Statistica Sinica **27** (2017), 1731-1756
doi:https://doi.org/10.5705/ss.202016.0169

ADAPTIVE FALSE DISCOVERY RATE CONTROL FOR HETEROGENEOUS DATA

Joshua D. Habiger

Oklahoma State University

Abstract: Efforts to develop more efficient multiple hypothesis testing procedures for false discovery rate (FDR) control have focused on incorporating an estimate of the proportion of true null hypotheses (such procedures are called adaptive) or exploiting heterogeneity across tests via some optimal weighting scheme. This paper combines these approaches using a weighted adaptive multiple decision function (WAMDF) framework. Optimal weights for a flexible random effects model are derived and a WAMDF that controls the FDR for arbitrary weighting schemes when test statistics are independent under the null hypotheses is given. Asymptotic and numerical assessment reveals that, under weak dependence, the proposed WAMDFs provide more efficient FDR control even if optimal weights are misspecified. The robustness and flexibility of the proposed methodology facilitates the development of more efficient, yet practical, FDR procedures for heterogeneous data. To illustrate, two different weighted adaptive FDR methods for heterogeneous sample sizes are developed and applied to data.

Key words and phrases: Decision function, multiple testing, p-value, weighted p-value.

1. Introduction

High throughput technology routinely generates data sets that call for hundreds or thousands of null hypotheses to be tested simultaneously. For example, in Anderson and Habiger (2012), RNA sequencing technology was used to measure the abundance of bacteria living near the roots of wheat plants across $i = 1, 2, \ldots, 5$ treatment groups for each of $m = 1, 2, \ldots, M = 778$ bacteria, thereby facilitating the simultaneous testing of 778 null hypotheses. See Table 1 for a depiction of the data, or see Section 8 for more details. See also Efron (2008); Dudoit and van der Laan (2008); Efron (2010) for other, sometimes called, high-dimensional (HD) data sets.

In general, multiple null hypotheses are simultaneously tested with a multiple testing procedure which, ideally, rejects as many null hypotheses as possible subject to the constraint that some global type 1 error rate is controlled at a prespecified level α. The false discovery rate (FDR) is the most frequently considered error rate in the HD setting. It is loosely defined as the expected value of the false discovery proportion (FDP), where the FDP is the proportion of erroneously rejected null hypotheses, also called false discoveries, among rejected null hypotheses, or discoveries. See Sarkar (2007) for other related error rates. In their seminal paper, Benjamini and Hochberg (1995) showed that a step-up procedure based on the Simes (1986) line, henceforth referred to as the BH procedure, has FDR $= \alpha a_0 \leq \alpha$ under a certain dependence structure, where a_0 is the proportion of true null hypotheses. Since then, much research has focused on developing more efficient procedures for FDR control.

One approach seeks to control the FDR at a level nearer α, as opposed to αa_0. For example, adaptive procedures in Benjamini and Hochberg (2000); Storey, Taylor and Siegmund (2004); Benjamini, Krieger and Yekutieli (2006); Gavrilov, Benjamini and Sarkar (2009); Liang and Nettleton (2012) utilize an estimate of a_0 and typically have FDR that is greater than αa_0 yet still less than or equal to α. Finner, Dickhaus and Roters (2009) proposed nonlinear procedures that "exhaust the α" in that, loosely speaking, their FDR converges to α under some least favorable configuration as M tends to infinity.

Another approach aims to exploit heterogeneity across hypothesis tests. Genovese, Roeder and Wasserman (2006); Blanachar and Roquain (2008); Roquain and van de Wiel (2009); Peña, Habiger and Wu (2011) proposed a weighted BH-type procedure, where weights are allowed to depend on the power functions of the individual tests or prior probabilities for the states of the null hypotheses. Storey (2007) considered a "single thresholding procedure" which allowed for heterogeneous data generating distributions. Cai and Sun (2009) and Hu, Zhao and Zhou (2010) provided methods for clustered data, where test statistics are heterogeneous across clusters but homogeneous within clusters, while Sun and McLain (2012) considered heteroscedastic standard errors. Data in Table 1 are heterogeneous because sample sizes $n_1, n_2, \ldots, n_M$ vary from test to test, with n_m being as small as 6 and as large as 911.

Whatever the nature of the heterogeneity may be, recent literature suggests that it should not be ignored. Roeder and Wasserman (2009) showed that weighted multiple testing procedures generally perform favorably over their unweighted counterparts, especially when the employed weights efficiently exploit heterogeneity. Further, Sun and McLain (2012) showed that procedures which ignore heterogeneity can produce lists of discoveries that are of little scientific interest.

Table 1. Depiction of the data in Anderson and Habiger (2012). Shoot biomass x_i in grams for groups $i = 1, 2, \ldots, 5$ was 0.86, 1.34, 1.81, 2.37, and 3.00, respectively. Row totals are in the last column.

Bacteria (m)	Y_{1m}	Y_{2m}	Y_{3m}	Y_{4m}	Y_{5m}	Total (n_m)
1	0	1	1	0	5	7
2	9	2	0	0	3	14
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
778	16	10	29	18	13	81

The objective of this paper is to provide a general approach for exploiting heterogeneity without sacrificing efficient FDR control. The idea is to combine adaptive FDR methods for exhausting the α with weighted procedures for exploiting heterogeneity using a decision theoretic framework. Sections 2 - 5 provide the general framework. Section 2 introduces multiple decision functions (MDFs) and a random effects model that can accommodate many types of heterogeneity including, but not limited to, those mentioned above. Tools which facilitate easy implementation of MDFs, such as weighted p-values, are also developed. Section 3 derives optimal weights for the random effects model and Section 4 introduces an asymptotically optimal weighted adaptive multiple decision function (WAMDF) for asymptotic FDP control. Section 5 provides a WAMDF for exact (nonasymptotic) FDR control.

Assessment in Sections 6 and 7 reveals that, under a weak dependence structure, WAMDFs dominate other MDFs even when weights are misspecified. Specifically, Section 6 shows that the asymptotic FDP of a WAMDF is larger than the FDP of its unadaptive counterpart, yet less than or equal to the nominal level α. Sufficient conditions for "α-exhaustion" are provided and shown to be satisfied in a variety of settings. For example, unweighted adaptive MDFs in Storey, Taylor and Siegmund (2004) and certain asymptotically optimal WAMDFs are α-exhaustive. In fact, α-exhaustion is achieved even in a worst-case-scenario setting, where employed weights are generated independently of optimal weights. Simulation studies in Section 7 demonstrate that WAMDFs are more powerful than competing MDFs as long as the employed weights are positively correlated with optimal weights, and only slightly less powerful in the worse-case-scenario weighting scheme.

Section 8 provides two different routes for implementing WAMDFs in practice and compares them to one another. They are applied to the data in Table 1 and shown analytically and with simulation to perform better than competing unweighted procedures. Concluding remarks are in Section 9 and technical details are in the Supplemental Article.

2. Background

2.1. Data

Let $\boldsymbol{Z} = (Z_m, m \in \mathcal{M})$ for $\mathcal{M} = \{1, 2, \ldots, M\}$ be a random vector of test statistics with joint distribution function F and let $\mathcal{F}$ be a model for F. The basic goal is to test null hypotheses $\boldsymbol{H} = (H_m, m \in \mathcal{M})$ of the form $H_m : F \in \mathcal{F}_m$, where $\mathcal{F}_m \subseteq \mathcal{F}$ is a submodel for $\mathcal{F}$. For short, we often denote the state of H_m by $\theta_m = 1 - I(F \in \mathcal{F}_m)$, where $I(\cdot)$ is the indicator function, so that $\theta_m = 0(1)$ means that H_m is true(false), and denote the state of $\boldsymbol{H}$ by $\boldsymbol{\theta} = (\theta_m, m \in \mathcal{M})$. Let $\mathcal{M}_0 = \{m \in \mathcal{M} : \theta_m = 0\}$ and $\mathcal{M}_1 = \mathcal{M} \setminus \mathcal{M}_0$ index the set of true and false null hypotheses, respectively, and denote the number of true and false null hypotheses by $M_0 = |\mathcal{M}_0|$ and $M_1 = |\mathcal{M}_1|$, respectively.

To make matters concrete, we often consider a random effects model for $\boldsymbol{Z}$. For related models see Efron et al. (2001); Genovese and Wasserman (2002); Storey (2003); Genovese, Roeder and Wasserman (2006); Sun and Cai (2007); Cai and Sun (2009); Roquain and van de Wiel (2009). In Model 1, heterogeneity across the Z_m's is attributable to prior probabilities $\boldsymbol{p} = (p_m, m \in \mathcal{M})$ for the states of the H_m's and parameters $\boldsymbol{\gamma} = (\gamma_m, m \in \mathcal{M})$, which we refer to as effect sizes for ease of exposition, although each γ_m could merely index a distribution for Z_m when H_m is false. See, for example, Section 8.

Model 1. Let $(Z_m, \theta_m, p_m, \gamma_m), m \in \mathcal{M}$, be independent and identically distributed random vectors with support in $\Re \times \{0, 1\} \times [0, 1] \times \Re^+$ and with conditional distribution functions $F(z_m|\theta_m, p_m, \gamma_m) = (1 - \theta_m)F_0(z_m) + \theta_m F_1(z_m|\gamma_m)$ and $F(z_m|p_m, \gamma_m) = (1 - p_m)F_0(z_m) + p_m F_1(z_m|\gamma_m)$. Assume $F(\gamma_m, p_m) = F(\gamma_m)F(p_m)$, $Var(\gamma_m) < \infty$ and that p_m has mean $1 - a_0 \in (0, 1)$.

Observe that Z_m has distribution function $F_0(\cdot)$ given $H_m : \theta_m = 0$ and has distribution function $F_1(\cdot|\gamma_m)$ otherwise. Here, parameters $\boldsymbol{\theta}$, $\boldsymbol{p}$, and $\boldsymbol{\gamma}$ are assumed to be random variables to facilitate asymptotic analysis, as in Genovese, Roeder and Wasserman (2006); Blanachar and Roquain (2008); Blanchard and Roquain (2009); Roquain and van de Wiel (2009); Roquain and Villers (2011). Analysis under Model 1 focuses on conditional distribution functions $F(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\gamma}) = \prod_{m \in \mathcal{M}} F(z_m|\theta_m, p_m, \gamma_m)$ and $F(\boldsymbol{z}|\boldsymbol{p}, \boldsymbol{\gamma}) = \prod_{m \in \mathcal{M}} F(z_m|p_m, \gamma_m)$, and an expectation taken over $\boldsymbol{Z}$ with respect to these distributions is denoted by $E[\cdot|\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\gamma}]$ and $E[\cdot|\boldsymbol{p}, \boldsymbol{\gamma}]$, respectively.

