The FDR-Linking Theorem

Weijie J. Su*

December 21, 2018

Abstract

This paper introduces the `FDR-linking` theorem, a novel technique for understanding *non-asymptotic* FDR control of the Benjamini–Hochberg (BH) procedure under arbitrary dependence of the p-values. This theorem offers a principled and flexible approach to linking all p-values and the null p-values from the FDR control perspective, suggesting a profound implication that, to a large extent, the FDR of the BH procedure relies mostly on the null p-values. To illustrate the use of this theorem, we propose a new type of dependence only concerning the null p-values, which, while strictly *relaxing* the state-of-the-art PRDS dependence (Benjamini and Yekutieli, 2001), ensures the FDR of the BH procedure below a level that is independent of the number of hypotheses. This level is, furthermore, shown to be optimal under this new dependence structure. Next, we present a concept referred to as *FDR consistency* that is weaker but more amenable than FDR control, and the `FDR-linking` theorem shows that FDR consistency is completely determined by the joint distribution of the null p-values, thereby reducing the analysis of this new concept to the global null case. Finally, this theorem is used to obtain a sharp FDR bound under arbitrary dependence, which improves the log-correction FDR bound (Benjamini and Yekutieli, 2001) in certain regimes.

Keywords: False discovery rate (FDR), Benjamini–Hochberg procedure, `FDR-linking` theorem, positive regression dependence within nulls (PRDN), FDR consistency, Simes method, compliance, Bonferroni-masked adversary,

1 Introduction

In 1995, Benjamini and Hochberg introduced a new procedure, henceforth referred to as the BH procedure, to control a type I error called the false discovery rate (FDR) in large-scale multiple testing problems [BH95]. To describe their procedure, order the n observed p-values $p_1, \ldots, p_n$ from the most significant to the least as $p_{(1)} \leq \cdots \leq p_{(n)}$. Taking $0 < \alpha < 1$ as the nominal level throughout the paper, the BH procedure finds R that is the last time $p_{(j)}$ is below the critical value $\alpha j/n$, that is,

$$R = \max\left\{1 \leq j \leq n : p_{(j)} \leq \alpha j/n\right\},$$

with the convention $\max \emptyset = 0$. The BH procedure rejects the R hypotheses with their corresponding p-values $p_j \leq \alpha R/n$. Letting V denote the number of falsely rejected null

---

*Department of Statistics, the University of Pennsylvania. Email: `suw@wharton.upenn.edu`. Homepage: `http://stat.wharton.upenn.edu/~suw/`.

p-values[1], [BH95] proves that their procedure controls the FDR in the sense that

$$\text{FDR} := \mathbb{E}\left[\frac{V}{\max\{R, 1\}}\right] \leq \alpha$$

under independence of the p-values[2].

After two decades of vigorous development, today the BH procedure and many of its variants have been extensively applied in high-throughput sciences such as genomics, where the FDR is increasingly recognized as the right error rate to control when simultaneously testing tens of thousands of, for example, gene expression levels (see [TTC01, GLN02, ST03] for exemplary applications and [Ben10] for a review of the FDR concept).

In stark contrast to the enormous popularity of the FDR criterion and methodologies, however, it is generally *unclear* whether the BH procedure controls the FDR given a joint distribution of the p-values, though there are continued efforts to bridge this unsettling gap. To put it into perspective, the BH procedure has been proved to control the FDR under independence [BH95] or certain positive dependence [BY01] of the p-values. Surprisingly, *no* dependence structure is known to yield *meaningful* (non-asymptotic) FDR control while also being conceptually different from the two rather stringent dependence structures. That said, there is a long line of work concerning asymptotic FDR control of the BH procedure under various dependence assumptions (see Section 1.4 for literature review).

This paper considers *non-asymptotic* FDR control of the BH procedure under dependence of the p-values. Roughly speaking, our approach begins by decomposing the full dependence structure into the *null dependence* (joint distribution of the null p-values) and the *null-non-null dependence* (conditional distribution of the non-null p-values given the null p-values). One might imagine that each of the two distributions would have an arbitrarily large effect on the FDR. This is, surprisingly, *not true*. As shown next, our main finding is that the FDR of the BH procedure admits an upper bound that is determined *only* by the null dependence.

To state our main theorem, let $\pi_0 := n_0/n$ denote the true null proportion, where n_0 is the number of null p-values. For any $0 < x < 1$, write $\text{FDR}_0(x)$ for the false discovery rate of the BH procedure at level x supplied with the null p-values[3]. As such, the $\text{FDR}_0(x)$ is entirely determined by the null dependence.

Theorem 1 (The `FDR-linking` theorem). *Under arbitrary dependence of the p-values, the FDR of the BH procedure at level α satisfies*

$$\text{FDR} \leq \pi_0\alpha + \pi_0\alpha \int_{\pi_0\alpha}^{1} \frac{\text{FDR}_0(x)}{x^2}\mathrm{d}x. \tag{1.1}$$

This theorem makes a link from the nulls to the full set of the p-values from the FDR control standpoint, hence the name the `FDR-linking` theorem. As an attractive feature,

---

[1]A p-value is null if its null hypothesis is true, in which case the p-value is stochastically larger than or equal to the uniform variable on $(0, 1)$.

[2]Formally, [BH95] requires that (1) the null p-values are jointly independent (2) and, furthermore, are independent of the non-null p-values. The bound can be strengthened to $\text{FDR} \leq \pi_0\alpha$ [BH95].

[3]To clear up any confusion, we remark that the BH procedure applied to the nulls uses the critical values xj/n_0 for $j = 1, \ldots, n_0$.

Theorem 1 holds *unconditionally*, without requiring any distributional assumptions of the p-values. The FDR bound in the FDR-link inequality (1.1) is a non-decreasing function of $\pi_0 \alpha$ (see Proposition 2.2 in Section 2.1) and, therefore, a slightly weaker but simpler form of this result is

$$\text{FDR} \leq \alpha + \alpha \int_\alpha^1 \frac{\text{FDR}_0(x)}{x^2} \mathrm{d}x.$$

As a useful fact, $\text{FDR}_0(x)$ is just the type I error of the Simes method [Sim86] at level x on the nulls (see Lemma 2.1 in Section 2.1).

Importantly, Theorem 1 implies that the null dependence is *central* to FDR control of the BH procedure. To illustrate this point, recognize that while the FDR presumably depends on both the null dependence and the null-non-null dependence, the upper bound in (1.1) is solely determined by the null dependence, no matter how *adversarial/unfavorable* the dependence between the nulls and non-nulls is for FDR control. Put differently, the non-null p-values have a bounded effect on the FDR as opposed to the profound effect of the null p-values. Indeed, the nulls in the worst case $(\text{FDR}_0(x) = 1)$ would turn the FDR bound in (1.1) into

$$\pi_0 \alpha + \pi_0 \alpha \int_{\pi_0 \alpha}^1 \frac{1}{x^2} \mathrm{d}x = 1,$$

which is possibly the largest and least useful (see Section 4 for the feasibility of $\text{FDR}_0(x) = 1$).

The FDR bound in Theorem 1 is *optimal* for certain null dependence in the sense that (1.1) can reduce to an equality (see Theorem 4). To obtain an upper bound on the FDR, therefore, the `FDR-linking` theorem suggests focusing on $\text{FDR}_0(x)$ on the nulls, at least as a worthwhile starting point. This flexibility is particularly useful in the case where the null dependence can be proven to yield a small value of $\text{FDR}_0(x)$, which will be demonstrated through two applications of this theorem in Section 1.1 and Section 1.2, respectively.

For completeness, Theorem 1 holds for all compliant procedures. In [DSZ18], a multiple testing procedure is said to be compliant at level α if, denoting by R the number of rejections, every rejected p-value p_j satisfies

$$p_j \leq \frac{\alpha R}{n}.$$

This condition is a special instance of the self-consistency condition first proposed in [BR08], where the threshold $\alpha R/n$ is replaced by a general function of R and prior information about hypotheses. As is clear, compliant procedures include the step-down BH procedure, the generalized step-up-step-down procedures [TLD98, Sar02] and, importantly, the (step-up, the one in Theorem 1) BH procedure, which is the most powerful in this family of procedures in the sense that the rejection set of any compliant procedure is a subset of that of the BH procedure. That said, although being the most powerful, the BH procedure in general does *not* maximize the FDR among compliant procedures (see discussion in Section 2.2). Interestingly, compliant procedures maintain the `FDR-linking` property as shown in Theorem 2, of which Theorem 1 is a corollary.

Theorem 2 (The `FDR-linking` theorem for compliant procedures). *Theorem 1 holds for any compliant procedures.*

To avoid any confusion, the FDR in (1.1) in the context of Theorem 2 corresponds to the compliant procedure at level α, while the $\text{FDR}_0(x)$ is for the BH procedure supplied with the nulls.

1.1 Application: FDR control under a new condition

As noted earlier, the BH procedure ensures FDR $\leq \alpha$ if the null p-values exhibit certain positive dependence and, furthermore, are positively dependent on the non-null p-values[BY01]. This type of dependence of the p-values is referred to as the *positive regression dependence on a subset* property, or PRDS for short. To formally define this property, let $\mathcal{N}_0$ denote the set of nulls and call $D \subset [0,1]^n$ an increasing set if $x \in D$ and $x \preceq y \preceq 1$ (in the coordinate-wise sense) together imply $y \in D$.

Definition 1.1 (PRDS[4], [BY01])**.** A set of p-values $(p_1, \ldots, p_n)$ is said to satisfy the PRDS property, if for any increasing set $D \subset [0,1]^n$ and each null index $i \in \mathcal{N}_0$, the probability $\mathbb{P}((p_1, \ldots, p_n) \in D | p_i \leq t)$ is non-decreasing in $t \in (0,1]$.

In relating to the literature, this property is implied by a condition termed the multivariate total positivity of order 2 (MTP$_2$) [KR80]. The PRDS property is perhaps the most popular condition used for proof of FDR control of BH-type procedures (see, for example, the use of PRDS in super-uniformity lemmas for proving FDR control [BR08, BR17, RBWJ17]). Although having relaxed the independence assumption in [BH95], the PRDS remains to assume certain positive dependence between the nulls and non-nulls. This feature is *undesirable* because, while some information of the null distribution is often available in applications, we typically lack knowledge of how the non-nulls depend on the nulls.

Going back to the `FDR-linking` theorem, however, we learn that the FDR of any compliant procedure is largely contingent upon the null dependence. In this spirit, we ask if the PRDS property can be replaced by a more general condition. To address this need, we propose the following definition for FDR control, where, for convenience, a set of p-values is said to satisfy the PRD property if all p-values are null and obey the PRDS property.

Definition 1.2 (PRDN)**.** A set of p-values is said to satisfy the positive regression dependence within nulls (PRDN) if the set of all null p-values satisfies the PRD property.

Put differently, the PRDN property boils down to the monotone non-decreasing of $\mathbb{P}(p_{\mathcal{N}_0} \in D_{\mathcal{N}_0} | p_i \leq t)$ as a function $t \in (0,1]$ for any $i \in \mathcal{N}_0$ and any increasing subset $D_{\mathcal{N}_0} \subset [0,1]^{|\mathcal{N}_0|}$. Implied by the PRDS property, this new property imposes distributional assumptions *only* on the null p-values. As such, the null-non-null dependence can be arbitrary in the PRDN property.

