

# Multiple Knockoffs



Wentong Zhang  
Harvard College

*Submitted in partial fulfillment of the requirements for the degree of Bachelor of  
Arts with Honors in Statistics April 2018*

April 1, 2018

# Abstract

In this thesis, we present an extension of the knockoffs procedure in multiple testing. In particular, extending the work done on model-X knockoffs, we investigate the effects of generating numerous knockoffs simultaneously.

# Acknowledgements

---

fill this in

# Contents

<b>1</b>	<b>An Introduction to Multiple Comparison Procedures</b>	<b>4</b>
1.1	The Family-Wise Error Rate . . . . .	4
1.1.1	The Bonferroni Correction . . . . .	5
1.1.2	Holm's Procedure . . . . .	5
1.1.3	Hochberg's Procedure . . . . .	7
1.2	The False Discovery Rate . . . . .	7
1.2.1	The Benjamini-Hochberg Procedure . . . . .	8
<b>2</b>	<b>Introducing (Model-X) Knockoffs</b>	<b>10</b>
<b>3</b>	<b>Multiple Knockoffs</b>	<b>11</b>
3.1	Definition . . . . .	11
3.2	Exchangeability of Null Covariates and Knockoffs . . . . .	11
3.3	Feature Statistics and $p$ -values . . . . .	12
3.4	The Procedure . . . . .	14

# Chapter 1

## An Introduction to Multiple Comparison Procedures

In hypothesis testing, the *multiple comparison problem* is a problem that arises when one wishes to test numerous hypotheses simultaneously. In particular, we wish to accept or reject each hypothesis. In this case, as the number of inferences that we make increase, we should also expect the number of errors to increase; for instance, when null  $p$ -values are distributed uniformly between 0 and 1, if 1000 null hypotheses are tested, an average of 50 will still be rejected at the 0.05 significance level, whereas if only 20 are tested, then an average of only 1 will be rejected at the 0.05 significance level.

When we test multiple hypotheses simultaneously, we will end up with four possible outcomes for each hypothesis, presented in the table below:

	accepted	rejected	total
true	$U$	$V$	$n_0$
false	$T$	$S$	$n - n_0$
total	$n - R$	$R$	$n$

In the table above, note that the quantities  $R$  and  $n_0$  are observed, but  $U, V, T, S$  are actual unobserved random variables. *Type I error* is concerned with the quantity  $V$ , which is the number of null hypotheses that are incorrectly rejected, and will be what we wish to control under various metrics. Note that although we will be providing bounds on measures of Type I error, power will ultimately become the metric by which we wish to judge our procedures, as we will seek to maximize power of a procedure given a bound on Type I error.

### 1.1 The Family-Wise Error Rate

In this section, we will first develop background on some classical procedures that control a measure of Type I error called the *family-wise error rate*. In doing so, we will see that for the same control on the *family-wise error rate*, the Hochberg procedure is more powerful than the Holms procedure (which is more powerful than the Bonferroni correction). To begin, we will the *family-wise error rate* as follows.

**Definition 1.1.1.** The *family-wise error rate*, abbreviated FWER, is defined as

$$\text{FWER} = \mathbb{P}(V \geq 1) \tag{1.1.1}$$

where  $V$  is the number of erroneous rejections of null hypotheses.

Note that bounding the FWER can be a rather stringent constraint on a procedure, as the procedure must then control the probability of any false rejections at all.

### 1.1.1 The Bonferroni Correction

One of the earliest methods presented in controlling the FWER is known as the *Bonferroni correction* (alternatively the Bonferroni method). First used in [Dun59] in 1959, this procedure draws its name from the Bonferroni inequalities, which are used in the proof of its control of the FWER.

**Procedure 1.1.2** (Bonferroni Correction). Suppose we are given a collection of  $n$  null hypotheses, denoted  $H_{0,1}, \dots, H_{0,n}$ , with associated  $p$ -values denoted  $p_1, \dots, p_n$  respectively. Fix a desired level  $\alpha$ . Then, reject all hypotheses  $H_{0,i}$  for which

$$p_i \leq \alpha/n. \quad (1.1.2)$$

In other words, test all hypotheses  $H_{0,i}$  at level  $\alpha/n$ .

