Generalized estimators for multiple testing: proportion of true nulls and false discovery rate

by

Xiongzhi Chen and R.W. Doerge

Department of Statistics, Purdue University, West Lafayette, USA.

Technical Report #12-04

Department of Statistics
Purdue University

November 2012

Generalized estimators for multiple testing: proportion of true nulls and false discovery rate

Xiongzhi Chen and R.W. Doerge †

Department of Statistics, Purdue University, West Lafayette, USA.

Summary. Two new estimators are proposed: one for the proportion of true null hypotheses and the other for the false discovery rate (FDR) of one-step multiple testing procedures (MTPs). They outperform existing such estimators when applied to discrete p-values whose null distributions dominate the uniform distribution and reduce to leading such estimators when applied to continuous p-values. For the new estimator of the FDR, we establish its simultaneous asymptotic conservativeness and justify formally the stopping time property of its threshold for p-values not necessarily independent or continuous. The superior performance of the new estimators is demonstrated theoretically and by simulation studies and an application to next-generation sequencing count data.

Keywords: Multiple testing; Discrete p-values; Generalized estimators; Stopping time property

1. Introduction

In typical simultaneous testing, there are m null hypotheses H_i with associated p-values p_i, $i = 1, ..., m$, such that only m_0 of them are true nulls but the rest m_1 false nulls. Further, the proportion $\pi_0 = m_0/m$ is unknown. To better balance the overall statistical error and power in multiple testing, Benjamini and Hochberg (1995) proposed the concept of the false discovery rate (FDR) and a linear step-up MTP (BH procedure) whose FDR is no larger than a prespecified level. Let I_0 be the set of indices i for the true nulls and I_1 that for the false nulls. A one-step MTP that uses a threshold $t \in [0, 1]$ to decide the status of each null hypothesis such that H_i is false if and only if $p_i \leq t$ induces

$$\begin{cases} V(t) = \#\{i \in I_0 : p_i \leq t\}, \\ S(t) = \#\{i \in I_1 : p_i \leq t\}, \\ R(t) = V(t) + S(t), \end{cases}$$

†Corresponding Author. Email: doerge@purdue.edu. Address: Department of Statistics, Purdue University, West Lafayette, IN 47907-1399, USA

and its FDR is defined as

$$FDR(t) = E\left[\frac{V(t)}{R(t)} \middle| R(t) > 0 \right].$$

Theorem 5.1 of Benjamini and Yekutieli (2001) shows that the FDR of the BH procedure, FDR_{BH}, satisfies

$$FDR_{BH} \leq \pi_0 \alpha \tag{1}$$

for independent p-values, where $\alpha \in [0,1]$ is the prespecified FDR level. The upper bound in (1) shows the role of the unknown π_0 in quantifying FDR_{BH}. Specifically, using an estimator $\hat{\pi}_0$ of π_0, such that $\pi_0 \leq \hat{\pi}_0 < 1$ when $\pi_0 < 1$, gives a tighter upper bound on FDR_{BH} than using $\pi_0 = 1$, and this in turn yields a rejection threshold no smaller than that induced by the BH procedure at the same FDR level. Consequently, incorporation of a proper $\hat{\pi}_0$ potentially increases the power of the BH procedure. This observation has produced various estimators (denoted by $\hat{\pi}_0$) of π_0 and resulted in many adaptive FDR control and/or estimation procedures (Storey et al., 2004; Benjamini et al., 2006; Blanchard and Roquain, 2009). However, almost all $\hat{\pi}_0$'s and adaptive FDR procedures are developed for continuous p-values whose cumulative distribution functions (c.d.f.'s) under true nulls (null p-value distributions) are uniform on the unit interval.

Besides continuous p-values, discrete p-values of various exact tests, such as the Fisher's exact test (FET, Lehmann and Romano, 2005) and the exact negative binomial test (ENT, Robinson and Smyth, 2008), are widely used when conditional inference is conducted with discrete test statistics in the analysis of data produced by next-generation technologies (NGS). The null p-value distributions induced by these tests are discrete and heterogeneous. Without proper adjustment, these characteristics are very likely to make more conservative $\hat{\pi}_0$ and FDR procedures that were originally designed for continuous p-values. Pounds and Cheng (2006) discusses the effects of discrete, non-uniform p-value distributions on $\hat{\pi}_0$ and FDR procedures. For permutation p-values, they showed that the BH procedure and the procedure in Storey (2002) can be erroneous, and proposed two robust estimators $\hat{\pi}_0^{PC}$ and $\tilde{\pi}_0^{PC}$ of π_0 whose resultant FDR estimators outperform the previous two.

Despite these results and the ubiquity of discrete data, little progress has been made in developing better $\hat{\pi}_0$'s for discrete p-values. In Section 2, we show that Storey's estimator $\hat{\pi}_0^{St}$ (Storey, 2002), $\hat{\pi}_0^{PC}$ and $\tilde{\pi}_0^{PC}$ all have excessive upward biases when applied to discrete p-values. In Section 3, we develop a new estimator $\hat{\pi}_0^{CD}$ of π_0 for discrete p-values such that $\hat{\pi}_0^{CD}$ inherits all excellent properties of $\hat{\pi}_0^{St}$ for continuous p-values. Unlike existing estimators of π_0, $\hat{\pi}_0^{CD}$

explicitly accounts for the deviations of discrete p-value distributions from the uniform distribution and thus has less upward bias. We provide conditions to ensure that $\hat{\pi}_0^{CD}$ is less conservative than some of its competitors and discuss the choice of tuning parameters for $\hat{\pi}_0^{CD}$. In Section 4, we propose a generalized FDR estimator for a one-step MTP, establish its simultaneous asymptotic conservativeness, and show the stopping time property of the thresholds of Storey-type FDR estimators. In Section 5, we compare $\hat{\pi}_0^{CD}$ with $\hat{\pi}_0^{St}$, $\hat{\pi}_0^{PC}$ and $\tilde{\pi}_0^{PC}$ using two-sided p-values of FETs and ENTs from simulated data. Our simulation studies demonstrate the superior performance of $\hat{\pi}_0^{CD}$ in terms of accuracy, conservativeness and stability. The generalized estimators are applied to NGS count data for *Arabidopsis thaliana* and shown to result in improvement. Section 6 concludes the paper with discussion. All proofs are relegated into the Appendix.

