FWER controlled sets

As pointed by Janson and Su (2016), in the case of $|S(v)| < k-1$, one can always continue rejecting positive W_i and try to make $k-1$ rejections without violating the k-FWER control guarantee (6). Formally, this procedure results in the set

$$S^*(v,k) = S(v) \cup S^{(k-1)}, \tag{22}$$

where $S(v)$ is defined by (5), and $S^{(k-1)} = \{i \in [p] : W_i \geq T^{(k-1)}\}$, and $T^{(k-1)} = \max\{|W_i| : \#\{j \in [p] : W_j \geq |W_i|\} = k-1\}$ with $T^{(k-1)} = \min_{i\in[p]}\{|W_i|\}$ if $\#\{i \in [p] : W_i > 0\} < k-1$.

When $|S(v)| \geq k-1$, we have that $S^{(k-1)} \subseteq S(v) = S^*(v,k)$. When $|S(v)| < k-1$, we have that $S(v) \subseteq S^{(k-1)} = S^*(v,k)$. Note that $|S^{(k-1)}| \leq k-1$.

It is clear that $S^*(v,k)$ satisfies the k-FWER control guarantee, and it is uniformly better than $S(v)$ in the sense that one obtain more rejections using $S^*(v,k)$ as $S(v) \subseteq S^*(v,k)$. Therefore, one may think that using $S^*(v,k)$ can obtain better simultaneous FDP bounds (10) than using $S(v)$. In the following, we show that the simultaneous FDP bounds based on $S(v)$ and $S^*(v,k)$ are actually equivalent.

Specifically, it is enough to show that

$$\min\{|R|, k-1+|R\backslash S^*(v,k)|\} = \min\{|R|, k-1+|R\backslash S(v)|\}. \tag{23}$$

In the case of $|S(v)| \geq k-1$, we have that $S^*(v,k) = S(v)$, so equality (23) holds.

In the case of $|S(v)| < k-1$, we have that $|S^*(v,k)| = |S^{(k-1)}| \leq k-1$. For the left hand side of (23), we have that

$$k-1+|R\backslash S^*(v,k)| \geq k-1+|R| - |S^*(v,k)| \geq k-1+|R| - (k-1) = |R|,$$

so

$$\min\{|R|, k-1+|R\backslash S^*(v,k)|\} = |R|.$$

For the right hand side of (23), we have that

$$k-1+|R\backslash S(v)| \geq k-1+|R|-|S(v)| > k-1+|R|-(k-1) = |R|,$$

so

$$\min\{|R|, k-1+|R\backslash S(v)|\} = |R|.$$

Hence equality (23) holds.

Therefore, the simultaneous FDP bounds (10) based on $S^*(v,k)$ and $S(v)$ are the same. So we will use $S(v)$ to get simultaneous FDP bounds for simplicity.

B Validity of Algorithm 3

Let $R \subseteq [p]$ be of size s_R and let $I \subseteq R$ be of size s_I. To obtain the false discovery bound of R using closed testing, we need to find its largest subset whose related null hypothesis is not rejected by closed testing. To check whether the null hypothesis H_I is rejected by closed testing, we need to check whether all superset of I are locally rejected. In the following, we show that if conditions (C1) and (C2) in Section 4.2 are satisfied, then the following two claims are true. Therefore, we only need to check one set of the same size when: (i) searching for the largest subset whose related null hypothesis is not rejected by closed testing; (ii) checking whether the null hypothesis H_I is rejected by closed testing. These then guarantees that Algorithm 3 is valid.

(a) Let $\pi^{I^c}_{(1)} < \cdots < \pi^{I^c}_{(p-s_I)}$ be the ordered values of $\{\pi(i), i \in I^c\}$, where $I^c = [p]\backslash I$. For any $t \in \{0, \ldots, p-s_I\}$, to know whether there exists a superset of I with size $s_I + t$ that is not locally rejected, we only need to check one superset $U_{s+t} = I \cup I^c_t$ of size $s_I + t$, where $I^c_t = \{\pi^{-1}(\pi^{I^c}_{(1)}), \ldots, \pi^{-1}(\pi^{I^c}_{(t)})\}$ (i.e., the index set of the t smallest $\pi(i)$ with $i \in I^c$) for $t > 0$ and $I^c_t = \emptyset$ for $t = 0$.

(b) Let $\pi^R_{(1)} < \cdots < \pi^R_{(s_R)}$ be the ordered values of $\{\pi(i), i \in R\}$. For any $t \in \{1, \ldots, s_R\}$, to know whether there exists a subset of R with size t that is not rejected by closed testing, we only need to check one subset $B = \{\pi^{-1}(\pi^R_{(1)}), \ldots, \pi^{-1}(\pi^R_{(t)})\} \subseteq R$ (i.e., the index set of the t smallest $\pi(i)$ with $i \in R$).

For (a): the claim holds when $t = 0$, now assume $t \in \{1, \ldots, p - s_I\}$.

- In the case of $L_i^{U_{s_I+t}} < z_i^{s_I+t}$ for all $i \in [m]$, we can conclude that there exists a superset of I with size $s_I + t$ that is not locally rejected.

- In the case of $L_{i^*}^{U_{s_I+t}} \geq z_{i^*}^{s_I+t}$ for some $i^* \in [m]$, we show that there is no superset of I with size $s_I + t$ that is not locally rejected. Let $V_{s_I+t} = I \cup J$ be any other superset of I with size $s_I + t$, where $J \subseteq I^c$ is of size t. Denote $U_{s_I+t} = \{u_1, \ldots, u_{s_I+t}\}$ and $V_{s_I+t} = \{v_1, \ldots, v_{s_I+t}\}$. By the definition of U_{s_I+t} and V_{s_I+t}, it is clear that

$\pi(u_i) \le \pi(v_i)$ for all $i = 1\dots, s_I + t$. Let $L_i^{U_{s_I+t}}$ and $L_i^{V_{s_I+t}}$ be the test statistics of $H_{U_{s_I+t}}$ and $H_{V_{s_I+t}}$, respectively. Then by (C2), we have $L_{i^*}^{V_{s_I+t}} \ge L_{i^*}^{U_{s_I+t}} \ge z_{i^*}^{s_I+t}$, which implies that $H_{V_{s_I+t}}$ is locally rejected by (C1).

For (b): the claim holds when $t = S_R$, now assume $t \in \{1, \dots, s_R - 1\}$.

- If B is not rejected by closed testing, then we can conclude that there exists a subset of R with size t that is not rejected by closed testing.

- If B is rejected by closed testing, we show that there is no subset of R with size t that is not rejected by closed testing.

  Let $J \subseteq R$ be any subset of R with size t, we show that it is rejected by closed testing, that is, all its supersets are locally rejected. By claim (a), it is sufficient to show that for any $r \in \{0, \dots, p-t\}$, $V_{t+r} = J \cup J_r^c$ is locally rejected, where $J^c = [p]\backslash J$, $J_r^c = \{\pi^{-1}(\pi_{(1)}^{J^c}), \dots, \pi^{-1}(\pi_{(r)}^{J^c})\}$ for $r > 0$, and $J_r^c = \emptyset$ for $r = 0$.

  B is rejected by closed testing means that all supersets of B are locally rejected. So for any $r \in \{0, \dots, p-t\}$ and $U_{t+r} = B \cup B_r^c$, there exists some $j_r \in [m]$ such that $L_{j_r}^{U_{t+r}} \ge z_{j_r}^{t+r}$. Here $B^c = [p]\backslash B$, $B_r^c = \{\pi^{-1}(\pi_{(1)}^{B^c}), \dots, \pi^{-1}(\pi_{(r)}^{B^c})\}$ for $r > 0$, and $B_r^c = \emptyset$ for $r = 0$.

  Denote $U_{t+r} = \{u_1, \dots, u_{t+r}\}$ and $V_{t+r} = \{v_1, \dots, v_{t+r}\}$ such that $\pi(u_1) < \dots < \pi(u_{t+r})$ and $\pi(v_1) < \dots < \pi(v_{t+r})$. By the definition of U_{t+r} and V_{t+r}, it is clear that $\pi(u_i) \le \pi(v_i)$ for all $i \in [t+r]$. Hence by (C2), we have $L_{j_r}^{V_{t+r}} \ge L_{j_r}^{U_{t+r}} \ge z_{j_r}^{t+r}$, which then implies that V_{t+r} is locally rejected by (C1).

C A fact about early-stopped negative binomial and truncated binomial statistics

We present the following fact which will be used in the proofs to connect early-stopped negative binomial and truncated binomial tests. Specifically, for positive integers v, k and a sequence $X_1, \dots, X_p$, where $X_i \in \{-1, +1\}$, let

$$t_v = \min\{i \in [p] : \#\{j \le i : X_j = -1\} = v\},$$

and $t_v = p$ if $\#\{i \le [p] : X_i = -1\} < v$. Let

$$NB^p(v) = \#\{i \le t_v : X_i = 1\}$$

be the early-stopped negative binomial variable, and let

$$B^p(k+v-1) = \#\{i \le \min(p, k+v-1) : X_i = 1\}$$

be the truncated binomial variable. Then

$$NB^p(v) \ge k \Longleftrightarrow B^p(k+v-1) \ge k \tag{24}$$

and

$$NB^p(v) \le k-1 \Longleftrightarrow B^p(k+v-1) \le k-1. \tag{25}$$

D Supplementary materials relate to Goeman et al. (2021)

In this section, we first give a brief review of Goeman et al. (2021), then we show two lemmas which will be used to obtain the interpolation version of the FDP bound in Katsevich and Ramdas (2020). At last, we connect this FDP bound to closed testing. Some results in this section will also be used in the main proofs in Appendix E.

D.1 A brief recap of Goeman et al. (2021)

We follow Goeman et al. (2021) and present the results in terms of the true discovery guarantee.