2.2. Multiple decision functions

A multiple decision function (MDF) framework is used to formally define a multiple testing procedure. For similar frameworks see Genovese and Wasserman (2004); Storey, Taylor and Siegmund (2004); Sun and Cai (2007); Peña, Habiger and Wu (2011). Let $\delta_m(Z_m; t_m)$ denote a decision function taking values in $\{0, 1\}$, where $\delta_m = 1(0)$ means that H_m is rejected(retained). A decision function depends functionally on data Z_m and (possibly random) "size threshold" $t_m \in [0, 1]$. To illustrate, suppose that large values of Z_m are evidence against $H_m : \theta_m = 0$ under Model 1. Then we may define

$$\delta_m(Z_m; t_m) = I(Z_m \geq F_0^{-1}(1 - t_m)). \tag{2.1}$$

Observe that $E[\delta_m(Z_m; t_m)|\theta_m = 0] = 1 - F_0(F_0^{-1}(1 - t_m)) = t_m$ so that t_m indeed represents the size of δ_m, hence the terminology "size threshold". An MDF is denoted $\boldsymbol{\delta}(\boldsymbol{Z}; \boldsymbol{t}) = [\delta_m(Z_m; t_m), m \in \mathcal{M}]$, where $\boldsymbol{t} = (t_m, m \in \mathcal{M})$ is called a threshold vector. If $t_m = \alpha/M$ for each m then $\boldsymbol{\delta}(\boldsymbol{Z}; \boldsymbol{t})$ represents the well-known Bonferroni procedure.

Assume that, for each m, $t_m \mapsto \delta_m(Z_m; t_m)$ is nondecreasing and right continuous with $\delta_m = 0(1)$ whenever $t_m = 0(1)$, almost surely, and that $t_m \mapsto E[\delta_m(Z_m; t_m)]$ is continuous and strictly increasing for $t_m \in (0, 1)$, with $E[\delta_m(Z_m; t_m)] = t_m$ whenever $m \in \mathcal{M}_0$. These assumptions are referred to as the nondecreasing-in-size (NS) assumptions and are satisfied, for example, under Model 1 for decision functions defined as in (2.1). For additional details and examples see Habiger and Peña (2011); Peña, Habiger and Wu (2011); Habiger (2012).

2.3. Tools for implementation

We break $\boldsymbol{t}$ down into the product of a positive valued weight vector $\boldsymbol{w} = (w_m, m \in \mathcal{M})$ satisfying $\bar{w} = M^{-1} \sum_{m \in \mathcal{M}} w_m = 1$ and an overall or average threshold t, $\boldsymbol{t} = t\boldsymbol{w}$. First, weights are specified and then data $\boldsymbol{Z} = \boldsymbol{z}$ are collected, the overall threshold t is computed, and the MDF $\boldsymbol{\delta}(\boldsymbol{z}; t\boldsymbol{w})$ is computed. If weights are based on Model 1, for example, then they are allowed to depend functionally on $\boldsymbol{p}$ and $\boldsymbol{\gamma}$. The overall threshold is allowed to depend functionally on $\boldsymbol{z}$ and $\boldsymbol{w}$.

It is useful to exploit the link between weighted p-values and decision functions. Define the (unweighted) *p-value* statistic corresponding to δ_m by

$$P_m = \inf\{t_m \in [0, 1] : \delta_m(Z_m; t_m) = 1\}.$$

This definition, see Habiger and Peña (2011); Peña, Habiger and Wu (2011), has the usual interpretation that P_m is the smallest size t_m allowing for H_m to be rejected, and ensures that $\delta_m(Z_m; t_m) = I(P_m \leq t_m)$ almost surely under the NS assumptions. For example, it can be verified that the p-value statistic corresponding to (2.1) is $P_m = 1 - F_0(Z_m)$ and that $I(Z_m \leq F_0^{-1}(1 - t_m)) = I(P_m \leq t_m)$ almost surely. See Habiger (2012); Habiger and Peña (2014) for more details or for derivations of more complex p-values, such as the p-value for the local FDR statistic in Efron et al. (2001); Sun and Cai (2007) or for the optimal discovery procedure in Storey (2007). Define the *weighted p-value* statistic by

$$Q_m = \inf\{t : \delta_m(Z_m; tw_m) = 1\}.$$

For w_m fixed, and writing $t_m = tw_m$,

$$P_m = \inf\{tw_m : \delta_m(Z_m; tw_m) = 1\} = w_m \inf\{t : \delta_m(Z_m; tw_m) = 1\} = w_m Q_m$$

almost surely. Thus, a weighted p-value can be computed by $Q_m = P_m/w_m$. Hence, we have established the almost surely equivalent expressions for a decision function under the NS assumptions:

$$\delta_m(Z_m; t_m) = \delta_m(Z_m; tw_m) = I(P_m \leq tw_m) = I(Q_m \leq t). \tag{2.2}$$

3. Optimal Weights

Though results regarding exact FDR control in Section 5 or asymptotic FDP control in Section 6.1 apply more generally (see assumptions (A3) and (A4) - (A6), respectively), optimal weights in this paper are developed for Model 1. We first derive optimal weights assuming that t is fixed/known.

3.1. Optimal fixed-t weights

We consider $\boldsymbol{\delta}(\boldsymbol{Z}; \boldsymbol{t})$ and the constraint that $\bar{w} = 1$ is replaced with the constraint that $\bar{t} = t$, where $\bar{t} = M^{-1} \sum_{m \in \mathcal{M}} t_m$. As weights are allowed to depend on $\boldsymbol{p}$ and $\boldsymbol{\gamma}$ under Model 1, the focus is on the conditional expectation of $\delta_m(Z_m; t_m)$ denoted by $G_m(t_m) \equiv E[\delta_m(Z_m; t_m)|\boldsymbol{p}, \boldsymbol{\gamma}] = (1-p_m)t_m + p_m \pi_{\gamma_m}(t_m)$, where $\pi_{\gamma_m}(t_m) = E[\delta_m(Z_m; t_m)|\theta_m = 1, \gamma_m]$ is the power function for δ_m. As in Genovese, Roeder and Wasserman (2006); Roquain and van de Wiel (2009); Peña, Habiger and Wu (2011), assume power functions (as a function of t_m) are concave.

(A1) For each $m \in \mathcal{M}$, $t_m \mapsto \pi_{\gamma_m}(t_m)$ is concave and twice differentiable for $t_m \in (0, 1)$, with $\lim_{t_m \uparrow 1} \pi'_{\gamma_m}(t_m) = 0$ and $\lim_{t_m \downarrow 0} \pi'_{\gamma_m}(t_m) = \infty$ almost surely, where $\pi'_{\gamma_m}(t_m)$ is the derivative of $\pi_{\gamma_m}(t_m)$ with respect to t_m.

This concavity condition is satisfied, for example, under monotone likelihood ratio considerations (Peña, Habiger and Wu (2011)) and under the generalized monotone likelihood ratio (GMLR) condition in Cao, Sun and Kosorok (2013).

Given $\boldsymbol{p}$, $\boldsymbol{\gamma}$, and t, the goal is to maximize the expected number of correctly rejected null hypotheses

$\pi(\boldsymbol{t}, \boldsymbol{p}, \boldsymbol{\gamma}) \equiv E\left[\sum_{m\in\mathcal{M}} \theta_m \delta_m(Z_m; t_m) \Big| \boldsymbol{\gamma}, \boldsymbol{p}\right] = \sum_{m\in\mathcal{M}} p_m \pi_{\gamma_m}(t_m)$ subject to the constraint that $\bar{t} = t$.

Theorem 1. *Suppose that* (A1) *is satisfied, and fix* $t \in (0,1)$. *Then under Model 1 the maximum of* $\pi(\boldsymbol{t}, \boldsymbol{p}, \boldsymbol{\gamma})$ *with respect to* $\boldsymbol{t}$ *subject to constraint* $\bar{t} = t$ *exists, is unique, and satisfies*

$$\pi'_{\gamma_m}(t_m) = k/p_m \tag{3.1}$$

for every $m \in \mathcal{M}$ *and some* $k > 0$.

Spjøtvoll (1972) and Storey (2007) also derived expressions for optimal fixed-t thresholds, but did not allow for the states of the H_m's to be random. Specifically, Spjøtvoll (1972) proposed maximizing $\sum_{m\in\mathcal{M}} \pi_{\gamma_m}(t_m)$ (see Roeder and Wasserman (2009) for an illustration in the normal distribution setting) while Storey (2007) proposed maximizing $\sum_{m\in\mathcal{M}} \theta_m \pi_{\gamma_m}(t_m)$.

The important quantity in (3.1) is the constant k. In particular it suffices to find the unique value of k, say k^*, that satisfies $\bar{t} = t$. For any value of k denote the (unique) solution to (3.1) in terms of t_m as $t_m(k/p_m, \gamma_m)$, and take $\boldsymbol{t}(k, \boldsymbol{p}, \boldsymbol{\gamma}) = [t_m(k/p_m, \gamma_m), m \in \mathcal{M}]$. Then to compute weights

1. find the k^* satisfying $\bar{t}_M(k^*, \boldsymbol{p}, \boldsymbol{\gamma}) = t$, where $\bar{t}_M(k, \boldsymbol{p}, \boldsymbol{\gamma}) = M^{-1} \sum_{m\in\mathcal{M}} t_m (k/p_m, \gamma_m)$,

2. compute each optimal fixed-t weight
$$w_m(k^*, \boldsymbol{p}, \boldsymbol{\gamma}) = \frac{t_m(k^*/p_m, \gamma_m)}{\bar{t}_M(k^*, \boldsymbol{p}, \boldsymbol{\gamma})}. \tag{3.2}$$

We sometimes denote $w_m(k^*, \boldsymbol{p}, \boldsymbol{\gamma})$ by w_m^* and the vector of optimal fixed-t weights $\boldsymbol{w}(k^*, \boldsymbol{p}, \boldsymbol{\gamma}) = [w_m(k^*, \boldsymbol{p}, \boldsymbol{\gamma}), m \in \mathcal{M}]$ by $\boldsymbol{w}^* = (w_m^*, m \in \mathcal{M})$.

To better understand how the solution is found and related to the values of p_m, γ_m and t consider an example.