2. Excessive bias of three existing estimators

Let the probability space be $(\Omega, \mathcal{A}, P)$ with Ω the sample space, $\mathcal{A}$ the sigma-algebra on Ω, and P the probability measure on $\mathcal{A}$. Without loss of generality (WLOG), we assume $\pi_0 < 1$. The conservativeness of adaptive FDR procedures relies on a crucial property of an estimator $\hat{\pi}_0$ of π_0, i.e., $E[\hat{\pi}_0] \geq \pi_0$ or $\hat{\pi}_0 \geq \pi_0$, P-a.s., called "conservativeness" or "total conservativeness" of the estimator, respectively. Without (total) conservativeness, an adaptive FDR procedure may fail to control or conservatively estimate the FDR of an MTP. We show that some existing estimators of π_0 can be too conservative (i.e., too upwardly biased) for discrete p-values.

To estimate π_0, Storey (2002) introduces

$$\hat{\pi}_0^{St}(\lambda) = \frac{\sum_{i=1}^{m} \mathbf{1}_{\{p_i > \lambda\}}}{(1-\lambda)\, m}, \ \lambda \in [0,1). \tag{2}$$

Let $U(a_1, b_1)$ denote the uniform distribution on a non-empty interval $[a_1, b_1]$. When $p_i \sim U(0,1)$ for all $i \in I_0$, $E\left[\hat{\pi}_0^{St}(\lambda)\right] = \pi_0 + \dfrac{\sum_{i \in I_1} \{1 - F_i(\lambda)\}}{(1-\lambda)\, m} = \pi_0 + b_1$ and $b_1 \geq 0$ (usually positive) is the bias caused by the unknown F_i, $i \in I_1$, where F_i, $i = 1, .., m$, is the the c.d.f. of p_i and any c.d.f. is taken to be right continuous with left limits (i.e., càdlàg). However, when $F_i(t) \leq t$ for $t \in [0,1]$ Lebesgue almost surely (denoted by "$F_i \preceq U(0,1)$" and referred to as "F_i dominates the uniform distribution"), $i \in I_0$,

$$E\left[\hat{\pi}_0^{St}(\lambda)\right] = \pi_0 + \frac{\sum_{i \in I_0} \{\lambda - F_i(\lambda)\}}{(1-\lambda)\, m} + b_1 = \pi_0 + b_0 + b_1$$

and $b_0 \geq 0$ (usually positive) is the extra bias.

Pounds and Cheng (2006) justified that, for randomized tests, FDR procedures that do not account for the discreteness or non-uniformity of the p-values are usually unreliable. They proposed two estimators of π_0 as

$$\tilde{\pi}_0^{PC} = \begin{cases} \min(1, 2\bar{p}) & \text{for 2-sided } p\text{-values}, \\ \min(1, 2\bar{a}) & \text{for 1-sided, continuous } p\text{-values}, \\ \min(1, 8\bar{a}) & \text{for 1-sided, discrete } p\text{-values}, \end{cases}$$

where $\bar{p} = m^{-1} \sum_{i=1}^{m} p_i$, $a_i = 2\min(p_i, 1-p_i)$, $\bar{a} = m^{-1} \sum_{i=1}^{m} a_i$, and

$$\hat{\pi}_0^{PC} = \min\left(1, \frac{1}{m} \sum_{i=1}^{m} \frac{p_i}{E[p_i | H_i = 0]}\right),$$

where $H_i = 0$ or 1 is set depending on whether H_i is a true or false null. They claimed that $\tilde{\pi}_0^{PC}$ is very robust to various types of p-value distributions, and demonstrated its improvement over $\hat{\pi}_0^{St}(\lambda)$. They further claimed $\hat{\pi}_0^{PC}$ to be better than $\tilde{\pi}_0^{PC}$. Interestingly, these two estimators still may have excessive upward bias when $F_i \preceq U(0,1)$, $i \in I_0$ as we show next. WLOG, we consider two-sided discrete p-values and assume neither of $\tilde{\pi}_0^{PC}$ nor $\hat{\pi}_0^{PC}$ is 1. For $\tilde{\pi}_0^{PC}$, we see

$$E\left[\tilde{\pi}_0^{PC}\right] \geq 2m^{-1} \sum_{i \in I_0} E[p_i] + 2m^{-1} \sum_{i \in I_1} E[p_i] \geq \pi_0 + 2m^{-1} \sum_{i \in I_1} E[p_i].$$

So $\tilde{\pi}_0^{PC}$ has extra (usually positive) bias $2m^{-1} \sum_{i \in I_0} E[p_i] - \pi_0 \geq 0$. Similarly,

$$E\left[\hat{\pi}_0^{PC}\right] = \pi_0 + \frac{1}{m} \sum_{i \in I_1} \frac{E[p_i | H_i = 1]}{E[p_i | H_i = 0]} = \pi_0 + b_3,$$

and if some ratios in the summation are large, so will be its positive bias b_3.

3. A generalized estimator of π_0

Let F_i^* denote the null p-value distribution for p_i, $i = 1, ..., m$. Unless otherwise stated, it is assumed that F_i^* dominates $U(0,1)$. To accommodate the discreteness of p-values, and to reduce the upward bias of $\hat{\pi}_0$ as much as possible whiling preserving its conservativeness, we define

$$\hat{\pi}_0^{CD}(\lambda, \epsilon) = \max\left[0, \min\left\{1, \hat{\pi}_0^{St}(\lambda) - \frac{\delta(\lambda, \epsilon)}{(1-\lambda)m}\right\}\right], \tag{3}$$

where

$$\delta(\lambda, \epsilon) = \sum_{i=1}^{m} \varepsilon_i \{\lambda - F_i^*(\lambda)\},$$

and $\epsilon = (\varepsilon_1, ..., \varepsilon_m)$ is a vector of prespecified or adaptively estimated constants (usually in $[0, 1]$).