Assume that we have data X from some distribution P_θ with $\theta \in \Theta$, and we formulate hypothesis for θ to get $H \subseteq \Theta$. We are interested in testing hypotheses $(H_i)_{i \in I}$, where $I \subseteq C \subseteq \mathbb{N}$ is finite. Here C plays the role of maximal set and will be important later. For any rejection set $R \subseteq I$, let $R_1 \subseteq R$ be the index set of false hypotheses (i.e., true discoveries). We call a random function (depends on data X) $\boldsymbol{d}^I : 2^I \to \mathbb{R}$ has *true discovery guarantee* on I if for all $\theta \in \Theta$, we have

$$\mathbb{P}_\theta(\boldsymbol{d}^I(R) \leq |R_1| \text{ for all } R \subseteq I) \geq 1 - \alpha. \tag{26}$$

True discovery guarantee and simultaneous FDP bound are equivalent, in the sense that if we have simultaneous FDP bound $\overline{\mathrm{FDP}}(R)$ satisfying (2), then

$$\boldsymbol{d}^I(R) = (1 - \overline{\mathrm{FDP}}(R))|R|$$

satisfies the true discovery guarantee (26). And if we have $\boldsymbol{d}^I(R)$ satisfies the true discovery guarantee (26), then

$$\overline{\mathrm{FDP}}(R) = \frac{|R| - \boldsymbol{d}^I(R)}{\max\{1, |R|\}}$$

satisfies (2). Hence we can work with $\boldsymbol{d}^I(R)$ to derive desired results, then equivalently transfer the results back to $\overline{\mathrm{FDP}}(R)$.

The *interpolation* $\bar{\boldsymbol{d}}^I$ of $\boldsymbol{d}^I$ is defined as

$$\bar{\boldsymbol{d}}^I(R) = \max_{U \in 2^I} \left\{\boldsymbol{d}^I(U) - |U \backslash R| + \boldsymbol{d}^I(R \backslash U)\right\}. \tag{27}$$

Lemma 2 in Goeman et al. (2021) shows that if $\boldsymbol{d}^I$ has true discovery guarantee (26), then also $\bar{\boldsymbol{d}}^I$. It is clear that $\bar{\boldsymbol{d}}^I \geq \boldsymbol{d}^I$, so the interpolation $\bar{\boldsymbol{d}}^I$ of $\boldsymbol{d}^I$ will never be worse than $\boldsymbol{d}^I$.

We call $\boldsymbol{d}^I$ *coherent* if $\boldsymbol{d}^I = \bar{\boldsymbol{d}}^I$. Interpolation may not be a one-off process, the following lemma from Goeman et al. (2021) provides a convenient way to check coherence.

Lemma D.1 (Lemma 3 in Goeman et al. (2021))**.** *$\bar{\boldsymbol{d}}^I$ is coherent if and only if for every disjoint $U, V \subseteq I$, we have*

$$\boldsymbol{d}^I(U) + \boldsymbol{d}^I(V) \leq \boldsymbol{d}^I(U \cup V) \leq \boldsymbol{d}^I(U) + |V|.$$

We can embed procedure $\boldsymbol{d}^I$ into a stack of procedures $\boldsymbol{d} = (\boldsymbol{d}^I)_{I\subseteq C,|I|\leq\infty}$, where we may have some maximal family $C \subseteq \mathbb{N}$. For example, when $\boldsymbol{d}^I$ is our Algorithm 1 applying to $(W_i)_{i\in I}$, then we can get a stack of procedures $(\boldsymbol{d}^I)_{I\subseteq[p]}$ (in this case $C = [p]$).

We call a stack of procedures $\boldsymbol{d}$ a *monotne procedure* if the following three criteria are fulfilled.

1. *true discovery guarantee*: $\boldsymbol{d}^I$ has true discovery guarantee for all finite $I \subseteq C$.
2. *coherent*: $\boldsymbol{d}^I$ is coherent for all finite $I \subseteq C$.
3. *monotonicity*: $\boldsymbol{d}^I(R) \geq \boldsymbol{d}^J(R)$ for every finite $R \subseteq I \subseteq J \subseteq C$

For closed testing procedure, let $\phi_R \in \{0, 1\}$ be the local test for testing H_R, with $\phi_R = 1$ indicating rejection. Choosing a local test for every finite $R \subseteq C$ will yield a suite of local tests $\phi = (\phi_R)_{R\subseteq C,|R|<\infty}$. The true discovery guarantee procedure based on closed testing (cf. Section 2.3) is

$$\boldsymbol{d}^I_\phi(R) = \min_{U\in 2^R}\{|R\backslash U| : \phi^I_U = 0\},$$

where

$$\phi^I_R = \min\{\phi_U : R \subseteq U \subseteq I\}$$

indicates whether H_R is rejected by closed testing.

Finally, we present the main result. This is a very powerful result, which shows that any monotone procedure $\boldsymbol{d}$ is either equivalent to or can be uniformly improved by a closed testing procedure with the explicit local test defined in (28).

Theorem D.1 (Theorem 1 in Goeman et al. (2021))**.** *Let $\boldsymbol{d}$ be a monotone procedure. Then for every finite $R \subseteq C$,*

$$\phi_R = \mathbb{1}\{\boldsymbol{d}^R(R) > 0\} \tag{28}$$

is a valid local test of H_R. That is,

$$\sup_{\theta\in H_R} \mathbb{P}_\theta(\phi_R = 1) \leq \alpha.$$

In addition, for the suite $\phi = (\phi_R)_{R\subseteq C,|R|<\infty}$ and all $R \subseteq I \subseteq C$ with $|I| < \infty$, we have

$$\boldsymbol{d}^I_\phi(R) \geq \boldsymbol{d}^I(R).$$

D.2 Two lemmas about interpolation and coherence

The interpolation formula in Goeman et al. (2021) and Blanchard et al. (2020) are defined in different forms. We will use the interpolation formula (27) from Goeman et al. (2021), so we show the following lemma which gives the simplified formula of the interpolation for certain true discovery guarantee procedure. The results in Lemma D.2 is equivalent to the results in Blanchard et al. (2020) (see equation (9) and Proposition 2.5 in their paper). We present Lemma D.2 here for clarity because the results of Blanchard et al. (2020) is obtained based on a different form of interpolation rather than formula (27).

Lemma D.2. *For $m \geq 1$ and a set I, let $R \subseteq I$, $K_1 \subseteq K_2 \subseteq \cdots \subseteq K_m \subseteq I$ be nested sets, $k_i \in \mathbb{Z}_{>0}$ and $|K_i| \geq k_i$, $i \in [m]$. For a true discovery guarantee procedure of the form*

$$\boldsymbol{d}^I(R) = \begin{cases} |K_i| - k_i, & \text{if } R = K_i \text{ for some } 1 \leq i \leq m, \\ 0, & \text{otherwise,} \end{cases} \tag{29}$$

its interpolation is

$$\bar{\boldsymbol{d}}^I(R) = \max_{i \in [m]} \{|R \cap K_i| - k_i\} \vee 0. \tag{30}$$

Proof. By definition of $\bar{\boldsymbol{d}}^I(R)$ (see (27)), we need to show that

$$\max_{U \in 2^I} \left\{ \boldsymbol{d}^I(U) - |U \backslash R| + \boldsymbol{d}^I(R \backslash U) \right\} = \max_{i \in [m]} \{|R \cap K_i| - k_i\} \vee 0.$$

For any $R \subseteq I$. We consider three cases of $U \in 2^I$.

(1) $U \in 2^I$ such that for all $i \in [m]$, $U \neq K_i$ and $R \backslash U \neq K_i$, then we have

$$\boldsymbol{d}^I(U) - |U \backslash R| + \boldsymbol{d}^I(R \backslash U) = -|U \backslash R|.$$

(2) $U \in 2^I$ such that for some $i \in [m]$, $R \backslash U = K_i$. Then $U \neq K_j$ for any $j \in [m]$ because $K_1 \subseteq \cdots \subseteq K_m$. And we have that $K_i = R \cap K_i$ as $K_i \subseteq R$. Hence

$$\begin{aligned} \boldsymbol{d}^I(U) - |U \backslash R| + \boldsymbol{d}^I(R \backslash U) &= -|U \backslash R| + \boldsymbol{d}^I(K_i) \\ &= -|U \backslash R| + |K_i| - k_i \\ &= -|U \backslash R| + |R \cap K_i| - k_i \end{aligned}$$

(3) $U \in 2^I$ such that for some $i \in [m]$, $U = K_i$. Then $R \backslash U \neq K_j$ for any $j \in [m]$ because $K_1 \subseteq \cdots \subseteq K_m$. Hence

$$\begin{aligned} \boldsymbol{d}^I(U) - |U \backslash R| + \boldsymbol{d}^I(R \backslash U) &= \boldsymbol{d}^I(K_i) - |K_i \backslash R| \\ &= |K_i| - k_i - |K_i \backslash R| \\ &= |R \cap K_i| - k_i \end{aligned}$$

In addition, since $\bar{\boldsymbol{d}}^I(R) \geq \boldsymbol{d}^I(R) \geq 0$, we have

$$\bar{\boldsymbol{d}}^I(R) = \max_{U \in 2^I} \left\{ \boldsymbol{d}^I(U) - |U \backslash R| + \boldsymbol{d}^I(R \backslash U) \right\} = \max_{i \in [m]} \{|R \cap K_i| - k_i\} \vee 0.$$

□

Next, we give the following lemma showing that the above interpolation $\bar{\boldsymbol{d}}^I$ of $\boldsymbol{d}$ (see (30)) is coherent.

Lemma D.3. *For $m \geq 1$ and a set I, let $R \subseteq I$, $K_1 \subseteq K_2 \subseteq \cdots \subseteq K_m \subseteq I$ be nested sets, and $k_i \in \mathbb{Z}_{>0}$, $i \in [m]$. Let*

$$\bar{\boldsymbol{d}}^I(R) = \max_{i \in [m]} \{|R \cap K_i| - k_i\} \vee 0.$$

Then $\bar{\boldsymbol{d}}^I(R)$ is coherent.

Proof. We show that $\bar{\boldsymbol{d}}^I$ is coherent by applying Lemma D.1. Specifically, for every disjoint $U, V \subseteq I$, we have

$$\begin{aligned}
\bar{\boldsymbol{d}}^I(U) &= \max_{i\in[m]}\{|U\cap K_i| - k_i\} \vee 0, \\
\bar{\boldsymbol{d}}^I(V) &= \max_{i\in[m]}\{|V\cap K_i| - k_i\} \vee 0, \\
\bar{\boldsymbol{d}}^I(U\cup V) &= \max_{i\in[m]}\{|U\cap K_i| + |V\cap K_i| - k_i\} \vee 0.
\end{aligned}$$

We first show that

$$\boldsymbol{d}^I(U) + \boldsymbol{d}^I(V) \leq \boldsymbol{d}^I(U\cup V).$$

In the case of $\bar{\boldsymbol{d}}^I(U) = \bar{\boldsymbol{d}}^I(V) = 0$ or one of them is 0, it is clear that $\boldsymbol{d}^I(U) + \boldsymbol{d}^I(V) \leq \boldsymbol{d}^I(U\cup V)$.