**Proposition 1.1.3.** *Procedure 1.1.2 controls the FWER at level  $\alpha$ .*

*Proof.* The proof of FWER control is a simple application of the union bound (or Bonferroni inequalities). In particular, without loss of generality, assume that  $H_{0,i}$  for  $i \leq n_0$  are the only true null hypotheses (where  $n_0 \leq n$ ). Then, the FWER may be bounded, as desired, by

$$\text{FWER} = \mathbb{P}(V \geq 1) = \mathbb{P}\left[\bigcup_{i=1}^{n_0} \left(p_i \leq \frac{\alpha}{n}\right)\right] \leq \sum_{i=1}^{n_0} \mathbb{P}\left(p_i \leq \frac{\alpha}{n}\right) = n_0 \cdot \frac{\alpha}{n} \leq \alpha. \quad (1.1.3)$$

□

The Bonferroni correction controls the FWER at level  $\alpha$ , but the cost in power can be quite severe. Consider a concrete example: suppose we have 1000 null hypotheses that we wish to test at the  $\alpha = 0.05$  level. In this case, we would only reject the hypotheses for which the  $p$ -values are less than  $5 \times 10^{-5}$ , or extremely significant  $p$ -values. As such, this test may be most suitable for situations in which we expect certain  $p$ -values to be extremely significant, as otherwise, it is possible that the Bonferroni procedure yields very low power.

### 1.1.2 Holm's Procedure

Another procedure that controls the FWER in multiple hypothesis testing is Holm's procedure. Previously, we saw that Procedure 1.1.2 successfully controlled the FWER at level  $\alpha$  for tests at level  $\alpha$ , but a simple numerical example illustrates how restrictive the Bonferroni procedure can be, in particular when considering a large number of hypotheses. Holm's procedure, introduced in [Hol79], provides less strict criteria for rejecting the null hypotheses, and hence will have uniformly more power. The procedure is as follows.

**Procedure 1.1.4** (Holm's Procedure). Suppose we are given a collection of  $n$  null hypotheses, denoted  $H_{0,1}, \dots, H_{0,n}$ , with associated  $p$ -values denoted  $p_1, \dots, p_n$  respectively. Fix a desired level  $\alpha$ . Then, order the  $p$ -values so that

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$$

and let  $H_{(1)}, \dots, H_{(n)}$  be the corresponding hypotheses. Then, define

$$k = \min \left\{ 1 \leq i \leq n : p_{(i)} > \frac{\alpha}{n - i + 1} \right\}$$

and reject hypotheses  $H_{(1)}, \dots, H_{(k-1)}$ . In particular, if it is the case that

$$p_{(i)} \leq \frac{\alpha}{n - i + 1}, \quad \forall 1 \leq i \leq n,$$

then we reject all  $H_1, \dots, H_n$ .

**Note.** A sequential way to think about Holm's procedure is as follows. We begin by considering  $p_{(1)}$ : if  $p_{(1)} \leq \alpha/n$ , then reject  $H_{(1)}$  and continue to  $p_{(2)}$ . Then, for each  $p_{(i)}$ , reject  $H_{(i)}$  if  $p_{(i)} \leq \alpha/(n - i + 1)$ , and continue to  $p_{(i+1)}$ ; otherwise, end the procedure. It is clearly seen that this is equivalent to finding  $k$  as defined in [Procedure 1.1.4](#) and rejecting hypotheses up to  $H_{(k-1)}$ .

In particular, Holm's procedure is a *step-down procedure*, as the procedure functions by rejecting hypothesis sequentially until one is accepted. On the other hand, the Hochberg procedure, which we examine in the next section, is actually an extremely similar step-up procedure. In terms of power, step-up procedures can be significantly more powerful than step-down procedures; this is a phenomenon we will elaborate on further in the next section.

From the sequential view of Holm's procedure, it now becomes evident that Holm's procedure is uniformly more powerful than Bonferroni's procedure as follows. We can translate Bonferroni's procedure into a sequential procedure as well: we may order the  $p$ -values in order, and sequentially go through each of the  $p$ -values with the fixed threshold  $p_{(i)} \leq \alpha/n$  and rejecting until  $p_{(i)} > \alpha/n$ . Hence, we can put Bonferroni's procedure into the same form as Holm's procedure, but the cutoffs are always stricter (or just as strict) as Holm's procedure for rejection; hence, Holm's procedure will be uniformly more powerful since it simply makes more rejections. We now provide a proof of FWER control under Holm's procedure.

**Proposition 1.1.5.** *[Procedure 1.1.4](#) controls the FWER at level  $\alpha$ .*

*Proof.* Let  $\mathcal{H}_0$  denote the set of true null hypotheses, and let  $n_0 = |\mathcal{H}_0|$  be the number of true null hypotheses. Then, let  $i_0$  be the rank of the smallest null  $p$ -value corresponding to a true null hypothesis. This means that the first true null hypothesis is encountered when examining  $p_{(i_0)}$  in the sequential interpretation of Holm's procedure. Now, note that we must have

$$i_0 \leq n - n_0 + 1.$$

This is true by inspection, since there are only  $n - n_0$  false null hypotheses.

