Adaptive False Discovery Rate Control for Heterogeneous Data

Joshua D. Habiger
Department of Statistics
Oklahoma State University
301G MSCS
Stillwater, OK 74078
June 11, 2015

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, a simple adaptive FDR method for heterogeneous sample sizes is developed and applied to real data.

Keywords: Multiple Testing; P-value; Weighted P-value; Decision Function

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 prevalence of bacteria living near the roots of wheat plants across $i = 1, 2, ..., 5$ treatment groups for each of $m = 1, 2, ..., M = 778$ 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 con-
straint that some global type 1 error rate is controlled at a prespecified level $\alpha$. 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 proce-
dure 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 $\alpha$, as opposed to $\alpha a_0$. For
example, adaptive procedures in Benjamini and Hochberg (2000); Storey et al. (2004); Ben-
jamini et al. (2006); Gavrilov et al. (2009); Liang and Nettleton (2012) utilize an estimate
of $a_0$ and typically have FDR that is greater than $\alpha a_0$ yet still less than or equal to $\alpha$.
Finner et al. (2009) proposed nonlinear procedures that “exhaust the $\alpha$” in that, loosely
speaking, their FDR converges to $\alpha$ under some least favorable configuration as $M$ tends
to infinity.

Another approach aims to exploit heterogeneity across hypothesis tests. Genovese et al.
(2006); Blanachar and Roquain (2008); Roquain and van de Wiel (2009); Peña et al. (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 et al. (2010) provide 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, for example, are heterogeneous because sample sizes $n_1, n_2, ..., n_M$ vary
from test to test, with $n_m$ being as small as 6 and as large as 911.
Table 1: Depiction of the data in Anderson and Habiger (2012). Shoot biomass $x_i$ in grams for groups $i = 1, 2, ..., 5$ was 0.86, 1.34, 1.81, 2.37, and 3.00, respectively. Row totals are in the last column.

<table>
<thead>
<tr>
<th>Bacteria ($m$)</th>
<th>$Y_{1m}$</th>
<th>$Y_{2m}$</th>
<th>$Y_{3m}$</th>
<th>$Y_{4m}$</th>
<th>$Y_{5m}$</th>
<th>Total ($n_m$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>778</td>
<td>16</td>
<td>10</td>
<td>29</td>
<td>18</td>
<td>13</td>
<td>81</td>
</tr>
</tbody>
</table>

Whatever the nature of the heterogeneity may be, recent literature suggests that it should not be ignored. Specifically, 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. In fact, Sun and McLain (2012) illustrated that procedures which ignore heterogeneity are not only inefficient, but can produce lists of discoveries that are of little scientific interest. However, one of the challenges with heterogeneous data is that, while more efficient procedures can readily be developed if parameters governing the distribution of the data are known, these Oracle parameters are typically not known in practice. Further, Oracle parameter estimation can be involved and limited to specific types of heterogeneity, such as the cluster-type heterogeneity mentioned above. See Habiger and Peña (2014) for more details on Oracle parameter estimation.

The objective of this paper is to provide a general approach for exploiting heterogeneity in a simple manner without sacrificing FDR control. The idea is to combine adaptive FDR methods for exhausting the $\alpha$ with weighted procedures for exploiting heterogeneity using a robust and flexible decision theoretic framework. The flexibility allows for simple weighting schemes to be developed using reasonable assumptions for the specific type of heterogeneity at hand, while the robustness of the proposed framework ensures that more efficient FDR control is provided as long as the said assumptions are indeed reasonable.

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 asym-
totically 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 $\alpha$. Sufficient conditions for “$\alpha$-exhaustion” are provided and shown to be satisfied in a variety of settings. For example, unweighted adaptive MDFs in Storey et al. (2004) and certain asymptotically optimal WAMDFs are $\alpha$-exhaustive. In fact, $\alpha$-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 illustrates how simple procedures, inspired by the asymptotically optimal WAMDF, can be developed. Specifically, it provides a WAMDF for testing multiple null hypotheses with heterogeneous sample sizes and applies it to the data in Table 1. A modest simulation study illustrates that it controls the FDR and is generally more efficient than its unweighted counterpart. Concluding remarks are in Section 9 and technical details are in the Supplemental Article.

2 Background

2.1 Data

Let $Z = (Z_m, m \in M)$ for $M = \{1, 2, ..., M\}$ be a random vector of test statistics with joint distribution function $F$ and let $F$ be a model for $F$. The basic goal is to test null hypotheses $H = (H_m, m \in M)$ of the form $H_m : F \in F_m$, where $F_m \subseteq F$ is a submodel for $F$. For short, we often denote the state of $H_m$ by $\theta_m = 1 - I(F \in F_m)$, 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 $H$ by $\theta = (\theta_m, m \in M)$. Let $M_0 = \{m \in M : \theta_m = 0\}$ and $M_1 = M \setminus M_0$ index the set of true and false null hypotheses, respectively, and denote the number of true and false null hypotheses by $M_0 = |M_0|$ and $M_1 = |M_1|$, respectively.
To make matters concrete, we often consider the following random effects model for $Z$. For related models see Efron et al. (2001); Genovese and Wasserman (2002); Storey (2003); Genovese et al. (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 $p = (p_m, m \in \mathcal{M})$ for the states of the $H_m$s and parameters $\gamma = (\gamma_m, m \in \mathcal{M})$, which we refer to as effect sizes, although each $\gamma_m$ could merely index a distribution for $Z_m$ when $H_m$ is false.

**Model 1** Let $(Z_m, \theta_m, p_m, \gamma_m), m \in \mathcal{M}$ be independent and identically distributed random vectors each with support in $\mathbb{R} \times \{0, 1\} \times [0, 1] \times \mathbb{R}^+$ and with conditional distribution functions $F(z_m|\theta_m, p_m, \gamma_m) = (1 - \theta_m)F_0(z_m) + \theta_mF_1(z_m|\gamma_m)$ and $F(z_m|p_m, \gamma_m) = (1 - p_m)F_0(z_m) + p_mF_1(z_m|\gamma_m)$. Assume further that $F(\gamma_m, p_m) = F(\gamma_m)F(p_m)$, $\text{Var}(\gamma_m) < \infty$ 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 $\theta$, $p$, and $\gamma$ are assumed to be random variables to facilitate asymptotic analysis, as in Genovese et al. (2006); Blanchar and Roquain (2008); Blanchard and Roquain (2009); Roquain and van de Wiel (2009); Roquain and Villers (2011). Analysis under Model 1 will focus on conditional distribution functions $F(z|\theta, p, \gamma) = \prod_{m \in \mathcal{M}} F(z_m|\theta_m, p_m, \gamma_m)$ and $F(z|p, \gamma) = \prod_{m \in \mathcal{M}} F(z_m|p_m, \gamma_m)$, and an expectation taken over $Z$ with respect to these distributions will be denoted $E[\cdot|\theta, p, \gamma]$ and $E[\cdot|p, \gamma]$, 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 et al. (2004); Sun and Cai (2007); Peña et al. (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). Note that 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

$$\delta_m(Z_m; t_m) = I(Z_m \geq F_0^{-1}(1 - t_m)). \quad (1)$$
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 $\delta_m$, hence the terminology “size threshold”. An MDF is denoted $\delta(Z; t) = [\delta_m(Z_m; t_m), m \in M]$, where $t = (t_m, m \in M)$ is called a threshold vector. Note that if $t_m = \alpha/M$ for each $m$ then $(Z; t)$ represents the well-known Bonferroni procedure.