Now consider the case of $\bar{\boldsymbol{d}}^I(U) = |U\cap K_{i_1}| - k_{i_1}$ and $\bar{\boldsymbol{d}}^I(V) = |V\cap K_{i_2}| - k_{i_2}$, in which we have that $\bar{\boldsymbol{d}}^I(U\cup V) = \max_{i\in[m]}\{|U\cap K_i| + |V\cap K_i| - k_i\}$.

For the case of $i_1 = i_2$, $\boldsymbol{d}^I(U) + \boldsymbol{d}^I(V) \leq \boldsymbol{d}^I(U\cup V)$ clearly holds.

Now assume without loss of generality that $i_1 < i_2$, we have

$$|U\cap K_{i_2}| \geq |U\cap K_{i_1}| \geq |U\cap K_{i_1}| - k_{i_1}$$

as K_i's are nested. So

$$\boldsymbol{d}^I(U\cup V) \geq |U\cap K_{i_2}| + |V\cap K_{i_2}| - k_{i_2} \geq |U\cap K_{i_1}| - k_{i_1} + |V\cap K_{i_2}| - k_{i_2} = \bar{\boldsymbol{d}}^I(U) + \bar{\boldsymbol{d}}^I(V)$$

Now we show that

$$\boldsymbol{d}^I(U\cup V) \leq \boldsymbol{d}^I(U) + |V|.$$

Since

$$|U\cap K_i| + |V\cap K_i| - k_i \leq |U\cap K_i| - k_i + |V|,$$

we have

$$(|U\cap K_i| + |V\cap K_i| - k_i) \vee 0 \leq ((|U\cap K_i| - k_i) \vee 0) + |V|,$$

which implies the desired result. So $\bar{\boldsymbol{d}}^I$ is coherent by Lemma D.1. $\square$

D.3 Interpolation version of the bound in Katsevich and Ramdas (2020)

For knockoff statistic vector $W = (W_1, \dots, W_p)$ satisfying the coin-flip property and $|W_1| > |W_2| > \cdots > |W_p| > 0$, let $S_i = \{j \leq i : W_j > 0\}$ and $S_i^- = \{j \leq i : W_j < 0\}$. For nested sets S_i, $i \in [p]$, Katsevich and Ramdas (2020) proposed simultaneous FDP bound $\frac{\bar{V}(|S_i^-|)}{\max\{1,|S_i|\}}$, where

$$\bar{V}(|S_i^-|) = \lfloor c(\alpha)\cdot(1 + |S_i^-|)\rfloor \quad \text{with} \quad c(\alpha) = \frac{\log(\alpha^{-1})}{\log(2-\alpha)}. \tag{31}$$

The above bound can be improved and extended to all sets using interpolation (Goeman et al., 2021; Blanchard et al., 2020). Katsevich and Ramdas (2020) mentioned, but didn't present the formula for the interpolation version of the bound. Here we give the explicit formula.

Specifically, following the equation (4) of Goeman et al. (2021), we first transform the false discovery bound (31) to its equivalent true discovery guarantee procedure

$$\boldsymbol{d}^{[p]}(R) = \begin{cases} |S_i| - \bar{V}(|S_i^-|), & \text{if } R = S_i \text{ for some } 1 \le i \le p, \\ 0, & \text{otherwise,} \end{cases} \tag{32}$$

Then, by Lemma D.2, the interpolated of $\boldsymbol{d}^{[p]}$ is

$$\bar{\boldsymbol{d}}^{[p]}(R) = \max_{i \in [p]} \{|R \cap S_i| - \bar{V}(|S_i^-|)\} \vee 0, \tag{33}$$

which is equivalent to the following simultaneous FDP bound

$$\overline{\mathrm{FDP}}^{\mathrm{KR}}(R) = \frac{\min_{i \in [p]}\{|R \backslash S_i| + \bar{V}(|S_i^-|), |R|\}}{\max\{1, |R|\}}, \quad \forall R \subseteq [p].$$

D.4 Apply Goeman et al. (2021) to Katsevich and Ramdas (2020)

In this section, we present a proof using the results from Goeman et al. (2021) (see Appendix D.1 for a brief review) to connect $\overline{\mathrm{FDP}}^{\mathrm{KR}}(R)$, the interpolation version of the simultaneous FDP bound in Katsevich and Ramdas (2020), and the FDP bound $\overline{\mathrm{FDP}}^{ct}(R)$ based on closed testing. Similar result for our proposed simultaneous FDP bound (12) can be obtained similarly.

Proposition D.1. *Given $\alpha \in (0, 1)$ and knockoff statistic vector $W = (W_1, \dots, W_p)$ satisfying the coin-flip property and $|W_1| > |W_2| > \cdots > |W_p| > 0$. For any $R \subseteq [p]$, let $\overline{FDP}^{KR}(R)$ be defined by* (17) *and $\overline{FDP}^{ct}(R)$ (see* (8)*) be the FDP bound based on closed testing using local test statistic* (21) *with $m = |I|$, $b_i^I = k_i^{raw} + i - 1$ and critical values $z_i^I = k_i^{raw}$ (for null hypothesis H_I), where $k_i^{raw} = \min_{j \ge 1}\{c_j : j - c_j + 1 = i\}$ and c_j is defined by* (15)*. Then $\overline{FDP}^{ct}(R) \le \overline{FDP}^{KR}(R)$.*

Proposition D.1. For any $I \subseteq [p]$ and $i \le |I|$, let $S_i^I = \{j \in I : W_j = W_l^I \text{ such that } l \le i \ \&\ W_l^I > 0\}$, where W_l^I denotes the W_j whose absolute value $|W_j|$ is the l-th largest in $\{|W_j| : j \in I\}$.

Based on $\{W_i : i \in I\}$, the corresponding true discovery control procedure (cf. (33) in Appendix D.3) of Katsevich and Ramdas (2020) is

$$\bar{\boldsymbol{d}}^I(R) = \max_{i=1,\dots,|I|} \{|R \cap S_i^I| - \bar{V}(S_i^I)\} \vee 0, \quad R \subseteq I,$$

where $\bar{V}(S_i^I) = \lfloor c(\alpha) \cdot (1 + i - |S_i^I|) \rfloor$ and $c(\alpha) = \frac{\log(\alpha^{-1})}{\log(2-\alpha)}$.

Now we show that the stack of procedures $\bar{\boldsymbol{d}} = (\bar{\boldsymbol{d}}^I)_{I \subseteq [p]}$ is monotone. For every $I \subseteq [p]$, it is clear that $\bar{\boldsymbol{d}}^I$ has true discovery guarantee, and $\bar{\boldsymbol{d}}^I$ is coherent by Lemma D.3, so it is enough to show the monotonicity.

For any $R \subseteq I \subseteq J \subseteq [p]$ and any $j_1 \in [|J|]$. In the case of $S_{j_1}^J \cap I = \emptyset$, we have $|R \cap S_{j_1}^J| - \bar{V}(S_{j_1}^J) \leq 0$. In the case of $S_{j_1}^J \cap I \neq \emptyset$, let $W_{j_2}^I = W_{i^*}$, where $i^* = \max\{S_{j_1}^J \cap I\}$. So for index j_2, we have $S_{j_2}^I \cap I = S_{j_1}^J \cap I$, $S_{j_2}^I \subseteq S_{j_1}^J$ and $j_2 - |S_{j_2}^I| \leq j_1 - |S_{j_1}^J|$ because $I \subseteq J$. Hence $|R \cap S_{j_2}^I| = |R \cap S_{j_2}^I \cap I| = |R \cap S_{j_1}^J \cap I| = |R \cap S_{j_1}^J|$ and $\bar{V}(S_{j_2}^I) \leq \bar{V}(S_{j_1}^J)$, which implies that $|R \cap S_{j_2}^I| - \bar{V}(S_{j_2}^I) \geq |R \cap S_{j_1}^J| - \bar{V}(S_{j_1}^J)$. Therefore,

$$\bar{\boldsymbol{d}}^J(R) = \max_{j_1=1,\dots,|J|} \{|R \cap S_{j_1}^J| - \bar{V}(S_{j_1}^J)\} \vee 0 \leq \max_{j_2=1,\dots,|I|} \{|R \cap S_{j_2}^I| - \bar{V}(S_{j_2}^I)\} \vee 0 = \bar{\boldsymbol{d}}^I(R),$$

which shows that $\bar{\boldsymbol{d}} = (\bar{\boldsymbol{d}}^I)_{I \subseteq [p]}$ is monotone.

By Theorem D.1, for the suite $\phi = (\phi_U)_{U \subseteq [p]}$ with local test

$$\phi_U = \mathbb{1}\{\bar{\boldsymbol{d}}^U(U) > 0\},$$

we have

$$\boldsymbol{d}_\phi^I(R) \geq \bar{\boldsymbol{d}}^I(R), \quad \forall R \subseteq I \subseteq [p],$$

where $\boldsymbol{d}_\phi^I(R)$ is the true discovery guarantee procedure based on closed testing with suite ϕ. Taking $I = [p]$, the above is equivalent to

$$\overline{\mathrm{FDP}}^{ct}(R) \leq \overline{\mathrm{FDP}}^{\mathrm{KR}}(R),$$

where $\overline{\mathrm{FDP}}^{ct}(R)$ is the corresponding simultaneous FDP bound of the closed testing using local test ϕ_U for local hypothesis H_U.

At last, we show that the local test ϕ_U is equivalent to the one described in the claim.

The local test $\phi_U = \mathbb{1}\{\bar{\boldsymbol{d}}^U(U) > 0\}$ means that we reject H_U if and only if

$$\begin{aligned}
\bar{\boldsymbol{d}}^U(U) > 0 \iff & \max_{i=1,\dots,|U|} \{|U \cap S_i^U| - \bar{V}(S_i^U)\} > 0 \\
& \iff \text{there exists some } i = 1, \dots, |U| \text{ such that } |S_i^U| > \bar{V}(S_i^U) \text{ (since } S_i^U \subseteq U).
\end{aligned}$$

Let $L^U(i)$ be defined as in (21), then we have

$$\begin{aligned}
|S_i^U| > \bar{V}(S_i^U) &\iff |S_i^U| > \lfloor c(\alpha) \cdot (1 + i - |S_i^U|) \rfloor \iff |S_i^U| > c(\alpha) \cdot (1 + i - |S_i^U|) \\
&\iff |S_i^U| > \frac{c(\alpha) \cdot (1+i)}{1 + c(\alpha)} \iff |S_i^U| \geq \left\lfloor \frac{c(\alpha) \cdot (1+i)}{1 + c(\alpha)} \right\rfloor + 1 \\
&\iff L^U(i) \geq \left\lfloor \frac{c(\alpha) \cdot (1+i)}{1 + c(\alpha)} \right\rfloor + 1 = c_i,
\end{aligned}$$

where the last equivalence follows from $|S_i^U| = L^U(i)$.