We denote $\hat{\pi}_0^{CD}(\lambda, \epsilon)$ by $\hat{\pi}_0^{CD}$. When $\varepsilon_i = \varepsilon$ for some $\varepsilon \in [0, 1]$ for all $1 \leq i \leq m$, we write $\hat{\pi}_0^{CD}(\lambda, \epsilon)$ as $\hat{\pi}_0^{CD}(\lambda, \varepsilon)$ and $\hat{\pi}_0^{CD}(\lambda, 1)$ as $\hat{\pi}_0^{CD}(\lambda)$. Clearly $\hat{\pi}_0^{CD}(\lambda, \epsilon) \leq \hat{\pi}_0^{St}(\lambda)$. Since $\hat{\pi}_0^{St}(\lambda) \in [0, 1]$ has to hold even though (2) does not ensure this, we see either $\hat{\pi}_0^{CD}(\lambda, \epsilon) = \hat{\pi}_0^{St}(\lambda)$ when each F_i^*, $i = 1, ..., m$ is $U(0, 1)$ or $\hat{\pi}_0^{CD}(\lambda, 0) = \hat{\pi}_0^{St}(\lambda)$. Hence, $\hat{\pi}_0^{CD}(\lambda, \epsilon)$ generalizes $\hat{\pi}_0^{St}(\lambda)$. Moreover, it distinguishes itself from all existing estimators of π_0 by explicitly subtracting the excessive upward bias caused by discrete p-values in its expectation, thus giving highly accurate and less conservative estimate.

3.1. Conservativeness of $\hat{\pi}_0^{CD}$

We present conditions to ensure the conservativeness of $\hat{\pi}_0^{CD}(\lambda, \varepsilon)$, and the relationship between $\hat{\pi}_0^{CD}(\lambda)$, $\hat{\pi}_0^{St}(\lambda)$ and $\hat{\pi}_0^{PC}$ below.

THEOREM 1. *Suppose* $\max\left\{\hat{\pi}_0^{CD}(\lambda), \hat{\pi}_0^{CD}(\lambda, \varepsilon), \hat{\pi}_0^{St}(\lambda), \hat{\pi}_0^{PC}\right\} < 1$. *Then the following hold.*

(a) $\pi_0 \leq E\left[\hat{\pi}_0^{CD}(\lambda, \varepsilon)\right]$ *if and only if*

$$\frac{\sum_{i=1}^{m} \{F_i(\lambda) - \varepsilon F_i^*(\lambda)\}}{(1 - \lambda) m} \leq \frac{1 - \varepsilon\lambda}{1 - \lambda} - \pi_0. \tag{4}$$

(b) Let $\kappa_i = E[p_i | H_i = 1] - E[p_i | H_i = 0]$. *Then* $E\left[\hat{\pi}_0^{CD}(\lambda)\right] \leq E\left[\hat{\pi}_0^{PC}\right]$ *if and only if*

$$\sum_{i \in I_1} \frac{\kappa_i}{E[p_i | H_i = 0]} \leq \frac{1}{1 - \lambda} \sum_{i \in I_1} \{F_i(\lambda) - F_i^*(\lambda)\}.$$

The proof of Theorem 1 follows from the definitions and is omitted. From this, we immediately have

COROLLARY 1. *Suppose* $\max\left\{\hat{\pi}_0^{CD}(\lambda), \hat{\pi}_0^{St}(\lambda), \hat{\pi}_0^{PC}\right\} < 1$. *Then*

$$\pi_0 \leq E\left[\hat{\pi}_0^{CD}(\lambda)\right] \leq \min\left\{E\left[\hat{\pi}_0^{St}(\lambda)\right], E\left[\hat{\pi}_0^{PC}\right]\right\}$$

if and only if

$$m^{-1} \sum_{i \in I_1} \frac{\kappa_i}{E[p_i | H_i = 0]} \leq m^{-1} \sum_{i \in I_1} \frac{F_i(\lambda) - F_i^*(\lambda)}{1 - \lambda} \leq \pi_1. \tag{5}$$

The proof of Corollary 1 is also omitted. Though Corollary 1 gives the necessary and sufficient condition to ensure that $\hat{\pi}_0^{CD}(\lambda)$ is the least conservative among the three, both (4) and (5) are not verifiable for $\lambda \neq 0$ due to the lack of information on F_i, $i \in I_1$. Nevertheless, surprisingly our simulation studies in Section 5 show that $\pi_0 \leq E\left[\hat{\pi}_0^{CD}(0.5)\right] \leq E\left[\hat{\pi}_0^{PC}\right]$ for relatively large values of π_0.

3.2. Computing $\hat{\pi}_0^{CD}(\lambda, \varepsilon)$ and choice of tuning parameters

The definition of $\hat{\pi}_0^{CD}$ implies the following algorithm for its computation, which we state only for discrete p-values. Assume each F_i^* is discontinuous and let its support be

$$S_i = \left\{ t \in \mathbb{R} : \int_{\{t\}} dF_i^*(t) > 0 \right\}.$$

Clearly, for each λ, there must be a unique $t_{i,\lambda} \in S_i$ such that

$$t_{i,\lambda} = \sup\{t \in S_i : \lambda \geq t\}.$$

Since $F_i^*(t) = t$ if $t \in S_i$ and that F_i^* is càdlàg, it follows that

$$\int_{(\lambda,1]} dF_i^*(t) = 1 - F_i^*(\lambda) = 1 - t_{i,\lambda} \geq 1 - \lambda.$$

Thus

$$\hat{\pi}_0^{CD}(\lambda, \epsilon) = \frac{\sum_{i=1}^m \left[\mathbf{1}_{\{p_i > \lambda\}} - \varepsilon_i (\lambda - t_{i,\lambda})\right]}{(1-\lambda)\, m},$$

where we have omitted the min, max operations in (3) for notational simplicity.

When m is large but there is little or no information on p-value distributions under the alternative hypotheses, it is more feasible to set all $\varepsilon_i = \varepsilon$. So we focus on $\hat{\pi}_0^{CD}(\lambda, \varepsilon)$ and discuss choice of tuning parameters λ, ε. In the ideal case where (4) is satisfied for (λ, ε) in a set $K_0 \subseteq [0,1) \times [0,1]$ composed of finitely many connected components, bootstrap method similar to that in Storey et al. (2004) can be implemented to find the "best" tuning pair $(\lambda_*, \varepsilon_*)$. For completeness this method is described below:

Step 1: for two sufficiently small d^*, $d_1^* > 0$, form a finite, compact set $K \subseteq K_0$ such that

$$\min\{|\lambda - \lambda'| + |\varepsilon - \varepsilon'| : (\lambda, \varepsilon) \in K, (\lambda', \varepsilon') \in K_0\} \leq d^*,$$

and

$$\max\{|\lambda - \lambda_1| + |\varepsilon - \varepsilon_1| : (\lambda, \varepsilon), (\lambda_1, \varepsilon_1) \in K\} \leq d_1^*;$$

compute $\hat{\pi}_0^{CD}(\lambda, \varepsilon)$ for each $(\lambda, \varepsilon) \in K$.