Assume that, for each $m$, $t_m \rightarrow \delta_m(Z_m; t_m)$ is nondecreasing and right continuous with $\delta_m = 0(1)$ whenever $t_m = 0(1)$, almost surely, and that $t_m \rightarrow E[\delta_m(Z_m; t_m)]$ is continuous and strictly increasing for $t_m \in (0, 1)$ with $E[\delta_m(Z_m; t_m)] = t_m$ whenever $m \in M_0$. 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 (1). For additional details and examples see Habiger and Peña (2011); Peña et al. (2011); Habiger (2012).

### 2.3 Tools for implementation

To simplify computation of $\delta(Z; t)$ we break $t$ down into the product of a positive valued weight vector $w = (w_m, m \in M)$ satisfying $\bar{w} = M^{-1} \sum_{m \in M} w_m = 1$ and an overall or average threshold $t$. That is, we write $t = tw$. The basic methodology is then operationally implemented in two steps. In step 1, weights are specified and in step 2 data $Z = z$ are collected, the overall threshold $t$ is computed and the MDF $\delta(z; tw)$ is computed. If weights are based on Model 1, for example, then they will be allowed to depend functionally on $p$ and $\gamma$ in the first step. The overall threshold will be allowed to depend functionally on $z$ and $w$.

It will also be useful to exploit the link between weighted $p$-values and decision functions. First, define the (unweighted) $p$-value statistic corresponding to $\delta_m$ by $P_m = \inf\{t_m \in [0, 1] : \delta_m(Z_m; t_m) = 1\}$. This definition, also considered in Habiger and Peña (2011); Peña et al. (2011), has the usual interpretation that $P_m$ is the smallest size $t_m$ allowing for $H_m$ to be rejected and also 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 (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 weighted $p$-value statistic by $Q_m = \inf\{t : \delta_m(Z_m; tw_m) = 1\}$.
Observe that 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 following almost surely equivalent expressions for a decision function under the NS assumptions:

$$\delta_m(Z_m; t_m) = \delta_m(Z_m; tw_m) = I(P_m \leq tw_m) = I(Q_m \leq t). \quad (2)$$

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

Recall that $\delta(Z; t) = \delta(Z; tw)$ almost surely. In this subsection, the former expression is considered for simplicity and the constraint that $\bar{w} = 1$ is replaced with the constraint that $\bar{t} = t$, where $\bar{t} = M^{-1} \sum_{m \in \mathcal{M}} t_m$. Recall that weights are allowed to depend on $p$ and $\gamma$ under Model 1. Thus, the focus is on the conditional expectation of $\delta_m(Z_m; t_m)$ denoted by $G_m(t_m) \equiv E[\delta_m(Z_m; t_m)| p, \gamma] = (1 - p_m)t_m + p_m\pi_{\gamma_m}(t_m)$, where $\pi_{\gamma_m}(t_m) = E[\delta_m(Z_m; t_m)| \theta_m = 1, \gamma_m]$ is the power function for $\delta_m$. As in Genovese et al. (2006); Roquain and van de Wiel (2009); Peña et al. (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 et al., 2011) and under the generalized monotone likelihood ratio (GMLR)
condition in Cao et al. (2013). To see this, note that the GMLR condition states that $g_i(t_m)/g_0(t_m)$ is monotonically decreasing in $t_m$, where $g_{im}(t_m)$ is the derivative of $G_{im}(t_m) = E[\delta_m(Z_m; t_m)|\theta_m = i]$ with respect to $t_m$ for $i = 0, 1$. In our notation, $g_i(t_m) = \pi'_i(t_m)$ and, under the NS conditions, $g_0(t_m) = 1$. Hence, the GMLR condition stipulates that $\pi'_i(t_m)$ is monotonically decreasing, i.e. that $\pi'_i(t_m)$ is concave.

Given $p$, $\gamma$ and $t$, the goal is to maximize the expected number of correctly rejected null hypotheses $\pi(t, p, \gamma) \equiv E[\sum_{m \in M} \theta_m \delta_m(Z_m; t_m)|\gamma, p] = \sum_{m \in M} p_m \pi_{\gamma_m}(t_m)$ subject to the constraint that $\bar{t} = t$. Theorem 1 describes the form of the solution and states that it exists and is unique.

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

$$\pi'_i(t_m) = k/p_m$$

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

In general, the optimal fixed-t thresholds above and their corresponding weights can be numerically found as follows. First, for a fixed value of $k$, denote the solution to equation (3) in terms of $t_m$ by $t_m(k/p_m, \gamma_m)$ and denote the vector of solutions by $t(k, p, \gamma) = [t_m(k/p_m, \gamma_m), m \in \mathcal{M}]$. In Example 1 below a closed form expression for $t_m(k/p_m, \gamma_m)$ exists, but in general we may employ any single root finding algorithm to compute each $t_m(k/p_m, \gamma_m)$ because $\pi'_i(t_m)$ is continuous and monotone in $t_m$ by (A1). Then to find the optimal fixed-t threshold vector, first find the unique $k^*$ satisfying $\bar{t}_M(k^*, p, \gamma) = t$, where $\bar{t}_M(k, p, \gamma) = M^{-1} \sum_{m \in M} t_m(k/p_m, \gamma_m)$, and then compute $t(k^*, p, \gamma)$. Because $t_m = tw_m$ in our setup and because $\bar{t}_M(k^*, p, \gamma) = t$, each optimal fixed-t weight is recovered via

$$w_m(k^*, p, \gamma) = \frac{t_m(k^*/p_m, \gamma_m)}{t_M(k^*, p, \gamma)}.$$  

We shall sometimes denote $w_m(k^*, p, \gamma)$ by $w_m^*$ for short. The vector of optimal fixed-t weights is denoted by $w(k^*, p, \gamma) = [w_m(k^*, p, \gamma), m \in \mathcal{M}]$ or by $w^* = (w^*_m, m \in \mathcal{M})$.