Note that $i - c_i + 1 = \left\lceil \frac{1}{1+c(\alpha)}(1+i) \right\rceil - 1$ and $\frac{1}{1+c(\alpha)} \in \left(0, \frac{1}{2}\right)$ for any $\alpha \in (0,1)$, we have that for any positive integer l, there exists some i such that $i - c_i + 1 = l$. Let

$k_i^{raw} = \min_{j\geq 1}\{c_j : j - c_j + 1 = i\}$ and $b_i^I = k_i^{raw} + i - 1$. By the fact (24) in Appendix C, we have

$$
\begin{aligned}
&\text{there exists some } i = 1, \ldots, |U| \text{ such that } L^U(i) \geq c_i \\
\Longleftrightarrow\ &\text{there exists some } i = 1, \ldots, |U| \text{ such that } NB^{|U|}(i - c_i + 1) \geq c_i \\
\Longleftrightarrow\ &\text{there exists some } i = 1, \ldots, |U| \text{ such that } NB^{|U|}(i) \geq k_i^{raw} \\
\Longleftrightarrow\ &\text{there exists some } i = 1, \ldots, |U| \text{ such that } L^U(b_i^I) \geq k_i^{raw}.
\end{aligned}
$$

Therefore, the local test ϕ_U is equivalent to using test statistic (21) with $m = |U|$, $b_i^U = k_i^{raw} + i - 1$ and critical values $z_i^U = k_i^{raw}$. □

E Main proofs

E.1 Proof of Lemma 3.1

Lemma 3.1. The proof follows similarly to the proof of Lemma 2 in Janson and Su (2016).

Let $m_0 = |\mathcal{N}|$ be the number of null variables. Denote the random variables obtained by taking the signs of W_i, $i \in \mathcal{N}$, by $X_1, \ldots, X_{m_0}$. By the coin-slip property, $X_1, \ldots, X_{m_0}$ are i.i.d. random variables with distribution $\mathbb{P}(X_i = +1) = \mathbb{P}(X_i = -1) = 1/2$. In the same probability space, we can generate i.i.d. $X_{m_0+1}, \ldots$ from the same distribution to form an i.i.d. Bernoulli sequence $X_1, \ldots, X_{m_0}, X_{m_0+1}, \ldots$ taking values $\{+1, -1\}$.

Let $U(v) = \min\{i \geq 1 : \#\{j \leq i : X_j = -1\} = v\}$, $NB^{m_0}(v) = \#\{j \leq \min(U(v), m_0) : X_j = 1\}$, and $NB^p(v) = \#\{j \leq \min(U(v), p) : X_j = 1\}$. For any $v \geq 1$, we have

$$|S(v) \cap \mathcal{N}| \leq NB^{m_0}(v) \leq NB^p(v),$$

where the first inequality is due to the fact that some W_i, $i \in [p]\backslash\mathcal{N}$, might be negative, so the threshold $T(v)$ might stop at a place that leads to less false discoveries $|S(v) \cap \mathcal{N}|$ than $NB^{m_0}(v)$. Hence the event $\{NB^p(v_1) \leq k_1 - 1, \ldots, NB^p(v_m) \leq k_m - 1\}$ implies the event $\{|S(v_1) \cap \mathcal{N}| \leq k_1 - 1, \ldots, |S(v_m) \cap \mathcal{N}| \leq k_m - 1\}$, which leads to the desired inequality.

□

E.2 Proof of Proposition 3.1

For the second claim of Proposition 3.1, we need to use the following Lemma E.1.

Lemma E.1. *For any $\alpha \in (0, 1)$ and $i \geq 1$,*

$$\min_{j\geq 1}\{c_j : j - c_j + 1 = i\} = \lfloor c(\alpha)i \rfloor + 1,$$

where $c_j = \left\lfloor \frac{c(\alpha)\cdot(1+j)}{1+c(\alpha)} \right\rfloor + 1$ and $c(\alpha) = \frac{\log(\alpha^{-1})}{\log(2-\alpha)}$.

We give the proof of Lemma E.1.

Lemma E.1. For convenience of notation, let

$$\beta = \frac{c(\alpha)}{1 + c(\alpha)},$$

so

$$j - c_j + 1 = j - \left\lfloor \frac{c(\alpha) \cdot (1 + j)}{1 + c(\alpha)} \right\rfloor = j - \lfloor \beta(1 + j) \rfloor = \lceil j - \beta(1 + j) \rceil = \lceil j(1 - \beta) - \beta \rceil,$$

and

$$\min_{j \geq 1} \{c_j : j - c_j + 1 = i\} = \min_{j \geq 1} \{c_j : \lceil j(1 - \beta) - \beta \rceil = i\}.$$

Note that

$$\lceil j(1 - \beta) - \beta \rceil = i \Longleftrightarrow i - 1 < j(1 - \beta) - \beta \leq i \Longleftrightarrow \frac{i + \beta - 1}{1 - \beta} < j \leq \frac{i + \beta}{1 - \beta},$$

and $\frac{i+\beta-1}{1-\beta}$ must not be an integer, we have that

$$\operatorname{argmin}\{j \geq 1 : \lceil j(1 - \beta) - \beta \rceil = i\} = \left\lceil \frac{i + \beta - 1}{1 - \beta} \right\rceil =: j^*,$$

so

$$\begin{aligned}
&\min_{j \geq 1} \{c_j : \lceil j(1 - \beta) - \beta \rceil = i\} = c_{j^*} = \lfloor \beta\,(1 + j^*) \rfloor + 1 \\
&= \left\lfloor \beta \left(1 + \left\lceil \frac{i + \beta - 1}{1 - \beta} \right\rceil \right) \right\rfloor + 1 = \left\lfloor \beta \left\lceil \frac{i}{1 - \beta} \right\rceil \right\rfloor + 1 \\
&= \left\lfloor \frac{c(\alpha)}{1 + c(\alpha)} \lceil (1 + c(\alpha)) i \rceil \right\rfloor + 1
\end{aligned}$$

Therefore, to show that

$$\min_{j \geq 1} \{c_j : j - c_j + 1 = i\} = \lfloor c(\alpha) i \rfloor + 1,$$

it is sufficient to show that

$$\left\lfloor \frac{c(\alpha)}{1 + c(\alpha)} \lceil (1 + c(\alpha)) i \rceil \right\rfloor = \lfloor c(\alpha) i \rfloor. \tag{34}$$

When $c(\alpha)i$ is an integer, (34) clearly holds.

Now assume that $c(\alpha)i$ is not an integer. For convenience of notation, let

$$\gamma = \frac{c(\alpha)}{1 + c(\alpha)} \lceil (1 + c(\alpha)) i \rceil,$$

note that we have $c(\alpha)i \le \gamma$. To show (34), it is equivalent to show that γ and $c(\alpha)i$ have the same integer part, which is equivalent to

$$\begin{aligned}\gamma < \lceil c(\alpha)i \rceil &\iff \frac{c(\alpha)}{1+c(\alpha)} \lceil (1+c(\alpha))i \rceil < \lceil c(\alpha)i \rceil \\ &\iff \frac{c(\alpha)}{1+c(\alpha)} i + \frac{c(\alpha)}{1+c(\alpha)} \lceil c(\alpha)i \rceil < \lceil c(\alpha)i \rceil \\ &\iff \frac{c(\alpha)}{1+c(\alpha)} i < \lceil c(\alpha)i \rceil \left(1 - \frac{c(\alpha)}{1+c(\alpha)}\right) \\ &\iff \frac{c(\alpha)}{1+c(\alpha)} i < \lceil c(\alpha)i \rceil \frac{1}{1+c(\alpha)} \\ &\iff c(\alpha)i < \lceil c(\alpha)i \rceil,\end{aligned}$$

which clearly holds. $\square$

Now we prove Proposition 3.1.

Proposition 3.1. We first prove claim (i). Note that

$$i - c_i + 1 = i - \left\lfloor \frac{c(\alpha)\cdot(1+i)}{1+c(\alpha)} \right\rfloor = \left\lceil i - \frac{c(\alpha)}{1+c(\alpha)}(1+i) \right\rceil = \left\lceil \frac{1}{1+c(\alpha)}(1+i) \right\rceil - 1. \quad (35)$$

Hence for any positive integer l, there exists some i such that $i - c_i + 1 = l$ as $\frac{1}{1+c(\alpha)} \in \left(0, \frac{1}{2}\right)$ for any $\alpha \in (0,1)$. So we can define two integers $i_1 = \min\{i \ge 1 : i - c_i + 1 = 1\}$ and $i_m = \min\{i \ge 1 : i - c_i + 1 = v_m\}$.

Let $q = \max\{i_m, p\}$ and $W = (W_1, W_2, \ldots, W_q)$ be knockoff statistic vector satisfying the coin-flip property and $|W_1| > |W_2| > \cdots > |W_q| > 0$. For any $I \subseteq [q]$ and $i \le |I|$, let $S_i^I = \{j \in I : W_j = W_l^I \text{ such that } l \le i \;\&\; W_l^I > 0\}$, where W_l^I denotes the W_j whose absolute value $|W_j|$ is the l-th largest in $\{|W_j| : j \in I\}$.