Now, consider the situation in which Holm's procedure falsely rejects a true null hypothesis. This happens if and only if the first true null hypothesis is rejected, so we may simply think about when the first true null hypothesis is rejected. This happens precisely when for all  $i \leq i_0$ , we have that  $p_{(i)} \leq \alpha/(n - i + 1)$ , and so in particular  $p_{(i_0)} \leq \alpha/(n - i_0 + 1) \leq \alpha/n_0$ . Hence,

$$\text{FWER} = \mathbb{P}(V \geq 1) = \mathbb{P}(p_{(i_0)} \leq \alpha/n_0) = \mathbb{P}\left(\min_{i \in \mathcal{H}_0} p_i \leq \alpha/n_0\right) \leq \sum_{i \in \mathcal{H}_0} \mathbb{P}(p_i \leq \alpha/n_0) = \alpha.$$

This completes the proof as desired.  $\square$

**Note.** Another perspective on Holm's procedure is as a *closure* of the Bonferroni global test. Though this is an interesting way to derive the procedure and lends itself to the construction of Hochberg's procedure as a more conservative and simpler procedure than the closure of the Simes global test, it is not particularly pertinent to our discussion of multiple comparison procedures, so we will omit a discussion and leave the details to [\[Can\]](#).

### 1.1.3 Hochberg's Procedure

Hochberg's procedure is the final procedure that provides control of the FWER that we present. At a glance, Hochberg's procedure is quite similar to Holm's procedure, as it uses the same modified  $p$ -value cutoffs.

**Procedure 1.1.6** (Hochberg's Procedure). Suppose we are given a collection of  $n$  null hypotheses, denoted  $H_{0,1}, \dots, H_{0,n}$ , with associated  $p$ -values denoted  $p_1, \dots, p_n$  respectively. Fix a desired level  $\alpha$ . Then, order the  $p$ -values so that

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$$

and let  $H_{(1)}, \dots, H_{(n)}$  be the corresponding hypotheses. Then, define

$$k = \max \left\{ 1 \leq i \leq n : p_{(i)} \leq \frac{\alpha}{n - i + 1} \right\}$$

and reject hypotheses  $H_{(1)}, \dots, H_{(k-1)}$ . In particular, if it is the case that

$$p_{(i)} \leq \frac{\alpha}{n - i + 1}, \quad \forall 1 \leq i \leq n,$$

then we reject all  $H_1, \dots, H_n$ .

**Proposition 1.1.7.** Assuming independence of the  $p$ -values, [Procedure 1.1.6](#) controls the FWER at level  $\alpha$ .

*Proof.* Omitted, see [\[Can\]](#) for details. □

**Note.** The sequential way to think about Hochberg's procedure is as follows. We begin by consider  $p_{(n)}$ , and if  $p_{(n)} \leq \alpha/n$ , then reject all hypotheses. Otherwise, consider  $p_{(n-1)}$ , and proceed similarly: if  $p_{(n-1)} \leq \alpha/(n-1)$ , then reject hypotheses  $H_{(1)}, \dots, H_{(n-1)}$ . In a sense, Hochberg's procedure operates opposite Holm's procedure as a *step-up procedure* rather than a *step-down procedure*: this allows for significantly increased power in some cases. For an extreme example, consider the case where all  $p$ -values are precisely  $\alpha$ . Then, Holm's procedure will not reject any hypotheses, whereas Hochberg's procedure will reject all hypotheses.

## 1.2 The False Discovery Rate

In the 1990s, a new metric for error control called the *false discovery rate* (FDR) was introduced. Colloquially, the FDR is a more relaxed metric to use than the FWER, which refers to the probability of there being any false discoveries at all.

**Definition 1.2.1.** The *false discovery proportion* is defined as

$$\text{FDP} = \frac{V}{R \vee 1} \tag{1.2.1}$$

where  $R$  is the total number of rejections we make, while  $V$  is the number of rejections of true null hypotheses (ie. erroneous rejections). Note that  $R$  is an observed value, but  $V$  is not, so the FDP is actually an unobserved random variable. As such, we strive for control of its expectation, and define

$$\text{FDR} = \mathbb{E}[\text{FDP}]. \tag{1.2.2}$$

The question of controlling FWER versus FDR should be answered by using context for what a false discovery means in particular situations. For instance, studies regarding cheating may want to control the



FWER (as false discoveries are serious false accusations), whereas studies regarding gene expression may want to control the FDR (since the consequences of discovering false connections are likely not as severe).