To better understand how the solution is found and related to the values of $p_m$, $\gamma_m$ and $t$ consider the following example.
Example 1 Suppose $Z_m | \gamma_m, \theta_m \sim N(\theta_m \gamma_m, 1)$ for $\gamma_m > 0$, where $N(a, b)$ represents a Normal distribution with mean $a$ and variance $b$, 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 denote $\Phi(\cdot) = 1 - \Phi(\cdot)$. Define $\delta_m(Z_m; t_m) = I(Z_m \geq \Phi^{-1}(t_m))$. The power function is $\pi_m(t_m) = \Phi(\Phi^{-1}(t_m) - \gamma_m)$ and has derivative $\pi'_m(t_m) = \frac{\phi(\Phi^{-1}(t_m) - \gamma_m)}{\Phi(\Phi^{-1}(t_m))}$. Setting the derivative equal to $k/p_m$ and solving yields

$$t_m(k/p_m, \gamma_m) = \Phi(0.5\gamma_m + \log(k/p_m)/\gamma_m).$$

(5)

The optimal fixed-$t$ threshold vector is computed by $t(k^*, p, \gamma)$, where $k^*$ satisfies $\bar{t}_M(k^*, p, \gamma) = t$, and the optimal fixed-$t$ weights are computed as in (4).

First, observe in expression (5) 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 $1$ 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

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

is increasing in $p_m$, as we might expect.
The relationship between \( w_m(k^*, p, \gamma) \) and \( \gamma_m \) 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 \( t = 0.01, \tilde{t}_M(k^*, p, \gamma) = 0.01 \) 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 \( w_2^* = 0.017/0.01 = 1.7. \) Note that because \( p_1 = p_2, \) the slopes of the power functions evaluated at 0.003 and 0.017, respectively, are equal; see equation (3). Now consider fixed threshold \( t = 0.05. \) Here \( k^* = 1.7 \) and which leads to weights \( w_1^* = 0.059/0.05 = 1.18 \) and \( w_2^* = 0.041/0.05 = 0.82. \) That is, 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 et al. (2011) or Section 8 of the current manuscript. The important point is that the optimal fixed-\( t \) weights depend on the choice of \( t. \)

### 3.2 Asymptotically optimal weights

The overall threshold \( t \) in Section 4 is chosen using an estimator of the FDP, which depends functionally on data \( Z; \) see expressions (6) and (7). Hence, optimal fixed-\( t \) weights, which are not allowed to functionally depend on \( Z, \) are not readily implementable. A simple solution, which will be illustrated in detail in Section 8, is to utilize a reasonable guess for \( t, \) say \( t = \alpha/2. \) However, to ensure asymptotic optimality (see Theorem 8), we approximate the FDP estimator using \( p \) and \( \gamma \) so that the data dependent threshold can be approximated and optimal fixed-\( t \) weights can be utilized.

The FDP “approximator” plugs \( G_m(t_m(k/p_m, \gamma_m)) = E[\delta_m(Z_m; t_m(k/p_m, \gamma_m)) | p, \gamma] \) in for each \( \delta_m \) in (6) and (7). Formally, denote \( \tilde{G}_M(t(k, p, \gamma)) = M^{-1} \sum_{m \in M} G_m(t_m(k/p_m, \gamma_m)) \) and define the FDP approximator by

\[
\tilde{FDP}_M(t(k, p, \gamma)) = 1 - \tilde{G}_M(t(k, p, \gamma) / \tilde{t}_M(k, p, \gamma)).
\]

Now, the asymptotically optimal weights are computed as follows.

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

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

\( b. \) for each \( m \in M, \) compute \( w_m^* = w_m(k^*_M, p, \gamma) \) as in (4).
In Theorem 2 below, we see that the restriction \( 0 < \alpha \leq 1 - p(M) \) ensures that a solution to \( \widetilde{FDP}_M(t(k, p, \gamma)) = \alpha \) exists. In practice, this restriction amounts to choosing \( \alpha \) and \( p \) so that \( 0 < \alpha \leq 1 - p_m \) for each \( m \). That is, the prior probability that the null hypothesis is true should be at least \( \alpha \). To see why this condition is reasonable, suppose that \( 1 - p_m < \alpha \) for each \( m \). Then we need not consider a weighting scheme or even collect data in the first place because even if all \( M \) null hypotheses are rejected, the model stipulates that the expected proportion of false discoveries among the \( M \) discoveries is \( M^{-1} \sum_{m \in M} 1 - p_m < \alpha \). Hence, if the conditions of Theorem 2 are not satisfied, then the model should be reconsidered.

**Theorem 2** Under (A1) and Model 1, \( k^*_M \) exists for \( 0 < \alpha \leq 1 - p(M) \).

Observe that \( \bar{t}_M(k^*_M, p, \gamma) = t \) 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 \( \widetilde{FDP}_M(t(k^*_M, p, \gamma)) = \alpha \). 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 \( \delta(Z; tw) \), yields the asymptotically optimal WAMDF.

#### 4.1 Threshold selection

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

We make use of an “adaptive” estimator of the FDP, i.e. it utilizes an estimator of \( M_0 \) defined by

\[
\hat{M}_0(\lambda w) = \frac{M - R(\lambda w) + 1}{1 - \lambda}
\]

for some fixed tuning parameter \( \lambda \in (0, 1) \). This estimator is essentially the weighted version of the estimator in Storey (2002) defined by \( \hat{M}_0(\lambda 1) = [M - R(\lambda 1)]/[1 - \lambda] \). For
earlier work on the estimation of \( M_0 \) see Schweder and Spjotvoll (1982). The idea, in the unweighted setting, is that for \( m \in M_1 \), \( E[\delta_m(\lambda)] \leq 1 \) but the inequality is relatively sharp if all tests have reasonable power and \( \lambda \) is chosen sufficiently large. Hence

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

and \( E[\hat{M}_0(\lambda 1)] \geq M_0 \). That is, \( \hat{M}_0 \) is positively biased but the bias is minor for suitably chosen \( \lambda \) and reasonably powerful tests. Similar intuition applies for \( \hat{M}_0(\lambda w) \). As in Storey et al. (2004), we add 1 to the numerator in expression (6) to ensure that \( \hat{M}_0(\lambda w) > 0 \) for finite sample results.