Based on $\{W_i : i \in I\}$, the corresponding true discovery control procedure (cf. (33) in Appendix D.3) of Katsevich and Ramdas (2020) is

$$\bar{\boldsymbol{d}}^I(R) = \max_{i=1,\ldots,|I|} \{|R \cap S_i^I| - \bar{V}(S_i^I)\} \vee 0, \quad R \subseteq I,$$

where $\bar{V}(S_i^I) = \lfloor c(\alpha)\cdot(1 + i - |S_i^I|)\rfloor$. By Theorem D.1, for any $R \subseteq [q]$, the local test

$$\phi_R = \mathbb{1}\{\bar{\boldsymbol{d}}^R(R) > 0\}$$

is valid, which means that under H_R,

$$\mathbb{P}(\phi_R = 0) = \mathbb{P}(|R \cap S_1^R| \le \bar{V}(S_1^R), \ldots, |R \cap S_{|R|}^R| \le \bar{V}(S_{|R|}^R)) \ge 1 - \alpha.$$

Because

$$\begin{aligned}|R \cap S_i^R| \le \bar{V}(S_i^R) &\iff |S_i^R| \le \lfloor c(\alpha)\cdot(1 + i - |S_i^R|)\rfloor \iff |S_i^R| \le c(\alpha)\cdot(1 + i - |S_i^R|) \\ &\iff |S_i^R| \le \frac{c(\alpha)\cdot(1+i)}{1+c(\alpha)} \iff |S_i^R| \le \left\lfloor \frac{c(\alpha)\cdot(1+i)}{1+c(\alpha)} \right\rfloor = c_i - 1,\end{aligned}$$

we have

$$\mathbb{P}(|S_1^R| \leq c_1 - 1, \ldots, |S_{|R|}^R| \leq c_{|R|} - 1) \geq 1 - \alpha.$$

Hence, by taking $R = [q]$, we have

$$\begin{aligned}
&\mathbb{P}(|S_{i_1}^{[q]}| \leq c_{i_1} - 1, |S_{i_1+1}^{[q]}| \leq c_{i_1+1} - 1, \ldots, |S_q^{[q]}| \leq c_q - 1) \\
\geq &\mathbb{P}(|S_1^{[q]}| \leq c_1 - 1, |S_2^{[q]}| \leq c_2 - 1, \ldots, |S_q^{[q]}| \leq c_q - 1) \\
\geq &1 - \alpha.
\end{aligned}$$

Let

$$NB^q(v) = \#\{i \in [q] : W_i \geq T^q(v)\},$$

where $T^q(v) = \max\{|W_i| : \#\{j \in [q] : |W_j| \geq |W_i| \text{ and } W_j < 0\} = v\}$ with $T^q(v) = \min_{i \in [q]}\{|W_i|\}$ if $\#\{i \in [q] : W_i < 0\} < v$. For $i \geq i_1$, we have that $i - c_i \geq 0$ because the sequence $i - c_i$ is monotonically increasing. Thus for $i \geq i_1$,

$$|S_i^{[q]}| \leq c_i - 1 \Longleftrightarrow NB^q(i - c_i + 1) \leq c_i - 1$$

by the fact (25) in Appendix C. Therefore, we have

$$\begin{aligned}
&\mathbb{P}(NB^q(i_1 - c_{i_1} + 1) \leq c_{i_1} - 1, \ldots, NB^q(q - c_q + 1) \leq c_q - 1) \\
= &\mathbb{P}(|S_{i_1}^{[q]}| \leq c_{i_1} - 1, \ldots, |S_q^{[q]}| \leq c_q - 1) \\
\geq &1 - \alpha,
\end{aligned}$$

Recall that for any positive integer l, there exists some i such that $i - c_i + 1 = l$, so we can define $k_i^{raw} = \min_{j \geq 1}\{c_j : j - c_j + 1 = i\}$ for $i \in [v_m]$. By the definition of q, for $1 \leq v_1 < \cdots < v_m$, we have

$$\begin{aligned}
&\mathbb{P}(NB^p(v_1) \leq k_{v_1}^{raw} - 1, NB^p(v_2) \leq k_{v_2}^{raw} - 1, \ldots, NB^p(v_m) \leq k_{v_m}^{raw} - 1) \\
\geq &\mathbb{P}(NB^p(1) \leq k_1^{raw} - 1, NB^p(2) \leq k_2^{raw} - 1, \ldots, NB^p(v_m) \leq k_{v_m}^{raw} - 1) \\
\geq &\mathbb{P}(NB^q(1) \leq k_1^{raw} - 1, NB^q(2) \leq k_2^{raw} - 1, \ldots, NB^q(v_m) \leq k_{v_m}^{raw} - 1) \\
\geq &\mathbb{P}(NB^q(i_1 - c_{i_1} + 1) \leq c_{i_1} - 1, NB^q(i_2 - c_{i_2} + 1) \leq c_{i_2} - 1, \ldots, NB^q(q - c_q + 1) \leq c_q - 1) \\
\geq &1 - \alpha.
\end{aligned}$$

Now we show claim (ii), which is equivalent to show that

$$\min_{i=1,\ldots,p}\{k_i^{raw} - 1 + |R \backslash S(i)|, |R|\} = \min_{j=1,\ldots,p}\{\bar{V}(|S_j^-|) + |R \backslash S_j|, |R|\}. \tag{36}$$

Note that $\bar{V}(i) \geq 1$ is an increasing sequence and $k_i^{raw} \geq 1$ is also increasing because $i - c_i$ and c_i are both increasing. Let $n^- = |S_p^-|$ be the number of negatives in $(W_1, W_2, \ldots, W_p)$.

In the case of $n^- = p$,

$$\min_{i=1,\ldots,p}\{k_i^{raw} - 1 + |R \backslash S(i)|, |R|\} = \min\{k_1^{raw} - 1 + |R|, |R|\} = |R|$$

and

$$\min_{j=1,\ldots,p}\{\bar{V}(|S_j^-|)+|R\backslash S_j|,|R|\}=\min_{j=1,\ldots,p}\{\bar{V}(1)+|R|,|R|\}=|R|,$$

so (36) holds.

In the case of $n^- < p$,

$$\min_{i=1,\ldots,p}\{k_i^{raw}-1+|R\backslash S(i)|,|R|\}=\min_{i=1,\ldots,n^-+1}\{k_i^{raw}-1+|R\backslash S(i)|,|R|\}.$$

Let $S_0 = S_0^- = \emptyset$. For $l = 0,\ldots,n^-$, let $a_l = \max\{j \geq 0 : |S_j^-| = l\}$, then

$$\min_{j=1,\ldots,p}\{\bar{V}(|S_j^-|)+|R\backslash S_j|,|R|\}=\min_{l=0,\ldots,n^-}\{\bar{V}(l)+|R\backslash S_{a_l}|,|R|\}.$$

For $u = 1,\ldots,n^-+1$, we have $S(u) = S_{a_{u-1}}$ by definition, so to prove (36), it is sufficient to show that for any $u = 1,\ldots,n^-+1$,

$$k_u^{raw}-1=\bar{V}(u-1),$$

that is,

$$\min_{v\geq 1}\{c_v : v-c_v+1=u\}-1=\lfloor c(\alpha)u\rfloor,$$

which holds by Lemma E.1. □

E.3 Proof of Theorem 4.1

The proof of Theorem 4.1 relies on the following result, which gives the explicit formula (or shortcut) for the false discovery bound of the closed testing procedure with certain local test.

Theorem E.1. *For $m \geq 1$, $R \subseteq I \subseteq [p]$, $S_1^I \subseteq \cdots \subseteq S_m^I \subseteq I$, $N^I = I\backslash S_m^I$ and $k_1 \leq \cdots \leq k_m$, let $t_{ct}^I(R)$ be the false discovery upper bound by closed testing (see (7)) with local test*

$$\phi_R=\max_{i\in[m]}\mathbb{1}\{L_i(R)\geq k_i\},$$

where $L_i(R) = |S_i^R|$. Let

$$t^I(R)=\min_{i\in[m]}\{|R\backslash S_i^I|+k_i-1,|R|\}.$$

If

$$R\cap S_i^I=S_i^{R\cup N^I}\subseteq S_i^R, \tag{37}$$

then

$$t_{ct}^I(R)=t^I(R).$$

We first give the proof of Theorem 4.1, and then give the proof of Theorem E.1.

Theorem 4.1. For any $I \subseteq [p]$, let $S^I(v_i) = \{j \in I : W_j \geq T^I(v_i)\}$ and $T^I(v_i) = \max\{|W_j| : j \in I, \#\{l \in I : |W_l| \geq |W_j| \text{ and } W_l < 0\} = v_i\}$ with $T^I(v_i) = \min_{j \in I}\{|W_j|\}$ if $\#\{j \in I : W_j < 0\} < v_i$.

Let $S_i^I = S^I(v_i)$ and $N^I = I \backslash S_m^I$, so we have $S_1^I \subseteq \cdots \subseteq S_m^I \subseteq I$. By the definition of S_i^I, we have

$$R \cap S_i^I = S_i^{R \cup N^I} \subseteq S_i^R.$$

Let $t_{ct}^I(R)$ be the false discovery upper bound by closed testing (see (7)) with local test

$$\phi_R = \max_{i \in [m]} \mathbb{1}\{|S_i^R| \geq k_i\},$$

and let

$$t^I(R) = \min_{i \in [m]}\{|R \backslash S_i^I| + k_i - 1, |R|\},$$

by Theorem E.1, we have

$$t_{ct}^I(R) = t^I(R). \tag{38}$$

Note that $\overline{\mathrm{FDP}}_{(k,v)}^m(R) = \frac{t^I(R)}{\max\{1,|R|\}}$ and the local test ϕ_R is equivalent to using local test statistic (21) with $b_j^I = k_j + v_j - 1$ and critical values $z_j^I = k_j$ (see fact (24) in Appendix C), we have

$$\overline{\mathrm{FDP}}_{(k,v)}^m(R) = \overline{\mathrm{FDP}}^{ct}(R)$$

by taking $I = [p]$ and dividing $\max(1, |R|)$ on both sides of (38). $\square$

Now we give the proof of Theorem E.1.

Theorem E.1. By condition (37), we have

$$L_i(R) = |S_i^R| \geq |S_i^{R \cup N^I}| = L_i(R \cup N^I) = |R \cap S_i^I|. \tag{39}$$

Let

$$t_i^I(R) = \min\{|R \backslash S_i^I| + k_i - 1, |R|\} \quad \text{ and } \quad a = \operatorname*{argmin}_{i \in [m]} t_i^I(R),$$

so $t^I(R) = \min_{i \in [m]} t_i^I(R) = t_a^I(R)$.

In the case that H_R is not rejected by closed testing, there exists some $R_{\text{sup}} \supseteq R$ such that $H_{R_{\text{sup}}}$ is not rejected by the local test. That is, for all $i \in [m]$, we have $L_i(R) \leq k_i - 1$. Hence for a, we have $|R \cap S_a^I| \leq |R_{\text{sup}} \cap S_a^I| \leq L_a(R_{\text{sup}}) \leq k_a - 1$, where the second last inequality holds due to (39). By how closed testing works, we have $t_{ct}^I(R) = |R| = |R \cap S_a^I| + |R \backslash S_a^I| = \min\{|R \cap S_a^I|, k_a - 1\} + |R \backslash S_a^I| = \min\{|R|, |R \backslash S_a^I| + k_a - 1\} = t^I(R)$.