### 1.2.1 The Benjamini-Hochberg Procedure

The most well-known procedure for controlling the FDR is the Benjamini-Hochberg procedure. Indeed, in the original paper, Benjamini and Hochberg proposed control the FDR in lieu of the FWER, and introduced the following procedure that controls the FDR. Note that similar to [Procedure 1.1.6](#), the Benjamini-Hochberg procedure is a *step-up procedure*; the selected cutoff  $k$  is a maximum index in a similar form.

**Procedure 1.2.2.** Suppose we are given a collection of  $n$  null hypotheses, denoted  $H_{0,1}, \dots, H_{0,n}$ , with associated  $p$ -values denoted  $p_1, \dots, p_n$  respectively. Fix a desired level of FDR control,  $q^*$ . Then, order the  $p$ -values so that

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$$

and let  $H_{(1)}, \dots, H_{(n)}$  be the corresponding hypotheses. Then, define

$$k = \max \left\{ 1 \leq i \leq n : p_{(i)} \leq \frac{i}{n} \cdot q^* \right\}$$

and reject all hypotheses  $H_{(1)}, \dots, H_{(k)}$ .

**Proposition 1.2.3.** Assume independence of the  $p$ -values. Then, [Procedure 1.2.2](#) controls the FDR at level  $q^*$ .

Instead of presenting the original proof of FDR control that Benjamini and Hochberg provide in their paper, we will provide a crisper proof from [\[Can\]](#).

*Proof.* Here, we are just rehearsing the original proof presented in [\[Can\]](#). Let  $\mathcal{H}_0$  denote the set of true null hypotheses, and let  $n_0 = |\mathcal{H}_0|$  be the number of true null hypotheses. First, we may assume that  $n_0 \geq 1$  (otherwise the claim is vacuously true). Then, recall that the FDP is defined as

$$\text{FDP} = \frac{V}{R \vee 1}$$

where  $V$  is the number of incorrect rejections and  $R$  is the total number of rejections. Now, for each null  $i \in \mathcal{H}_0$ , define the indicator variable  $V_i = \mathbb{1}_{H_{0,i} \text{ rejected}}$ . This allows us to re-express the FDP as

$$\text{FDP} = \frac{V}{R \vee 1} = \sum_{i \in \mathcal{H}_0} \frac{V_i}{R \vee 1}.$$

Next, note that if we can show  $\mathbb{E}[V_i/(R \vee 1)] = q/n$ , then the rest of the proof follows easily. To show that this is true, first note that we can rewrite

$$\frac{V_i}{R \vee 1} = \sum_{k=1}^n \frac{V_i \mathbb{1}_{\{R=k\}}}{k} \tag{1.2.3}$$

since the sum simply runs over the possible values of the number of rejections,  $R$ ; in addition, note that the above formula still holds when  $R = 0$  since in that case, we have that  $V_i = 0$  by definition.

Next, we will note some properties of the Benjamini-Hochberg procedure, which we will use later in the proof. First, note that when the procedure makes  $k$  rejections, then a hypothesis  $H_{0,i}$  is rejected if and only if  $p_i \leq q^* \cdot k/n$ . As such, we have that

$$V_i = \mathbb{1}_{H_{0,i} \text{ rejected}} = \mathbb{1}_{\{p_i \leq q^* \cdot k/n\}}. \tag{1.2.4}$$

Second, consider the case where  $H_{0,i}$  is rejected, so  $p_i \leq q \cdot k/n$ . Then, let  $R'_i$  denote the number of rejections by the Benjamini-Hochberg procedure if we were to set  $p_i \rightarrow 0$ : when  $H_{0,i}$  is rejected,  $R'_i = R$  since we are simply reordering the first  $k$   $p$ -values, which will not change the number of hypotheses rejected. On the other hand, consider the case where  $H_{0,i}$  is not rejected, so  $p_i > q \cdot k/n$ . Then,  $V_i = 0$ . In either case, we will have that the value of  $V_i \cdot \mathbb{1}_{\{R=k\}} = V_i \cdot \mathbb{1}_{\{R'_i=k\}}$ .