The adaptive FDP estimator is defined by

\[
\hat{FDP}^\lambda(t w) = \frac{\hat{M}_0(\lambda w)t}{\max\{R(t w), 1\}}.
\] (7)

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 \( \alpha \), is defined by

\[
\hat{t}_\alpha^\lambda = \sup\{0 \leq t \leq u : \hat{FDP}^\lambda(t w) \leq \alpha\}.
\] (8)

We assume that \( u \), the upper bound for \( \hat{t}_\alpha^\lambda \), and the tuning parameter \( \lambda \) satisfy assumption (A2) \( \lambda \leq u \leq 1/w(M) \),

where \( w(M) \equiv \max\{w\} \). This ensures that \( \hat{t}_\alpha^\lambda w_m \leq 1 \) and \( \lambda w_m \leq 1 \) for every \( m \). It should be noted that for \( w = 1 \) and \( u = \lambda \) (which implies \( \hat{t}_\alpha^\lambda \leq \lambda \)), we recover the unweighted adaptive MDF for finite FDR control in Storey et al. (2004).

In practice \( \hat{t}_\alpha^\lambda \) can be difficult to compute. Alternatively, we may apply the original BH procedure to the weighted \( p \)-values at level \( \alpha M/\hat{M}_0(\lambda w) \). Note that due to (2) we can also use weighted \( p \)-values to estimate \( M_0 \) via \( \hat{M}_0(\lambda w) = [M - \sum_{m \in M} I(Q_m \leq \lambda) + 1]/[1 - \lambda] \).

Formally, this threshold selection procedure can be implemented as follows.

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

a. compute \( Q_m = P_m/w_m \) and ordered weighted \( p \)-values via \( Q(1) \leq Q(2) \leq \ldots \leq Q(M) \).
b. If $Q_{(m)} > \alpha m/\hat{M}_0(\lambda w)$ for each $m$, set $j = 0$, otherwise take
\[ j = \max \left\{ m \in \mathcal{M} : Q_{(m)} \leq \alpha/\hat{M}_0(\lambda w) \right\}. \]

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

The WAMDF implemented above is equivalent to $\delta(Z; \hat{t}_\alpha^{\lambda*}w)$ in that
\[ \delta_m(Z_m; \hat{t}_\alpha^{\lambda*}w_m) = I(Q_m \leq \hat{t}_\alpha^{\lambda*}) = I(Q_m \leq \hat{t}_\alpha^{\lambda*}) \] almost surely for each $m$, i.e. both procedures reject the same set of null hypotheses. The first equality in (9) follows from (2) and the last equality in (9) is a consequence of Lemma 2 in Storey et al. (2004).

4.2 The asymptotically optimal WAMDF

The asymptotically optimal WAMDF is formally defined as $\delta(Z; \hat{t}_\alpha^{\lambda*}w^*)$ for $0 < \alpha \leq 1 - p_{(M)}$ and $\lambda = \tilde{t}_M(k_M^*, p, \gamma)$, and where $k_M^*$ and $w^*$ are defined as in the Weight Selection Procedure. It should be noted that this particular choice of $\lambda$ ensures that the employed weights are indeed “asymptotically optimal” (see Theorem 8) and additionally that (A2) is satisfied if we take $u = 1/w_{(M)}$. However, other values of $\lambda$ could be considered, as in Section 8. To implement the asymptotically WAMDF, we first compute $w^*$ using the Weight Selection Procedure. Then, we choose $\lambda = \tilde{t}_M(k_M^*, p, \gamma)$ and $u$ satisfying (A2), collect data $Z = z$ and compute $\delta(z; \hat{t}_\alpha^{\lambda*}w^*)$ 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 take $\alpha = 0.05$. Recall the goal is to test $H_m : \theta_m = 0$ with decision functions $\delta_m(Z_m; t_m) = I(Z_m \geq \Phi^{-1}(t_m))$ or their corresponding $p$-values $P_m = \Phi(Z_m)$ 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. As before, the Weight Selection Procedure is broken down into 2 sub-steps and the Threshold Selection Procedure is split into three sub-steps. Now, to test these null hypotheses we

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

13
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 \( Q_1, Q_2, \ldots, Q_M \).

<table>
<thead>
<tr>
<th>( \theta_m )</th>
<th>( \gamma_m )</th>
<th>( w_m^* )</th>
<th>( Z_m )</th>
<th>( P_m )</th>
<th>( Q_m )</th>
<th>( 0.05m/M_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>0.74</td>
<td>3.14</td>
<td>0.001</td>
<td>0.001</td>
<td>0.006</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1.26</td>
<td>2.55</td>
<td>0.005</td>
<td>0.005</td>
<td>0.012</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0.74</td>
<td>2.56</td>
<td>0.005</td>
<td>0.006</td>
<td>0.018</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1.26</td>
<td>1.47</td>
<td>0.070</td>
<td>0.062</td>
<td>0.024</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1.74</td>
<td>1.17</td>
<td>0.121</td>
<td>0.106</td>
<td>0.030</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0.74</td>
<td>-0.60</td>
<td>0.724</td>
<td>0.844</td>
<td>0.061</td>
</tr>
</tbody>
</table>

1b. Compute asymptotically optimal weights \( w_m^* = w_m(k^*_M, p, \gamma) \) as in (4). See column 3 in Table 2.

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

2b. Observe that \( Q_m \leq \alpha M_0(\lambda w^*) \) for \( m = 3 \) but not for \( m = 4, 5, \ldots, 10 \) and hence \( \alpha j/M_0(\lambda w^*) = 0.05 \frac{3}{23} = 0.013 \).

2c. Compute \( \hat{i}^\lambda \alpha = \min\{0.013, 0.79\} = 0.013 \) 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

Next 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 the following dependence structure for \( Z \):

(A3) \( (Z_m, m \in M_0) \) are mutually independent and independent of \( (Z_m, m \in M_1) \).

This structure has also utilized in Benjamini and Hochberg (1995); Genovese et al. (2006); Peña et al. (2011); Storey et al. (2004) to prove FDR control for unweighted un-adaptive, weighted unadaptive, and unweighted adaptive procedures. It is satisfied under Model 1 conditionally upon \( (\theta, p, \gamma) \), but it is not limited to this setting.

To define the FDR, let \( V(tw) = \sum_{m \in M_0} \delta_m(tw_m) \) denote the number of erroneously rejected null hypotheses (false discoveries) at \( tw \) and recall that \( R(tw) = \sum_{m \in M} \delta_m(tw) \)
is the number of rejected null hypotheses. Define the FDP at $tw$ by

$$FDP(tw) = \frac{V(tw)}{\max\{R(tw), 1\}}.$$  

(10)

The FDR at $tw$ is defined by $FDR(tw) = E[FDP(tw)]$, where the expectation is taken over $Z$ 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 et al. (2004), we force $\hat{t}_\alpha^\lambda \leq \lambda$ by taking $u = \lambda$ in (8). This facilitate 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 $u = \lambda$,

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

(11)

where $\bar{w}_0 = M_0^{-1} \sum_{m \in M_0} w_m$ 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 $w = 1$ then $\bar{w}_0 = 1$ and we recover Theorem 3 in Storey et al. (2004) as a corollary.