Example 1. Suppose $Z_m | \gamma_m, \theta_m \sim N(\theta_m \gamma_m, 1)$ for $\gamma_m > 0$ and consider testing $H_m : \theta_m = 0$. Denote the standard normal cumulative distribution function and density function by $\Phi(\cdot)$ and $\phi(\cdot)$, respectively, and let $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$. Take $\delta_m(Z_m; t_m) = I(Z_m \geq \bar{\Phi}^{-1}(t_m))$. The power function is $\pi_{\gamma_m}(t_m) = \bar{\Phi}(\bar{\Phi}^{-1}(t_m) - \gamma_m)$ and has derivative $\pi'_{\gamma_m}(t_m) = (\phi(\bar{\Phi}^{-1}(t_m) - \gamma_m))/(\phi(\bar{\Phi}^{-1}(t_m)))$. Setting the

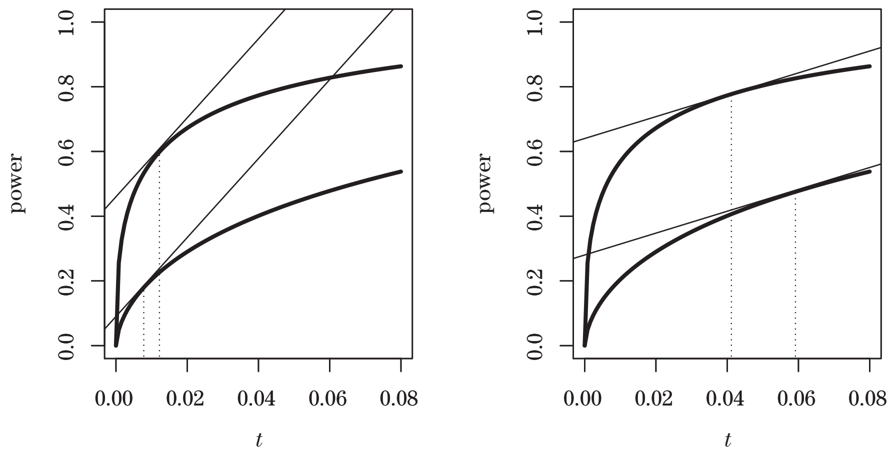

Figure 1. A depiction of the optimal thresholds for $M = 2$ hypotheses tests when power functions vary under constraint $0.5(t_1 + t_2) = 0.01$ (left) and $0.5(t_1 + t_2) = 0.05$ (right).

derivative equal to k/p_m and solving yields

$$t_m\left(\frac{k}{p_m}, \gamma_m\right) = \bar{\Phi}\left(0.5\gamma_m + \frac{\log(k/p_m)}{\gamma_m}\right). \tag{3.3}$$

The optimal fixed-t threshold vector is computed as $\boldsymbol{t}(k^*, \boldsymbol{p}, \boldsymbol{\gamma})$, where k^* satisfies $\bar{t}_M(k^*, \boldsymbol{p}, \boldsymbol{\gamma}) = t$, and the optimal fixed-$t$ weights are computed as in (3.2).

Observe in (3.3) that $t_i(k/p_i, \gamma_i) = t_j(k/p_j, \gamma_j)$ if $\gamma_i = \gamma_j$ and $p_i = p_j$ regardless of k and, consequently, the optimal fixed-t weight vector is $\mathbf{1}$ for any t when data are homogeneous. On the other hand, we see that $t_m(k/p_m, \gamma_m)$ is increasing in p_m and hence

$$w_m(k^*, \boldsymbol{p}, \boldsymbol{\gamma}) = M\frac{t_m(k^*/p_m, \gamma_m)}{t_m(k^*/p_m, \gamma_m) + \sum_{j\neq m} t_j(k^*/p_j, \gamma_j)}$$

is increasing in p_m, as we might expect.

The relationship between $w_m(k^*, \boldsymbol{p}, \boldsymbol{\gamma})$ and γ_m is more complex. To illustrate, consider testing $M = 2$ null hypotheses and suppose $\gamma_1 = 1.5$, $\gamma_2 = 2.5$, and $p_1 = p_2 = 0.5$. In Figure 1, observe that for $t = 0.01$, $\bar{t}_M(k^*, \boldsymbol{p}, \boldsymbol{\gamma}) = 0.01$ when $k^* = 6.1$, which gives $t_1(k^*/p_1, \gamma_1) = 0.003$, $t_2(k^*/p_2, \gamma_2) = 0.017$, $w_1^* = 0.003/0.01 = 0.3$ and $w_2^* = 0.017/0.01 = 1.7$. Because $p_1 = p_2$, the slopes of the power functions evaluated at 0.003 and 0.017, respectively, are equal; see equation (3.1). Now consider the fixed threshold $t = 0.05$. Here $k^* = 1.7$, which leads to weights $w_1^* = 0.059/0.05 = 1.18$ and $w_2^* = 0.041/0.05 = 0.82$. Thus, when $t = 0.01$, the hypothesis with the larger effect size is given more weight, but when $t = 0.05$ it is given less weight. For a more detailed discussion on this phenomenon see Peña, Habiger and Wu (2011). The important point is that the optimal fixed-t weights are only implementable if t is fixed or specified before data collection.

3.2. Asymptotically optimal weights

The overall threshold t in Section 4 depends on data $\boldsymbol{Z}$ because it depends on the FDP estimator, which depends functionally on $\boldsymbol{Z}$; see (4.1) and (4.2). The idea in this subsection is to approximate the FDP estimator using $\boldsymbol{p}$ and $\boldsymbol{\gamma}$. This allows t to be approximated before data collection so that the optimal fixed-t weights can be utilized.

The FDP "approximator" plugs $G_m(t_m(k/p_m,\gamma_m)) = E[\delta_m(Z_m; t_m(k/p_m,\gamma_m))|\boldsymbol{p},\boldsymbol{\gamma}]$ in for each δ_m in (4.1) and (4.2). Formally, write $\bar{G}_M(\boldsymbol{t}(k,\boldsymbol{p},\boldsymbol{\gamma})) = M^{-1}\sum_{m\in\mathcal{M}} G_m(t_m(k/p_m,\gamma_m))$ and define the FDP approximator by

$$\widetilde{FDP}_M(\boldsymbol{t}(k,\boldsymbol{p},\boldsymbol{\gamma})) = \frac{1-\bar{G}_M(\boldsymbol{t}(k,\boldsymbol{p},\boldsymbol{\gamma}))}{1-\bar{t}_M(k,\boldsymbol{p},\boldsymbol{\gamma})}\frac{\bar{t}_M(k,\boldsymbol{p},\boldsymbol{\gamma})}{\bar{G}_M(\boldsymbol{t}(k,\boldsymbol{p},\boldsymbol{\gamma}))}.$$

Now, the asymptotically optimal weights are computed as follows.

Weight selection procedure: *For* $0 < \alpha \le 1 - p_{(M)}$, *where* $p_{(M)} = \max\{\boldsymbol{p}\}$,

a. get $k_M^* = \inf\left\{k : \widetilde{FDP}_M(\boldsymbol{t}(k,\boldsymbol{p},\boldsymbol{\gamma})) = \alpha\right\}$, *and*

b. for each $m \in \mathcal{M}$, *compute* $w_m^* = w_m(k_M^*,\boldsymbol{p},\boldsymbol{\gamma})$ *as in* (3.2).

In Theorem 2 we find that the restriction $0 < \alpha \le 1 - p_{(M)}$ ensures that a solution to $\widetilde{FDP}_M(\boldsymbol{t}(k,\boldsymbol{p},\boldsymbol{\gamma})) = \alpha$ exists. In practice, this restriction amounts to choosing α and $\boldsymbol{p}$ so that $0 < \alpha \le 1 - p_m$ for each m. That is, the prior probability that the null hypothesis is true should be at least α, which is reasonable in practice.

Theorem 2. *Under* (A1) *and Model 1,* k_M^* *exists for* $0 < \alpha \le 1 - p_{(M)}$.

Observe that $\bar{t}_M(k_M^*,\boldsymbol{p},\boldsymbol{\gamma}) = t$ for some $t \in (0,1)$ so that indeed these weights could be viewed as optimal fixed-t weights. However, here weight computation is based on the constraint $\widetilde{FDP}_M(\boldsymbol{t}(k_M^*,\boldsymbol{p},\boldsymbol{\gamma})) = \alpha$. These weights are henceforth referred to as asymptotically optimal for reasons that will be formalized later.

4. The Procedure

Now we are now in position to formally define the proposed adaptive threshold which, when used in conjunction with asymptotically optimal weights in $\boldsymbol{\delta}(\boldsymbol{Z}; t\boldsymbol{w})$, yields the asymptotically optimal WAMDF.

4.1. Threshold selection

For the moment, let $\boldsymbol{w}$ be any fixed vector of positive weights satisfying $\bar{w} = 1$. For brevity, we sometimes suppress the Z_m in each δ_m and write $\delta_m(tw_m)$ and denote $\boldsymbol{\delta}(\boldsymbol{Z}; t\boldsymbol{w})$ by $\boldsymbol{\delta}(t\boldsymbol{w})$. Further, denote the number of discoveries at $t\boldsymbol{w}$ by $R(t\boldsymbol{w}) = \sum_{m\in\mathcal{M}} \delta_m(tw_m)$.

We make use of an "adaptive" estimator of the FDP that utilizes an estimator of M_0 defined by

$$\hat{M}_0(\lambda\boldsymbol{w}) = \frac{M - R(\lambda\boldsymbol{w}) + 1}{1-\lambda} \tag{4.1}$$

for some fixed tuning parameter $\lambda \in (0,1)$. This estimator is essentially the weighted version of the estimator in Storey (2002) defined by $\hat{M}_0(\lambda\mathbf{1}) = [M - R(\lambda\mathbf{1})]/[1-\lambda]$. For earlier work on the estimation of M_0, see Schweder and Spjotvoll (1982). As outlined in Storey, Taylor and Siegmund (2004) in the unweighted setting, the idea is that for $m \in \mathcal{M}_1$, $E[\delta_m(\lambda)] \leq 1$, but the inequality is relatively sharp if all tests have reasonable power, which should be the case for large enough λ. Hence

$$E[M - R(\lambda\mathbf{1})] = \sum_{m\in\mathcal{M}} E[1-\delta_m(\lambda)] \geq \sum_{m\in\mathcal{M}_0} E[1-\delta_m(\lambda)] = (1-\lambda)M_0$$

and $E[\hat{M}_0(\lambda\mathbf{1})] \geq M_0$. That is, $\hat{M}_0$ is positively biased but the bias is minor. Similar intuition applies for $\hat{M}_0(\lambda\boldsymbol{w})$. As in Storey, Taylor and Siegmund (2004), we add 1 to the numerator in (4.1) to ensure that $\hat{M}_0(\lambda\boldsymbol{w}) > 0$ for finite sample results.