We will now piece the proof together as follows. Instead of looking at the expectation, we will examine a conditional expectation upon all the other  $p$ -values as follows. Let  $\mathcal{F}_i = \{p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_n\}$ . Then,

$$\mathbb{E} \left[ \frac{V_i}{R \vee 1} \middle| \mathcal{F}_i \right] = \sum_{k=1}^n \frac{\mathbb{E} [V_i \mathbb{1}_{\{R=k\}} | \mathcal{F}_i]}{k} = \sum_{k=1}^n \frac{\mathbb{E} [V_i \mathbb{1}_{\{R'_i=k\}} | \mathcal{F}_i]}{k} = \sum_{k=1}^n \frac{\mathbb{E} [\mathbb{1}_{\{p_i \leq q \cdot k/n\}} \mathbb{1}_{\{R'_i=k\}} | \mathcal{F}_i]}{k}, \quad (1.2.5)$$

where the last two equalities used the above observations about the Benjamini-Hochberg procedure to simplify the expression. Next, noting that the  $p$ -values are independent and that  $p_i \sim \text{Unif}[0, 1]$ , we can further simplify that

$$\sum_{k=1}^n \frac{\mathbb{E} [\mathbb{1}_{\{p_i \leq q \cdot k/n\}} \mathbb{1}_{\{R'_i=k\}} | \mathcal{F}_i]}{k} = \sum_{k=1}^n \frac{(q \cdot k/n) \cdot \mathbb{E} [\mathbb{1}_{\{R'_i=k\}} | \mathcal{F}_i]}{k} = \frac{q}{n} \cdot \sum_{k=1}^n \mathbb{1}_{\{R'_i=k\}}. \quad (1.2.6)$$

Above, the first equality comes from separating the expectation via independence and using the distribution of the  $p$ -values, while the second equality is given by the fact that  $\mathbb{1}_{\{R'_i=k\}}$  is deterministic conditional on  $\mathcal{F}_i$ . However, we claim that

$$\sum_{k=1}^n \mathbb{1}_{\{R'_i=k\}} = 1. \quad (1.2.7)$$

This is true as follows. Note that  $R_i \geq 1$ , as when we set  $p_i \rightarrow 0$ , we must make at least one rejection of  $H_{0,i}$  at the least. On the other hand, we know that  $R'_i$  will take on one value between 1 and  $n$ , which means that the sum of the indicators will be exactly 1. Hence, we have shown that

$$\mathbb{E} \left[ \frac{V_i}{R \vee 1} \middle| \mathcal{F}_i \right] = \frac{q}{n}.$$

Then, the law of total expectation allows to write the FDR as

$$\text{FDR} = \mathbb{E}[\text{FDP}] = \sum_{i \in \mathcal{H}_0} \mathbb{E} \left[ \frac{V_i}{R \vee 1} \right] = \sum_{i \in \mathcal{H}_0} \mathbb{E} \left[ \mathbb{E} \left[ \frac{V_i}{R \vee 1} \middle| \mathcal{F}_i \right] \right] = \sum_{i \in \mathcal{H}_0} \frac{q}{n} = \frac{n_0}{n} \cdot q \leq q.$$

This bounds the FDR as desired and completes the proof.  $\square$

further  
discussion  
on BH  
procedure?

## Chapter 2

# Introducing (Model-X) Knockoffs

Recently, Barber and Candès have proposed a new procedure that also controls the FDR for linear Gaussian models in [BC15]. The procedure, known as the *knockoff filter*, creates “knockoff” variables, which have the same covariance structure as the covariates, but are independent of the output, and uses these variables to control the false discovery rate. [CFJL17] then provided a sweeping generalization of the procedure beyond linear models.

not sure how much detail should be given to the background on knockoffs: should just cite the paper and move on?

## Chapter 3

# Multiple Knockoffs

### 3.1 Definition

The concept of multiple knockoffs is a direct extension of the work done in [BC15] and [CFJL17] that we revisited in Chapter 2. We define multiple model-X knockoffs as follows. First, recall the definition of a null covariate.

**Definition 3.1.1.** A variable  $X_j$  is said to be “null” if and only if  $Y$  is independent of  $X_j$  conditionally on the other variables  $X_{-j} = \{X_1, \dots, X_p\} \setminus \{X_j\}$ . The subset of null variables is denoted by  $\mathcal{H}_0$  and we call a variable  $X_j$  “non-null” or relevant if  $j \notin \mathcal{H}_0$ .

**Definition 3.1.2.** Multiple model-X knockoffs for a family of random variables  $X = (X_1, \dots, X_p)$  are a collection of  $n$  families of new random variables  $X^{(i)} = (X_1^{(i)}, \dots, X_p^{(i)})$  constructed satisfying the following two properties, analogous to those of knockoffs: (1) *Exchangeability*. Let  $[[n]]$  denote the set  $\{0, \dots, n\}$ . For each  $1 \leq j \leq p$ , let  $\sigma_j : [[n]] \rightarrow [[n]]$  be any permutation of the indices 0 to  $n$ . Then, letting  $X^{(0)} = X$ ,

$$(X^{(0)}, \dots, X^{(n)}) \stackrel{d}{=} (X_1^{(\sigma_1(0))}, \dots, X_p^{(\sigma_p(0))}, \dots, X_1^{(\sigma_1(n))}, \dots, X_p^{(\sigma_p(n))}) \quad (3.1.1)$$