In the case that H_R is rejected by closed testing, there must exist some $b \in [m]$ such that $|R \cap S_b^I| \geq k_b$. Otherwise $H_{R\cup N^I}$ will not be rejected by the local test because $L_i(R \cup N^I) = |R \cap S_i^I| \leq k_i - 1$ for all $i \in [m]$, which contradicts to the fact that H_R is reject by closed testing. Therefore, we can define a non-empty set

$$B = \{i \in [m] : |R \cap S_i^I| \geq k_i\} := \{b_1, \ldots, b_{|B|}\}$$

with $S_{b_1}^I \subseteq S_{b_2}^I \subseteq \cdots \subseteq S_{b_{|B|}}^I$. Let $B^c = \{i \in [m] : |R \cap S_i^I| \leq k_i - 1\}$ and $S_B^I = \cup_{i\in B} S_i^I$, so $S_{b_{|B|}}^I = S_B^I$ and $R \cap S_{b_1}^I \subseteq R \cap S_{b_2}^I \subseteq \cdots \subseteq R \cap S_{b_{|B|}}^I = R \cap S_B^I$.

We construct the largest subset of R that is not rejected by closed testing. Let $E_1 = R \cap S_{b_1}^I$ and $E_i = (R \cap S_{b_i}^I) \backslash (R \cap S_{b_{i-1}}^I)$, $i = 2, \ldots, |B|$, so they are disjoint and $\cup_{i\leq j} E_i = R \cap S_{b_j}^I$, $j = 1, \ldots, |B|$. We then construct set F_i, $i = 1, \ldots, |B|$, in the following way: we take F_1 as a subset of E_1 such that $|F_1| = k_{b_1} - 1$, such F_1 must exist because $|E_1| = |R \cap S_{b_1}^I| \geq k_{b_1}$. We take F_2 as a subset of E_2 with size $\min\{k_{b_2} - |F_1| - 1, |E_2|\}$, so $|F_1 \cup F_2| = |F_1| + |F_2| \leq k_{b_2} - 1$. We take F_3 as a subset of E_3 with size $\min\{k_{b_3} - |\cup_{i\leq 2} F_i| - 1, |E_3|\}$, so $|\cup_{i\leq 3} F_i| = |\cup_{i\leq 2} F_i| + |F_3| \leq k_{b_3} - 1$. Recursively, for $4 \leq j \leq |B|$, we take F_j as a subset of E_j with size $\min\{k_{b_j} - |\cup_{i\leq j-1} F_i| - 1, |E_j|\}$, so $|\cup_{i\leq j} F_i| = |\cup_{i\leq j-1} F_i| + |F_j| \leq k_{b_j} - 1$. Note that $k_{b_j} - |\cup_{i\leq j-1} F_i| - 1 \geq k_{b_j} - k_{b_{j-1}} \geq 0$ for $2 \leq j \leq |B|$, and we have that $F_i \subseteq E_i \subseteq S_{b_i}^I \subseteq S_B^I$, $F_i \subseteq E_i \subseteq R$ and F_i, $i = 1, \ldots, |B|$, are disjoint.

Let

$$R_{sub} = \cup_{i\leq |B|} F_i \cup (R \backslash S_B^I).$$

Since F_i's and $R \backslash S_B^I$ are disjoint, we have

$$|R_{sub}| = \sum_{i=1}^{|B|} |F_i| + |R \backslash S_B^I|.$$

We claim that R_{sub} is the largest subset of R that is not rejected by the closed testing procedure. To prove this, we first show that $H_{R_{sub}}$ is not rejected by closed testing, then we show that any strict subset of it must be rejected by closed testing.

Specifically, consider the superset $R_{sub} \cup N^I$ of R_{sub}. For $i \in B^c$, we have $L_i(R_{sub} \cup N^I) = |R_{sub} \cap S_i^I| \leq |R \cap S_i^I| \leq k_i - 1$. For $i \in B$, say $i = b_l$. If $l = |B|$, then we have

$$L_i(R_{sub} \cup N^I) = |R_{sub} \cap S_i^I| = |R_{sub} \cap S_{b_{|B|}}^I| = |\cup_{j\leq |B|} F_j| \leq k_{b_{|B|}} - 1 = k_i - 1.$$

If $1 \leq l < |B|$, for $l < j \leq |B|$, we have

$$F_j \cap S_{b_l}^I \subseteq E_j \cap S_{b_l}^I = E_j \cap S_{b_l}^I \cap R = (R \cap S_{b_j}^I) \backslash (R \cap S_{b_{j-1}}^I) \cap (R \cap S_{b_l}^I) = \emptyset,$$

where the first equality holds because E_j is a subset of R, and the last equality holds because $R \cap S_{b_l}^I \subseteq R \cap S_{b_{j-1}}^I$. Hence for the test statistic we have

$$\begin{aligned} L_i(R_{sub} \cup N^I) &= |R_{sub} \cap S_i^I| = |R_{sub} \cap S_{b_l}^I| = |(\cup_{j\leq |B|} F_j) \cap S_{b_l}^I| = |\cup_{j\leq l} (F_j \cap S_{b_l}^I)| \\ &\leq |\cup_{j\leq l} F_j| \leq k_{b_l} - 1 = k_i - 1. \end{aligned}$$

Thus for any $i \in [m]$, we have $L_i(R_{sub} \cup N^I) \leq k_i - 1$, so $H_{R_{sub} \cup N^I}$ is not rejected by the local test, which then implies that $H_{R_{sub}}$ is not rejected by closed testing.

Next, we show that for any $R' \subseteq R$ with $|R'| \geq |R_{sub}| + 1$, $H_{R'}$ must be rejected by closed testing. To see this, we write $R' = \cup_{i \leq |B|} R'_i \cup R'_c \subseteq R$, where $R'_i \subseteq E_i$ and $R'_c \subseteq R \backslash S^I_B$. Since $|R_{sub}| = \sum_{i=1}^{|B|} |F_i| + |R \backslash S^I_B|$, $|R'| = \sum_{i=1}^{|B|} |R'_i| + |R'_c|$ and $|R'_c| \leq |R \backslash S^I_B|$, the set $\{j \leq |B| : \sum_{i=1}^{j} |R'_i| \geq \sum_{i=1}^{j} |F_i| + 1\} \neq \emptyset$ because $|B|$ must in this set. Otherwise $|R'| \leq \sum_{i=1}^{|B|} |F_i| + |R \backslash S^I_B| = |R_{sub}|$, which contradicts to $|R'| \geq |R_{sub}| + 1$.

Let $u = \operatorname{argmin}\{j \leq |B| : \sum_{i=1}^{j} |R'_i| \geq \sum_{i=1}^{j} |F_i| + 1\}$, so $| \cup_{i \leq u} R'_i| \geq | \cup_{i \leq u} F_i| + 1$ and $| \cup_{i \leq u-1} R'_i| \leq | \cup_{i \leq u-1} F_i|$. For this u, we must have $|F_u| = k_{b_u} - | \cup_{i \leq u-1} F_i| - 1$. Otherwise if $|F_u| = |E_u|$, we have $R'_u \subseteq E_u = F_u$. Then by $| \cup_{i \leq u-1} R'_i| \leq | \cup_{i \leq u-1} F_i|$, we have $| \cup_{i \leq u} R'_i| = | \cup_{i \leq u-1} R'_i| + |R'_u| \leq | \cup_{i \leq u-1} F_i| + |F_u| = | \cup_{i \leq u} F_i|$, which contradicts to $| \cup_{i \leq u} R'_i| \geq | \cup_{i \leq u} F_i| + 1$.

Then by $|F_u| = k_{b_u} - | \cup_{i \leq u-1} F_i| - 1$, we have $| \cup_{i \leq u} R'_i| \geq | \cup_{i \leq u} F_i| + 1 = k_{b_u}$. Then, for any $R'_{sup} \supseteq R'$, we have $L_{b_u}(R'_{sup}) \geq |R'_{sup} \cap S^I_{b_u}| \geq |R' \cap S^I_{b_u}| = | \cup_{i \leq |B|} R'_i \cap S^I_{b_u}| \geq | \cup_{i \leq u} R'_i \cap S^I_{b_u}| = | \cup_{i \leq u} R'_i \cap R \cap S^I_{b_u}| = | \cup_{i \leq u} R'_i| \geq k_{b_u}$, where the second last equality is due to $\cup_{i \leq u} R'_i \subseteq \cup_{i \leq u} E_i = R \cap S^I_{b_u}$. Hence $H_{R'}$ is rejected by closed testing.

Therefore, in the case that H_R is rejected by the closed testing procedure, R_{sub} is the largest subset of R that is not rejected by closed testing, so we have

$$t^I_{ct}(R) = |R_{sub}| = \sum_{i=1}^{|B|} |F_i| + |R \backslash S^I_B|.$$

Based on the above explicit formula of $t^I_{ct}(R)$, we are now ready to show that $t^I_{ct}(R) = t^I(R)$. Before that, we first show a result that we will use many times later.