The adaptive FDP estimator is defined by

$$\widehat{FDP}^{\lambda}(t\boldsymbol{w}) = \frac{\hat{M}_0(\lambda\boldsymbol{w})t}{\max\{R(t\boldsymbol{w}), 1\}}. \tag{4.2}$$

The adaptive threshold, which essentially chooses t as large as possible subject to the constraint that the estimate of the FDP is less than or equal to α, is defined by

$$\hat{t}_{\alpha}^{\lambda} = \sup\{0 \leq t \leq u : \widehat{FDP}^{\lambda}(t\boldsymbol{w}) \leq \alpha\}. \tag{4.3}$$

We assume that u, the upper bound for $\hat{t}_{\alpha}^{\lambda}$, and the tuning parameter λ satisfy

(A2) $\lambda \leq u \leq 1/w_{(M)}$,

where $w_{(M)} \equiv \max\{\boldsymbol{w}\}$. This ensures that $\hat{t}_{\alpha}^{\lambda} w_m \leq 1$ and $\lambda w_m \leq 1$ for every m. For $\boldsymbol{w} = \mathbf{1}$ and $u = \lambda$ (which implies $\hat{t}_{\alpha}^{\lambda} \leq \lambda$), we recover the unweighted adaptive MDF for finite FDR control in Storey, Taylor and Siegmund (2004).

In practice $\hat{t}_{\alpha}^{\lambda}$ can be difficult to compute. Alternatively, we can apply the original BH procedure to the weighted p-values at level $\alpha M/\hat{M}_0(\lambda\boldsymbol{w})$. Due to (2.2), we can also use weighted p-values to estimate M_0 via $\hat{M}_0(\lambda\boldsymbol{w}) = [M - \sum_{m\in\mathcal{M}} I(Q_m \leq \lambda) + 1]/[1-\lambda]$. This threshold selection procedure can be implemented as follows.

Threshold selection procedure: *Fix* λ *and* u *satisfying* (A2). *Then*

a. *compute* $Q_m = P_m/w_m$ *and ordered weighted p-values via* $Q_{(1)} \leq Q_{(2)} \leq \cdots \leq Q_{(M)}$.

b. *If* $Q_{(m)} > \alpha m/\hat{M}_0(\lambda\boldsymbol{w})$ *for each* m, *set* $j = 0$, *otherwise take*

$$j = \max\left\{m \in \mathcal{M} : Q_{(m)} \leq \frac{\alpha m}{\hat{M}_0(\lambda\boldsymbol{w})}\right\}.$$

c. *Get* $\hat{t}_\alpha^{\lambda *} = \min\{j\alpha/\hat{M}_0(\lambda\boldsymbol{w}), u\}$ *and reject* H_m *if* $Q_m \leq \hat{t}_\alpha^{\lambda *}$.

The WAMDF implemented above is equivalent to $\boldsymbol{\delta}(\boldsymbol{Z}; \hat{t}_\alpha^\lambda \boldsymbol{w})$ in that

$$\delta_m(Z_m; \hat{t}_\alpha^\lambda w_m) = I(Q_m \leq \hat{t}_\alpha^\lambda) = I(Q_m \leq \hat{t}_\alpha^{\lambda *}) \tag{4.4}$$

almost surely for each m, so both procedures reject the same set of null hypotheses. The first equality in (4.4) follows from (2.2) and the last equality in (4.4) is a consequence of Lemma 2 in Storey, Taylor and Siegmund (2004).

4.2. The asymptotically optimal WAMDF

The asymptotically optimal WAMDF is formally defined as $\boldsymbol{\delta}(\boldsymbol{Z}; \hat{t}_\alpha^\lambda \boldsymbol{w}^*)$ for $0 < \alpha \leq 1 - p_{(M)}$ and $\lambda = \bar{t}_M(k_M^*, \boldsymbol{p}, \boldsymbol{\gamma})$, where k_M^* and $\boldsymbol{w}^*$ are defined as in the Weight Selection Procedure. This particular choice of λ ensures that the employed weights are indeed "asymptotically optimal" (see Theorem 8) and additionally that (A2) is satisfied if we take $u = 1/w_{(M)}$. Other values of λ could be considered, as in Section 8. To implement the the asymptotically optimal WAMDF, we compute $\boldsymbol{w}^*$ using the Weight Selection Procedure, then choose $\lambda = \bar{t}_M(k_M^*, \boldsymbol{p}, \boldsymbol{\gamma})$ and u satisfying (A2), collect data $\boldsymbol{Z} = \boldsymbol{z}$, and compute $\boldsymbol{\delta}(\boldsymbol{z}; \hat{t}_\alpha^\lambda \boldsymbol{w}^*)$ using the Threshold Selection Procedure.

To illustrate, consider testing $M = 10$ null hypotheses under the setting outlined in Example 1, with $p_m = 0.5$ for $m = 1, 2, \ldots, 10$, $\gamma_m = 2$ for $m = 1, 2, \ldots, 5$, $\gamma_m = 3$ for $m = 6, 7, \ldots, 10$, and $\alpha = 0.05$. The goal is to test $H_m : \theta_m = 0$ with decision functions $\delta_m(Z_m; t_m) = I(Z_m \geq \bar{\Phi}^{-1}(t_m))$ or their corresponding p-values $P_m = \bar{\Phi}(Z_m)$ and weighted p-values $Q_m = P_m/w_m$. See Table 2 for summaries of parameters, weights, simulated data, p-values and weighted p-values. The Weight Selection Procedure is broken down into 2 sub-steps and the Threshold Selection Procedure is split into three sub-steps. To test these null

Table 2. A portion of the parameters, data, weights, p-values, and weighted p-values in columns 1 - 5, respectively. Each row is sorted in ascending order according to $Q_1, Q_2, \ldots, Q_M$.

θ_m	γ_m	w_m^*	Z_m	P_m	Q_m	$0.05m/\hat{M}_0$
1	3	0.74	3.14	0.001	0.001	0.006
1	2	1.26	2.55	0.005	0.005	0.012
1	3	0.74	2.56	0.005	0.006	0.018
1	2	1.26	1.47	0.070	0.062	0.024
0	2	1.74	1.17	0.121	0.106	0.030
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
0	3	0.74	-0.60	0.724	0.844	0.061

hypotheses we

1a. specify $\boldsymbol{\gamma}$ (see column 2 of Table 2), $\boldsymbol{p}$ and α and find $k_M^* = 2.52$.

1b. Compute asymptotically optimal weights $w_m^* = w_m(k_M^*, \boldsymbol{p}, \boldsymbol{\gamma})$ as in (3.2). See column 3 in Table 2.

2a. Take $\lambda = \bar{t}_M(k_M^*, \boldsymbol{p}, \boldsymbol{\gamma}) = 0.028$ and $u = 1/1.26 = 0.79$. Collect data $\boldsymbol{Z} = \boldsymbol{z}$ and compute and order weighted p-values (see columns 4 - 6 in Table 2).

2b. Observe that $Q_{(m)} \leq \alpha m/\hat{M}_0(\lambda \boldsymbol{w}^*)$ for $m = 3$ but not for $m = 4, 5, \ldots, 10$ and hence $\alpha j/\hat{M}_0(\lambda \boldsymbol{w}^*) = 0.05(3/8.23) = 0.013$.

2c. Compute $\hat{t}_\alpha^{\lambda *} = \min\{0.013, 0.79\} = 0.013$ and reject null hypotheses with weighted p-values 0.001, 0.005 and 0.006 because they are less than 0.013.

5. Finite FDR Control

An upper bound for the FDR is given for arbitrary weights satisfying $w_m > 0$ for each m and $\bar{w} = 1$. The bound is computed under a dependence structure for $\boldsymbol{Z}$:

(A3) $(Z_m, m \in \mathcal{M}_0)$ are mutually independent and independent of $(Z_m, m \in \mathcal{M}_1)$.

This structure has been utilized in Benjamini and Hochberg (1995); Genovese, Roeder and Wasserman (2006); Peña, Habiger and Wu (2011); Storey, Taylor and Siegmund (2004) to prove FDR control for unweighted unadaptive, weighted unadaptive, and unweighted adaptive procedures. It is satisfied under Model 1 conditionally upon $(\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\gamma})$, but it is not limited to this setting.

To define the FDR, let $V(t\boldsymbol{w}) = \sum_{m\in\mathcal{M}_0} \delta_m(tw_m)$ denote the number of erroneously rejected null hypotheses (false discoveries) at $t\boldsymbol{w}$, with $R(t\boldsymbol{w}) = \sum_{m\in\mathcal{M}} \delta_m(t\boldsymbol{w})$ the number of rejected null hypotheses. Define the FDP at $t\boldsymbol{w}$ by

$$FDP(t\boldsymbol{w}) = \frac{V(t\boldsymbol{w})}{\max\{R(t\boldsymbol{w}), 1\}}. \tag{5.1}$$

The FDR at $t\boldsymbol{w}$ is defined by $FDR(t\boldsymbol{w}) = E[FDP(t\boldsymbol{w})]$, where the expectation is taken over $\boldsymbol{Z}$ with respect to an arbitrary $F \in \mathcal{F}$.

The bound is presented in Lemma 1. The focus is on the setting when $M_0 \geq 1$ because the FDR is trivially 0 if $M_0 = 0$. As in Storey, Taylor and Siegmund (2004), we force $\hat{t}^{\lambda}_{\alpha} \leq \lambda$ by taking $u = \lambda$ in (4.3). This facilitates the use of the Optional Stopping Theorem in the proof.

Lemma 1. *Suppose $M_0 \geq 1$ and that* (A2) *and* (A3) *are satisfied. Then for $u = \lambda$,*

$$FDR(\hat{t}^{\lambda}_{\alpha}\boldsymbol{w}) \leq \alpha\bar{w}_0 \frac{1-\lambda}{1-\lambda\bar{w}_0}[1-(\lambda\bar{w}_0)^{M_0}] \leq \alpha\bar{w}_0\frac{1-\lambda}{1-\lambda\bar{w}_0}, \tag{5.2}$$

where $\bar{w}_0 = M_0^{-1}\sum_{m\in\mathcal{M}_0} w_m$ is the mean of the weights from true null hypotheses.