For convenience, we write

$$\{\sigma_i\}_{i=1}^p (X^{(0)}, \dots, X^{(n)}) \stackrel{\text{def}}{=} (X_1^{(\sigma_1(0))}, \dots, X_p^{(\sigma_p(0))}, \dots, X_1^{(\sigma_1(n))}, \dots, X_p^{(\sigma_p(n))}). \quad (3.1.2)$$

(2) *Conditional Independence*.  $(X^{(1)}, \dots, X^{(n)}) \perp\!\!\!\perp Y \mid X$  if there is a response variable  $Y$ . This property is guaranteed if the  $X^{(i)}$  are constructed without looking at  $Y$ .

Here, the two conditions are analogous to the ones presented in Definition 3.1 of [CFJL17]. In fact, the second condition is identical, while the first only seeks to generalize the idea of pairwise exchangeability.

### 3.2 Exchangeability of Null Covariates and Knockoffs

First, we provide a generalization of Lemma 3.2 in [CFJL17]. In particular, we extend the proof that we can permute null covariates with their knockoffs without changing the joint distribution of  $X$  and the knockoffs  $X^{(i)}$ , conditional on  $Y$ .

omitted detail with rows, add in here?

**Lemma 3.2.1.** Let  $S \subseteq \mathcal{H}_0$  be a subset of nulls. Consider a set of permutations  $\sigma_j : [[n]] \rightarrow [[n]]$  such that if  $j \notin S$ , then  $\sigma_j = \text{id}_{[[n]]}$ . Then,

$$(X^{(0)}, \dots, X^{(n)}) \mid Y \stackrel{d}{=} \{\sigma_i\}_{i=1}^p \left( (X^{(0)}, \dots, X^{(n)}) \mid Y \right).$$

*Proof.* The proof is quite similar to that of the original lemma. Without loss of generality, we can assume that  $S = \{1, \dots, m\}$ . Then, since the marginal distribution of  $Y$  is the same on both sides of the equation, it is equivalent to show that the joint distributions are the same. Then, in the same way as the original lemma, by the exchangeability condition that

$$(X^{(0)}, \dots, X^{(n)}) \stackrel{d}{=} \{\sigma_i\}_{i=1}^p \left( (X^{(0)}, \dots, X^{(n)}) \right),$$

so the only thing we need to show is that

$$Y \mid (X^{(0)}, \dots, X^{(n)}) \stackrel{d}{=} Y \mid \{\sigma_i\}_{i=1}^p \left( (X^{(0)}, \dots, X^{(n)}) \right). \quad (3.2.1)$$

To see this, let  $p_{Y|X}(y|x)$  be the conditional distribution of  $Y$  given  $X$ . Then, note that

$$\begin{aligned} p_{Y|\{\sigma_i\}_{i=1}^p(X^{(0)}, \dots, X^{(n)})}(y|(x^{(0)}, \dots, x^{(n)})) &= p_{Y|(X^{(0)}, \dots, X^{(n)})}(y|\{\sigma_i^{-1}\}_{i=1}^p(x^{(0)}, \dots, x^{(n)})) \\ &= p_{Y|X^{(0)}}(y|x'), \end{aligned}$$

where  $x'_i = x_i^{(\sigma_i^{-1}(0))}$  if  $i \in S$  and  $x'_i = x_i$  otherwise. In particular, the second equality above comes from the fact that  $Y$  is conditionally independent of the knockoffs  $(X^{(1)}, \dots, X^{(n)})$  given  $X$  by definition of multiple knockoffs.

Next, note that we assumed earlier that  $S = \{1, \dots, m\}$  is the subset of nulls. Then, by definition,  $Y$  and  $X_1$  will be conditionally independent given  $X_{2:p}$ , we may further simplify that

$$p_{Y|X^{(0)}}(y|x') = p_{Y|X_{1:p}^{(0)}}(y|x_1^{(\sigma_1^{-1}(0))}, x'_{2:p}) = p_{Y|X_{2:p}^{(0)}}(y|x'_{2:p}) = p_{Y|X_{1,p}^{(0)}}(y|x_1^{(0)}, x'_{2:p})$$

This shows that the conditional distributions of  $Y$  are the same:

$$Y \mid \{\sigma_i\}_{i=1}^p \left( (X^{(0)}, \dots, X^{(n)}) \right) = Y \mid \{\sigma'_i\}_{i=1}^p \left( (X^{(0)}, \dots, X^{(n)}) \right),$$

where  $\sigma'_1 = \text{id}_{[[n]]}$  and  $\sigma'_i = \sigma_i$  for  $i > 1$ . This reduces  $S$  to the case where  $1 \notin S$ , so we may repeat this argument until  $S$  is empty, and this proves the claim.  $\square$