For $l \leq |B| - 1$, recall that $R \cap S^I_B = \cup_{i \leq |B|} E_i$, $R \cap S^I_{b_l} = \cup_{i \leq l} E_i$ and E_i's are disjoint, so we have

$$\begin{aligned}(R \backslash S^I_{b_l}) \backslash (R \backslash S^I_B) &= (R \cap (S^I_{b_l})^c) \cap (R \cap (S^I_B)^c)^c = R \cap (S^I_{b_l})^c \cap S^I_B \\ &= (R \cap S^I_B) \cap (R^c \cup (S^I_{b_l})^c) = (R \cap S^I_B) \backslash (R \cap S^I_{b_l}) = \cup_{l < i \leq |B|} E_i,\end{aligned}$$

and so

$$|R \backslash S^I_{b_l}| - |R \backslash S^I_B| = |(R \backslash S^I_{b_l}) \backslash (R \backslash S^I_B)| = | \cup_{l < i \leq |B|} E_i| = \sum_{l < i \leq |B|} |E_i|. \tag{40}$$

Now we show that $t^I_{ct}(R) = t^I(R)$, we first show that $t^I_{ct}(R) \leq t^I_j(R)$ for all $j \in [m]$. In the case of $j \in B^c$, we have $t^I_j(R) = |R| \geq t^I_{ct}(R)$. In the case of $j \in B$, we have $t^I_j(R) = k_j - 1 + |R \backslash S^I_j|$. Let $j = b_l$. If $l = |B|$, we have

$$t^I_{ct}(R) - t^I_j(R) = \sum_{i=1}^{|B|} |F_i| + |R \backslash S^I_B| - k_{b_{|B|}} + 1 - |R \backslash S^I_{b_{|B|}}| = \sum_{i=1}^{|B|} |F_i| - k_{b_{|B|}} + 1 \leq 0,$$

where the last inequality follows from $\sum_{i=1}^{|B|} |F_i| = |\cup_{i\leq|B|} F_i| \leq k_{b_{|B|}} - 1$. If $l \leq |B|-1$, we have

$$\begin{aligned}
t^I_{ct}(R) - t^I_j(R) &= \sum_{i=1}^{|B|} |F_i| + |R\backslash S^I_B| - k_{b_l} + 1 - |R\backslash S^I_{b_l}| \\
&= \sum_{i=1}^{|B|} |F_i| - k_{b_l} + 1 - \sum_{l<i\leq|B|} |E_i| \\
&\leq \sum_{i\leq l} |F_i| + \sum_{l<i\leq|B|} |E_i| - k_{b_l} + 1 - \sum_{l<i\leq|B|} |E_i| \\
&\leq \sum_{i\leq l} |F_i| - k_{b_l} + 1 \leq 0,
\end{aligned}$$

where the second equality follows from (40). Therefore, we have

$$t^I_{ct}(R) \leq \min_{i\in[m]} t^I_i(R) = t^I(R).$$

Next, we show that there exists some $j \in [m]$ such that $t^I_{ct}(R) = t^I_j(R)$. If $|F_i| = |E_i|$ for all $1 < i \leq |B|$, let $l = b_1 \in B$, so $t^I_l(R) = k_{b_1} - 1 + |R\backslash S_{b_1}|$. Then we have

$$\begin{aligned}
t^I_{ct}(R) - t^I_l(R) =& \sum_{i=1}^{|B|} |F_i| + |R\backslash S_B| - k_{b_l} + 1 - |R\backslash S_{b_l}| \\
=& |F_1| + \sum_{1<i\leq|B|} |F_i| - k_{b_1} + 1 - \sum_{1<i\leq|B|} |E_i| = 0,
\end{aligned}$$

where we used (40) and $|F_1| = k_{b_1} - 1$.

If there exists $1 < j \leq |B|$ such that $|F_j| = k_{b_j} - |\cup_{i\leq j-1} F_i| - 1$. Let j' be the largest among such j and let $l = b_{j'} \in B$, so $t^I_l(R) = k_{b_{j'}} - 1 + |R\backslash S_{b_{j'}}|$ and $|\cup_{i\leq j'} F_i| = \sum_{i=1}^{j'} |F_i| = k_{b_{j'}} - 1$. If $j' = |B|$, we have

$$\begin{aligned}
t^I_{ct}(R) - t^I_l(R) &= \sum_{i=1}^{|B|} |F_i| + |R\backslash S_B| - k_{b_{|B|}} + 1 - |R\backslash S_{b_{|B|}}| \\
&= \sum_{i=1}^{|B|} |F_i| - k_{b_{|B|}} + 1 = 0.
\end{aligned}$$

If $j' < |B|$, we have $|F_i| = |E_i|$ for all $j' < i \leq |B|$. Thus

$$\begin{aligned}
t^I_{ct}(R) - t^I_l(R) &= \sum_{i=1}^{|B|} |F_i| + |R\backslash S_B| - k_{b_{j'}} + 1 - |R\backslash S_{b_{j'}}| \\
&= \sum_{i\leq j'} |F_i| + \sum_{j'<i\leq|B|} |F_i| - k_{b_{j'}} + 1 - \sum_{j'<i\leq|B|} |E_i| \\
&= 0.
\end{aligned}$$

Therefore we have

$$t_{ct}^I(R) = \min_{i\in[m]} t_i^I(R) = t^I(R).$$

□

E.4 Proof of Corollary 4.1

Corollary 4.1. By applying Proposition 3.1 and Theorem 4.1, we have that $\overline{\mathrm{FDP}}^{\mathrm{KR}}(R)$ is equal to the FDP bound based on closed testing using local test statistic (21) with $m = p$, $b_i^I = k_i^{raw} + i - 1$ and critical values $z_i^I = k_i^{raw}$ (for null hypothesis H_I). Because $k_i^{raw} \geq 1$, we have that $b_i^I \geq i$, so using $m = |I|$ and $m = p$ give the same test result. □

E.5 Proof of Proposition 4.1

Proposition 4.1. For $m = 1$ and $b_1^I = k + v - 1$, the local test statistic is

$$L^I(k+v-1) = \sum_{i\in I} \mathbb{1}_{r_i^I > |I|-(k+v-1)} \mathbb{1}_{D_i=1}.$$

In the case of $|S(v)| \leq k-1$, we have $L^{S(v)}(k+v-1) = |S(v)| \leq k-1 < z_1^I$, so $S(v)$ is not locally rejected, which implies that $t_\alpha(S(v)) = |S(v)|$. In particular, $t_\alpha(S(v)) = k-1$ when $|S(v)| = k-1$.

In the case of $|S(v)| \geq k$, for any superset I of $S(v)$, we have that the number of positive W_i before v negative W_i for $i \in I$ is larger than or equal to k, which is equivalent to $L^I(k+v-1) \geq k = z_1^I$ (see (24) in Appendix C). So $S(v)$ is rejected by closed testing. Let $S' \subseteq S(v)$ with $|S'| = k-1$, then we have $L^{S'}(k+v-1) = |S'| = k-1 < z_1^I$, so $S(v)$ is not locally rejected, which implies that $t_\alpha(S(v)) = k-1$. □

F Supplementary materials relate to simulations

F.1 Simulation details of Figure 1

For all settings, the knockoff statistic vector W is generated based on the linear Gaussian model as described in Section 5.1. In particular,

- Setting 1: $p = 200$, $n = 500$, $s = 0.1$, $a = 10$.
- Setting 2: $p = 200$, $n = 500$, $s = 0.3$, $a = 10$.
- Setting 3: $p = 1000$, $n = 2500$, $s = 0.1$, $a = 8$.
- Setting 4: $p = 1000$, $n = 2500$, $s = 0.3$, $a = 8$.

For the simultaneous FDP bounds (10), we use $S(v)$ obtained by using the randomized version of the method in Janson and Su (2016) for better power.

F.2 More simulation results for k^{raw}, k^{step1} and k^{step2} in Algorithm 2 using four types of v

In Figure 8, we present the values of k^{raw}, k^{step1} and k^{step2} by Algorithm 2 for four types of v described in (18). One can see that for all four types of v, $k^{step2} \leq k^{step1} \leq k^{raw}$, which implies that using k^{step2} will lead to better FDP bounds.

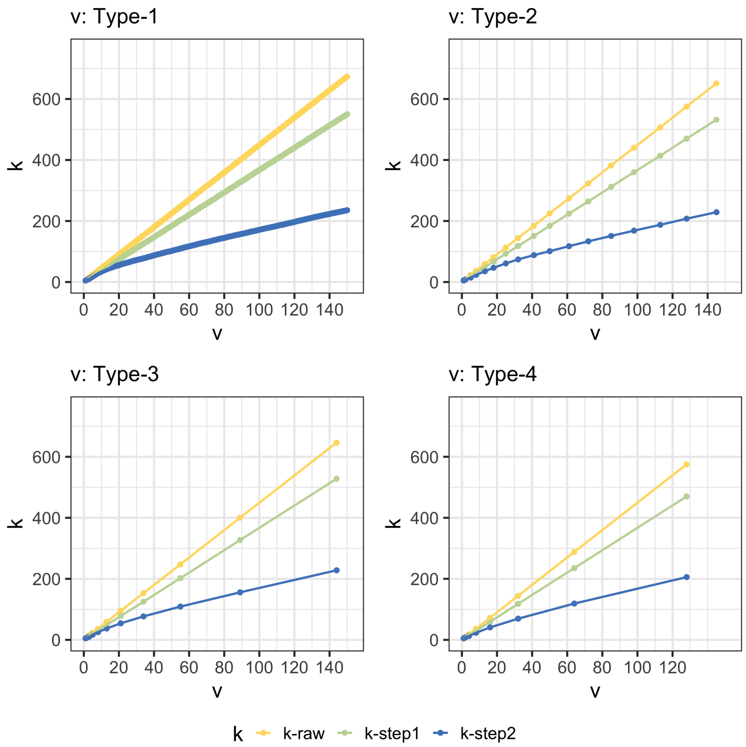

Figure 8: The k^{raw}, k^{step1} and k^{step2} in Algorithm 2 for the four types of v in (18).

F.3 More simulation results for KR and KF-JKI

To check the simultaneous FDP guarantee (2), it is equivalent to check the type-I error guarantee

$$\mathbb{P}(\mathrm{FDP}(R) > \overline{\mathrm{FDP}}(R), \exists R \subseteq [p]) \leq \alpha,$$

where $\alpha = 0.05$ in our case. The type-I error plots of KR and KF-JKI are presented in Figure 9. Here different settings in the x-axis correspond to different combinations of the setting parameters a and s. As expected, all methods control the type-I error.

In Figures 10 and 11, we present the FDP bounds of KR and KF-JKI in the linear and logistic regression model settings, respectively. Here we take $p = 1000$ and $n = 2500$ and use the tuning parameter vector v with $v_m < 300$.

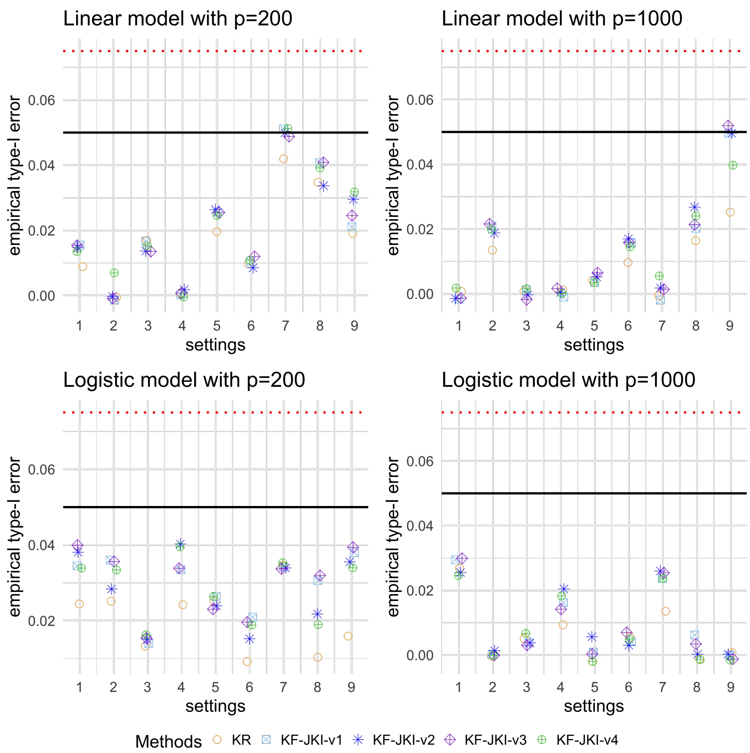

Figure 9: Empirical type-I errors of KR and KF-JKI in different settings. The solid black line indicates the nominal level 0.05. The dotted red line indicate the 95% confidence upper bound of the true type-I error over 200 replications.