Step 2: for each $(\lambda, \varepsilon) \in K$, form B bootstrap versions $\hat{\pi}_{0,b}^{CD}(\lambda, \varepsilon)$ of $\hat{\pi}_0^{CD}(\lambda, \varepsilon)$, $b = 1, ..., B$ by taking bootstrap samples of the p-values.

Step 3: for each $(\lambda, \varepsilon) \in K$, estimate its respective mean-squared error (MSE) as

$$\widehat{MSE}(\lambda, \varepsilon) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\pi}_{0,b}^{CD}(\lambda, \varepsilon) - \min_{(\tilde{\lambda}, \tilde{\varepsilon}) \in K}\left\{\hat{\pi}_0^{CD}\left(\tilde{\lambda}, \tilde{\varepsilon}\right)\right\}\right]^2.$$

Step 4: set $(\lambda_*, \varepsilon_*) = \arg\min_{(\lambda,\varepsilon)\in K}\left\{\widehat{MSE}(\lambda, \varepsilon)\right\}$, and the overall estimate of π_0 is $\hat{\pi}_0^{CD}(\lambda_*, \varepsilon_*)$.

When $\varepsilon = 0$ or the p-values are continuously distributed with null distribution being $U(0, 1)$, $\hat{\pi}_0^{CD}(\lambda, \epsilon) = \hat{\pi}_0^{St}(\lambda)$ and the above bootstrap method reduces to that in Storey et al. (2004) when $K = \{0, 0.05, ..., 0.95\} \times \{0\}$ is used as the searching grid. However, knowing the impracticality of condition (4), we will not compute $\hat{\pi}_0^{CD}(\lambda_*, \varepsilon_*)$. Instead, we will empirically show in Section 5 that $\hat{\pi}_0^{CD}(0.5)$ is conservative, and less upwardly biased than $\hat{\pi}_0^{St}(0.5)$, $\hat{\pi}_0^{PC}$ and $\tilde{\pi}_0^{PC}$ for relatively large values of π_0. We provide a heuristic explanation for the conservativeness of $\hat{\pi}_0^{CD}(0.5)$. Let $\pi = 1 - \pi_0$. Consider the choice of $\lambda = 1/2$ and $0 \le \varepsilon \le 1$. Then $E\left[\hat{\pi}_0^{CD}(\lambda, \varepsilon)\right] = \pi_0 + \beta_0 + \beta_1$, where $\beta_0 = 2m^{-1}(1 - \varepsilon)\sum_{i\in I_0}\{1/2 - F_i(1/2)\} \ge 0$ by definition and

$$\begin{aligned}\beta_1 &= 2(1 - \varepsilon/2)\pi - 2m^{-1}\sum_{i\in I_1}\{F_i(1/2) - \varepsilon F_i^*(1/2)\} \\ &> 2(1 - \varepsilon/2)\pi - 2\pi = -\varepsilon\pi.\end{aligned}$$

When π_0 is large and m is not too small, $\beta_0 \ge \varepsilon\pi_1$ for $\varepsilon < 1$; when $\varepsilon = 1$, $\beta_1 \ge 0$ can happen since F_i, $i \in I_1$ does not have to dominate $U(0, 1)$. Therefore, under such circumstance, choosing ε close to or equal to 1 usually yields $\beta_0 + \beta_1 \ge 0$ and $E\left[\hat{\pi}_0^{CD}(\lambda, \varepsilon)\right] \ge \pi_0$.

4. FDR estimation using $\hat{\pi}_0^{CD}$

With $\hat{\pi}_0^{CD}$ it is natural to update some existing adaptive FDR procedures. Storey's FDR estimator (Storey et al., 2004) is defined as

$$\widehat{FDR}_\lambda(t) = \frac{\hat{\pi}_0^{St}(\lambda)\, t}{m^{-1}\{R(t) \vee 1\}},$$

whose variant is $\widehat{FDR}^*_\lambda(t) = \dfrac{\tilde{\pi}_0^{St}(\lambda)\, t}{m^{-1}\{R(t) \vee 1\}} 1(t \leq \lambda) + 1(t > \lambda)$, where $\tilde{\pi}_0^{St}(\lambda) = \hat{\pi}_0^{St}(\lambda) + (1-\lambda)^{-1} m^{-1}$. We remark that $\tilde{\pi}_0^{St}(\lambda)$ may have non-negligible upward bias when m is moderate because of its extra summand $(1-\lambda)^{-1} m^{-1}$. Additionally, we point out that setting $\widehat{FDR}^*_\lambda(t) = 1$ for $t > \lambda$ in the definition of $\widehat{FDR}^*_\lambda(t)$ is inappropriate in certain situations. For example, we can easily construct $m > 3$ p-values such that $R(1/2) > 2(m+1)/3$. For $1 \geq t > 1/2$ this implies

$$\frac{t\tilde{\pi}_0^{St}(1/2)}{m^{-1}\{R(t) \vee 1\}} = \frac{2t\{m - R(1/2) + 1\}}{R(t)} < 1.$$

Consequently setting $\widehat{FDR}^*_{1/2}(t) = 1$ in this case ignores the truth and may lead to erroneous conclusions.

In view of the above, we propose the generalized FDR estimator as

$$\widetilde{FDR}_\lambda(t) = \min\left[1, \frac{\hat{\pi}_0^{CD}(\lambda, \epsilon)\, t}{m^{-1}\{R(t) \vee 1\}}\right].$$

Further, for a function f with domain $[0,1]$, we define the threshold of f at level α as

$$t_\alpha(f) = \sup\{t \in [0,1] : f(t) \leq \alpha\},$$

and let

$$\mathcal{F}_t = \sigma\left(\mathbf{1}_{\{p_i \leq s\}}, t \leq s \leq 1, i = 1, ..., m\right).$$

4.1. Stopping time property of the threshold of $\widetilde{FDR}_\lambda(t)$

Recently, martingale methods have been applied to obtain the conservativeness of FDR procedures using the stopping time property of relevant thresholds with respect to certain filtrations. For example, the stopping time property of $t_\alpha\left(\widehat{FDR}^*_\lambda\right)$ is claimed in Lemma 4 of Storey et al. (2004) without a formal proof, so is that of $\alpha^*(q^*)$ in (7.2) of Pena et al. (2011). These claims have not been proved in the setting of independent, uniform null p-values, and it is not clear whether this property holds for the generalized FDR estimator and for general p-values. We show that $t_\alpha\left(\widetilde{FDR}_\lambda\right)$ is a stopping time with respect to the backward filtration

$$\mathcal{G} = \{\mathcal{F}_{t \wedge \lambda} : 1 \geq t \geq 0\}$$

for general p-values, discrete or continuous, independent or dependent. Further, we claim that $\widetilde{FDR}_\lambda$ is exhausted at level α at $t_\alpha\left(\widetilde{FDR}_\lambda\right)$ under one simple condition.