### 3.3 Feature Statistics and $p$ -values

In the original work on knockoffs, statistics denoted  $W_j$  are computed for each of the  $X_j$  based on both  $X_j$  and its knockoff, where a large value of  $W_j$  indicates evidence for significance of the covariate. Furthermore,  $W_j$  is required to satisfy the flip-sign property, which says that swapping  $X_j$  with its knockoff has the effect of reversing the sign of  $W_j$ . These statistics are then used in a sequential process when running the knockoffs procedure.

Returning to multiple knockoffs, we now wish to generalize the concept of the statistic  $W_j$ . To be informal, there are two properties of  $W_j$  that are critical to the knockoffs procedure: the sign of  $W_j$ , which determines the  $p$ -value that is used in the selection process, and the magnitude of  $|W_j|$ , which determines the ordering of the  $W_j$  in the selection process. As such, to generalize to multiple knockoffs, we just need

to ensure that we have generalizations of the ideas of obtaining a  $p$ -value from our statistic and obtaining a magnitude of our statistic.

To be more precise, we may refer to the original knockoffs paper [BC15] and view the knockoffs procedure as a sequential selection procedure. A key idea in the proof of the main theorems, Theorem 1 and Theorem 2, in [BC15] is the conversion of the sign of  $W_j$  to 1-bit  $p$ -values: covariates such that  $W_j > 0$  are assigned a  $p$ -value of 0.5, whereas the covariates such that  $W_j < 0$  are given a  $p$ -value of 1. This conversion allows us to prove the desired statements via Theorem 3 of [BC15], which provides proof of control of FDR for generalized sequential testing procedures, the FSTP and SSTP. Note here that the  $W_j$  statistics for both the covariate and its knockoff are in correspondence with generating statistics  $Z_j$  and  $\tilde{Z}_j$  for the covariate and its knockoff respectively via the formula

$$W_j = f_j(Z_j, \tilde{Z}_j) \quad (3.3.1)$$

where  $f_j$  is an anti-symmetric function.

In the case of multiple knockoffs, instead of thinking about  $W_j$ , we will consider  $Z_j^{(i)}$  for each knockoff  $i$  and retaining the convention that the case of  $i = 0$  is the statistic corresponding to the original covariate. As such, we can generalize the definition of the statistic  $T$  to become

$$T \stackrel{\text{def}}{=} (Z^{(0)}, \dots, Z^{(n)}) = t([X^{(0)}, \dots, X^{(n)}], y) \quad (3.3.2)$$

Here, the  $Z_j^{(k)}$  are intended to measure the significance of the covariate (or knockoff)  $X^{(k)}_j$ .

Now, we must consider the other critical component to the knockoff procedure: the magnitude of  $W_j$ , which determines the order in which we look at the covariates in the sequential selection procedure. Here, similar to the choice of anti-symmetric function we make for  $f_j(Z_j, \tilde{Z}_j) = W_j$  in the original knockoffs case, a choice must be made to determine the “magnitude” of the statistic, denoted  $M_j$ , which can also write as  $g_j(Z_j^{(0)}, \dots, Z_j^{(n)})$ . In addition, noting that  $f_j$  is anti-symmetric in the original case, we will require that  $g_j$  is a symmetric function in its arguments (the analog here is that the magnitude of  $f_j$  would be symmetric).

Now, we present a lemma which is the desired generalization of Lemma 3.3 in [CFJL17].

**Lemma 3.3.1.** *For the null  $X_j$ , where  $j \in \mathcal{H}_0$ , let  $A_{j,(i)}$  denote the  $i$ -th order statistic of  $Z_j^{(0)}, \dots, Z_j^{(n)}$ . Then, conditional on  $(M_1, \dots, M_p)$ ,*

$$\mathbb{P}(Z_j^{(0)} = A_{j,(i)}) = \frac{1}{n+1}, \quad 1 \leq i \leq n+1. \quad (3.3.3)$$

*Proof.*

slightly confused here

Like Lemma 3.3 in [CFJL17], this is a direct consequence of Lemma 3.2.1, and we will provide an analogous proof. From the lemma, we may deduce that  $T \stackrel{d}{=} \{\sigma_i\}_{i=1}^p(T)$  where the  $\sigma_i$  satisfy the condition of the lemma (ie. are non-trivial only for null  $j$ ). Then, by symmetry, the statistic  $Z_j^{(0)}$  must be equally likely to be  $i$ -th largest statistic among the  $Z_j^{(i)}$ , as desired.  $\square$

The above lemma allows us to obtain the analogous version of  $p$ -values for multiple knockoffs. Using the same notation as in Lemma 3.3.1 for the covariate  $X_j$ , we define the  $p$ -value

$$p_j = 1 - \frac{i-1}{n+1}, \quad Z_j^{(0)} = A_{j,(i)}. \quad (3.3.4)$$

Note that the special case where  $n = 1$  simplifies directly into the original knockoffs framework.