Observe that $1-(\lambda\bar{w}_0)^{M_0} \leq 1$ due to (A2). Further, if $\boldsymbol{w} = \boldsymbol{1}$ then $\bar{w}_0 = 1$ and we recover Theorem 3 in Storey, Taylor and Siegmund (2004) as a corollary.

If $\boldsymbol{w} \neq \boldsymbol{1}$, the bound in Lemma 1 is not immediately applicable because $\mathcal{M}_0$, and consequently $\bar{w}_0$, is unobservable. One solution is to use an upper bound for $\bar{w}_0$ and adjust the "α" at which the procedure is applied. This adjustment is described below.

Theorem 3. *If*

$$\alpha^* = \alpha\frac{1}{w_{(M)}}\frac{1-\lambda w_{(M)}}{1-\lambda},$$

then under the conditions of Lemma 1*, $FDR(\hat{t}^{\lambda}_{\alpha^*}\boldsymbol{w}) \leq \alpha$.*

As $\bar{w}_0$ is typically less than or equal to 1, asymptotically, this α adjustment is not needed for large M.

6. Asymptotic Results

We show that WAMDFs always reject more null hypotheses than their unadaptive counterparts, and provide sufficient conditions for asymptotic FDP control and α-exhaustion. These results are then used in the asymptotic analysis of the asymptotically optimal WAMDF.

To facilitate asymptotic analysis, denote weight vectors of length M by $\boldsymbol{w}_M$ and the mth element of $\boldsymbol{w}_M$ by $w_{m,M}$. Write the mean of the weights from true null hypotheses as $\bar{w}_{0,M}$. Denote the adaptive FDP estimator in (4.2) by $\widehat{FDP}^{\lambda}_M(t\boldsymbol{w}_M)$ and the FDP in (5.1) by $FDP_M(t\boldsymbol{w}_M)$. We also consider an unadaptive FDP estimator that uses M in the place of an estimate of M_0, defined by

$$\widehat{FDP}^{0}_M(t\boldsymbol{w}_M) = \frac{Mt}{\max\{R(t\boldsymbol{w}_M), 1\}}.$$

When necessary, we denote the tuning parameter in (4.1) by λ_M because, as in the asymptotically optimal WAMDF where $\lambda_M = \bar{t}_M(k^*_M, \boldsymbol{p}, \boldsymbol{\gamma})$, it may depend on M.

For asymptotic analysis, (A2) is redefined:

(A2) $\lambda_M \to \lambda \leq u = 1/k$ almost surely, where k satisfies $\lim_{M\to\infty} w_{(M)} \leq k$ almost surely.

The adaptive threshold in (4.3) is denoted $\hat{t}^{\lambda}_{\alpha,M}$. We find that (A2) is satisfied, for example, under Model 1 and (A1) for the asymptotically optimal WAMDF. The unadaptive threshold is defined by

$$\hat{t}^{0}_{\alpha,M} = \sup\{0 \leq t \leq u : \widehat{FDP}^{0}_M(t\boldsymbol{w}_M) \leq \alpha\}.$$

6.1. Arbitrary weights

Convergence criteria considered here are similar to criteria in Storey, Taylor and Siegmund (2004); Genovese, Roeder and Wasserman (2006) and allow for weak dependence structures. See Billingsley (1999), Storey (2003), or see Theorem 7 for examples. For u defined as in (A2) and $t \in (0, u]$, we assume the following.

(A4) $R(t\boldsymbol{w}_M)/M \to G(t)$ almost surely.

(A5) $V(t\boldsymbol{w}_M)/M \to a_0\mu_0 t$ almost surely, for $0 < \mu_0 < \infty$ and $0 < a_0 < 1$, where $\bar{w}_{0,M} \to \mu_0$ and $M_0/M \to a_0$.

(A6) $t/G(t)$ is strictly increasing and continuous over (0,u) with $\lim_{t\downarrow 0} t/G(t) = 0$ and $\lim_{t\uparrow u} u/G(u) \leq 1$.

Here μ_0 is the asymptotic mean of the weights corresponding to true null hypotheses and a_0 is the asymptotic proportion of true null hypotheses. The last condition is natural as it ensures that, asymptotically, the FDP is continuous and increasing in t and takes on value 0, thereby ensuring that it can be controlled. Writing $R(t\boldsymbol{w}_M)/M = \sum_{m\in\mathcal{M}} I(Q_m \leq t)/M$ via (2.2), we see that (A4) corresponds to the assumption that the empirical process of the weighted p-values converges pointwise to $G(t)$ almost surely.

Asymptotic analysis for arbitrary weights focuses on comparing random thresholds $\hat{t}^{\lambda}_{\alpha,M}$ and $\hat{t}^{0}_{\alpha,M}$ to their corresponding asymptotic (nonrandom) thresholds, which are based on the limits of the unadaptive and adaptive FDP estimators. Denote the pointwise limits of the unadaptive FDP estimator, the adaptive FDP estimator, and the FDP by

$$FDP^0_\infty(t) = \frac{t}{G(t)}, \quad FDP^\lambda_\infty(t) = \frac{1-G(\lambda)}{1-\lambda}\frac{t}{G(t)}, \quad \text{and} \quad FDP_\infty(t) = \frac{a_0\mu_0 t}{G(t)},$$

respectively (see Lemma S1 in the Supplemental Article for verification and details). Define asymptotic unadaptive and asymptotic adaptive thresholds by, respectively,

$$t^0_{\alpha,\infty} = \sup\{0 \leq t \leq u : FDP^0_\infty(t) \leq \alpha\},$$
$$\text{and } t^\lambda_{\alpha,\infty} = \sup\{0 \leq t \leq u : FDP^\lambda_\infty(t) \leq \alpha\}.$$

The unadaptive and adaptive thresholds converge to their asymptotic (nonrandom) counterparts, with the asymptotic adaptive threshold larger than the asymptotic unadaptive threshold. As $E[\delta_m(tw_m)]$ is strictly increasing in t for each m, it follows that the adaptive procedure leads to a higher proportion of rejected null hypotheses, asymptotically. Our result generalizes Corollary 2 in Storey, Taylor and Siegmund (2004), which focused on the unweighted setting.

Theorem 4. *Fix $\alpha \in (0,1)$. Then under* (A2) *and* (A4) - (A6)*, almost surely,*

$$\lim_{M\to\infty} \hat{t}^0_{\alpha,M} = t^0_{\alpha,\infty} \leq \lim_{M\to\infty} \hat{t}^\lambda_{\alpha,M} = t^\lambda_{\alpha,\infty}. \tag{6.1}$$

It is useful to formally describe the notion of an α-exhaustive MDF. Loosely speaking, Finner, Dickhaus and Roters (2009) referred to an unweighted multiple decision function, say $\boldsymbol{\delta}(\hat{t}^*_{\alpha,M}\mathbf{1}_M)$, as "asymptotically optimal" (we will use the terminology α-exhaustive) if $FDR(\hat{t}^*_{\alpha,M}\mathbf{1}_M) \to \alpha$ under some least favorable distribution. A Dirac Uniform (DU) distribution was shown to often be least favorable for the FDR in that, among all Fs that satisfy $E[\delta_m(t)] = t$ for every $t \in [0,1]$ when $m \in \mathcal{M}_0$ and dependency structure (A3), $FDR(\hat{t}^*_{\alpha,M}\mathbf{1}_M)$ is the largest under a DU distribution. In our notation, a DU distribution is any distribution satisfying $E[\delta_m(t)] = t$ if $m \in \mathcal{M}_0$ and $E[\delta_m(t)] = 1$ otherwise. If (A4) - (A5) are satisfied, then $G(t) = a_0\mu_0 t + (1 - a_0)$ under a DU distribution for $t \leq u$. Write this $G(t)$ as $G^{DU}(t)$.

To study the FDP of WAMDFs consider

$$\lim_{M\to\infty} FDP_M(\hat{t}^0_{\alpha,M}\boldsymbol{w}_M) \leq \lim_{M\to\infty} FDP_M(\hat{t}^\lambda_{\alpha,M}\boldsymbol{w}_M) \leq \alpha \tag{6.2}$$

and three claims regarding these inequalities.

(C1) The first inequality in (6.2) is satisfied almost surely.

(C2) The second inequality in (6.2) is satisfied almost surely.

(C3) The second inequality in (6.2) is an equality almost surely under a DU distribution.

Informally, Claim (C1) states that the FDP of the WAMDF is asymptotically always larger than the FDP of its unadaptive counterpart and is referred to as the *asymptotically less conservative* claim. Claim (C2) states that the WAMDF has asymptotic FDP that is less than or equal to α and is referred to as the *asymptotic FDP control* claim. Claim (C3) is the α-*exhaustive* claim and states that the asymptotic FDP of the WAMDF is equal to α under a DU distribution. Theorem 5 provides sufficient conditions for each claim.

Theorem 5. *Fix* $\alpha \in (0,1)$ *and suppose that* (A2) *and* (A4) - (A6) *are satisfied. Then Claim* (C1) *holds. Claim* (C2) *holds if, additionally,* $\mu_0 \leq 1$. *Claim* (C3) *holds for* $0 < \alpha \leq FDP_\infty(u)$ *if, additionally,* $\mu_0 = 1$.

Asymptotic FDP control (C2) and α-exhaustion (C3) depend on the unobservable value of μ_0, which necessarily depends on the weighting scheme at hand. The next theorem is useful for verifying (C2) and/or (C3).

Theorem 6. *Suppose that* $(W_{m,M}, \theta_{m,M}), m \in \mathcal{M}$, *are identically distributed random vectors with support* $\Re^+ \times \{0,1\}$, *and with* $E[W_{m,M}] = 1$ *and* $E[\theta_{m,M}] \in (0,1)$. *Take*

$$\bar{W}_{0,M} = \frac{\sum_{m\in\mathcal{M}}(1-\theta_{m,M})W_{m,M}}{\sum_{m\in\mathcal{M}}(1-\theta_{m,M})}$$

whenever $\boldsymbol{\theta}_M \neq \mathbf{1}_M$ *and* $\bar{W}_{0,M} = 1$ *otherwise. If* $\bar{W}_{0,M} \to \mu_0$ *almost surely, then* $\mu_0 \leq 1$ *if* $Cov(W_{m,M}, \theta_{m,M}) \geq 0$ *and* $\mu_0 = 1$ *if* $Cov(W_{m,M}, \theta_{m,M}) = 0$.