For example, consider an n-dimensional Gaussian vector $X \sim \mathcal{N}(\mu, \Sigma)$ with known Σ for testing $\mu_i = 0$ against $\mu_i > 0, i = 1, \ldots, n$. The one-sided p-values for this testing problem obey the PRDN property if $\Sigma_{ij} \geq 0$ for all $i, j \in \mathcal{N}_0 = \{1 \leq l \leq n : \mu_l = 0\}$. In contrast, the PRDS property additionally requires the nonnegativity of Σ_{ij} for all $i \in \mathcal{N}_0$ and $j \notin \mathcal{N}_0$ [BY01]. See the Appendix for a proof of this fact and an PRDN example of two-sided normal p-values [KR81].

---

[4]In [BY01], the PRDS property is defined for conditioning indices from any subset of all hypotheses. But the subset is almost always the nulls in the use of this property. Recognizing this fact, we assume the null set $\mathcal{N}_0$ in the definition for simplicity.

As shown below, any compliant procedure approximately controls the FDR for p-values satisfying this new property.

Theorem 3. *Assume that the p-values satisfy the PRDN property. Then, any compliant procedure at level α obeys*

$$\text{FDR} \le \alpha + \alpha \log \frac{1}{\alpha}. \tag{1.2}$$

In particular, this FDR bound applies to the BH procedure.

The proof of this theorem follows from a simple application of the `FDR-linking` theorem in conjunction with the fact $\text{FDR}_0(x) \le x$ (Lemma 2.1 in Section 2) under the PRDN property. In fact, the proof shows that Theorem 3 can be strengthened to

$$\text{FDR} \le \pi_0 \alpha + \pi_0 \alpha \log \frac{1}{\pi_0 \alpha},$$

which implies (1.2) since $\alpha + \alpha \log \frac{1}{\alpha}$ is increasing in $\alpha \in (0, 1]$.

Theorem 3 is the first result that maintains the FDR of the (original) BH procedure at a *meaningful* level under a condition that is *strictly less* stringent than PRDS. By "meaningful," we mean that the FDR bound (1.2) can be set to 10%, say, by taking the nominal level $\alpha = 2.05\%$ regardless of n, as opposed to the log-correction bound [BY01], which becomes too large to be useful as n increases (see Section 4). The FDR bound (1.2) and particularly the logarithmic factor $\log \frac{1}{\alpha}$ *cannot* be improved under the PRDN property, as seen from the lower bound below.

Theorem 4. *Fix any $\epsilon > 0$. If α is sufficiently small (depending on ϵ) and n is sufficiently large (depending on both ϵ and α), then the BH procedure applied to certain PRDN p-values satisfies*

$$\text{FDR} > (1 - \epsilon) \left(\alpha + \alpha \log \frac{1}{\alpha} \right).$$

The lower bound on the FDR is attained in an example where the null p-values are i.i.d. uniform random variables on $(0, 1)$, with certain least favorable non-null p-values constructed from an adversarial standpoint.

For more elaboration on this application of the `FDR-linking` theorem, see Section 2. This section also proves the `FDR-linking` theorem.

1.2 Application: FDR consistency

To further explore the use of the `FDR-linking` theorem, we introduce *FDR consistency*, which is a relaxation of FDR control in a certain sense. This new concept is inspired in part by Theorem 3, from which it follows that the FDR of the BH procedure working on any PRDN p-values converges to 0 as the nominal level $\alpha \to 0$ by recognizing the fact

$$\lim_{\alpha \to 0} \alpha + \alpha \log \frac{1}{\alpha} = 0.$$

The convergence is *uniform*, regardless of the specific distribution of the PRDN p-values and, in particular, the number of p-values.

To formulate the definition of FDR consistency, let $\boldsymbol{\mathcal{P}}$ denote a class of p-value dependence structures. For example, $\boldsymbol{\mathcal{P}}$ can be all PRDN distributions on $[0,1]^n$ for $n = 2, 4, 6, \ldots$. Next, for any p-value distribution element $P \in \boldsymbol{\mathcal{P}}$, let $\mathrm{FDR}(\alpha; P)$ denote the FDR of the BH procedure at level α supplied with p-values sampled from the distribution P.

Definition 1.3 (FDR consistency). A dependence class $\boldsymbol{\mathcal{P}}$ is said to be FDR-consistent if

$$\limsup_{\alpha \to 0} \; \sup_{P \in \boldsymbol{\mathcal{P}}} \mathrm{FDR}(\alpha; P) = 0.$$

Our discussion above immediately reveals that the following result is true.

Corollary 1.4. *All probability distributions of p-values satisfying the PRDN property form an FDR-consistent dependence class.*

In words, instead of specifying the exact FDR level, FDR consistency only requires the FDR to tend to zero at a rate uniformly over all distribution elements in the class. In particular, FDR control at the nominal level (or up to a multiplicative constant) implies FDR consistency. This seemingly weak concept is, however, by no means trivial as there exists certain p-value distribution that violates FDR consistency (see Section 4). To appreciate this new concept, recognize that to ensure the FDR below a target level over an FDR-consistent class $\boldsymbol{\mathcal{P}}$, we only need to pick a certain nominal level for the BH procedure regardless of the number of p-values, albeit perhaps small. As such, the BH procedure controls the FDR at a "meaningful" level uniformly over the dependence class $\boldsymbol{\mathcal{P}}$.

In light of the `FDR-linking` theorem, a natural question is whether FDR consistency of the null dependence implies any useful properties of FDR control on all p-values. Remarkably, this is true: the FDR consistency is entirely a property of the null p-values. Loosely speaking, if the null dependence structures induced by a class $\boldsymbol{\mathcal{P}}$ form an FDR-consistent class, then the original class $\boldsymbol{\mathcal{P}}$ is ensured to be FDR-consistent.

Now we formally state this result. For any p-value distribution P in $[0,1]^n$ for some n, let P_0 be the induced probability distribution of the null components of P in $[0,1]^{n_0}$ for some $n_0 \le n$. Moving on to a dependence class $\boldsymbol{\mathcal{P}}$, we define the induced class

$$\boldsymbol{\mathcal{P}}_0 := \{P_0 : P \in \boldsymbol{\mathcal{P}}\},$$

that is, the class of all null dependence induced by $\boldsymbol{\mathcal{P}}$.

Theorem 5. *If $\boldsymbol{\mathcal{P}}_0$ is FDR-consistent, then $\boldsymbol{\mathcal{P}}$ must also be FDR-consistent.*

This theorem suggests that to prove FDR consistency, it is sufficient to only consider the null p-values. This profound implication significantly reduces the problem complexity. For comparison, FDR control in the strict sense does not enjoy such desired complexity reduction because, as noted in Theorem 4, a loss of a multiplicative factor is unavoidable in FDR control. Perhaps this is part of the reason why proving exact FDR control under general dependence is notoriously challenging. In light of the above, FDR consistency might serve as a target for future theoretical FDR research, at least being a reasonable starting point for proving exact FDR control.

In Section 3, we prove Theorem 5 using the `FDR-linking` theorem and present several FDR-consistent dependence classes.

1.3 Overview of other contributions

In addition to the two aforementioned contributions, this paper applies the `FDR-linking` theorem to arbitrary dependence in Section 4 and considers certain mildly adversarial non-null p-values for FDR control in Section 5. In short, Section 4 uses the log-correction bound under the global null [BY01] for $\text{FDR}_0(x)$ in Theorem 2, which, together with the `FDR-linking` theorem, allows us to obtain state-of-the-art FDR guarantee under arbitrary dependence in certain regimes. In Section 5, we ask if the logarithmic factor in Theorem 4 can be dropped by imposing some constraints on the null-non-null dependence and, interestingly, this is true if the non-null p-values are constructed without knowing the smallest null p-value.

1.4 Related work

We give a brief review of the literature concerning FDR control of the BH procedure. In [Yek08], FDR control under non-PRDS dependence is examined through several examples in the non-asymptotic setting. In addition to introducing the PRDS property, [BY01] shows that the FDR is always bounded by α multiplied by a factor about $\log n$ under arbitrary dependence (the so-called log-correction). This correction can be extended to incorporate prior knowledge such as weights of hypotheses via shape functions [BR08, RBWJ17]. A dependence property called PRDSS is proposed in [BR08], which generalizes the PRDS property but is applied to a step-down procedure related to the Benjamini–Liu procedure [BL99]. In [HJ15], FDR control is ensured under reverse martingale dependence, but this dependence is mainly motivated by the convenience of the martingale technique for proving FDR control [STS04]. Recently, [DSZ18, KR18] demonstrate that the FDR is maintained at a reasonable level as long as the nulls are independent.

An integral part of the `FDR-linking` theorem is FDR control under the global null. This is equivalent to the validity of the Simes method. This method is known to control the type I error under many types of positive dependence structures [HR95, SC96, SC97, Sar98], for example, under the PRD property [BY01]. Convincing evidence from [Rød06, FRS17] suggests that the Simes method is likely to be approximately valid in a broader setting except for certain pathological dependence structures.

On a less related note, much more effort has been devoted to understanding asymptotic FDR control under dependence, using tools from empirical processes, factor models, and stochastic processes, among others [YB99, FR01, STS04, GW04, FDR07, Far07, RSW08, Wu08, RV11, FHG12, LLS17]. Under (strong) dependence, the FDP is well-recognized to exhibit high variability [Owe05, FZ06, Efr10, SL11] and, accordingly, many methods have been developed to offer more reliable FDR control in the asymptotic setting [Efr07, SC09, FKC09, BvdBS⁺15, FH17, FKSZ18].

2 FDR Control Under PRDN

In this section, we detail the first application of the `FDR-linking` theorem introduced in Section 1.1, proving Theorems 3 and 4 in Sections 2.1 and 2.2, respectively. Notably, the treatment of Section 2.1 is unusual in that we first prove the `FDR-linking` theorem and Theorem 3 in fact follows as a corollary.

2.1 Proving the `FDR-linking` theorem

We present an upper bound on the false discovery proportion (FDP)

$$\frac{V}{\max\{R, 1\}}$$

for compliant procedures, including the BH procedure. Denote by $p_1^{\text{null}}, \ldots, p_{n_0}^{\text{null}}$ the null p-values and thus $V \leq n_0$ records the number of rejections made from the n_0 null p-values[5]. Let $p_{(1)}^{\text{null}} \leq p_{(2)}^{\text{null}} \leq \cdots \leq p_{(n_0)}^{\text{null}}$ be the ordered null p-values. As will be shown later, any compliant procedure must satisfy

$$\text{FDP} \leq \min\left\{ \max_{1 \leq j \leq n_0} \frac{j}{\lceil np_{(j)}^{\text{null}}/\alpha \rceil}, 1 \right\}, \tag{2.1}$$

where the ceiling function $\lceil x \rceil$ maps x to the least integer that is greater than or equal to x.