### 3.4 The Procedure

Now that we have developed the analog of the desired test statistic in the setting of multiple knockoffs, we may briefly describe the procedure itself. The parallel here is rather direct, and is simply using the machinery that we have developed in the framework of [BC15] under the FSTP and SSTP described there. We will present these results in a similar fashion to [CFJL17].

**Theorem 3.4.1.** *Fix a threshold  $c \in (0, 1)$ . Then, choose a threshold  $\tau > 0$  by setting*

$$\tau = \min \left\{ t > 0 : \frac{\#\{j : M_j \geq t, p_j > c\}}{\#\{j : M_j \geq t, p_j \leq c\}} \leq q \right\} \quad (\text{MultipleKnockoffs})$$

where  $q$  is the target FDR level (or  $\tau = +\infty$  if the set above is empty). Then, the procedure that selects the variables

$$\hat{S} = \{j : M_j \geq \tau, p_j \leq c\}$$

controls the modified FDR defined as

$$\text{mFDR} = \mathbb{E} \left[ \frac{|\{j \in \hat{S} \cap \mathcal{H}_0\}|}{|\hat{S}| + 1/q} \right] \leq q.$$

Similarly, the more conservative procedure given by choosing the threshold  $\tau_+ > 0$  where

$$\tau_+ = \min \left\{ t > 0 : \frac{1 + \#\{j : M_j \geq t, p_j > c\}}{\#\{j : M_j \geq t, p_j \leq c\}} \leq q \right\} \quad (\text{MultipleKnockoffs+})$$

and setting  $\hat{S} = \{j : M_j \geq \tau, p_j \leq c\}$  controls the usual FDR by

$$\mathbb{E} \left[ \frac{|\{j \in \hat{S} \cap \mathcal{H}_0\}|}{|\hat{S}| \vee 1} \right] \leq q.$$

*Proof.* The proof of the theorem above follows from Theorem 3 of [BC15]. In particular, the connection with knockoffs in Section 5.2 of [BC15] explains the connection between knockoffs and the SSTP. Here, the connection is even more explicit: we replace  $|W_j|$  with  $M_j$ , and store the information of the  $p$ -values directly. In particular, from the work that we have done, it is evident that  $p_j \geq \text{Unif}[0, 1]$ , and we assume that they are independent from the non-null  $p$ -values. Finally, note that we are doing something analogous to ordering our covariates in order of decreasing  $M_j$  in the procedure by looking at cutoffs for  $M_j$ . Hence, it is clear that the procedure outlined above is a special case of the SSTP, and we are done.  $\square$

A noteworthy difference between multiple knockoffs and knockoffs is the addition of the cutoff parameter  $c$  for  $p$ -values. Indeed, since knockoffs only using extremely rough 1-bit  $p$ -values, a choice for  $c$  in the SSTP is not particularly enlightening. In particular, selecting any  $c \in (1/2, 1)$  is strictly worse than selecting  $c = 1/2$ , whereas selecting  $c \in (0, 1/2)$  leads to zero power; it is evident that the choice  $c = 1/2$  is optimal.

For multiple knockoffs, however, there is no canonical choice for  $c$  in the SSTP description. For instance, consider the case where  $n = 2$ . Then, two candidates for  $c$  exist:  $1/3$  and  $2/3$  are both possible. Later, we will investigate the effects of choosing different  $c$  on the power of the procedure in various situations for different  $n$ .

# Bibliography

- [BC15] Rina Foygel Barber and Emmanuel Candès. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43, 2015.
- [Can] Emmanuel Candès. Stats 300c: Lectures. <https://statweb.stanford.edu/~candes/stats300c/lectures.html>.
- [CFJL17] Emmanuel Candès, Yingying Fan, Lucas Janson, and Jinchi Lv. Panning for gold: Model-x knock-offs for high-dimensional controlled variable selection. *Journal of the Royal Statistical Society Series B*, 2017.
- [Dun59] Olive Jean Dunn. Estimation of the medians for dependent variables. *The Annals of Mathematical Statistics*, 30, 1959.
- [Hol79] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 1979.