Corollary 1. *Suppose that* (A4) - (A6) *are satisfied and take* $\boldsymbol{w}_M = \mathbf{1}_M$. *Then for any fixed* $\lambda \in (0,1)$ *and* $0 < \alpha \leq a_0$, *Claims* (C1) - (C3) *hold.*

This corollary suggests that the procedure in Storey, Taylor and Siegmund (2004) is competitive with the α-exhaustive nonlinear procedures in Finner, Dickhaus and Roters (2009). That a DU distribution is the least favorable among such (unweighted) adaptive linear step-up procedures under our weak dependence structure is interesting; the search for least favorable distributions remains a challenging problem. See Finner, Dickhaus and Roters (2007); Roquain and Villers (2011); Finner, Gontscharuk and Dickhaus (2012).

6.2. Asymptotically optimal weights

We verify that the conditions allowing for the WAMDF to provide less conservative asymptotic FDP control are satisfied under Model 1, even if the asymptotically optimal weights are perturbed or "noisy". Weight vectors and elements of weight vectors are indexed by M to facilitate asymptotic arguments, and, we sometimes write $\bar{t}_M(k_M^*) = \bar{t}_M(k_M^*, \boldsymbol{p}, \boldsymbol{\gamma})$ for brevity.

Perturbed weights are simulated by multiplying each asymptotically optimal weight by a positive random variable U_m,

$$\tilde{w}_{m,M}(k_M^*, \boldsymbol{p}, \boldsymbol{\gamma}) = U_m w_{m,M}(k_M^*, \boldsymbol{p}, \boldsymbol{\gamma}) \tag{6.3}$$

for each m. A perturbed weight is often denoted by $\tilde{w}_{m,M}$ and the vector of perturbed weights is denoted by $\tilde{\boldsymbol{w}}_M(k_M^*, \boldsymbol{p}, \boldsymbol{\gamma})$ or $\tilde{\boldsymbol{w}}_M$. To allow for (A2) to be satisfied, assume each triplet (U_m, γ_m, p_m) has a joint distribution satisfying $0 \leq U_m t_m(k_M^*/p_m, \gamma_m) \leq 1$ almost surely, and that $E[U_m|\boldsymbol{p}, \boldsymbol{\gamma}] = 1$ for each m so that perturbed weights have mean 1. Here $\tilde{\boldsymbol{w}}_M = \boldsymbol{w}_M^*$ if $U_m = 1$ for each m (almost surely). Hence, results regarding perturbed weights immediately carry over to asymptotically optimal weights.

Theorem 7. *Suppose that* $\Pr(p_m \leq 1-\alpha) = 1$, *take* $\lambda_M = \bar{t}_M(k_M^*)$, *and consider the perturbed weights* $\tilde{\boldsymbol{w}}_M$. *Under Model* 1 *and* (A1), (A2) *and* (A4) - (A6) *are satisfied and* $\mu_0 \leq 1$. *Hence the conditions of Theorem* 4 *are satisfied and* (C1) *and* (C2) *hold.*

Next the notion of "asymptotically optimal" is formalized and some examples of α-exhaustive weighting schemes are provided. Asymptotically optimal weights are equivalent to optimal fixed-t weights with $t = \bar{t}_M(k_M^*)$, while the asymptotically optimal WAMDF utilizes the asymptotic threshold $t_{\alpha,\infty}^{\lambda}$ (see Theorem 4).

Theorem 8. *Suppose that* $\Pr(p_m \leq 1-\alpha) = 1$ *and take* $\lambda_M = \bar{t}_M(k_M^*)$. *Then under Model* 1 *and* (A1), $\bar{t}_M(k_M^*) \to t_{\alpha,\infty}^{\lambda}$ *almost surely.*

Two corollaries show that asymptotic α-exhaustive FDP control is provided for a variety of weighing schemes.

Corollary 2. *Under Model* 1 *and* (A1) - (A2), *if* $\boldsymbol{w}_M$ *are mutually independent* *weights and independent of* $\boldsymbol{\theta}_M$ *with* $E[w_{m,M}] = 1$*, then* (C1) - (C2) *hold for* $\alpha \in (0,1)$ *and* (C3) *holds for* $0 < \alpha \leq FDP_\infty(u)$*.*

The next setting arises in practice whenever the distributions of the Z_m's from false nulls are heterogeneous, but heterogeneity attributable to prior probabilities for the states of the null hypotheses either does not exist or is not modeled. For an illustration see Section 8. See also Spjøtvoll (1972); Storey (2007); Peña, Habiger and Wu (2011) for more on this type of heterogeneity.

Corollary 3. *Suppose that the conditions of Theorem* 7 *are satisfied and consider perturbed weights* $\tilde{\boldsymbol{w}}_M$*. If* $p_i = p_j$ *for every* i, j*, then* (C3) *holds for* $0 < \alpha \leq FDP_\infty(u)$*.*

The fact that α-exhaustion need not be achieved when $p_i \neq p_j$ in Model 1 for the asymptotically optimal WAMDF, even though it is more powerful than competing MDFs, is noteworthy. A similar phenomenon was observed in Genovese, Roeder and Wasserman (2006) in the unadaptive setting, and it was suggested that one potential route for improvement is to incorporate an estimate of μ_0 into the procedure. However, it is not clear how this objective could be accomplished without sacrificing FDP control, especially when weights may be perturbed.

7. Simulation

This section compares weighted adaptive MDFs to other MDFs in terms of power and FDP control via simulation. In particular, for each of $K = 1{,}000$ replications, we generate $Z_m \overset{i.i.d.}{\sim} N(\theta_m\gamma_m, 1)$ for $m = 1, 2, \ldots,$ 1,000 and compute $\boldsymbol{\delta}(\hat{t}^{\lambda}_{\alpha,M}\boldsymbol{w}_M)$, $\boldsymbol{\delta}(\hat{t}^{0}_{\alpha,M}\boldsymbol{w}_M)$, $\boldsymbol{\delta}(\hat{t}^{\lambda}_{\alpha,M}\mathbf{1}_M)$, and $\boldsymbol{\delta}(\hat{t}^{0}_{\alpha,M}\mathbf{1}_M)$ as in Example 1, where $\alpha = 0.05$ and $\lambda_M = \bar{t}_M(k^*_M, \boldsymbol{p}, \boldsymbol{\gamma})$. The average FDP and average correct discovery proportion (CDP) was computed over the K replications for each procedure, where CDP $= \sum_{m\in\mathcal{M}_1}\delta_m/\max\{M_1, 1\}$.

In each simulation experiment, $\gamma_m \overset{i.i.d.}{\sim} Un(1, a)$ for $a = 1, 3, 5$, $Un(1, a)$ the uniform distribution over $(1, a)$. When $a = 1$ the effect sizes were identical, while when $a = 3$ or $a = 5$ they varied. In Simulation 1, $p_m = 0.5$ for each m and weighted procedures utilized asymptotically optimal weights. In Simulation 2, weighted procedures used asymptotically optimal weights as before and the effect sizes varied as before, but $p_m \overset{i.i.d.}{\sim} Un(0, 1)$. Thus, though the procedure was optimally weighted and asymptotic FDP control was provided, the conditions of (C3) are no longer satisfied. In Simulation 3, data were generated according to the same mechanism as in Simulation 2, but asymptotically optimal weights were perturbed via $U_m w_{m,M}(k_M^*, \boldsymbol{p}, \boldsymbol{\gamma})$, where $U_m \overset{i.i.d.}{\sim} Un(0,2)$. Simulation 4 represents a worst case scenario weighting scheme, in which weights were generated as $w_{m,M} \overset{i.i.d.}{\sim} Un(0,2)$.

Detailed results and discussions of simulations are in the supplemental materials. The main point is that the WA procedure dominates all other procedures as long as the employed weights are at least positively correlated with the optimal weights, and it performs nearly as well as other procedures otherwise. In particular, its FDP was less than or equal to 0.05 in all simulations, as Theorem 7 stipulates. Further, its average CDP was as large as or larger than the CDP of all other procedures in the first three simulations. The WA procedure did have a slightly smaller average CDP than the UA procedure in the worst case scenario (Simulation 4), as one might expect.

8. Implementation

In practical applications parameters $\boldsymbol{p}$ and $\boldsymbol{\gamma}$ in Model 1 are not (at least fully) observable and hence the asymptotically optimal WAMDF is not readily implementable. However, these parameters can be estimated or specified based on reasonable assumptions if the nature of the heterogeneity is at least partially observable. This section illustrates these two implementation approaches on the data in Table 1 and discusses strengths and limitations of each.

8.1. The setup

The goal is to test $H_m : \beta_m = 0$ for each m, where β_m is the regression coefficient for regressing $\boldsymbol{Y}_m = (Y_{1m}, Y_{2m}, \ldots, Y_{5m})^T$ on $\boldsymbol{x} = (x_1, x_2, \ldots, x_5)^T$ with the log-linear model $\log(\mu_{im}) = \alpha_m + \beta_m x_i$ and where Y_{im} are independent Poisson random variables with mean μ_{im}. Let $N_m = \sum_{i=1}^5 Y_{im}$ and $T_m = \sum_{i=1}^5 x_i Y_{im}$. As per McCullagh and Nelder (1989), we focus on the conditional distribution of $T_m | N_m = n_m$, which is free of the nuisance parameter α_m. Given $N_m = n_m$, $\boldsymbol{Y}_m$ has a multinomial distribution with mean $n_m \boldsymbol{p}(\beta_m)$ and covariance $n_m[diag(\boldsymbol{p}(\beta_m)) - \boldsymbol{p}(\beta_m)\boldsymbol{p}(\beta_m)^T]$, where $\boldsymbol{p}(a) = [\exp(x_1 a)/\sum_i \exp\{x_i a\}, \exp(x_2 a) /\sum_i \exp\{x_i a\}, \ldots, \exp(x_5 a)/\sum_i \exp\{x_i a\}]^T$. Thus, the Z-score for $T_m = \boldsymbol{x}^T \boldsymbol{Y}_m$ is

$$Z_m = \left(\frac{T_m - n_m \boldsymbol{x}^T \boldsymbol{p}(0)}{\sqrt{n_m \boldsymbol{x}^T [diag(\boldsymbol{p}(0)) - \boldsymbol{p}(0)\boldsymbol{p}(0)^T] \boldsymbol{x}}} \right).$$