Of course, if $w \neq 1$, the bound in Lemma 1 is not immediately applicable because $M_0$ and consequently $\bar{w}_0$ is unobservable. One solution is to use an upper bound for $\bar{w}_0$ and adjust the “$\alpha$” at which the procedure is applied. This adjustment is described below.

Theorem 3 Define

$$\alpha^* = \alpha \frac{1}{w(M)} \frac{1 - \lambda w(M)}{1 - \lambda}.$$  

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

In the next section, we see that $\bar{w}_0$ is typically less than or equal to 1, asymptotically, so that this $\alpha$ adjustment is not needed for large $M$.

6 Asymptotic results

The first subsection shows that WAMDFs always reject more null hypotheses than their unadaptive counterparts and provides sufficient conditions for asymptotic FDP control.
and $\alpha$-exhaustion. These results are used in the asymptotic analysis of the asymptotically optimal WAMDF in the second subsection.

To facilitate asymptotic analysis, denote weight vectors of length $M$ by $w_M$ and the $m$th element of $w_M$ by $w_{m,M}$. Further, denote the mean of the weights from true null hypotheses by $\bar{w}_0,M$. Denote the adaptive FDP estimator in (7) by $\widehat{FDP}_M^\lambda(tw_M)$ and the FDP in (10) by $FDP_M(tw_M)$. We will also consider an unadaptive FDP estimator, which uses $M$ in the place of an estimate of $M_0$, defined by

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

When necessary, we also denote the tuning parameter in (6) by $\lambda_M$ because, as in the asymptotically optimal WAMDF where $\lambda_M = \bar{t}_M(k^*_M,p,\gamma)$, it may depend on $M$.

Recall that we assumed $\lambda \leq u \leq 1/w(M)$ in (A2) to ensure that every individual threshold was bounded above by 1. In asymptotic analysis, (A2) is redefined as follows:

(A2) $\lambda_M \rightarrow \lambda \leq u = 1/k$ almost surely, where $k$ satisfies $\lim_{M \rightarrow \infty} w(M) \leq k$ almost surely, and the adaptive threshold in (8) is denoted $\bar{t}_M^\lambda$. It will be verified that assumption (A2) is satisfied, for example, under Model 1 and (A1) for the asymptotically optimal WAMDF. The unadaptive threshold is defined by

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

### 6.1 Arbitrary weights

Convergence criteria considered here are similar to criteria in Storey et al. (2004); Genovese et al. (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 $t \in (0,u]$, assume that

(A4) $R(tw_M)/M \rightarrow G(t)$ almost surely,

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

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

16
Here $\mu_0$ 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 will ensure that, asymptotically, the FDP is continuous and increasing in $t$ and takes on value 0, thereby ensuring that it can be controlled. Note that writing $R(tw_M)/M = \sum_{m \in M} I(Q_m \leq t)/M$ via (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 will focus on comparing random thresholds $\hat{t}_{\alpha,M}^\lambda$ and $\hat{t}_{\alpha,M}^0$ 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,

$$t_{0,\infty}^\alpha = \sup\{0 \leq t \leq u : FDP_{0,\infty}(t) \leq \alpha\} \quad \text{and} \quad t_{\lambda,\infty}^\lambda = \sup\{0 \leq t \leq u : FDP_{\lambda,\infty}(t) \leq \alpha\}.$$

In Theorem 4 we see that both the unadaptive and adaptive thresholds converge to their asymptotic (nonrandom) counterparts, with the asymptotic adaptive threshold being larger than the asymptotic unadaptive threshold. By virtue of $E[\delta_m(tw_m)]$ being strictly increasing in $t$ for each $m$, it follows that the adaptive procedure will lead to a higher proportion of rejected null hypotheses, asymptotically.

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

$$\lim_{M \to \infty} t_{0,\infty}^0 = t_{\alpha,\infty}^0 \leq \lim_{M \to \infty} t_{\alpha,\infty}^\lambda = t_{\alpha,\infty}^\lambda$$

almost surely.

Before focusing on the FDP it is useful to formally describe the notion of an $\alpha$-exhaustive MDF. Loosely speaking, Finner et al. (2009) referred to an unweighted multiple decision function, say $\delta(\hat{t}_{0,M}^*1_M)$, as “asymptotically optimal” (we will use the terminology $\alpha$-exhaustive) if $FDR(\hat{t}_{0,M}^*1_M) \to \alpha$ 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), $FDR(i_{\alpha,M}^\star 1_M)$ 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. Observe that if (A4) - (A5) are satisfied, then $G(t) = a_0\mu_0 t + (1 - a_0)$ under a DU distribution for $t \leq u$. Denote this particular $G(t)$ by $G^{DU}(t)$.

To study the FDP of WAMDFs consider:

$$\lim_{M \to \infty} FDP_M(i_{\alpha,M}^0 \mathbf{w}_M) \leq \lim_{M \to \infty} FDP_M(i_{\alpha,M}^\lambda \mathbf{w}_M) \leq \alpha \quad (13)$$

and the following three claims regarding the inequalities:

(C1) the first inequality in (13) is satisfied almost surely,

(C2) the second inequality in (13) is satisfied almost surely, and

(C3) the second inequality in (13) 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 $\alpha$ and is referred to as the *asymptotic FDP control* claim. Claim (C3) is the *$\alpha$-exhaustive* claim and states that the asymptotic FDP of the WAMDF is equal to $\alpha$ 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, $\mu_0 = 1$.

Observe that asymptotic FDP control (C2) and $\alpha$-exhaustion (C3) depend on the unobservable value of $\mu_0$, which will necessarily depend on the weighting scheme at hand. The next theorem will be useful for verifying asymptotic FDP control and/or $\alpha$-exhaustion. It states that a WAMDF provides asymptotic FDP control if the states of the null hypotheses (the $\theta_m$s) are positively correlated with the weights, and $\alpha$-exhaustion is provided if weights are uncorrelated with the $\theta_m$s.
Theorem 6 Suppose that \((W_{m,M}, \theta_{m,M}), m \in \mathcal{M}\) are identically distributed random vectors each with support \(\mathbb{R}^+ \times \{0, 1\}\) and with \(E[W_{m,M}] = 1\) and \(E[\theta_{m,M}] \in (0, 1)\). Define,

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

whenever \(\theta_M \neq 1_M\) and take \(\bar{W}_{0,M} = 1\) otherwise. Further assume that \(\bar{W}_{0,M} \rightarrow \mu_0\) almost surely. Then \(\mu_0 \leq 1\) if \(\text{Cov}(W_{m,M}, \theta_{m,M}) \geq 0\) and \(\mu_0 = 1\) if \(\text{Cov}(W_{m,M}, \theta_{m,M}) = 0\).