To achieve this, we start with $L(t) = t\{R(t) \vee 1\}^{-1}$, $t \in [0,1]$, the *scaled inverse rejection process*. Order the p-values into $p_{(1)} < p_{(2)} < ... < p_{(n)}$ distinctly, where the multiplicity of $p_{(i)}$ is n_i for $i = 1, ..., n$. Let $p_{(n+1)} = \max\left\{p_{(n)}, 1\right\}$ and $p_{(0)} = 0$. Define $T_j = \sum_{l=1}^{j} n_l$ for $j = 1, ..., n$.

Lemma 1. *The process $\{L(t), t \in [0,1]\}$ is such that*

$$L(t) = \begin{cases} t & \text{if} \quad t \in [0, p_{(1)}), \\ tT_j^{-1} & \text{if} \quad t \in [p_{(j)}, p_{(j+1)}) \text{ for } j = 1, .., n-1, \\ tm^{-1} & \text{if} \quad t \in [p_{(n)}, p_{(n+1)}]. \end{cases} \tag{6}$$

Moreover, it can only be discontinuous at $p_{(i)}$, $1 \leq i \leq n$, where it can only have a downward jump with size

$$L\left(p_{(i)}-\right) - L\left(p_{(i)}\right) = \frac{p_{(i)} n_i}{R\left(p_{(i)}\right)\left\{R\left(p_{(i)}\right) - n_i\right\}} > 0.$$

The conclusion of Lemma 1 is right the contrary to the claim in the proof of Theorem 2 in Storey et al. (2004) that "the process $mt/R(t)$ has only upward jumps and has a final value of 1", since it says "the process $mt/R(t)$ has only downward jumps". We construct a counterexample to their claim as follows. For a small increase c in t which results increase a_c in $R(t)$, we see that

$$L(t+c) - L(t) = \frac{t+c}{R(t)+a_c} - \frac{t}{R(t)} = \frac{cR(t) - ta_c}{\{R(t)+a_c\}R(t)} < 0$$

if and only if $\dfrac{c}{a_c} < \dfrac{t}{R(t)}$. Construct m p-values with $n \geq 4$ such that there exists some $1 \leq j_0 < n-2$ with $n_{j_0+1} > T_{j_0}$ but $p_{(j_0+1)} < 1$. Choose c_1 and c_2 such that $0 < c_1 < \left\{p_{(j_0+1)} - p_{(j_0)}\right\}/2$ and $0 < c_2 < p_{(j_0+1)} - 2c_1$. Let $t_0 = p_{(j_0+1)} - c_1$ and $c = c_1 + c_2$. Then $p_{(j_0)} < t_0 < p_{(j_0+1)}$, $R(t_0) = T_{j_0}$ and $R(t_0 + c) = T_{j_0} + n_{j_0+1}$. Further, $0 < c < t_0$ and $a_c = n_{j_0+1}$. So $\dfrac{c}{n_{j_0+1}} < \dfrac{t_0}{T_{j_0}}$ and $L(t_0+c) - L(t_0) < 0$. Letting $c \to 0$ gives $p_{(j_0+1)}$ as a point of downward jump for $L(t)$.

Equipped with the properties of $L(\cdot)$, we have

Theorem 2. *$t_\alpha\left(\widetilde{FDR}_\lambda\right)$ is a stopping time for $\mathcal{G}$. Further,*

$$\widetilde{FDR}_\lambda\left\{t_\alpha\left(\widetilde{FDR}_\lambda\right)\right\} = \alpha$$

whenever $\hat{\pi}_0^{CD}(\lambda, \epsilon) > \alpha$ and $R(t) > 0$.

The conclusion of Theorem 2 is illustrated by an application of $\hat{\pi}_0^{CD}$ and $\widetilde{FDR}_\lambda$ in Section 5. Define

$$\hat{\pi}_0(\lambda, \varsigma) = \max\left[0, \min\left\{1, \hat{\pi}_0^{St}(\lambda) - \frac{\varsigma}{(1-\lambda)m}\right\}\right]$$

and

$$\widetilde{FDR}_\lambda^\varsigma(t) = \min\left[1, \frac{\hat{\pi}_0(\lambda,\varsigma)\,t}{m^{-1}\{R(t) \vee 1\}}\right],$$

where $\varsigma \in \left[0, (1-\lambda)\, m\hat{\pi}_0^{St}(\lambda)\right]$ is deterministic and depends only on the p-values and λ. We call $\widetilde{FDR}_\lambda^\varsigma$ a "Storey-type FDR estimator". Using the same arguments in the proof of Theorem 2, we can show

THEOREM 3. *$t_\alpha\left(\widetilde{FDR}_\lambda^\varsigma\right)$ is a stopping time for $\mathcal{G}$. Further,*

$$\widetilde{FDR}_\lambda^\varsigma\left\{t_\alpha\left(\widetilde{FDR}_\lambda^\varsigma\right)\right\} = \alpha$$

whenever $\hat{\pi}_0(\lambda,\varsigma) > \alpha$ and $R(t) > 0$.

The proof of Theorem 3 is omitted. Theorem 3 implies, regardless of whether the p-values are independent or continuous, that the stopping time property of the thresholds of Storey-type FDR estimators is generic, and that the FDR estimator reaches the prespecified FDR level at its threshold whenever the estimated proportion of true nulls is larger than this FDR level and there is at least one rejection. It provides the first formal and most general support for the use of the stopping time property of the thresholds of a certain family of adaptive FDR estimators.