To facilitate Model 1 we consider the mixture model introduced in Habiger, Watts and Anderson (2016), that assumes apriori that $\Pr(\beta_m = 0) = \pi_0$, $\Pr(\beta_m =$

$\eta_1) = \pi_1$, and $\Pr(\beta_m = \eta_2) = \pi_2$ for some $\eta_1 \neq \eta_2 \neq 0$ and $\pi_0 + \pi_1 + \pi_2 = 1$. Denote the mixing proportions by $\boldsymbol{\pi}$ and take $\boldsymbol{\eta} = (\eta_1, \eta_2)$. Utilizing a normal approximation for the distribution of Z_m results in normal mixture density for $Z_m|N_m = n_m$:

$$f(z_m|n_m;\boldsymbol{\pi},\boldsymbol{\eta}) = \pi_0\phi\left(z_m;0,1\right) + \pi_1\phi\left(z_m;\mu(\eta_1,n_m),\sigma^2(\eta_1)\right) + \pi_2\phi\left(z_m;\mu(\eta_2,n_m),\sigma^2(\eta_2)\right), \tag{8.1}$$

where

$$\mu(a,n_m) = \frac{\sqrt{n_m}\boldsymbol{x}^T[\boldsymbol{p}(a) - \boldsymbol{p}(0)]}{\sqrt{\boldsymbol{x}^T\left[diag(\boldsymbol{p}(0)) - \boldsymbol{p}(0)\boldsymbol{p}(0)^T\right]\boldsymbol{x}}},$$

$$\text{and } \sigma^2(a) = \frac{\boldsymbol{x}^T\left[diag(\boldsymbol{p}(a)) - \boldsymbol{p}(a)\boldsymbol{p}(a)^T\right]\boldsymbol{x}}{\boldsymbol{x}^T\left[diag(\boldsymbol{p}(0)) - \boldsymbol{p}(0)\boldsymbol{p}(0)^T\right]\boldsymbol{x}}.$$

In the context of Model 1, $F_0 = \Phi$, $p_m = 1 - a_0 = \pi_1 + \pi_2$, $\gamma_m = n_m$ and

$$F_1(z_m|\gamma_m) = F_1(z_m|n_m;\boldsymbol{\pi},\boldsymbol{\eta}) = \frac{\pi_1}{\pi_1+\pi_2}\Phi\left(\frac{z_m - \mu(\eta_1,n_m)}{\sigma(\eta_1)}\right) + \frac{\pi_2}{\pi_1+\pi_2}\Phi\left(\frac{z_m - \mu(\eta_2,n_m)}{\sigma(\eta_2)}\right).$$

Here $\gamma_m = n_m$ is not an unobservable effect size. It is observable and indexes a mixture distribution for Z_m when H_m is false, which depends on the parameters $\boldsymbol{\pi}$ and $\boldsymbol{\eta}$.

The uniformly most powerful unbiased decision function is $\delta_m(Z_m;t_m) = I(|Z_m| \geq \Phi^{-1}(1 - t_m/2))$, with power function

$$\pi_{n_m}(t_m) = F_1(\Phi^{-1}(\frac{t_m}{2})|n_m;\boldsymbol{\pi},\boldsymbol{\eta}) + [1 - F_1(\Phi^{-1}(1 - \frac{t_m}{2})|n_m;\boldsymbol{\pi},\boldsymbol{\eta})].$$

To compute optimal fixed-t weights, first note that $\phi(\Phi^{-1}(t_m/2)) = \phi(\Phi^{-1}(1 - t_m/2))$ so that the derivative of $\pi_{n_m}(t_m)$ with respect to t_m is

$$\pi'_{n_m}(t_m) \propto \frac{f_1(\Phi^{-1}(t_m/2)|n_m;\boldsymbol{\pi},\boldsymbol{\eta}) + f_1(\Phi_0^{-1}(1 - t_m/2)|n_m;\boldsymbol{\pi},\boldsymbol{\eta})}{\phi(\Phi^{-1}(t_m/2))}. \tag{8.2}$$

Setting this derivative equal to $k/p_m = k/(\pi_1 + \pi_2)$ and solving for t_m gives a collection of optimal fixed-t thresholds. Denote each such t_m by $t_m(k,\boldsymbol{\pi},\boldsymbol{\eta},n_m)$. Then, optimal fixed-t weights are computed as in $w_m(k^*,\boldsymbol{\pi},\boldsymbol{\eta},n_m) = (t_m(k^*,\boldsymbol{\pi},\boldsymbol{\eta}, n_m))/(\bar{t}_M(k^*,\boldsymbol{\pi},\boldsymbol{\eta},\boldsymbol{n}))$ where k^* satisfies $\bar{t}_M(k^*,\boldsymbol{\pi},\boldsymbol{\eta},\boldsymbol{n}) \equiv M^{-1}\sum_{m\in\mathcal{M}} t_m(k^*,\boldsymbol{\pi}, \boldsymbol{\eta},n_m) = t$.

The five steps for implementing the WAMDF are:

1a. get $(\boldsymbol{\pi},\boldsymbol{\eta},n_m)$ for each m;

1b. compute $w_m^* = w_m(k_M^*,\boldsymbol{\pi},\boldsymbol{\eta},n_m)$ as in (4.3);

Table 3. Maximum likelihood estimates for the model in (8.1).

$\hat{\pi}_0$	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\eta}_1$	$\hat{\eta}_2$
0.66	0.17	0.17	-1.09	0.71

2a. specify λ and compute $Q_m = P_m/w_m^* = 2\bar{\Phi}(|z_m|)/w_m^*$;

2b. get $j = \max\{m : Q_{(m)} \leq \alpha m/\hat{M}_0(\lambda \boldsymbol{w}^*)\}$;

2c. get $\hat{t}_\alpha^{\lambda *} = \min\{j\alpha/\hat{M}_0(\lambda \boldsymbol{w}), \lambda\}$ and reject H_m if $Q_m \leq \hat{t}_\alpha^{\lambda *}$.

The parameters $\boldsymbol{\pi}$ and $\boldsymbol{\eta}$ are unobservable and hence must be estimated or specified.

8.2. Parameter estimation

Parameters are estimated via maximum likelihood. Specifically, assuming that $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_M$ are independent conditionally upon $N_1, N_2, \ldots, N_M$, then under (8.1), the log likelihood is

$$l(\boldsymbol{\pi}, \boldsymbol{\eta}) = \sum_{m=1}^{M} log(f(z_m|n_m; \boldsymbol{\pi}, \boldsymbol{\eta}))$$

and maximum likelihood estimates are found using the EM algorithm (Dempster, Laird and Rubin (1977)). Results are summarized in Table 3. For more details on the EM algorithm and finite mixtures of normal distributions, see McLachlan and Peel (2000) and see, for example, Benaglia et al. (2009) for available software. For $\alpha = 0.05$ and $\lambda = 0.5$, the unweighted adaptive procedure resulted in 86 discoveries. The weighted adaptive procedure with estimated weights as above (but modified via $\tilde{w}_m = [w_m^* + 0.1]/[M^{-1}\sum_m (w_m^* + 0.1)]$ to avoid impractically small weights) was applied for $\alpha = 0.05$ and $\lambda = 0.5$ and resulted in 85 discoveries. Of course, we cannot know the average power or FDR for the weighted and unweighed adaptive procedures based on this run of the experiment.

Some asymptotic results are readily available. In particular, because $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\eta}}$ are maximum likelihood estimates, $\hat{\boldsymbol{\pi}} \to \boldsymbol{\pi}$ and $\hat{\boldsymbol{\eta}} \to \boldsymbol{\eta}$ as $M \to \infty$ almost surely. Consequently, $w_m(k^*, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\eta}}, n_m) \to w_m(k^*, \boldsymbol{\pi}, \boldsymbol{\eta}, n_m)$ as $M \to \infty$ almost surely. See for example Serfling (1980), pg. 145 - 150. Thus, this WAMDF is α-exhaustive and asymptotically optimal under (8.1). A limitation of this approach is that it can be computationally intense, especially when M is large. Here parameters $\boldsymbol{\pi}$ and $\boldsymbol{\eta}$ must be estimated with an iterative procedure, a root finding algorithm is necessary to compute $t_m(k, \boldsymbol{\pi}, \boldsymbol{\eta}, n_m)$ for each m and each value of k, and a root-finding algorithm is necessary to find the k^* corresponding to the asymptotically optimal weights.

8.3. Parameter specification

One of the advantages of the WAMDF is that computationally simpler versions can be utilized with potentially little loss in efficiency and without sacrificing FDR control. To illustrate, consider weights computed $w_m^* = t_m/\bar{t}$ where

$$t_m = 2\bar{\Phi}\left(0.5\bar{\Phi}^{-1}(\frac{\alpha}{4})\left[\frac{\sqrt{n_m}}{\sqrt{n.}/M} + \frac{\sqrt{n.}/M}{\sqrt{n_m}}\right]\right) \tag{8.3}$$

and where $\sqrt{n.} = \sum_m \sqrt{n_m}$. The WAMDF, with $\alpha = 0.05$, $\lambda = 0.5$, and $\tilde{w}_m = [w_m^*+0.1]/[M^{-1}\sum_m(w_m^*+0.1)]$ to safeguard against impractically small weights, was applied and resulted in 87 discoveries.

These weights utilized were justified as in (3.3), and by assuming that the average power and prior probability of H_m being false is 1/2. Specifically, $\mu(a, n_m)/\sigma(a) \propto \sqrt{n}_m$ and leads to approximate power functions as in Example 1 via $\pi_{\gamma_m}(t_m) = \bar{\Phi}(\bar{\Phi}^{-1}(t_m/2) - \gamma_m) = \bar{\Phi}(\bar{\Phi}^{-1}(t_m/2) - \gamma\sqrt{n_m})$ for γ some tuning parameter. Then, assume $p_m = 0.5$ and $\pi_{\bar{\gamma}}(t) = \bar{\Phi}(\bar{\Phi}^{-1}(t/2)-\bar{\gamma}) = 0.5$. Approximating the FDR at t when $p_m = 1/2$ and $\pi_{\bar{\gamma}}(t) = 0.5$ with $FDR(t) = 0.5t/[0.5t+(1-0.5)\pi_{\bar{\gamma}}(t)]$, solving $FDR(t) = \alpha$ and $\pi_{\bar{\gamma}}(t) = 1/2$ simultaneously gives approximate fixed-t threshold $t = \alpha/[2(1-\alpha)] \approx \alpha/2$ and $\bar{\gamma} = \bar{\Phi}^{-1}(t/2) \approx \bar{\Phi}^{-1}(\alpha/4)$. Taking the derivative of $\pi_{\gamma_m}(t_m) = \bar{\Phi}(\bar{\Phi}^{-1}(t_m/2) - \gamma_m)$ with respect to t_m and setting it equal to k/p and solving yields $log(k/p) = \bar{\Phi}^{-1}(t_m/2)\gamma_m - 0.5\gamma_m^2$, and

$$t_m = 2\bar{\Phi}\left(0.5\gamma_m + \frac{log(k/p)}{\gamma_m}\right).$$

Plugging $\bar{\Phi}^{-1}(t/2)\bar{\gamma} - 0.5\bar{\gamma}^2 = \bar{\gamma}^2 - 0.5\bar{\gamma}^2 = 0.5\bar{\gamma}^2$ in for $log(k/p)$, $\gamma_m = \sqrt{n_m}\gamma$, and $\bar{\gamma} = \gamma\sqrt{n.}/M$ here, we recover (8.3).