Before focusing on the asymptotically optimal WAMDF, we first provide Corollary 1, which is now easily established using Theorems 5 and 6. It states that the unweighted adaptive (linear step-up) procedure defined in Storey et al. (2004), which is denoted \(\hat{t}_{\alpha}^{\lambda,1_M}\) here, is \(\alpha\)-exhaustive.

Corollary 1 Suppose that (A4) - (A6) are satisfied and take \(w_M = 1_M\). 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 et al. (2004) is competitive with \(\alpha\)-exhaustive nonlinear procedures in Finner et al. (2009). The fact that a DU distribution is the least favorable among such (unweighted) adaptive linear step-up procedures under our weak dependence structure is also interesting as the search for least favorable distributions remains a challenging problem. See Finner et al. (2007); Roquain and Villers (2011); Finner et al. (2012).

6.2 Asymptotically optimal weights

We first 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” in Theorem 7. As in the previous subsection, weight vectors and elements of weight vectors are indexed by \(M\) to facilitate asymptotic arguments. Further, we sometimes write \(\bar{t}_M(k^*_M) = \bar{t}_M(k^*_M, p, \gamma)\) for brevity.

Perturbed weights are simulated by multiplying each asymptotically optimal weight by a positive random variable \(U_m\) so that, by the law of iterated expectation, perturbed weights are positively correlated with asymptotically optimal weights. They are formally defined by

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

Theorem 7 is formally stated below. It assumes \( \Pr(p_m \leq 1 - \alpha) = 1 \) in Model 1 so that \( \alpha \leq 1 - p(M) \) for every \( M \) with probability 1, which implies that asymptotically optimal weights and their perturbed versions exist; see Theorem 2.

**Theorem 7** Suppose that \( \Pr(p_m \leq 1 - \alpha) = 1 \), take \( \lambda_M = \bar{\lambda}_M(k^*_M) \), and consider perturbed weights \( \tilde{\mathbf{w}}_M \). 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 Claims (C1) and (C2) hold.

Next the notion of “asymptotically optimal” is formalized and some examples of \( \alpha \)-exhaustive weighting schemes are provided. To motivate the first theorem, recall that the asymptotically optimal weights are equivalent to optimal fixed-\( t \) weights with \( t = \bar{\lambda}_M(k^*_M) \). However, the asymptotically optimal WAMDF utilizes asymptotic threshold \( t^\lambda_{\alpha,\infty} \) (see Theorem 4). Theorem 8 states that \( \bar{\lambda}_M(k^*_M) \to t^\lambda_{\alpha,\infty} \) almost surely, i.e. the asymptotically optimal weights are asymptotically equivalent to the optimal fixed-\( t \) weights corresponding to Theorem 1.

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

The other notion of optimality considered in this paper is the notion of \( \alpha \)-exhaustion. The next two corollaries illustrate that asymptotic \( \alpha \)-exhaustive FDP control is provided for a variety of weighing schemes. Specifically, Corollary 2 states that a WAMDF is \( \alpha \)-exhaustive in a worst case scenario weighting scheme, i.e. when weights are generated independent of \( \mathbf{\theta}_M \). The fact that WAMDFs are so robust with respect to weight misspecification, in terms of FDP control, is perhaps not surprising in light of Corollary 1, which recall states that \( \alpha \)-exhaustion is achieved even if \( \mathbf{w}_M = \mathbf{1}_M \).
Corollary 2 Under Model 1 and (A1) - (A2), if $\mathbf{w}_M$ are mutually independent weights and independent of $\mathbf{\theta}_M$ with $E[w_{m,M}] = 1$, then Claims (C1) - (C2) hold for $\alpha \in (0, 1)$ and Claim (C3) holds for $0 < \alpha \leq FDP_\infty(u)$.

Corollary 3 states that if $p_i = p_j$ for every $i, j$ - or equivalently if weights depend only on $\gamma$ - then optimal weights and their perturbed versions allow for $\alpha$-exhaustion as well. This 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 doesn’t exist or is not modeled. For an illustration see Section 8. See also Spjøtvoll (1972); Storey (2007); Peña et al. (2011) for more on this type of heterogeneity.

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

The fact that $\alpha$-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 et al. (2006) in the unadaptive setting, and it was suggested that one potential route for improvement is to incorporate an estimate of $\mu_0$ 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 $K = 1000$ replications, we generate $Z_m \sim N(\theta_m \gamma_m, 1)$ for $m = 1, 2, ..., 1000$ and compute $\delta(\hat{i}_{\alpha,M}^{\lambda} \mathbf{w}_M)$, $\delta(\hat{i}_{\alpha,M}^{\mu} \mathbf{w}_M)$, $\delta(\hat{i}_{\alpha,M}^{\lambda} \mathbf{1}_M)$, and $\delta(\hat{i}_{\alpha,M}^{\mu} \mathbf{1}_M)$ as in Example 1, where $\alpha = 0.05$ and $\lambda_M = \tilde{t}_M(k_M^*, \mathbf{p}, \gamma)$. Recall these MDFs are referred to as weighted adaptive (WA), weighted unadaptive (WU), unweighted adaptive (UA), and unweighted unadaptive (UU) procedures, respectively. The average FDP and average correct discovery proportion (CDP) is computed over the $K$ replications for each procedure, where the CDP is defined by $\text{CDP} = \sum_{m \in \mathcal{M}_1} \delta_m / \max\{M_1, 1\}$.

In each simulation experiment, $\gamma_m \sim Un(1, a)$ for $a = 1, 3, 5$, where $Un(1, a)$ denotes a uniform distribution over $(1, a)$. Observe that when $a = 1$ the effect sizes are identical
while when $a = 3$ or $a = 5$ they vary. In Simulation 1, $p_m = 0.5$ for each $m$ and weighted procedures utilize asymptotically optimal weights. Here the WAMDF is both optimally weighted and $\alpha$-exhaustive; see Corollary 3. This setting also arises in the analysis of the data in Table 1 in the next section. In Simulation 2, weighted procedures use asymptotically optimal weights as before and the effect sizes vary as before, but $p_m \sim Un(0, 1)$. Thus, though the procedure is optimally weighted and asymptotic FDP control is provided, the conditions of Claim (C3) are no longer satisfied. Hence, the WA procedure is not $\alpha$-exhaustive. In Simulation 3, data are generated according to the same mechanism as in Simulation 2, but asymptotically optimal weights are perturbed via $U_m w_{m,M}(k^*, p, \gamma)$, where $U_m \sim Un(0, 2)$. Thus, the WA procedure is no longer optimally weighted but weights are positively correlated with optimal weights. Simulation 4 represents a worst case scenario weighting scheme, in which weights are generated $w_{m,M} \sim Un(0, 2)$.