4.2. Large sample properties of $\widetilde{FDR}_\lambda$

Even though Theorem 3 reveals that $\widetilde{FDR}_\lambda^\varsigma$ reaches the prespecified FDR level at its threshold, it does not tell whether $\widetilde{FDR}_\lambda^\varsigma$ conservatively estimates the FDR of a one-step MTP whose rejection threshold is this threshold. Surprisingly, under appropriate conditions, a much stronger property of $\widetilde{FDR}_\lambda^\varsigma$ can be obtained. We illustrate this using $\widetilde{FDR}_\lambda$ and assume that

A1) $m_0^{-1}V(t)$ and $m_1^{-1}S(t)$ converge almost surely to càdlàg functions $G_0(t)$ and $G_1(t)$ such that $0 < G_0(t) \le t$ for all $t \in (0,1]$ and $\lim_{m\to\infty} m_0/m = \pi_{0,\infty} \in (0,1)$ exists.

When $G_i(\cdot)$, $i = 0, 1$ in A1) are continuous, Storey et al. (2004) proved that $\widehat{FDR}_\lambda$ is asymptotically conservative, even simultaneously:

$$\begin{cases} \lim_{m\to\infty} \inf_{t\geq\delta} \left\{ \widehat{FDR}_\lambda(t) - FDR(t) \right\} \geq 0, \\ \lim_{m\to\infty} \inf_{t\geq\delta} \left\{ \widehat{FDR}_\lambda(t) - \dfrac{V(t)}{R(t)\vee 1} \right\} \geq 0. \end{cases} \tag{7}$$

We show that these properties are endowed to $\widetilde{FDR}_\lambda$. Noticing that (7) hinges on the total conservativeness of $\hat{\pi}_0^{St}(\lambda)$ and that $\hat{\pi}_0^{St}(\lambda) \geq \pi_{0,\infty}$, P-a.s. automatically once A1) holds, we assume

A2) $\lim_{m\to\infty} \hat{\pi}_0^{CD}(\lambda, \epsilon) \geq \pi_{0,\infty}$, P-a.s..

Thus we have

Theorem 4. *Suppose A1) and A2) hold. Then for each $\delta \in (0, 1]$,*

$$\liminf_{m\to\infty} \inf_{1\geq t\geq\delta} \left\{ \widetilde{FDR}_\lambda(t) - \frac{V(t)}{R(t)\vee 1} \right\} \geq 0 \tag{8}$$

and

$$\liminf_{m\to\infty} \inf_{1\geq t\geq\delta} \left\{ \widetilde{FDR}_\lambda(t) - FDR(t) \right\} \geq 0 \tag{9}$$

hold P-a.s..

Theorem 4 ensures the simultaneous conservativeness of $\widetilde{FDR}_\lambda$ when m is sufficiently large. On the other hand, we have

$$\widetilde{FDR}_\lambda(t) - \widehat{FDR}_\lambda(t) = \frac{\delta(\lambda, \epsilon)\, t}{m^{-1}\{R(t)\vee 1\}} \leq 0 \tag{10}$$

and $\sup_{1\geq t\geq 0} \dfrac{\delta(\lambda, \epsilon)\, t}{m^{-1}\{R(t)\vee 1\}} \geq \delta(\lambda, \epsilon)$. Hence, $\widetilde{FDR}_\lambda$ may fail to conservatively estimate the FDR of a one-step MTP. However, since (10) implies $t_\alpha\left(\widetilde{FDR}_\lambda\right) \geq t_\alpha\left(\widehat{FDR}_\lambda\right)$, it follows that, at the same FDR level, a one-step MTP using the threshold of $\widetilde{FDR}_\lambda$ is usually more powerful than one that uses the threshold of $\widehat{FDR}_\lambda$.

When more asymptotic uniformity of the p-value distributions is available, such as

A3) $\lim_{m\to\infty} m^{-1} \sum_{i=1}^{m} \varepsilon_i = \bar{\varepsilon}$ and $\lim_{m\to\infty} m_1^{-1} \sum_{i\in I_1} F_i^*(\lambda) = h_1(\lambda)$, then with A1), we have

$$\lim_{m\to\infty} m^{-1} \sum_{i=1}^{m} \varepsilon_i \left\{\lambda - F_i^* (\lambda)\right\} = h^* (\lambda),$$

where $h^*(\lambda) = t\bar{\varepsilon} - \pi_{0,\infty} G_0(\lambda) - \pi_{1,\infty} h_1(\lambda)$ and $\pi_{1,\infty} = 1 - \pi_{0,\infty}$. Further, the limit of $\widetilde{FDR}_\lambda(t)$ as $m \to \infty$ is

$$\widetilde{FDR}_\lambda^\infty (t) = \frac{\left[\{1 - G_0(\lambda)\}\pi_{0,\infty} + \{1 - G_1(\lambda)\}\pi_{1,\infty}\right] t - h^*(\lambda)\, t}{(1-\lambda)\{\pi_0 G_0(t) + \pi_1 G_1(t)\}}.$$

We show that a one-step MTP using $t_\alpha\left(\widetilde{FDR}_\lambda\right)$ as the rejection threshold maintains FDR control at level α asymptotically.

THEOREM 5. *Suppose A1) to A3) hold. If* $\widetilde{FDR}_\lambda^\infty(t) < \alpha$ *for some* $t \in [0,1]$*, then*

$$\limsup_{m\to\infty} FDR\left\{t_\alpha\left(\widetilde{FDR}_\lambda\right)\right\} \le \alpha.$$

We remark that the conclusions in Theorem 4 and Theorem 5 hold for Storey-type FDR estimators with $\hat{\pi}_0^{CD}$ replaced by any totally conservative estimator $\hat{\pi}_0$ of π_0 when the involved proportions and empirical processes converge appropriately. When $\hat{\pi}_0^{CD}$ is not conservative or totally conservative, $\widetilde{FDR}_\lambda$ may not conservatively estimate the FDR of an MTP. Fortunately, we can quantify the possible downward bias of $\widetilde{FDR}_\lambda$. Set $T_{n+1} = T_n$. For $\delta \in (0,1]$, define

$$L_{\inf}^\infty = \liminf_{m\to\infty} \min_{1\le j\le n+1} \left\{\frac{p_{(j)}}{m^{-1}T_j} : p_{(j)} \ge \delta\right\},$$

and

$$C_{\delta,\lambda}(t) = \liminf_{m\to\infty} \inf_{1\ge t\ge\delta} \left\{\widehat{FDR}_\lambda(t) - FDR(t)\right\}.$$

THEOREM 6. *Suppose A1) and A3) hold. Then for any* $\delta \in (0,1]$,

$$C_{\delta,\lambda}(t) - \frac{h^*(\lambda)}{1-\lambda} \le \lim_{m\to\infty} \inf_{1\ge t\ge\delta} \left\{\widetilde{FDR}_\lambda(t) - FDR(t)\right\} \le C_{\delta,\lambda}(t) - \frac{h^*(\lambda)\, L_{\inf}^\infty}{1-\lambda}$$

with probability 1.