These weights need not be asymptotically optimal. However, under (8.1) this WAMDF it is still α-exhaustive (Corollary 3) and simulation studies suggest that it is more efficient than its unweighted version even if these weights are only positively correlated with optimal weights. The main advantage of this approach is that weights still exploit heterogeneity attributable to the n_m's and are computationally simple.

The fact that weights are so simple allows for a simulation study to gauge the performance of the WAMDF. In Simulation 5, for each of 1,000 replications and $M = 1{,}000$, we sampled n_m's from the n_m's in Table 1 and generate $\theta_m \sim Bernoulli(p)$ and $Z_m \sim N(\gamma\sqrt{n_m}\theta_m, 1)$. We considered all p-γ combinations where γ is chosen so that $\bar{\gamma} = \gamma M^{-1}\sqrt{n.} = 1.75, 2, 2.25$ and $p = 0.2, 0.5, 0.8$. For each replication and setting, the unweighted adaptive MDF was applied and the WAMDF was applied with $\alpha = 0.01, 0.05, 0.10$. The average FDP and CDP were recorded over the 1,000 replications for each setting. Detailed results are in the supplemental materials.

Although the weights were based on some simplifying assumptions, the WAMDF was more powerful than in its unweighted counterpart even if $p = 0.2$ or $p = 0.8$, as long as the CDP was at least 0.2. Further, the average FDP was always less than α. Our simplifying assumptions were made merely because they were the least informative and lead to the simplest weights. Other weighting schemes could be considered. We leave more extensive methodological development of this nature as future work. The goal here was to demonstrate that the theory developed in the previous sections will be useful in developing WAMDFs that are simple and practical.

9. Concluding Remarks

Efforts to improve upon the original BH procedure have focused on controlling the FDR at a level nearer α, or exploiting heterogeneity across tests. We have combined these objectives using a weighted decision theoretic framework and showed that the resulting procedure is more powerful than procedures which only consider of them. We have provided weighted adaptive multiple decision functions that satisfy the α-exhaustive optimality criterion considered in Finner, Dickhaus and Roters (2009), but allow for further improvements via an optimal weighting scheme that incorporates heterogeneity.

The proposed WAMDFs are robust, and coupled with the flexibility of the WAMDF framework, allow for multiple testing procedures that exploit heterogeneity to be developed in a wide variety of settings, even when the nature and degree of heterogeneity is not fully observable or known.

The finite sample and asymptotic results here are valid under independence and weak dependence conditions, respectively. Benjamini and Yekutieli (2001) showed that the unweighted unadaptive BH procedure provides (finite) FDR control under a certain positive dependence structure, and that it can be modified to control the FDR for arbitrary dependence. One could study the performance of weighted adaptive procedures under other types of dependence, but obtaining finite sample analytical results for adaptive MDFs then appears to be very challenging. See Blanchard and Roquain (2009); Roquain and Villers (2011) for some results. As for large sample results, Fan, Han and Gu (2012) and Desai and Storey (2012) provide techniques for transforming test statistics so that they are weakly dependent, and our WAMDF framework facilitates weak dependence. Perhaps these transformed test statistics could be used in conjunction with our WAMDF, but this requires further development.

Other estimators for M_0 could be considered. For example, it is possible to use the unweighted estimator from Storey, Taylor and Siegmund (2004) in the WAMDF, or to consider data dependent choices of the tuning parameter λ as in Liang and Nettleton (2012). A more detailed assessment of $\hat{M}_0(\lambda \boldsymbol{w})$, though warranted, is beyond the scope of the present work.

Supplementary Materials

Additional details and further discussion regarding simulations referred to in Sections 7 and 8, and proofs of theorems, lemmas, and corollaries in Sections 3, 5, and 6 are in the supplemental materials.

Acknowledgment

The author would like to thank referees and the AE for helpful suggestions.

References

Anderson, M. and Habiger, J. (2012). Characterization and identification of productivity-associated rhizobacteria in wheat. *Appl. Environ. Microbiol.* **78**, 4434–444.

Benaglia, T., Chauveau, D., Hunter, D. and Young, D. (2009). Mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software* **32**.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57**, 289–300.

Benjamini, Y. and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Stat.* **25**, 60–83.

Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507.

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

Billingsley, P. (1999). *Convergence of Probability Measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 2nd edn. A Wiley-Interscience Publication.

Blanachar, G. and Roquain, E. (2008). Two simple sufficient conditions for FDR control. *Electron. J. Stat.* **2**, 963–992.

Blanchard, G. and Roquain, E. (2009). Adaptive fdr control under independence and dependence. *J. Mach. Learn. Res.* **10**, 2837–2831.

Cai, T. T. and Sun, W. (2009). Simultaneous testing of grouped hypotheses: finding needles in multiple haystacks. *J. Amer. Statist. Assoc.* **104**, 1467–1481.

Cao, H., Sun, W. and Kosorok, M. R. (2013). The optimal power puzzle: scrutiny of the monotone likelihood ratio assumption in multiple testing. *Biometrika* **100**, 495–502.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B. Methodological* **39**, 1–38. With discussion.

Desai, K. H. and Storey, J. D. (2012). Cross-dimensional inference of dependent high-dimensional data. *J. Amer. Statist. Assoc.* **107**, 135–151.

Dudoit, S. and van der Laan, M. J. (2008). *Multiple Testing Procedures with Applications to Genomics*. Springer Series in Statistics. Springer, New York.

Efron, B. (2008). Microarrays, empirical bayes and the two-group model. *Statistical Science* **23**, 1–22.

Efron, B. (2010). *Large-scale Inference*, vol. 1 of *Institute of Mathematical Statistics (IMS) Monographs*. Cambridge University Press, Cambridge. Empirical Bayes methods for estimation, testing, and prediction.

Efron, B., Tibshirani, R., Storey, J. D. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96**, 1151–1160.

Fan, J., Han, X. and Gu, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.* **107**, 1019–1035.

Finner, H., Dickhaus, T. and Roters, M. (2007). Dependency and false discovery rate: asymptotics. *Ann. Statist.* **35**, 1432–1455.

Finner, H., Dickhaus, T. and Roters, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.* **37**, 596–618.

Finner, H., Gontscharuk, V. and Dickhaus, T. (2012). False discovery rate control of step-up-down tests with special emphasis on the asymptotically optimal rejection curve. *Scand. J. Stat.* **39**, 382–397.

Gavrilov, Y., Benjamini, Y. and Sarkar, S. K. (2009). An adaptive step-down procedure with proven FDR control under independence. *Ann. Statist.* **37**, 619–629.

Genovese, C., Roeder, K. and Wasserman, L. (2006). False discovery control with p-value weighting. *Biometrika* **93**, 509–524.

Genovese, C. and Wasserman, L. (2002). Operating characteristic and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B Stat Methodol.* **64**, 499–517.

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32**, 1035–1061.

Habiger, J. (2012). A method for modifying multiple testing procedures. *J. Statist. Plann. Inference* **142**, 2227–2231.

Habiger, J. and Peña, E. (2011). Randomized p-values and nonparametric procedures in multiple testing. *J. Nonparametr. Stat.* **23**, 583–604.

Habiger, J., Watts, D. and Anderson, M. (2016). Multiple testing with heterogeneous multinomial distributions. *Biometrics* **in press**, arXiv:1511.01400.

Habiger, J. D. and Peña, E. A. (2014). Compound p-value statistics for multiple testing procedures. *J. Multivariate Anal.* **126**, 153–166.

Hu, J. X., Zhao, H. and Zhou, H. H. (2010). False discovery rate control with groups. *J. Amer.*

Statist. Assoc. **105**, 1215–1227.

Liang, K. and Nettleton, D. (2012). Adaptive and dynamic adaptive procedures for false discovery rate control and estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **74**, 163–182.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Monographs on Statistics and Applied Probability. Chapman & Hall, London. Second edition [of MR0727836].

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models.* Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York.

Peña, E. A., Habiger, J. D. and Wu, W. (2011). Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. *Ann. Statist.* **39**, 556–583.

Roeder, K. and Wasserman, L. (2009). Genome-wide significance levels and weighted hypothesis testing. *Statist. Sci.* **24**, 398–413.

Roquain, E. and van de Wiel, M. A. (2009). Optimal weighting for false discovery rate control. *Electron. J. Stat.* **3**, 678–711.

Roquain, E. and Villers, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.* **39**, 584–612.

Sarkar, S. (2007). Stepup procedures controlling generalized FWER and generalized FDR. *Ann. Statist.* **35**, 2405–2420.

Schweder, T. and Spjotvoll, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* **69**, pp. 493–502.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 751–754.

Spjøtvoll, E. (1972). On the optimality of some multiple comparison procedures. *Ann. Math. Statist.* **43**, 398–411.

Storey, J. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat Methodol.* **64**, 479–498.

Storey, J. (2003). The positive false discovery rate: a bayesian interpretation and the q-value. *Ann. Statist.* **31**, 2012–2035.

Storey, J. D. (2007). The optimal discovery procedure: a new approach to simultaneous significance testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69**, 347–368.

Storey, J. D., Taylor, J. E. and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66**, 187–205.

Sun, W. and Cai, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* **102**, 901–912.

Sun, W. and McLain, A. C. (2012). Multiple testing of composite null hypotheses in heteroscedastic models. *J. Amer. Statist. Assoc.* **107**, 673–687.

Department of Statistics, Oklahoma State University, 301 MSCS, Stillwater, OK 74078-1056, USA

E-mail: jhabige@okstate.edu

(Received April 2016; accepted August 2016)