Detailed results and discussions of simulations are in the Supplemental Article. The main important 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 is less than or equal to 0.05 in all simulations, as Theorem 7 stipulates. Further, its average CDP is as large as or larger than the CDP of all other procedures in the first three simulations. The WA procedure does have a slightly smaller average CDP than the UA procedure in the worst case scenario (Simulation 4), as we might expect.

8 A simple WAMDF for heterogeneous sample sizes

As mentioned in the Introduction, one of the challenges in exploiting heterogeneity across tests is that, while optimal weighting schemes can be developed when effect sizes and/or prior probabilities are known, in practice these “Oracle” parameters are rarely observable. Further, Oracle parameter estimation can be involved and cause FDR methods to fail; see Rubin et al. (2006); Habiger and Peña (2014). However, the WAMDF proposed in this paper allows for more efficient FDR control even when weights are not optimal. This robustness property allows one to avoid Oracle parameter estimation and develop practical and easily implemented WAMDFs based on reasonable assumptions about the nature and
degree of the heterogeneity at hand. This section illustrates the idea. Specifically, a simple WAMDF for settings when the only observable heterogeneity is attributable to the sample sizes across tests, as in Table 1, is developed. The main idea is that though an individual effect size $\gamma_m$ is not known or easily estimated, the sample size $n_m$ is known and $\gamma_m = O(n_m^{1/2})$. Hence, larger sample sizes should generally lead to larger effect sizes. The WAMDF is first developed and then illustrated on the data in Table 1. A brief simulation study is used to assess the method.

8.1 The WAMDF

Let $T_m$ be a test statistic with $E[T_m] = \mu_m$ and $Var(T_m) = \sigma_m/n_m$. To accommodate Model 1, let $\theta_m \sim \text{Bernoulli}(p)$ and assume $\mu_m = (1 - \theta_m)\mu_0 + \theta_m\mu_1$ where $\mu_0$ is known. Further assume $\sigma_m = (1 - \theta_m)\sigma_0 + \theta_m\sigma_1$ where $\sigma_0$ is known. It should be noted that $\mu_1$ and $\sigma_1$ need not be known or specified in what follows. They may merely represent marginal means and variances when $H_m$ is false. Regardless, the purpose of this model is to facilitate weight construction; it need not be satisfied for FDR control, for example.

Now, assume that the goal is to test null hypothesis $H_m : \mu_m = \mu_0$ against alternative hypothesis $K_m : \mu_m \neq \mu_0$ with test statistic $T_m$ based on sample size $n_m$.

Define $Z$-score for $T_m$ by

$$Z_m = \sqrt{n_m} \left( \frac{T_m - \mu_0}{\sigma_0} \right)$$

and define decision function $\delta_m(Z_m; t_m) = I(|Z_m| \geq \Phi^{-1}(t_m/2))$. Denote $\gamma_m = \sqrt{n_m} |\mu_1 - \mu_0|\sigma_1/\sigma_0 \equiv \sqrt{n_m} \gamma$. Observe $E[\delta_m(Z_m; t_m)|\theta_m = 1] = \Phi(\Phi^{-1}(t_m/2) - \gamma_m) + o(1)$. Thus, we focus on power function

$$\pi_{\gamma_m}(t_m) = \Phi(\Phi^{-1}(t_m/2) - \gamma_m) = \Phi(\Phi^{-1}(t_m/2) - \sqrt{n_m} \gamma).$$

(15)

Based on this power function approximation, we propose weights $w_m = t_m/\bar{t}$ where

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

(16)

where $\sqrt{n} = \sum_m \sqrt{n_m}$. Some justification is as follows. For the moment, assume that the proportion of true null hypotheses is $p = 0.5$ and that the average power is $\pi = \pi_{\gamma}.
\[ \Phi(\Phi^{-1}(t/2) - \gamma) = 0.5. \] Under these assumptions, solving \( pt/\left[p + (1 - p)\pi\right] = \alpha \) gives an overall threshold \( t = \alpha/[2(1 - \alpha)] \approx \alpha/2 \) for small \( \alpha \). Further solving \( \Phi(\Phi^{-1}(t/2) - \gamma) = 0.5 \) yields \( \gamma = \Phi^{-1}(t/2) = \Phi^{-1}(\alpha/4) \). As in Example 1, taking the derivative of \( \pi(m(t_m) = \Phi(\Phi^{-1}(t_m/2) - \gamma) \) with respect to \( t_m \) and setting it equal to \( k/p \) and solving yields \( log(k/p) = \Phi^{-1}(t_m/2)\gamma_m - 0.5\gamma_m^2 \) and also gives \( t_m = 2\Phi \left( 0.5\gamma_m + \frac{log(k/p)}{\gamma_m} \right) \).

Plugging \( \Phi^{-1}(t/2)\gamma - 0.5\gamma^2 = \gamma^2 - 0.5\gamma^2 = 0.5\gamma^2 \) in for \( log(k/p) \) in the above expression we get

\[ t_m = 2\Phi \left( 0.5\gamma_m + 0.5\gamma^2/\gamma_m \right). \] (17)

Finally, because \( \gamma_m = \sqrt{n_m}\gamma \) for some \( \gamma \) and \( \Phi^{-1}(\alpha/4) = \tilde{\gamma} = \gamma\sqrt{m}/M \), we get \( \gamma_m = M\gamma\sqrt{m}/\sqrt{n} \) and recall \( \tilde{\gamma} = \Phi^{-1}(\alpha/4) \). Plugging in for \( \gamma_m \) and \( \tilde{\gamma} \) in (17) gives the expression in (16).

Now, to test \( H_m : \mu_m = \mu_0 \) against \( H_m : \mu_m \neq \mu_0 \) with test statistic \( T_m \) based on sample size \( n_m \) for \( m = 1, 2, ..., M \) the following the following procedure is proposed. Note that below we choose \( \lambda = 0.5 \) for simplicity and because this choice was recommended in Storey et al. (2004) for unweighted adaptive MDFs.

1. Compute weights \( w_m = t_m/\ell \) where \( t_m \) is computed as in (16).

2. Compute weighted \( p \)-values \( Q_m = P_m/w_m = 2[\Phi(|Z_m|)]/w_m \) and order them via \( Q(1) \leq Q(2) \leq ... \leq Q(M) \).

3. If \( Q(m) > \alpha m/\tilde{M}_0(0.5w) \) for every \( m \) then retain all null hypotheses. Otherwise take \( j = \max\{m : Q(m) \leq \alpha m/\tilde{M}_0(0.5w)\} \) and reject the null hypotheses corresponding to \( Q(1), Q(2), ..., Q(j) \).