With some algebra, the proof of Theorem 6 follows from Lemma 1, (9) and (10), and is thus omitted.

5. Simulation studies for and application of $\hat{\pi}_0^{CD}$

Now we will show the superior performance of $\hat{\pi}_0^{CD}$ via simulation studies. Since $\hat{\pi}_0^{CD}$ reduces to $\hat{\pi}_0^{St}$ for continuous p-values whose null distribution is $U(0,1)$ and the performance of $\hat{\pi}_0^{St}$ has been well documented, we compare the competing estimators of π_0 based on discrete, two-sided p-values induced by FETs and ENTs. We do not assess $\hat{\pi}_0^{CD}$ for dependent discrete p-values, since the deterministic term $\delta(\lambda,\epsilon)$ as the difference between $\hat{\pi}_0^{St}$ and $\hat{\pi}_0^{CD}$ and the extensive study of $\hat{\pi}_0^{St}$ under dependence (Storey et al., 2004; Blanchard and Roquain, 2009) have already characterized the performance of $\hat{\pi}_0^{CD}$ in this situation. Finally, even though we have justified the simultaneous asymptotic conservativeness of $\widetilde{FDR}_\lambda$, we do not study $\widetilde{FDR}_\lambda$ using dependent discrete p-values due to the possible, practical violation of assumptions on asymptotic convergence of heterogeneous, discrete p-value distributions.

The simulation and estimation of the proportion of true nulls are set up as follows:

(a) Select an $m \in \mathbb{N}$. Pick π_0, π_{1s}, $\pi_{1l} \in \{0, 0.1, 0.2, ..., 0.9, 1\}$ compatibly, which gives the triples $\pi_a = (\pi_0, \pi_{1s}, \pi_{1l})$ with $a = 1, ..., 111$. Set $m_0 = [m\pi_0]$, $m_1 = m - m_0$, $m_{1s} = [m_1\pi_{1s}]$, $m_{1l} = m_1 - m_{1s}$; $m_{1s_1} = [0.5m_{1s}]$, $m_{1s_2} = m_{1s} - m_{1s_1}$, $m_{1l_1} = [0.5m_{1l}]$, $m_{1l_2} = m_{1l} - m_{1l_1}$, where $[x]$ is the integer part of $x \geq 0$.

(b) Let $H_{1s_1} = U(0.8, 0.98)$, $H_{1s_2} = U(1.02, 1.2)$, $H_{1l_1} = U(0.05, 0.5)$ and $H_{1l_2} = U(2, 20)$. Generate $\boldsymbol{\rho} = (\rho_1, ..., \rho_m)$ with subvectors in order as $\boldsymbol{\rho}_{m_0}$, $\boldsymbol{\rho}_{m_{1s_1}}$, $\boldsymbol{\rho}_{m_{1s_2}}$, $\boldsymbol{\rho}_{m_{1l_1}}$, $\boldsymbol{\rho}_{m_{1l_2}}$. Specifically, $\boldsymbol{\rho}_{m_0}$ is m_0 ones; $\boldsymbol{\rho}_{m_{1s_k}}$ contains m_{1s_k} realizations from H_{1s_k}, and $\boldsymbol{\rho}_{m_{1l_k}}$ are m_{1l_k} realizations from H_{1l_k} for $k = 1, 2$. The purpose of $\boldsymbol{\rho}_{m_{1s_k}}$, $k = 1, 2$ is to induce small effect signals.

(c) Define θ_{i1}. Set $\theta_{i2} = h(\theta_{i1}, \rho_i)$ for some measurable function $h : \Theta_1 \times \mathbb{R} \to \Theta_2$, where Θ_g, $g = 1, 2$ are non-empty subsets of $\mathbb{R}^d$ for some $d \in \mathbb{N}$. Pick non-empty $\Psi_g \subseteq \mathbb{R}^e$, $g = 1, 2$ for some $e \in \mathbb{N}$. Generate $n_{ig} \in \Psi_g$ from distribution G_{ig} and ξ_{ig} from parametric distribution $F_{ig}(\theta_{ig}|n_{ig})$.

(d) With ξ_{ig}, $g = 1, 2$, conduct the FET or ENT to test either

$$H_{i0} : \theta_{i2} = \theta_{i1} \text{ versus } H_{i1} : \theta_{i2} \neq \theta_{i1}$$

or

$$H_{i0} : \theta_{i2} = \theta_0 \text{ versus } H_{i1} : \theta_{i2} \neq \theta_0$$

for a given $\theta_0 \in \Theta_1$. Estimate $\pi_0 = \#\{1 \leq i \leq m : \theta_{i2} = \theta_{i1}\}/m$ or $\pi_0 = \#\{1 \leq i \leq m : \theta_{i2} = \theta_0\}/m$ based on the two sided p-values of these m tests.

(e) For each π_a, compute $\hat{\pi}_0^{St}(0.5)$, $\hat{\pi}_0^{CD}(0.5)$, $\tilde{\pi}_0^{PC}$ and $\hat{\pi}_0^{PC}$.
(f) For each π_a, repeat Steps (b) to (e) 1000 times to obtain the sample mean $\hat{E}(\cdot)$ and sample standard deviation $\hat{S}(\cdot)$ of each estimator $\hat{\pi}_0$.