This bound on the FDP serves as the basis for proving Theorem 2. To verify this simple inequality, our approach is to take an adversarial viewpoint as this will shed light on the development of the remainder of the paper. Fixing the null p-values, we aim to maximize the FDP (consequently maximizing FDR) by seeking the least favorable scenario of the non-null p-values for FDR control. To this end, consider a thought experiment where an adversary knows which are the n_0 null p-values and can adversely alter the non-null p-values to maximize the FDP using a compliant procedure. By the definition of V, there must be one (falsely) rejected p-value that is no smaller than $p_{(V)}^{\text{null}}$. This fact, together with the compliance property of the procedure, ensures that $p_{(V)}^{\text{null}} \leq \alpha R/n$ (recall that R is the total number of rejections). Rearranging this inequality yields

$$R \geq \left\lceil np_{(V)}^{\text{null}}/\alpha \right\rceil.$$

Therefore, if $V \geq 1$, we get

$$\text{FDP} = \frac{V}{\max\{R, 1\}} = \frac{V}{R} \leq \frac{V}{\lceil np_{(V)}^{\text{null}}/\alpha \rceil},$$

from which (2.1) follows since $1 \leq V \leq n_0$ and FDP is always no greater than 1. As is clear, the maximum FDP can be achieved if the adversary sets $\lceil np_{(j^\star)}^{\text{null}}/\alpha \rceil - j^\star$ non-null p-values to 0 and the rest to 1 (suppose $n_1 = n - n_0$ is sufficiently large), where $j^\star$ is the index maximizing (2.1). Note that (2.1) also holds for $V = 0$ because FDP $= 0$ in this case.

Next, we make a connection between (2.1) and the Simes method. Write

$$\max_{1 \leq j \leq n_0} \frac{j}{\lceil np_{(j)}^{\text{null}}/\alpha \rceil} \leq \max_{1 \leq j \leq n_0} \frac{\alpha j}{np_{(j)}^{\text{null}}} = \frac{\pi_0 \alpha}{\min_{1 \leq j \leq n_0} n_0 p_{(j)}^{\text{null}}/j}, \tag{2.2}$$

---

[5]Throughout the paper, we assume the number n_0 of null hypotheses are nonzero. Otherwise, FDP is always 0 by definition.

where recall that $\pi_0 = n_0/n$. The random variable

$$\min_{1 \leq j \leq n_0} \frac{n_0 p_{(j)}^{\text{null}}}{j} \tag{2.3}$$

is precisely the Simes p-value restricted to the nulls. Let $F(x)$ denote its CDF. A closer look reveals that, for any $0 < x < 1$, $F(x)$ is precisely the false discovery rate of the BH procedure at level x when restricted to the null p-values $p_1^{\text{null}}, \ldots, p_{n_0}^{\text{null}}$. In brief, when applied to the null p-values, the Simes method at level x rejects the global null hypothesis if and only if the Simes p-value $\min_{1 \leq j \leq n_0} n_0 p_{(j)}^{\text{null}}/j \leq x$. In the case of rejection, it must hold that $p_{(j)}^{\text{null}} \leq xj/n_0$ for some j, meaning that the BH procedure applied to the nulls rejects at least one p-value, in which case the FDP is equal to 1. This relationship between the two methods ensures

$$F(x) = \mathbb{P}\left(\min_{1 \leq j \leq n_0} n_0 p_{(j)}^{\text{null}}/j \leq x\right) = \text{FDR}_0(x) \tag{2.4}$$

for $0 < x < 1$. Taken together, the discussion above yields the following proof.

Proof of Theorem 2. Making use of (2.1) and (2.2), we get

$$
\begin{aligned}
\text{FDR} &\leq \mathbb{E}\left[\min\left\{\frac{\pi_0 \alpha}{\min_{1 \leq j \leq n_0} n_0 p_{(j)}^{\text{null}}/j}, 1\right\}\right] \\
&= \mathbb{P}\left(\min_{1 \leq j \leq n_0} n_0 p_{(j)}^{\text{null}}/j \leq \pi_0 \alpha\right) + \mathbb{E}\left[\frac{\pi_0 \alpha}{\min_{1 \leq j \leq n_0} n_0 p_{(j)}^{\text{null}}/j}; \min_{1 \leq j \leq n_0} n_0 p_{(j)}^{\text{null}}/j > \pi_0 \alpha\right] \\
&= F(\pi_0 \alpha) + \int_{\pi_0 \alpha}^{1} \frac{\pi_0 \alpha}{x} \mathrm{d}F(x),
\end{aligned}
\tag{2.5}
$$

where the integral is in the sense of the Riemann–Stieltjes integration by recognizing that the integrator F is càdlàg. Using integration by parts, we proceed with the simplification of (2.5) as

$$
\begin{aligned}
F(\pi_0 \alpha) + \int_{\pi_0 \alpha}^{1} \frac{\pi_0 \alpha}{x} \mathrm{d}F(x) &= F(\pi_0 \alpha) + \pi_0 \alpha F(1) - F(\pi_0 \alpha) + \pi_0 \alpha \int_{\pi_0 \alpha}^{1} \frac{F(x)}{x^2} \mathrm{d}x \\
&= \pi_0 \alpha + \pi_0 \alpha \int_{\pi_0 \alpha}^{1} \frac{F(x)}{x^2} \mathrm{d}x \\
&= \pi_0 \alpha + \pi_0 \alpha \int_{\pi_0 \alpha}^{1} \frac{\text{FDR}_0(x)}{x^2} \mathrm{d}x.
\end{aligned}
$$

This completes the proof for any compliance procedures, including the BH procedure. $\square$

Next, we turn to Theorem 3. We have the following lemma concerning p-values that are PRDN distributed.

Lemma 2.1. *Assume that the p-values satisfy the PRDN property. Then, the Simes p-value (2.3) on the nulls is stochastically larger than or equal to the uniform random variable on $(0, 1)$.*

9

This lemma is equivalent to

$$\mathbb{P}\left(\min_{1 \leq j \leq n_0} n_0 p_{(j)}^{\text{null}}/j \leq x\right) \leq x$$

for $0 < x < 1$. This is true by using (2.4) and the fact $\text{FDR}_0(x) \leq x$ ensured by [BY01]. As an aside, in the case of i.i.d. uniform null p-values, $\min_{1 \leq j \leq n_0} n_0 p_{(j)}^{\text{null}}/j$ is *exactly* uniformly distributed on $(0, 1)$ [Sim86].

Applying Theorem 2, any compliant procedure with PRDN p-values gives

$$\begin{aligned}
\text{FDR} &\leq \pi_0 \alpha + \pi_0 \alpha \int_{\pi_0 \alpha}^{1} \frac{x}{x^2} \mathrm{d}x \\
&= \pi_0 \alpha + \pi_0 \alpha \log \frac{1}{\pi_0 \alpha} \\
&\leq \alpha + \alpha \log \frac{1}{\alpha}.
\end{aligned}$$

This proves Theorem 3. It should be emphasized that the analysis and reasoning above apply to not only the BH procedure but also any compliant procedures. More broadly, if the BH procedure controls the FDR up to a constant c on the nulls or, equivalently, $\text{FDR}_0(x) \leq cx$, then the FDR-linking theorem concludes that

$$\text{FDR} \leq \pi_0 \alpha + c\pi_0 \alpha \log \frac{1}{\pi_0 \alpha}$$

for compliant procedures, roughly paying a factor of $\log \frac{1}{\alpha}$ in FDR control due to the least favorable non-nulls.

As additional background, we informally explain why the logarithmic factor $\log \frac{1}{\alpha}$ appears in the FDR bound in Theorem 3, though a rigorous treatment will be given in Section 2.2. Let all null p-values be i.i.d. uniform variables on $(0, 1)$ and all non-null p-values be 0 and, additionally, assume $n - n_0$ is sufficiently large. Consider a procedure that rejects $R - 1$ non-null p-values and the smallest null p-value $p_{(1)}^{\text{null}}$, where $R = \lceil np_{(1)}^{\text{null}}/\alpha \rceil$. This procedure is compliant. To see how the term $\alpha \log(1/\alpha)$ appears, first note that

$$\text{FDP} = \frac{1}{\lceil np_{(1)}^{\text{null}}/\alpha \rceil} \geq \frac{10}{11} \cdot \frac{\alpha}{np_{(1)}^{\text{null}}}$$

in the case of $np_{(1)}^{\text{null}}/\alpha \geq 10$. If $n_0 \approx n$ and n is large, $np_{(1)}^{\text{null}}$ is approximately distributed as an exponential random variable with mean 1. For sufficiently small α, the logarithmic

term $\alpha \log(1/\alpha)$ appears in

$$\mathbb{E}\left[\frac{10}{11} \cdot \frac{\alpha}{np_{(1)}^{\text{null}}}; np_{(1)}^{\text{null}}/\alpha \geq 10\right] \approx \mathbb{E}\left[\frac{10}{11} \cdot \frac{\alpha}{\text{Exp}(1)}; \text{Exp}(1) \geq 10\alpha\right]$$

$$= \frac{10\alpha}{11} \int_{10\alpha}^{\infty} \frac{1}{x} e^{-x} \, dx$$

$$= \frac{10\alpha}{11} \int_{10\alpha}^{\log \frac{12}{11}} \frac{1}{x} e^{-x} \, dx + \frac{10\alpha}{11} \int_{\log \frac{12}{11}}^{\infty} \frac{1}{x} e^{-x} \, dx \qquad (2.6)$$

$$\geq \frac{10\alpha}{11} \int_{10\alpha}^{\log \frac{12}{11}} \frac{1}{x} \frac{11}{12} \, dx + O(\alpha)$$

$$= \frac{5}{6} \cdot \alpha \log \frac{1}{\alpha} + O(\alpha).$$

The factor $5/6$ above can be made arbitrarily close to 1. As a crucial fact, the analysis above rests on knowing the most significant null p-value $p_{(1)}^{\text{null}}$ (see Section 5).

In passing, we discuss an important property of the FDR bound in Theorem 1. Intuitively, one would expect that the FDR bound given in Theorem 1 increases with the nominal level α. This is indeed true as we show next.

Proposition 2.2. *The function*

$$t + t \int_{t}^{1} \frac{\text{FDR}_0(x)}{x^2} \, dx$$

is non-decreasing in $0 < t < 1$.

Proof of Proposition 2.2. Write

$$f(t) = t + t \int_{t}^{1} \frac{\text{FDR}_0(x)}{x^2} \, dx.$$

This function is continuous. Thus, it suffices to show that the right derivative of f exists and is nonnegative everywhere. Interpreted as the CDF of the Simes p-value (2.3), $\text{FDR}_0(x)$ must be càdlàg. The càdlàg condition allows us to evaluate

$$\partial_+ f(t) = 1 + \int_{t}^{1} \frac{\text{FDR}_0(x)}{x^2} \, dx - \frac{\text{FDR}_0(t)}{t}$$

$$\geq 1 + \int_{t}^{1} \frac{\text{FDR}_0(t)}{x^2} \, dx - \frac{\text{FDR}_0(t)}{t}$$

$$= 1 + \text{FDR}_0(t)\left(\frac{1}{t} - 1\right) - \frac{\text{FDR}_0(t)}{t}$$

$$= 1 - \text{FDR}_0(t)$$

$$\geq 0.$$

For completeness, note that the first equality rests on the fact that $\text{FDR}_0(t+) = \text{FDR}_0(t)$, which is, again, ensured by the càdlàg condition. Hence, the proof is complete.