F.4 Simulations to compare the closed testing based approaches with different local tests

In this section, we examine the performance of the closed testing method with the rank-generalization local test proposed in Section 4.2, and compare it to KF-CT (see Section 5.2). Specifically, we consider the following rank-generalization approach:

- KF-CT-rank: Our proposed method to obtain simultaneous FDP bounds based on closed testing using local test statistic is (19) with $w_{i,j}^I = r_j^I \mathbb{1}_{r_j^I > |I| - b_i^I}$, where the four types of b_i^I are the same as KF-CT, and the critical value is approximated based on (20). We denote them as KF-CT-rank-v1, KF-CT-rank-v2, KF-CT-rank-v3, and KF-CT-rank-v4, respectively. We use the shortcut Algorithm 3 for the implementation of closed testing.

We first implement KF-CT and KF-CT-rank in the same linear and logistic regression model settings as in Section 5.1 with $p = 200$ and $n = 500$. The simulation results are shown in Figure 12 and 13. One can see that KF-CT is generally better than KF-CT-rank. Possible explanation for this observation is that KF-CT also makes use of the information of rank because $|W_i|$'s are ordered. So the rank-generalization version might not give much better FDP bounds when the information of rank is good. On the other hand, when

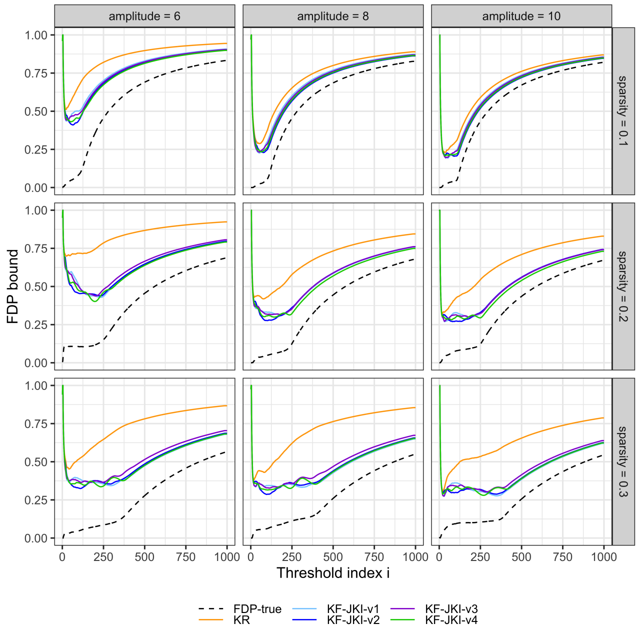

Figure 10: The FDP bounds of KR and KF-JKI with four types of (v, k) in the linear regression model setting with $p = 1000$ and $n = 2500$. The dashed black line indicates the true FDP. All FDP bounds are the average values over 200 replications.

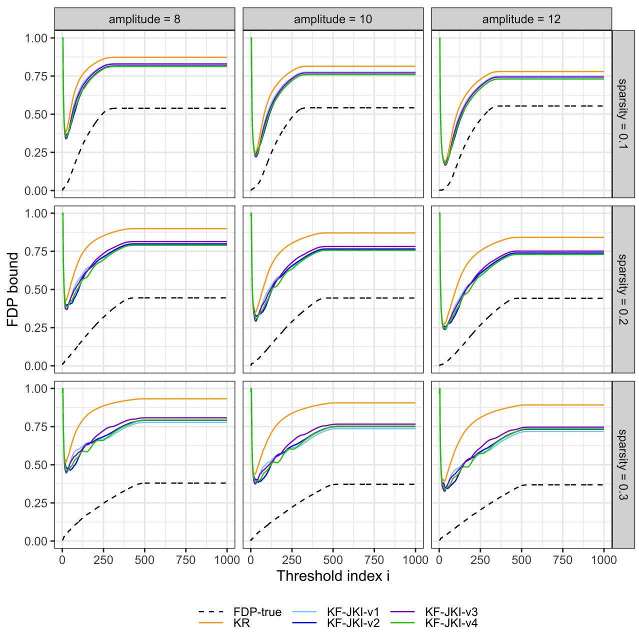

Figure 11: The FDP bounds of KR and KF-JKI with four types of (v, k) in the logistic regression model setting with $p = 1000$ and $n = 2500$. The dashed black line indicates the true FDP. All FDP bounds are the average values over 200 replications.

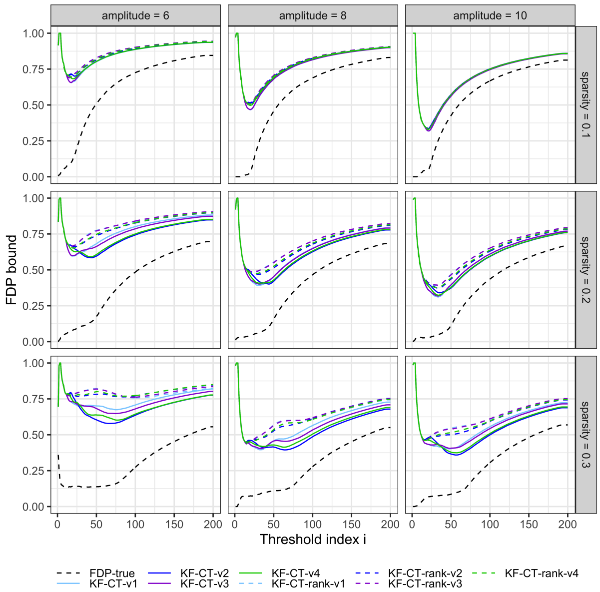

Figure 12: The FDP bounds of KF-CT and KF-CT-rank in the linear regression model setting with $p = 200$ and $n = 500$. The dashed black line indicates the true FDP. All FDP bounds are the average values over 200 replications.

the information of rank is bad, it will give worse FDP bounds than KF-CT because it is influenced more by using rank as weight in the local test statistic.

To show that KF-CT-rank can be better than KF-CT in certain cases, we consider a simulation setting directly generating knockoff statistic vector W. Specifically, we consider a setting with $p = 50$ and null variable set $\mathcal{N} = \{8, 12, 13, \ldots, 50\}$. For the knockoff statistics, we set $|W_i| = 50 - i + 1$, $\text{sign}(W_i) = 1$ for $i \notin \mathcal{N}$ and $\text{sign}(W_i) \overset{i.i.d.}{\sim} \{-1, 1\}$ with same probability $1/2$ for $i \in \mathcal{N}$. It is clear that this W satisfies the coin-flip property, so it is a valid knockoff statistic vector. Figure 14 shows the simulation results in this settings, and one can see that KF-CT-rank is better than KF-CT.

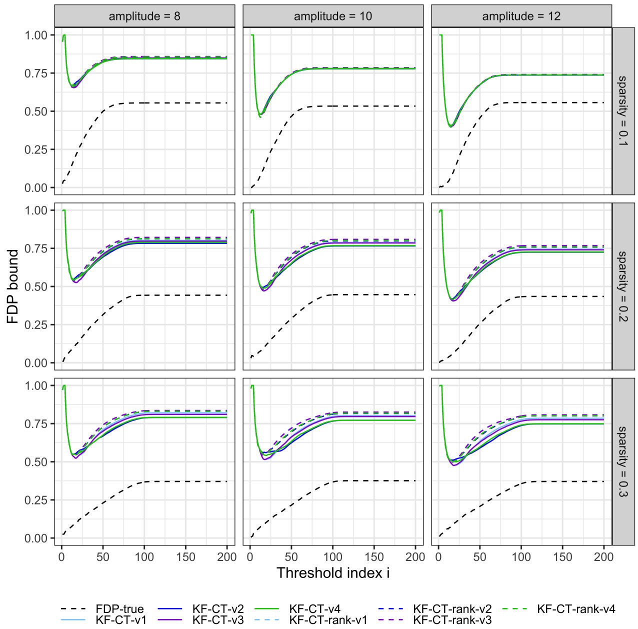

Figure 13: The FDP bounds of KF-CT and KF-CT-rank in the logistic regression model setting with $p = 200$ and $n = 500$. The dashed black line indicates the true FDP. All FDP bounds are the average values over 200 replications.

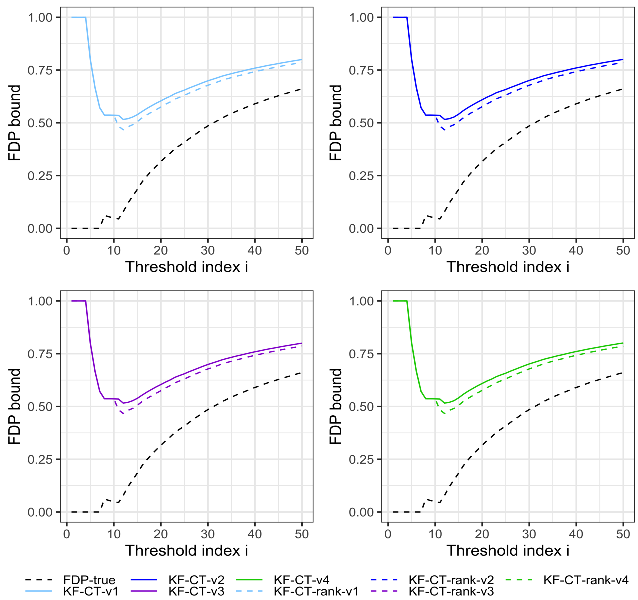

Figure 14: The FDP bounds of KF-CT and KF-CT-rank in the simulation setting where the knockoff statistic vector $W \in \mathbb{R}^{50}$ is directly generated. The dashed black line indicates the true FDP. All FDP bounds are the average values over 200 replications.