8.2 Illustration

To analyze the data in Table 1, the goal is to test \( H_m : \beta_{1m} = 0 \) vs. \( K_m : \beta_{1m} \neq 0 \) for each \( m \), where \( \beta_{1m} \) is the regression coefficient for regressing \( Y_m = (Y_{im}, i = 1, 2, ..., 5) \) on \( \mathbf{x} = (x_i, i = 1, 2, ..., 5) \) with a log-linear model defined by \( log(\mu_{im}) = \beta_{0m} + \beta_{1m}x_i \), where \( Y_{im} \) are independent Poisson random variables with mean \( \mu_{im} \). As per McCullagh and Nelder
Weights $w^*_m$ for $\alpha = 0.01(\circ), 0.05(\triangle)$, and $0.10(\triangledown)$ are on the left. Boxplots of $Q_m - P_m$ vs. $n_m$ are on the right for $\alpha = 0.05$.

(1989), we focus on the conditional distribution of $Y_m$ given $\sum_{i=1}^{5} Y_{im} = n_m$ and utilize sufficient statistic $T_m = x^T Y_m / n_m$. Observe that $\mu_0 = E[T_m|\theta_m = 0, \sum_{i=1}^{5} Y_{im} = n_m] = \bar{x}$ and $Var(T_m|\theta_m = 0, \sum_{i=1}^{5} Y_{im} = n_m) = \sigma_0/n_m$, where $\sigma_0 = x' [\frac{1}{5} diag(1) - \frac{1}{25} J] x$. Thus, the Z-score is

$$Z_m = \sqrt{n_m} \left( \frac{T_m - \bar{x}}{x' [\frac{1}{5} diag(1) - \frac{1}{25} J] x} \right).$$

Weights are computed as in (16) and modified via $w^*_m = M[w_m + 0.01]/\sum_m[w_m + 0.01]$ to safeguard against extremely small weights - for example $w_m < 10^{-16}$ when $n_m = 911$ which would be deemed impractical. Weights are depicted in Figure 2 for $\alpha = 0.01, 0.05$ and $0.10$ and weighted $p$-values vs. unweighted $p$-values are compared. Observe that weights are less than 1 when $n_m$ is small or large, in which case we may anticipate small or large power, respectively. Consequently, weighted $p$-values are stochastically greater than regular $p$-values. On the other hand, weights are greater than 1 for moderate sample sizes when we may anticipate moderate power, and weighted $p$-values are stochastically less than regular $p$-values. See Peña et al. (2011) for a discussion on this phenomenon.

For $\alpha = 0.01$ the weighted and unweighted adaptive procedures resulted in 39 and 40 discoveries, respectively. For $\alpha = 0.05$ both procedures resulted in 87 discoveries and for $\alpha = 0.10$, the weighted procedure yielded 125 discoveries and the unweighted procedure resulted in 122.
8.3 Simulation

A modest simulation study is used to assess the performance of the WAMDF applied in the previous subsection. In Simulation 5, for each of 1000 replications, we sample $M = 1000$ $n_m$s with replacement 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 consider all $p, \gamma$ combination where $\gamma$ is chosen so that $\bar{\gamma} = \gamma M^{-1} \sqrt{\pi} = 1.75, 2, 2.25$ and $p = 0.2, 0.5, 0.8$. For each replication and setting, the unweighted adaptive MDF is applied and the WAMDF is applied with $\alpha = 0.01, 0.05, 0.10$. The average FDP and CDP are recorded over the 1000 replications for each setting.

Detailed results are in the Supplemental Article. The important point is that, though the weights are based on the assumption that the average power (CDP) is 0.5 and the proportion of false nulls is 0.5, the WAMDF is more powerful than its unweighted counterpart even if $p = 0.2$ or $p = 0.8$ as long as the CDP is at least 0.2. Further, the average FDP of both procedures is always less than $\alpha$. In summary, the robustness properties of the proposed WAMDF indeed facilitate simple procedure for heterogeneity that perform better than their competitors even though some simplifying assumptions are made.

It should be noted that it is not necessary to assume that the average power is 0.5 and that the prior probabilities for the states of the null hypotheses are equal to 0.5, as was done in the development of the WAMDF in this section. These assumptions were made merely because they are the least informative, lead to the simplest weights and because the resulting WAMDF outperforms its unweighted counterpart in most scenarios. It is possible (and not difficult), to extend the reasoning in Subsection 8.1 to develop a WAMDF more suitable for low/high power settings and various sparsity settings. For example, solving for $t$ in $pt/[pt + (1 - p)\pi] = \alpha$ yields $t = [(1 - p)\pi]/[p(1 - \alpha)]$, which can be used in place of $t = \alpha/2$ for other specification of $p$ and $\pi$. 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 1) controlling the FDR at a level nearer $\alpha$ or 2) exploiting heterogeneity across tests. This paper combined these ideas using a weighted decision theoretic framework and showed that the resulting procedure is more powerful than procedures which only consider 1) or 2), but not both. Specifically, we have provided weighted adaptive multiple decision functions that satisfy the $\alpha$-exhaustive optimality criterion considered in Finner et al. (2009), but allow for further improvements via an optimal weighting scheme that incorporates heterogeneity.

Sections 6 and 7 demonstrated that the proposed WAMDFs are robust. This allows for simple procedures for more efficient FDR control to be developed even if the nature and degree of heterogeneity is not precisely known. To illustrate, Section 8 developed a simple FDR method for heterogeneous samples sizes. Certainly other weighting schemes, varying in complexity and efficiency, could be developed for other types of heterogeneity as well. For example, as mentioned in the previous section, different sparsity or power considerations could be considered for the heterogeneous sample size setting. In other applications heterogeneous prior probabilities for the states of the null hypotheses are available or estimable, and could be and be exploited using the proposed WAMDF framework. While such methods are warranted, and in fact motivated this work, they are also beyond the scope of the current manuscript. The goal here was to develop a general framework to accommodate the development of such methods.

Finite sample results and asymptotic results in this paper 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. It would be interesting to study the performance of weighted adaptive procedures under other types of dependence. However, obtaining finite sample analytical results for adaptive MDFs appears to be very challenging under dependence. See Blanchard and Roquain (2009); Roquain and Villers (2011) for some results. As for large sample results, Fan et al. (2012); Desai and Storey (2012) provide techniques for transforming test statistics so that they are weakly dependent, and recall 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 et al. (2004) in the WAMDF or to consider data dependent choices of the tuning parameter $\lambda$ as in Liang and Nettleton (2012). However, a more detailed assessment of $\hat{M}_0(\lambda_w)$, though warranted, is also beyond the scope of the current manuscript.

References