$\square$

2.2 Optimality

In this section, we state and prove Theorem 6, a stronger version of Theorem 4. In what follows, let $\{\xi_j\}_{j=1}^{\infty}$ be i.i.d. exponential random variables with mean 1.

Theorem 6 (A stronger version of Theorem 4). *For any $0 < \alpha < 1$, denote by*

$$D_\alpha := \mathbb{E}\left[\min\left\{\max_{1 \leq j < \infty} \frac{j}{\lceil (\xi_1 + \cdots + \xi_j)/\alpha \rceil}, 1\right\}\right].$$

For any $\nu > 0$, there exist certain PRDN p-values on which the BH procedure gives

$$\text{FDR} > D_\alpha - \nu.$$

To be complete, the number n of p-values in the example of this theorem should exceed some number depending on both α and ν. In relating to Theorem 3, we immediately conclude that

$$D_\alpha \leq \alpha + \alpha \log \frac{1}{\alpha}$$

for all $0 < \alpha < 1$. Alternatively, this inequality can be gleaned from the fact that

$$
\mathbb{E}\left[\min\left\{\max_{1 \leq j < \infty} \frac{j}{\lceil (\xi_1 + \cdots + \xi_j)/\alpha \rceil}, 1\right\}\right] \leq \mathbb{E}\left[\min\left\{\max_{1 \leq j < \infty} \frac{j}{(\xi_1 + \cdots + \xi_j)/\alpha}, 1\right\}\right]
$$
$$
= \mathbb{E}\left[\min\left\{\frac{\alpha}{\min_{1 \leq j < \infty}(\xi_1 + \cdots + \xi_j)/j}, 1\right\}\right]
$$
$$
= \alpha + \alpha \log \frac{1}{\alpha}.
$$

The last equality recognizes that the distribution of $\min_{1 \leq j < \infty}(\xi_1 + \cdots + \xi_j)/j$ is uniform on $(0, 1)$.

The following lemma shows that for small α, D_α is about as large as $\alpha + \alpha \log \frac{1}{\alpha}$. This explains why Theorem 6 implies Theorem 4 for sufficiently small α.

Lemma 2.3. *For any $\epsilon > 0$, if α is sufficiently small depending on ϵ, then*

$$D_\alpha > (1 - \epsilon)\left(\alpha + \alpha \log \frac{1}{\alpha}\right).$$

Proof of Theorem 4. In Lemma 2.3, use $\epsilon/2$ in place of ϵ. We have

$$\text{FDR} > D_\alpha - \nu > (1 - \epsilon/2)\left(\alpha + \alpha \log \frac{1}{\alpha}\right) - \nu$$

if α is sufficiently small. Set ν in Theorem 6 as $\nu = \frac{\epsilon}{2}\left(\alpha + \alpha \log \frac{1}{\alpha}\right)$. Thus, we get

$$\text{FDR} > (1 - \epsilon)\left(\alpha + \alpha \log \frac{1}{\alpha}\right),$$

as desired. $\qquad\square$

Next, we turn to the proof of Theorem 6. To construct certain PRDN p-values for Theorem 6, we start with the observation that the upper bound given in (2.1) is almost tight if the adversary is well-informed. More precisely, given that the adversary knows all null p-values, the adversary can find the index

$$j^\star := \underset{1 \leq j \leq n_0}{\operatorname{argmax}} \ \frac{j}{\lceil np_{(j)}^{\mathrm{null}}/\alpha \rceil}, \tag{2.7}$$

where by convention $j^\star$ takes the largest value in the presence of ties. Having defined $j^\star$, we consider the following *informed adversarial* non-null p-values: given n_0 null p-values, set

$$\min \left\{ \left(\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star \right)_+, \ n_1 \right\}$$

non-null p-values to 0 and the remaining non-null p-values to 1. Above, x_+ denotes $\max\{x, 0\}$. In the event that $\left(\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star \right)_+ \leq n_1$, we have the following fact concerning the BH procedure, where no distributional assumptions are made on the nulls.

Lemma 2.4. *Supplied with the informed adversarial non-null p-values, the BH procedure at level α in the case where $\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star \leq n_1$ obeys*

$$\mathrm{FDP} = \min \left\{ \underset{1 \leq j \leq n_0}{\max} \ \frac{j}{\lceil np_{(j)}^{\mathrm{null}}/\alpha \rceil}, 1 \right\}, \tag{2.8}$$

thereby attaining the equality in (2.1) for the FDP.

This lemma confirms that the informed adversary indeed yields (nearly) the least favorable p-value configuration for the BH procedure in terms of FDR control. For completeness, if $1 \leq \lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star \leq n_1$, the BH procedure exactly rejects the $j^\star$ smallest null p-values and the $\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star$ non-null p-values that are set to 0. On the flip side, if $np_{(j^\star)}^{\mathrm{null}}/\alpha \leq j^\star$, the BH procedure rejects at least the $j^\star$ smallest null p-values (possibly rejects more than $j^\star$ null p-values) but none of the non-null p-values, in which case the FDP is 1. This is in agreement with (2.8). Next is the proof of this lemma.

Proof of Lemma 2.4. In the case where $1 \leq \lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star \leq n_1$, observe that

$$\frac{\alpha(j^\star + \lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star)}{n} \geq \frac{\alpha np_{(j^\star)}^{\mathrm{null}}/\alpha}{n} = p_{(j^\star)}^{\mathrm{null}},$$

which implies that the BH procedure rejects at least the $j^\star$ smallest null p-values and the $\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star$ zero-valued non-null p-values. Now, by way of contradiction, suppose, in addition to the $\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star$ non-null p-values, the BH procedure rejects $j+l$ null p-values for some $l \geq 1$. By the compliance property of the BH procedure, it must hold that

$$p_{(j^\star+l)}^{\mathrm{null}} \leq \frac{\alpha(\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil + l)}{n},$$

which implies

$$\lceil np_{(j^\star+l)}^{\mathrm{null}}/\alpha \rceil \leq \lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil + l.$$

13

As such, it concludes that

$$\frac{j^\star + l}{\lceil np_{(j^\star+l)}^{\mathrm{null}}/\alpha \rceil} \geq \frac{j^\star + l}{\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil + l} > \frac{j^\star}{\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil},$$

where the second inequality rests on the assumption that $j^\star < \lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil$. This contradicts the definition of $j^\star$ in (2.7). Therefore, the BH procedure must reject exactly $j^\star$ null p-values and $\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star$ non-null p-values, thus yielding

$$\mathrm{FDP} = \frac{j^\star}{j^\star + \lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star} = \max_{1 \leq j \leq n_0} \frac{j}{\lceil np_{(j)}^{\mathrm{null}}/\alpha \rceil} < 1.$$

Next, we turn to the case where $np_{(j^\star)}^{\mathrm{null}}/\alpha \leq j^\star$. In this case, note that $p_{(j^\star)}^{\mathrm{null}} \leq \alpha j^\star/n$. Thus, the BH procedure rejects at least the $j^\star$ most significant null p-values, and none of the non-null p-values is rejected because non-nulls are all set to 1. $\qquad\square$

Taken together, the discussion above readily gives the proof of Theorem 6.

Proof of Theorem 6. Lemma 2.4 reveals that

$$\mathrm{FDP} \geq \min \left\{ \max_{1 \leq j \leq n_0} \frac{j}{\lceil np_{(j)}^{\mathrm{null}}/\alpha \rceil}, 1 \right\} - \mathbf{1}(\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star > n_1)$$

for the BH procedure supplied with the informed adversarial p-values, where $\mathbf{1}(\cdot)$ is the indicator function. Thus, we get[6]

$$\mathrm{FDR} \geq \mathbb{E}\left[\min \left\{ \max_{1 \leq j \leq n_0} \frac{j}{\lceil np_{(j)}^{\mathrm{null}}/\alpha \rceil}, 1 \right\} \right] - \mathbb{P}\left(\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star > n_1 \right). \qquad (2.9)$$

The next step is to find certain PRD null p-values such that the right-hand side of (2.9) tends to D_α as $n \to \infty$. To this end, we consider the case where the n_0 null-p-values are i.i.d. uniform variables $U_1, \ldots, U_{n_0}$ on $(0,1)$. The asymptotic regime considered below satisfies $n_0, n_1 \to \infty$ and $n_1/n_0 \to 0$. Under these assumptions, we can prove that

$$\mathbb{E}\left[\min \left\{ \max_{1 \leq j \leq n_0} \frac{j}{\lceil nU_{(j)}/\alpha \rceil}, 1 \right\} \right] \to \mathbb{E}\left[\min \left\{ \max_{1 \leq j < \infty} \frac{j}{\lceil (\xi_1 + \cdots + \xi_j)/\alpha \rceil}, 1 \right\} \right] \qquad (2.10)$$

(recall $\xi_1, \xi_2, \ldots$ are i.i.d. exponential random variables with mean 1) and

$$\mathbb{P}\left(\lceil nU_{(j^\star)}/\alpha \rceil - j^\star > n_1 \right) \to 0, \qquad (2.11)$$

where $j^\star$ is the same as in (2.7) using $U_{(j)}$ in place of $p_{(j)}^{\mathrm{null}}$. As is self-evident, (2.9) together with (2.10) and (2.11) finishes the proof of Theorem 6.

To complete the last step of the proof, we briefly explain why (2.10) and (2.11) are true. For the first inequality, note that $nU_{(j)}$ converges weakly to $\xi_1 + \cdots + \xi_j$ as $n \to \infty$ since

---

[6]The analysis shows that the BH is not the most adversarial in terms of maximizing the FDP given the null p-values.

$n_0/n \to 1$. This weak convergence holds simultaneously for $1 \le j \le J$ if J is fixed. To bridge the gap between taking maximum over $\{1, \ldots, J\}$ and $\{1, \infty\}$, use the strong law of large numbers to show that both

$$\max_{J+1 \le j \le n_0} \frac{j}{\lceil nU_{(j)}/\alpha \rceil} \quad \text{and} \quad \max_{J+1 \le j < \infty} \frac{j}{\lceil (\xi_1 + \cdots + \xi_j)/\alpha \rceil}$$

converge to α in probability as both $J, n_0 \to \infty$. For the second inequality, the index $j^\star$ is bounded in probability as $n \to \infty$ and, hence, one can slowly increase n_1 so that the probability of $\lceil nU_{(j^\star)}/\alpha \rceil - j^\star > n_1$ tends to zero.

$\square$

In passing, we briefly remark that, supplied with the informed adversarial p-values, the BH procedure is *not* the most anti-conservative compliant procedure in the sense of maximizing the FDP. To see this, define

$$j^\diamond := \underset{1 \le j \le n_0}{\operatorname{argmax}} \left\{ \frac{j}{\lceil np_{(j)}^{\mathrm{null}}/\alpha \rceil} : \lceil np_{(j)}^{\mathrm{null}}/\alpha \rceil - j \le n_1 \right\},$$

with the convention that $\operatorname{argmax} \emptyset = 0$ in the case where $\lceil np_{(j)}^{\mathrm{null}}/\alpha \rceil - j > n_1$ for all $1 \le j \le n_0$. With this definition in place, the most anti-conservative compliant procedure rejects the following p-values, depending on whether $j^\diamond \ge 1$ or not,

$$\begin{cases} p_{(1)}^{\mathrm{null}}, \ldots, p_{(j^\diamond)}^{\mathrm{null}}, \text{ and } \left(\lceil np_{(j^\diamond)}^{\mathrm{null}}/\alpha \rceil - j^\diamond \right)_+ \text{ zero-valued non-nulls} & \text{if } j^\diamond \ge 1 \\ \emptyset & \text{if } j^\diamond = 0. \end{cases}$$

On a related note, if $\lceil np_{(j^\star)}^{\mathrm{null}}/\alpha \rceil - j^\star \le n_1$, this procedure rejects the same set of p-values as the BH procedure.

3 FDR Consistency

In this section, we prove Theorem 5 and showcase a number of FDR-consistent dependence classes. From Definition 1.3, FDR consistency can be equivalently defined as follows: a dependence class $\mathcal{P}$ is FDR-consistent if the BH procedure ensures that, for all p-value distribution element $P \in \mathcal{P}$,

$$\mathrm{FDR}(\alpha; P) \le Q(\alpha)$$

holds for a function $Q(\alpha)$ obeying $Q(\alpha) \to 0$ as $\alpha \to 0$. With this equivalent definition in place, we turn to the proof of Theorem 5.

Proof of Theorem 5. By assumption, we know that

$$\mathrm{FDR}(\alpha; P_0) \le Q(\alpha)$$

for any $P_0 \in \mathcal{P}_0$ and a function $Q(\alpha) \to 0$ as $\alpha \to 0$. Without loss of generality, assume $Q(\alpha) \le 1$. From the `FDR-linking` theorem, it immediately follows that

$$\mathrm{FDR}(\alpha; P) \le \alpha + \alpha \int_\alpha^1 \frac{Q(x)}{x^2} \mathrm{d}x. \tag{3.1}$$

Above, P_0 is the null dependence by P.

Leveraging the fact $Q(\alpha) \to 0$ as $\alpha \to 0$, for any small $\epsilon > 0$, we can find $\delta > 0$ such that $Q(\alpha) < \epsilon$ if $\alpha < \delta$. By means of (3.1), for $\alpha < \delta$, we get

$$
\begin{aligned}
\mathrm{FDR}(\alpha; P) &\leq \alpha + \alpha \int_\alpha^\delta \frac{Q(x)}{x^2} \, \mathrm{d}\, x + \alpha \int_\delta^1 \frac{Q(x)}{x^2} \mathrm{d}x \\
&\leq \alpha + \alpha \int_\alpha^\delta \frac{\epsilon}{x^2} \, \mathrm{d}\, x + \alpha \int_\delta^1 \frac{1}{x^2} \mathrm{d}x \\
&= \epsilon + \frac{(1 - \epsilon)\alpha}{\delta}.
\end{aligned}
$$

Taking any $\alpha < \frac{\epsilon \delta}{1 - \epsilon}$, we get $\mathrm{FDR}(\alpha; P) \leq 2\epsilon$, which holds for all $P \in \mathcal{P}$. Since ϵ can be arbitrarily small, what we have established reveals that

$$
\lim_{\alpha \to 0} \sup_{P \in \mathcal{P}} \mathrm{FDR}(\alpha; P) = 0.
$$

This completes the proof.

$\square$

Next, we present several examples of FDR-consistent dependence classes. Using Theorem 5, it suffices to examine whether FDR consistency holds under the induced null dependence. For comparison, it is unclear whether the BH procedure controls the FDR at the nominal level or not in these examples, whereas FDR consistency is a much more feasible concept to deal with. That being said, a worthwhile point to make is that FDR consistency is no trivial concept in the sense that, for example, there exists a joint distribution of the p-values under the global null such that the BH procedure leads to [GR08]

$$
\mathrm{FDR} = \min\left\{\left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right)\alpha, 1\right\} \approx \min\{\alpha \log n, 1\}
$$

for each n, which forms a non-FDR-consistent dependence class because the FDR bound involves $\log n$.

3.1 FDR-consistent examples

Example 3.1 (Equicorrelated normal distribution)**.** Consider a p-value vector whose null components are calculated from a (possibly negatively) equicorrelated normal distribution. More precisely, suppose we observe $X^0 \sim \mathcal{N}(0, \Sigma^0)$, where $\Sigma_{ij}^0 = \rho$ if $i \neq j \leq n_0$, and otherwise $\Sigma_{ii}^0 = 1$. The correlation coefficient ρ obeys $-\frac{1}{n_0 - 1} \leq \rho < 0$. The null p-values are those for testing $\mathbb{E}\, X_i^0 = 0$ against the one-sided or two-sided alternatives, while the non-null p-values are arbitrary. Then, the dependence class $\mathcal{P}$ comprised of all such p-value distributions is FDR-consistent.

Proof. By Theorem 5, we only need to consider the global null case $n = n_0$. For simplicity, the proof concerns the case of one-sided p-values and the proof for the two-sided case is very similar. Furthermore, Corollary 1.4 shows that it is sufficient to only focus on $-\frac{1}{n-1} \leq \rho < 0$, the negative correlation regime.

To begin with, note that FDR consistency of this example is equivalent to the following problem: for any $\epsilon > 0$, there exists $\delta > 0$ such that for any nominal level $\alpha < \delta$, the BH procedure gives FDR $\leq \epsilon$ under any dependence element in $\mathcal{P}$. Let $X_1, \ldots, X_n$ be n standard normal variables that are ρ-equicorrelated for some $-\frac{1}{n-1} \leq \rho < 0$. Denote by

$$Y_i = \sqrt{1+\rho}X_i + \sqrt{-\rho}Z,$$

where Z is a standard normal variable independent of $X_1, \ldots, X_n$. As is clear, $Y_1, \ldots, Y_n$ are i.i.d. standard normal variables. Let $\mathcal{R}_{\alpha,X}$ denote the event that the Simes method at level α rejects the global null hypothesis on $X_1, \ldots, X_n$. Define $\mathcal{R}_{\alpha,Y}$ similarly for $Y_1, \ldots, Y_n$. By definition, in the event $\mathcal{R}_{\alpha,X}$ there must exist some $j \geq 1$ such that

$$\min\left\{X_{i_1}, \ldots, X_{i_j}\right\} \geq \Phi^{-1}\left(1 - \frac{\alpha j}{n}\right)$$

for j indices $i_1, \ldots, i_j$ from $\{1, 2, \ldots, n\}$. Let $\mathcal{A}$ denote the event that $Z \geq -\sqrt[4]{n-1}$. Assuming $\alpha < \frac{1}{2}$, we have

$$\min\left\{Y_{i_1}, \ldots, Y_{i_j}\right\} \geq \sqrt{1+\rho}\,\Phi^{-1}\left(1 - \frac{\alpha j}{n}\right) - \sqrt{-\rho}\sqrt[4]{n-1}$$

$$\geq \sqrt{1 - 1/(n-1)}\,\Phi^{-1}\left(1 - \frac{\alpha j}{n}\right) - \sqrt{1/(n-1)}\sqrt[4]{n-1}$$

$$\geq \frac{n-2}{n-1}\Phi^{-1}\left(1 - \frac{\alpha j}{n}\right) - (n-1)^{-\frac{1}{4}}.$$

Now, suppose we have

$$\frac{n-2}{n-1}\Phi^{-1}\left(1 - \frac{\alpha j}{n}\right) - (n-1)^{-\frac{1}{4}} \geq \Phi^{-1}\left(1 - \frac{\alpha_\epsilon j}{n}\right) \tag{3.2}$$

for some $\alpha_\epsilon \in (0,1)$ that will be specified later. Then, we get

$$\min\left\{Y_{i_1}, \ldots, Y_{i_j}\right\} \geq \Phi^{-1}\left(1 - \frac{\alpha_\epsilon j}{n}\right)$$

in the event $\mathcal{R}_{\alpha,X} \cap \mathcal{A}$. As such, $\mathcal{R}_{\alpha,X} \cap \mathcal{A} \subset \mathcal{R}_{\alpha_\epsilon,Y}$. Consequently, we have

$$\mathbb{P}(\mathcal{R}_{\alpha,X}) = \mathbb{P}(\mathcal{R}_{\alpha,X} \cap \mathcal{A}) + \mathbb{P}(\mathcal{R}_{\alpha,X} \cap \mathcal{A}^c)$$

$$\leq \mathbb{P}(\mathcal{R}_{\alpha_\epsilon,Y}) + \mathbb{P}(\mathcal{B}^c)$$

$$= \mathbb{P}(\mathcal{R}_{\alpha_\epsilon,Y}) + \Phi\left(-\sqrt[4]{n-1}\right).$$

From [BY01], it is known that $\mathbb{P}(\mathcal{R}_{\alpha_\epsilon,Y}) = \alpha_\epsilon$. Therefore, the FDR of the BH procedure applied to the p-values from $X_1, \ldots, X_n$ satisfies

$$\text{FDR} = \mathbb{P}(\mathcal{R}_{\alpha,X}) \leq \alpha_\epsilon + \Phi\left(-\sqrt[4]{n-1}\right). \tag{3.3}$$

To continue the proof, we need the following lemma.

Lemma 3.2. *For any $\epsilon > 0$, there exists a sufficiently large integer N such that*

$$\frac{n-2}{n-1}\Phi^{-1}\left(1 - \frac{\alpha j}{n}\right) - (n-1)^{-\frac{1}{4}} \geq \Phi^{-1}\left(1 - \frac{\epsilon j}{2n}\right) \tag{3.4}$$

for all $n \geq N$ and all $1 \leq j \leq n$ if $\alpha < \epsilon/4$.

By Lemma 3.2, we can always find sufficiently large N_1, depending only on ϵ, such that (3.2) holds for $\alpha_\epsilon = \epsilon/2$ and $\alpha < \epsilon/4$. In this case, we have

$$\mathbb{P}(\mathcal{R}_{\alpha,X}) \leq \frac{\epsilon}{2} + \Phi\left(-\sqrt[4]{n-1}\right).$$

Next, we choose N_2 such that $\Phi\left(-\sqrt[4]{n-1}\right) \leq \frac{\epsilon}{2}$ if $n \geq N_2$. Note that N_2 depends only on ϵ. Thus, if $n \geq \max\{N_1, N_2\}$, from (3.3) we get

$$\mathrm{FDR} = \mathbb{P}(\mathcal{R}_{\alpha,X}) \leq \epsilon \tag{3.5}$$

for any $\alpha < \epsilon/4$. Last, let δ_1 be sufficiently small such that

$$\sup_{\alpha < \delta_1} \mathrm{FDR}(\alpha) \leq \epsilon \tag{3.6}$$

for all $n < \max\{N_1, N_2\}$ and $-\frac{1}{n-1} \leq \rho < 0$.

Taken together, (3.5) and (3.6) show that $\mathrm{FDR} \leq \epsilon$ for all n and any $-\frac{1}{n-1} \leq \rho < 0$ if $\alpha < \min\{\epsilon/4, \delta_1\}$. This completes the proof. The proof of Lemma 3.2 is deferred to the Appendix.

$\square$

Example 3.3 (Arbitrary dependence with vanishing true null proportion). Consider a sequence of integer pairs $\{n^{(l)}, n_0^{(l)}\}_{l=1}^\infty$ such that[7] $n^{(l)} \geq n_0^{(l)}$ and

$$\sup_l \frac{n_0 \log n_0}{n} < \infty. \tag{3.7}$$

Let $\mathcal{P}$ consist of all $n^{(l)}$-dimensional p-value distributions with $n_0^{(l)}$ null components. Except for specified $n^{(l)}, n_0^{(l)}$, the null dependence and null-non-null dependence can be arbitrary. Then, $\mathcal{P}$ is FDR-consistent.

Proof. We give a direct proof of the FDR consistency of $\mathcal{P}$. By Theorem 7 in Section 4, it suffices to show that

$$\inf_l \frac{1}{\pi_0 S(n_0)} > 0$$

and

$$\lim_{\alpha \to 0} S(n_0)\pi_0\alpha \log \frac{\mathrm{e}}{S(n_0)\pi_0\alpha} = 0$$

uniformly over $P \in \mathcal{P}$. The first inequality is used to pick a sufficiently small α such that the FDR bound given in Theorem 7 is nontrivial for all l. To verify the two requirements, note that

$$\frac{1}{\pi_0 S(n_0)} = (1 + o(1))\frac{n}{n_0 \log n_0} > c$$

---

[7]Below, the dependence on l is often omitted for the sake of simplicity.

uniformly in l for some $c > 0$, as implied by (3.7). Likewise, we have

$$S(n_0)\pi_0\alpha \log \frac{e}{S(n_0)\pi_0\alpha} \le \frac{\alpha}{c} \log \frac{ec}{\alpha},$$

which tends to zero uniformly as $\alpha \to 0$. $\qquad\square$

Example 3.4 (Two-sided-PRDN dependence)**.** Going back to the example of one-sided normal p-values given right below Definition 1.2, observe that the PRDN property is not satisfied by the two-sided p-values, unless under certain additional assumptions [KR81]. In fact, as with the PRDS property, PRDN in generally does carry over from one-sided to two-sided p-values (see, for example, one-sided p-values for Studentized tests in [BY01]). Interestingly, these two-sided p-values maintain FDR consistency. To state precisely, let $p_1^{(o)}, \ldots, p_n^{(o)}$ be (one-sided) PRDN p-values and, in addition, the corresponding test statistics are distributed symmetrically about 0 under the null and are continuous, thereby yielding uniformly distributed null p-values. Define the induced two-sided p-values as

$$p_i^{(t)} = \begin{cases} 2p_i^{(o)}, & \text{if } p_i^{(o)} \le \frac{1}{2} \\ 2(1 - p_i^{(o)}), & \text{if } p_i^{(o)} > \frac{1}{2}. \end{cases}$$

In the case of normal distribution, the two-sided p-value $p_i^{(t)} = 2\Phi(-|z|)$ is given as above by the one-sided $p_i^{(o)} = \Phi(-z)$. Below, we show that all distributions of these two-sided p-values form an FDR-consistent dependence class.

Proof. As earlier, Theorem 5 allows us to only consider $p_1^{(o)}, \ldots, p_n^{(o)}$ that satisfy PRD under the global null. To start with, recognize that the BH procedure at level α controls the FDR at α on these p-values. Due to symmetry, this also applies to $1 - p_1^{(o)}, \ldots, 1 - p_n^{(o)}$. Denote by R the number of rejected two-sided p-values by the BH procedure. In the event that $R \ge 1$, it must be the case where at least $\lceil R/2 \rceil$ of them are given as $2p_i^{(o)}$ or the case where at least $\lceil R/2 \rceil$ are given as $2(1 - p_i^{(o)})$. In the former case, at least $\lceil R/2 \rceil$ from $p_1^{(o)}, \ldots, p_n^{(o)}$ are no greater than

$$\frac{1}{2} \cdot \frac{\alpha R}{n} \le \frac{1}{2} \cdot \frac{2\alpha\lceil R/2 \rceil}{n} = \frac{\alpha\lceil R/2 \rceil}{n}. \tag{3.8}$$

Otherwise, at least $\lceil R/2 \rceil$ from $p_1^{(o)}, \ldots, p_n^{(o)}$ satisfies

$$\max\left\{1 - p_{i_1}^{(o)}, \ldots, 1 - p_{i_{\lceil R/2 \rceil}}^{(o)}\right\} \le \frac{\alpha\lceil R/2 \rceil}{n}. \tag{3.9}$$

As noted earlier, each of the two events (3.8) and (3.9) happens with probability no more than α. Taking a union bound, we prove that the FDR on the two-sided p-values is controlled at level 2α. Therefore, it is FDR-consistent.

$\qquad\square$

Example 3.5 (Block dependence)**.** Fix a positive integer b. Consider a b-block structure $\{1, 2, \ldots, n\} = \cup_{i=1}^{m} B_i$, each of size $|B_i| = b_i \le b$ and $b_1 + \cdots + b_m = n$. Let $\mathcal{P}$ be the class of all p-value distributions with a b-block structure in the sense that p-values from different blocks are jointly independent while, within every block, the joint distribution can be arbitrary. This dependence class is FDR-consistent.

Proof. Consider the induced null set, which remains block-structured. The induced block sizes are no more than b. Thus, by Theorem 5, it suffices to consider the global null case.

Denote the p-values by $p_1^{(l)}, p_2^{(l)}, \ldots, p_{b_l}^{(l)}$ for the lth block, where $l = 1, \ldots, m$. Now, we consider the adjusted p-value

$$p^{(l)} := \min\left\{ b_l \min_i p_i^{(l)}, 1 \right\}.$$

Note that these are valid p-values. Now, in the case where the BH procedure rejects at least one p-value, that is, $R > 0$, we know that at least $\lceil R/b \rceil$ blocks are rejected. By assumption, the minimum p-values from at least $\lceil R/b \rceil$ blocks are all less than or equal to

$$\frac{\alpha R}{n} \leq \frac{b\alpha \lceil R/b \rceil}{n} \leq \frac{b\alpha \lceil R/b \rceil}{m}.$$

Thus, whenever the BH procedure rejects at least one on the full set of p-values, it also rejects at least one when applied to the adjusted p-values $p^{(1)}, \ldots, p^{(m)}$. Thus, due to the joint independence of the m blocks, we must have

$$\text{FDR} \leq b\alpha,$$

which tends to 0 uniformly over the dependence class. Thus, this dependence class is FDR-consistent.

$\square$

4 FDR Control Under Arbitrary Dependence

Under arbitrary dependence, [BY01] proves that the FDR of the BH procedure always satisfies

$$\text{FDR} \leq \min\{S(n)\pi_0\alpha, 1\}, \tag{4.1}$$

where $S(n)$ is a shorthand for $1 + \frac{1}{2} + \cdots + \frac{1}{n} \approx \log n + 0.577$. In the literature, to our knowledge, this log-correction bound is the best known unconditional FDR bound for the full range $0 < \alpha < 1$. As a matter of fact, if $(1 - \pi_0 + \pi_0 S(n_0))\alpha \leq 1$, this FDR bound is tight in the sense that there exists a joint distribution of the p-values such that (4.1) is an equality [GR08] (see also [Røterm06] for the case of $\pi_0 = 1$). Recognizing this fact, one would imagine that it is very challenging to improve on the log-correction bound.

Interestingly, a direct application of the `FDR-linking` theorem allows us to strictly improve the bound (4.1) for a certain range of the nominal level α.

Theorem 7. *Under arbitrary dependence of the p-values, any compliant procedure at level α satisfies*

$$\text{FDR} \leq \begin{cases} 1, & \text{if } \alpha \geq \frac{1}{\pi_0 S(n_0)} \\ S(n_0)\pi_0\alpha \log \frac{e}{S(n_0)\pi_0\alpha}, & \text{if } \alpha < \frac{1}{\pi_0 S(n_0)}. \end{cases}$$

This theorem applies to all compliant procedures, whereas it is unclear whether the bound (4.1) applies to general compliant procedures. To further appreciate this theorem, we take

$$\alpha = \frac{1}{\pi_0 S(n)}.$$

20

With this choice of α, (4.1) yields the trivial bound FDR ≤ 1, whereas Theorem 7 gives

$$\text{FDR} \leq \frac{S(n_0)}{S(n)} \log \frac{\mathrm{e}S(n)}{S(n_0)} < 1$$

under the very mild condition $n > n_0$ (that is, any configurations of the p-values except for the global null). More broadly, the new unconditional FDR bound strictly improves (4.1) for the following range of α:

$$\frac{1}{\pi_0 S(n_0)} \mathrm{e}^{1-\frac{S(n)}{S(n_0)}} < \alpha < \frac{1}{\pi_0 S(n_0)}. \tag{4.2}$$

To see this point, note that it suffices to verify

$$S(n_0)\pi_0\alpha \log \frac{\mathrm{e}}{S(n_0)\pi_0\alpha} < S(n)\pi_0\alpha$$

if α satisfies (4.2). This is clearly true.

We conclude this section by proving Theorem 7.

Proof of Theorem 7. Using the unconditional FDR bound [Hom83, BY01]

$$\text{FDR}_0(x) \leq \min\{S(n_0)x, 1\}$$

under arbitrary dependence of the null p-values, from Theorem 1 we get

$$\text{FDR} \leq \pi_0\alpha + \pi_0\alpha \int_{\pi_0\alpha}^1 \frac{\min\{S(n_0)x, 1\}}{x^2} \mathrm{d}x.$$

This FDR bound can be simplified as

$$\pi_0\alpha + \pi_0\alpha \int_{\pi_0\alpha}^1 \frac{\min\{S(n_0)x, 1\}}{x^2} \mathrm{d}x = \begin{cases} 1, & \text{if } \alpha \geq \frac{1}{\pi_0 S(n_0)} \\ S(n_0)\pi_0\alpha \log \frac{\mathrm{e}}{S(n_0)\pi_0\alpha}, & \text{if } \alpha < \frac{1}{\pi_0 S(n_0)}. \end{cases}$$

This concludes the proof.

$\square$

5 Bounded Adversariness

As shown in Theorem 4, the logarithmic factor $\log \frac{1}{\alpha}$ in the FDR bound is unavoidable provided the informed adversarial non-null p-values. A question arising from this observation is whether the logarithmic factor can be removed if the adversary is restricted from accessing certain information of the null p-values.

Going back to the informal explanation of the factor $\log \frac{1}{\alpha}$ surrounding (2.6), we learn that the smallest null p-value is essential for the adversary to construct non-null p-values that maximize the FDP. In light of this observation, we consider a *Bonferroni-masked* adversary that has access to all null p-values but the smallest. Thus, the available information is not sufficient to perform the Bonferroni correction on the nulls (unless $p_{(2)}^{\mathrm{null}}$ is below the

21

Bonferroni threshold), hence the name Bonferroni-masked adversary. To formally define this adversary, we use a function

$$T = T(p_{(2)}^{\mathrm{null}}, \ldots, p_{(n_0)}^{\mathrm{null}})$$

to denote the number of true rejections made by the BH procedure, which depends on all sorted null p-values except for the smallest one. This can be realized by setting T non-null p-values to 0 and the rest to 1.

The following theorem shows that the logarithmic factor can be dropped in the FDR bound with independent nulls and any Bonferroni-masked non-nulls. We assume $n_0 \geq 2$ to exclude trivial cases.

Theorem 8. *Let the null p-values be i.i.d. uniform variables on $(0,1)$ and the non-null p-values be provided by any Bonferroni-masked adversary. Then, any compliant procedure at level α ensures*

$$\mathrm{FDR} \leq 3.5\alpha.$$

As the proof of Theorem 4 also assumes i.i.d. uniform null p-values, the improvement of the FDR bound in Theorem 8 can only be attributed to the withholding of the most significant null p-value to the Bonferroni-masked adversary. As an aside, we remark that the independence of the null p-values is crucial to Theorem 8 since, in the presence of correlation, other null p-values would reveal much information about the smallest null p-value.

Now, we prove Theorem 8 as follows.

Proof of Theorem 8. As earlier, V denotes the number of false rejections. If $V \geq 2$, we have

$$\mathrm{FDP} \leq \max_{2 \leq j \leq n_0} \frac{\alpha j}{n p_{(j)}^{\mathrm{null}}},$$

which remains true if $V = 0$ because $\mathrm{FDP} = 0$ in this case. In the case of $V = 1$, the compliance condition shows that the rejected null p-value is below $\alpha(T+1)/n$, thus implying

$$p_{(1)}^{\mathrm{null}} \leq \frac{\alpha(T+1)}{n}.$$

Taken together, the two inequalities above give

$$\mathrm{FDR} = \mathbb{E}(\mathrm{FDP}; R \neq T+1) + \mathbb{E}(\mathrm{FDP}; R = T+1)$$

$$\leq \mathbb{E}\left[\max_{2 \leq j \leq n_0} \frac{\alpha j}{n p_{(j)}^{\mathrm{null}}}; R \neq T+1\right] + \mathbb{E}\left[\frac{1}{T+1}; R = T+1\right]$$

$$\leq \mathbb{E}\left[\max_{2 \leq j \leq n_0} \frac{\alpha j}{n p_{(j)}^{\mathrm{null}}}\right] + \mathbb{E}\left[\frac{\mathbf{1}\left(p_{(1)}^{\mathrm{null}} \leq \alpha(T+1)/n\right)}{T+1}\right]. \tag{5.1}$$

To bound the term contributed by the case of $R \neq T+1$ in (5.1), note that Lemma 3.1 of [DSZ18] ensures that

$$\mathbb{E}\left[\max_{2 \leq j \leq n_0} \frac{\alpha j}{n p_{(j)}^{\mathrm{null}}}\right] \leq C_2 \pi_0 \alpha \tag{5.2}$$

22

for a constant C_2 between 2.4 and 2.5. Now we turn to the term contributed by the case of $R = T + 1$. First, write

$$\mathbb{E}\left[\frac{\mathbf{1}\left(p_{(1)}^{\mathrm{null}} \leq \alpha(T+1)/n\right)}{T+1}\right] = \mathbb{E}\left\{\mathbb{E}\left[\frac{\mathbf{1}\left(p_{(1)}^{\mathrm{null}} \leq \alpha(T+1)/n\right)}{T+1}\bigg| p_{(2)}^{\mathrm{null}}, \ldots, p_{(n_0)}^{\mathrm{null}}\right]\right\}.$$

Recognizing that $p_{(1)}^{\mathrm{null}}$ is (conditionally) uniformly distributed on $(0, p_{(2)}^{\mathrm{null}})$ and T is measurable with respect to $p_{(2)}^{\mathrm{null}}, \ldots, p_{(n_0)}^{\mathrm{null}}$, we get

$$\mathbb{E}\left[\frac{\mathbf{1}\left(p_{(1)}^{\mathrm{null}} \leq \alpha(T+1)/n\right)}{T+1}\bigg| p_{(2)}^{\mathrm{null}}, \ldots, p_{(n_0)}^{\mathrm{null}}\right] \leq \frac{\frac{\alpha(T+1)/n}{p_{(2)}^{\mathrm{null}}}}{T+1} = \frac{\alpha}{np_{(2)}^{\mathrm{null}}},$$

where the inequality reduces to equality if $\alpha(T+1)/n \leq p_{(2)}^{\mathrm{null}}$. This yields

$$\mathbb{E}\left[\frac{\mathbf{1}\left(p_{(1)}^{\mathrm{null}} \leq \alpha(T+1)/n\right)}{T+1}\right] \leq \mathbb{E}\left[\frac{\alpha}{np_{(2)}^{\mathrm{null}}}\right]. \tag{5.3}$$

To proceed, note the fact that $p_{(2)}^{\mathrm{null}}$ follows the Beta distribution parametrized by $(2, n_0-1)$. By means of this fact, we have

$$\mathbb{E}\left[\frac{\alpha}{np_{(2)}^{\mathrm{null}}}\right] = \frac{\alpha}{n}\int_0^1 \frac{1}{x}\frac{x(1-x)^{n_0-2}}{\mathrm{Beta}(2, n_0-1)}\mathrm{d}x = \pi_0\alpha,$$

which, together with (5.2) and (5.3), reveals that

$$\mathrm{FDR} \leq C_2\pi_0\alpha + \pi_0\alpha < 3.5\pi_0\alpha \leq 3.5\alpha.$$

This completes the proof.

$\square$

6 Discussion

In this paper, we have demonstrated the use of the `FDR-linking` theorem in proving FDR control properties through examples for compliant procedures, including the BH procedure. In slightly more detail, this theorem shows that any compliant procedures control the FDR up to a factor that is independent of the number of hypotheses under the PRDN dependence of the p-values, which strictly relaxes the PRDS property by only imposing distributional assumptions on the nulls. Moreover, the FDR bound is optimal under this new dependence, shedding light on the sharpness of the `FDR-linking` theorem. In another application, we propose FDR consistency as a flexible, amenable concept to complement FDR control and, using the `FDR-linking` theorem, prove that this new property relies entirely on the null p-values, regardless of the non-nulls.

Loosely speaking, the `FDR-linking` theorem reveals that FDR control is basically a matter of the null p-values. This is a blessing for researchers in that an analysis of the

FDR would not lose much of its value by focusing only on the joint distribution of the null p-values. This is particularly convenient if some knowledge of the null dependence is available. From a different angle, this theorem unveils the robustness of the BH procedure in FDR control against any unfavorable or even adversarial dependence between the null p-values and non-null p-values. This viewpoint is in agreement with the widely observed phenomenon that the BH procedure seldom loses much control over the FDR beyond what has been proved [Sto03, RB07, GSS08, CH09].

We conclude this paper with several promising directions for future work. First, the simple bound on the FDP (2.1) can possibly be extended to variants of the FDR criterion and general BH-type multiple testing procedures. For instance, the false discovery exceedance [GW04, vdLDP04, RW07, GHS14], defined as the probability that the FDP exceeds a specified level, is a popular alternative to the FDR since the realized FDP can significantly deviate from its expectation (FDR) in certain settings [Efr07]. To control the false discovery exceedance, it follows from (2.1) that

$$\mathbb{P}(\text{FDP} \geq \gamma) \leq \mathbb{P}\left(\max_{1 \leq j \leq n_0} \frac{\alpha j}{n p_{(j)}^{\text{null}}} \geq \gamma\right) \leq \frac{\pi_0 \alpha}{\gamma}$$

for any $0 < \gamma < 1$ if the p-values satisfy the PRDN property. Moreover, the bound (2.1) can be used to derive a bound on $\mathbb{E}\,\text{FDP}^k$ for an integer $k \geq 2$ [FZ06] and to possibly maximize the variance of FDP, thereby delineating the variability of FDR control. It is also of interest to incorporate prior knowledge such as weights of hypotheses and the true null proportion into the FDP bound [GRW06, BKY06, STS04, Sto02, RBWJ17].

More broadly, the `FDR-linking` inequality (1.1) in Theorem 1 calls for further investigation. In addition to Theorem 4, we wonder if the tightness of (1.1) remains under general dependence of the nulls. Practically speaking, the non-null p-values are unlikely to be the least favorable for FDR control as in Theorem 4 and, therefore, it is worth trying to improve the FDR bound in the `FDR-linking` theorem by assuming more realistic conditions on the null-non-null dependence. Theorem 8 has made a step toward this goal. Due to the connection between $\text{FDR}_0(x)$ and the Simes method, a related direction is to explore null dependence under which the Simes method controls the type I error at a reasonably small level. For example, it is likely that for a ρ-equicorrelated multivariate normal distribution, the Simes p-value would be quite conservative for a range of ρ. If so, it would lead to an FDR bound of the form $c\alpha$ for some constant c using the `FDR-linking` theorem.

Finally, we present a challenging problem whether FDR consistency holds for many common distributions of p-values. Specifically, we ask:

Open Problem. *Prove or disprove that p-values for testing means in any multivariate normal distribution with known covariance form an FDR-consistent dependence class.*

By Theorem 5, we only need to focus on the global null case, and it does not matter whether the p-values are two-sided or one-sided as in Example 3.1. We believe that this problem is true given that the Simes method is empirically observed to be approximately valid under many types of dependence structures [RB07], though pathological counterexamples exist [Rød06, FRS17].

Acknowledgments

References

[Ben10]   Yoav Benjamini. Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):405–416, 2010.

[BH95]   Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate – A practical and powerful approach to multiple testing. *Journal of the Royal Statistics Society: Series B (Statistical Methodology)*, 57:289–300, 1995.

[BKY06]   Yoav Benjamini, Abba M Krieger, and Daniel Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, 2006.

[BL99]   Yoav Benjamini and Wei Liu. A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *Journal of Statistical Planning and Inference*, 82(1-2):163–170, 1999.

[BR08]   Gilles Blanchard and Etienne Roquain. Two simple sufficient conditions for FDR control. *Electronic Journal of Statistics*, 2:963–992, 2008.

[BR17]   Rina Foygel Barber and Aaditya Ramdas. The p-filter: multilayer false discovery rate control for grouped hypotheses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1247–1268, 2017.

[BvdBS+15] Małgorzata Bogdan, Ewout van den Berg, Chiara Sabatti, Weijie Su, and Emmanuel J Candès. SLOPE – adaptive variable selection via convex optimization. *The Annals of Applied Statistics*, 9(3):1103, 2015.

[BY01]   Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, 2001.

[CH09]   Sandy Clarke and Peter Hall. Robustness of multiple testing procedures against dependence. *The Annals of Statistics*, 37(1):332–358, 2009.

[DSZ18]   Cynthia Dwork, Weijie J Su, and Li Zhang. Differentially private false discovery rate control. *arXiv preprint arXiv:1807.04209*, 2018.

[Efr07]   Bradley Efron. Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association*, 102(477):93–103, 2007.

[Efr10]   Bradley Efron. Correlated z-values and the accuracy of large-scale statistical estimates. *Journal of the American Statistical Association*, 105(491):1042–1055, 2010.

[Far07]   Alessio Farcomeni. Some results on the control of the false discovery rate under dependence. *Scandinavian Journal of Statistics*, 34(2):275–297, 2007.

[FDR07]   Helmut Finner, Thorsten Dickhaus, and Markus Roters. Dependency and false discovery rate: Asymptotics. *The Annals of Statistics*, 35(4):1432–1455, 2007.

[FH17]   Jianqing Fan and Xu Han. Estimation of the false discovery proportion with unknown dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1143–1164, 2017.

[FHG12]   Jianqing Fan, Xu Han, and Weijie Gu. Estimating false discovery proportion under arbitrary covariance dependence. *Journal of the American Statistical Association*, 107(499):1019–1035, 2012.

[FKC09]   Chloé Friguet, Maela Kloareg, and David Causeur. A factor model approach to multiple testing under dependence. *Journal of the American Statistical Association*, 104(488):1406–1415, 2009.

[FKSZ18]   Jianqing Fan, Yuan Ke, Qiang Sun, and Wen-Xin Zhou. FarmTest: Factor-adjusted robust multiple testing with approximate false discovery control. *Journal of the American Statistical Association*, (just-accepted):1–36, 2018.

[FR01]   Helmut Finner and Markus Roters. On the false discovery rate and expected type i errors. *Biometrical Journal*, 43(8):985–1005, 2001.

[FRS17]   Helmut Finner, M Roters, and K Strassburger. On the Simes test under dependence. *Statistical Papers*, 58(3):775–789, 2017.

[FZ06]   JA Ferreira and AH Zwinderman. On the Benjamini–Hochberg method. *The Annals of Statistics*, 34(4):1827–1849, 2006.

[GHS14]   Wenge Guo, Li He, and Sanat K Sarkar. Further results on controlling the false discovery proportion. *The Annals of Statistics*, 42(3):1070–1101, 2014.

[GLN02]   Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, 2002.

[GR08]   Wenge Guo and M Bhaskara Rao. On control of the false discovery rate under no assumption of dependency. *Journal of Statistical Planning and Inference*, 138(10):3176–3188, 2008.

[GRW06]   Christopher R Genovese, Kathryn Roeder, and Larry Wasserman. False discovery control with p-value weighting. *Biometrika*, 93(3):509–524, 2006.

[GSS08]   Yongchao Ge, Stuart C Sealfon, and Terence P Speed. Some step-down procedures controlling the false discovery rate under dependence. *Statistica Sinica*, 18(3):881, 2008.

[GW04]   Christopher Genovese and Larry Wasserman. A stochastic process approach to false discovery control. *The Annals of Statistics*, 32(3):1035–1061, 2004.

[HJ15]   Philipp Heesen and Arnold Janssen. Inequalities for the false discovery rate (FDR) under dependence. *Electronic Journal of Statistics*, 9(1):679–716, 2015.

[Hom83]   G Hommel. Tests of the overall hypothesis for arbitrary dependence structures. *Biometrische Zeitschrift*, 25(5):423–430, 1983.

[HR95]   Yosef Hochberg and Dror Rom. Extensions of multiple testing procedures based on simes' test. *Journal of Statistical Planning and Inference*, 48(2):141–152, 1995.

[KR80]   Samuel Karlin and Yosef Rinott. Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis*, 10(4):467–498, 1980.

[KR81]   Samuel Karlin and Yosef Rinott. Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. *The Annals of Statistics*, 9(5):1035–1049, 1981.

[KR18]   Eugene Katsevich and Aaditya Ramdas. Towards "simultaneous selective inference": post-hoc bounds on the false discovery proportion. *arXiv preprint arXiv:1803.06790*, 2018.

[LLS17]   Weidong Liu, Dennis Leung, and Qi-Man Shao. False discovery control for pairwise comparisons—an asymptotic solution to Williams, Jones and Tukey's conjecture. *arXiv preprint arXiv:1712.03305*, 2017.

[Owe05]   Art B Owen. Variance of the number of false discoveries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(3):411–426, 2005.

[RB07]   Anat Reiner-Benaim. FDR control by the BH procedure for two-sided correlated tests with implications to gene expression data analysis. *Biometrical Journal*, 49(1):107–126, 2007.

[RBWJ17]   Aaditya Ramdas, Rina Foygel Barber, Martin J Wainwright, and Michael I Jordan. A unified treatment of multiple testing with prior knowledge. *arXiv preprint arXiv:1703.06222*, 2017.

[Rød06]   Einar Andreas Rødland. Simes' procedure is 'valid on average'. *Biometrika*, 93(3):742–746, 2006.

[RSW08]   Joseph P Romano, Azeem M Shaikh, and Michael Wolf. Control of the false discovery rate under dependence using the bootstrap and subsampling. *Test*, 17(3):417, 2008.

[RV11]   Etienne Roquain and Fanny Villers. Exact calculations for false discovery proportion with application to least favorable configurations. *The Annals of Statistics*, pages 584–612, 2011.

[RW07]   Joseph P Romano and Michael Wolf. Control of generalized error rates in multiple testing. *The Annals of Statistics*, 35(4):1378–1408, 2007.

[Sar98]   Sanat K Sarkar. Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *The Annals of Statistics*, pages 494–504, 1998.

[Sar02]   Sanat K Sarkar. Some results on false discovery rate in stepwise multiple testing procedures. *The Annals of Statistics*, pages 239–257, 2002.

[SC96]   Ester Samuel-Cahn. Is the simes improved bonferroni procedure conservative? *Biometrika*, 83(4):928–933, 1996.

[SC97]   Sanat K Sarkar and Chung-Kuei Chang. The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association*, 92(440):1601–1608, 1997.

[SC09]   Wenguang Sun and T Tony Cai. Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):393–424, 2009.

[Sim86]   R John Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.

[SL11]   Armin Schwartzman and Xihong Lin. The effect of correlation in false discovery rate estimation. *Biometrika*, 98(1):199–214, 2011.

[ST03]   John D Storey and Robert Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445, 2003.

[Sto02]   John D Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.

[Sto03]   John D Storey. The positive false discovery rate: a Bayesian interpretation and the q-value. *The Annals of Statistics*, pages 2013–2035, 2003.

[STS04]   John D Storey, Jonathan E Taylor, and David Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):187–205, 2004.

[TLD98]   Ajit C Tamhane, Wei Liu, and Charles W Dunnett. A generalized step-up-down multiple test procedure. *The Canadian Journal of Statistics*, pages 353–363, 1998.

[TTC01]   Virginia Goss Tusher, Robert Tibshirani, and Gilbert Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, 98(9):5116–5121, 2001.

[vdLDP04]   Mark J van der Laan, Sandrine Dudoit, and Katherine S Pollard. 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):1–25, 2004.

[Wu08]   Wei Biao Wu. On false discovery control under dependence. *The Annals of Statistics*, 36(1):364–380, 2008.

[YB99]   Daniel Yekutieli and Yoav Benjamini. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, 82(1-2):171–196, 1999.

[Yek08]   Daniel Yekutieli. False discovery rate control for non-positively regression dependent test statistics. *Journal of Statistical Planning and Inference*, 138(2):405–415, 2008.

A Appendix

A.1 The PRDN example in Section 1

The one-sided p-values given as an example right below Definition 1.2 satisfy the PRDN property. The simple proof below follows from that of Case 1 in Section 3.1 of [BY01].

We introduce some notation for the proof. Let $X_{\mathcal{N}_0}$ be the vector comprised of the n_0 null components of X. For any $i \in \mathcal{N}_0$, let $X_{\mathcal{N}_0}^{(i)}$ denote the entry corresponding to i and $X_{\mathcal{N}_0}^{(-i)}$ denote the vector derived from $X_{\mathcal{N}_0}$ by deleting $X_{\mathcal{N}_0}^{(i)}$. Let $\Sigma_{-i,-i}^0$ be the covariance matrix of $X_{\mathcal{N}_0}^{(-i)}$, $\Sigma_{i,-i}^0$ be the covariance of $X_{\mathcal{N}_0}^{(i)}$ and $X_{\mathcal{N}_0}^{(-i)}$, and last, $\Sigma_{i,i}^0$ be the variance of $X_{\mathcal{N}_0}^{(i)}$. In this setup, the distribution of $X_{\mathcal{N}_0}^{(-i)}$ given $X_{\mathcal{N}_0}^{(i)} = x$ is a normal distribution with mean and covariance given as

$$\Sigma_{i,-i}^0(\Sigma_{i,i}^0)^{-1}x, \ \Sigma_{-i,-i}^0 - \Sigma_{i,-i}^0(\Sigma_{i,i}^0)^{-1}(\Sigma_{i,-i}^0)^{\top},$$

respectively.

Above, while the conditional covariance $\Sigma_{-i,-i}^0 - \Sigma_{i,-i}^0(\Sigma_{i,i}^0)^{-1}(\Sigma_{i,-i}^0)^{\top}$ is fixed, the conditional mean $\Sigma_{i,-i}^0(\Sigma_{i,i}^0)^{-1}x$ increases with x since $\Sigma_{i,-i}^0$ is entrywise nonnegative. This fact, together with the observation that the one-sided normal p-value is a monotone transformation of the z-value, implies that the one-sided p-values satisfy the PRDN property.

For information, we give a sufficient condition for the two-sided normal p-values normal to satisfy the PRDN property. The setting is the same as above. To this end, we only need to focus on the covariance Σ^0 of the null part $X_{\mathcal{N}_0}$ of the full vector X. As shown in [KR81], the density of $|X_{\mathcal{N}_0}|$ is MTP$_2$ if and only if there exists a diagonal matrix B with diagonal entries ± 1 such that the off-diagonal entries of $-B(\Sigma^0)^{-1}B$ are all nonnegative. Thus, the existence of such a matrix B ensures that the two-sided p-values satisfy the PRDN property because the MTP$_2$ condition implies PRDS on any subset. As an aside, this sufficient condition might not be necessary.

A.2 Proof of Lemma 3.2 in Section 3

Below, we prove Lemma 3.2.

Proof of Lemma 3.2. The lemma is equivalent to

$$\Phi^{-1}\left(1 - \frac{\epsilon j}{4n}\right) - \Phi^{-1}\left(1 - \frac{\epsilon j}{2n}\right) \geq \frac{1}{n-1}\Phi^{-1}\left(1 - \frac{\epsilon j}{4n}\right) + (n-1)^{-\frac{1}{4}}. \qquad (A.1)$$

On the one hand, for sufficiently large n and any $1 \leq j \leq n$, we know

$$\frac{1}{n-1}\Phi^{-1}\left(1 - \frac{\epsilon j}{4n}\right) + (n-1)^{-\frac{1}{4}} = \frac{O(1)\sqrt{\log(4n/(\epsilon j))}}{n-1} + (n-1)^{-\frac{1}{4}}$$
$$= O(n^{-\frac{1}{4}}). \qquad (A.2)$$

On the other hand, we have

$$\Phi^{-1}\left(1 - \frac{\epsilon j}{4n}\right) - \Phi^{-1}\left(1 - \frac{\epsilon j}{2n}\right) \gtrsim \frac{2\log 2}{2\sqrt{2\log\frac{4n}{\epsilon j}}}$$

$$\gtrsim \frac{1}{\sqrt{\log n}} \qquad\qquad (\text{A.3})$$

if n is sufficiently large. Combining (A.2) and (A.3) proves (A.1) if $n \geq N$ for some N that only depends on ϵ.

$\square$