The simulations have three unique features that are more challenging to all competing estimators. First, we create different configurations of π_0 and π_{1s}, π_{1l} corresponding to the proportions of true nulls, small effect signals and large effect signals. This will examine the power of the test represented by its p-value and how small effect signals affect the $\hat{\pi}_0$'s. Second, when π_a remains fixed, for each run $\boldsymbol{\rho}_{m_{1s_k}}$ and $\boldsymbol{\rho}_{m_{1l_k}}$ are re-generated from H_{1s_k} and H_{1l_k} for $k = 1, 2$, respectively. These small perturbations to the alternative hypotheses will check the robustness of the tests and estimators. Third, in the simulations we do not verify condition (4), meaning that (4) is either satisfied or not. However, the results from our simulation studies will show that $\hat{\pi}_0^{CD}(0.5)$ is robust, stable, and is the least conservatively biased for relatively large π_0 and wide range of values of m.

5.1. Estimation of π_0 from two-sided p-values of FETs

FET has been widely used to compare two independent binomial populations and test independence between quantities from various discrete distributions in 2-by-2 tables. The test statistic of an FET follows the Fisher's non-central hypergeometric (FNH) distribution. Suppose three marginal quantities are observed and stored in an ordered triple $N^* = (N_1, N_2, M_1)$, then with the non-centrality parameter $\theta \in (0, \infty)$, an FNH distribution, $FNH(\theta, N^*)$, has density

$$f_\theta(x; N^*) = \binom{N_1}{x}\binom{N_2}{M_1 - x}\theta^x \Big/ \sum_{u=L}^{L_*}\binom{N_1}{u}\binom{N_2}{M_1 - u}\theta^u \qquad (11)$$

for $L \le x \le L_*$, where $L = \max(0, M_1 - N_2)$ and $L_* = \min(N_1, M_1)$. Following Agresti (2002), given N^* the two-sided p-value of an FET for an observation $X = x_0$ is defined as

$$p_{\theta,N^*}(x_0) = \int_{\{L \le y \le L_*: f_\theta(y;N^*) \le f_\theta(x_0;N^*)\}} dF_\theta(y; N^*),$$

where $F_\theta(\cdot; N^*)$ is the c.d.f. induced by $f_\theta(\cdot; N^*)$.

Simulation I: Set $m = 10,000$ and generate n_{ig} from a Poisson distribution with mean μ_g, where $\mu_1 = 20$ and $\mu_2 = 30$. Set $\theta_{i1} = 0.5$ for $1 \le i \le m$ and $\theta_{i2} = (\rho_i + 1)^{-1}$. Generate ξ_{ig} from $F_{ig}(\theta_{ig}|n_{ig}) = BIN(\theta_{ig}, n_{ig})$, where

Table 1. Partial digital gene expression data for gene i. For $j = 1, 2$, $n_{i,j}$ is the discrete measurement for gene i under Treatment j, respectively, and $n_i = n_{i1} + n_{i2}$. For $j = 1, 2$, $n_{(-i),j}$ is the total of the discrete measurements for all other genes under Treatment j, respectively, and $n_{(-i)} = n_{(-i),1} + n_{(-i),2}$. For $j = 1, 2$, m_j^* is the total of the discrete measurements for all genes under Treatment j, respectively, and $m^* = m_1^* + m_2^*$.

	Treatment1	Treatment2	Total
Gene i	n_{i1}	n_{i2}	n_i
All other genes	$n_{(-i),1}$	$n_{(-i),2}$	$n_{(-i)}$
Total	m_1^*	m_2^*	m^*

$BIN\left(p^*, \tilde{N}\right)$ denotes a binomial distribution with success probability p^* in $\tilde{N}$ independent trials.

Simulation II: Extract counts for $m = 10,000$ genes from an NGS RNA-Seq data set for American ash trees under two treatments, for gene i form a 2-by-2 table as illustrated in Table 1 and extract $\tilde{n}_{ig} = \left(n_i, n_{(-i)}, m_1^*\right) \in \mathbb{N}^3$. Set $\theta_{i1} = 1$ for all i but $\theta_{i2} = \rho_i \theta_{i1}$. Let $F_{ig}\left(\theta_{ig}|\tilde{n}_{ig}\right) = FNH\left(\theta_{ig}, \tilde{n}_{ig}\right)$ and generate ξ_{ig} from $FNH\left(\theta_{ig}, \tilde{n}_{ig}\right)$. For each i test

$$H_{i0} : \theta_{i2} = 1 \text{ versus } H_{i1} : \theta_{i2} \neq 1$$

via the FET, where H_{i0} means that the treatments do not have any effect on gene i.

5.2. *Estimation of π_0 from two-sides p-values of ENTs*

The ENT has been widely used in analysis of NGS count data that are modeled by negative binomial (NB) distributions. Di et al. (2011) proposed to use an over-parameterized version of the NB distribution (called "NBP distribution") to model the non-constant dispersion of such counts. Let $NBP\left(\mu, \phi, \beta\right)$ denote an NBP distribution with mean μ and dispersions (ϕ, β). The probability mass function for $Y \sim NBP\left(\mu, \phi, \beta\right)$ is

$$f\left(Y = y; \left(\mu, \phi, \beta\right)\right) = \frac{\Gamma\left(\gamma + y\right)}{\Gamma\left(\gamma\right)\Gamma\left(1 + y\right)} \left(1 - p\right)^{\gamma} p^y$$

where $p = \mu / \left(\mu + \gamma\right)$ and $\gamma = \phi^{-1}\mu^{2-\beta}$ for $y = 0, 1, ...$

Simulation III: Set $l = 1,294,326$ as the active library size, apply the R package *NBPSeq* to the RNA-Seq count data for *Arabidopsis thaliana* provided by Di et al. (2011), and extract $m = 6,000$ estimated frequencies $\tilde{p}_{i1}$, $i =$

$1, ..., 6,000$ for genes in the first observational unit whose estimated means are not zero. Set $\tilde{p}_{i2} = \rho_i \tilde{p}_{i1}$ and $\mu_{gi} = l\tilde{p}_{ig}$ but $\gamma_{ig} = \phi^{-1}\mu_{1i}^{2-\beta}$ with $\beta = 1.5$. Define $\tilde{\theta}_{ig} = \dfrac{\mu_{gi}}{\mu_{gi} + \gamma_{gi}}$ and $\theta_{ig} = (\mu_{ig}, \gamma_{ig})$. Set $F_{ig}(\theta_{ig}) = NBP(\mu_{ig}, \gamma_{ig})$ and generate counts ξ_{ig} from $F_{ig}(\theta_{ig})$, where the alternative parametrization for an NBP distribution has been used. The ENT is then conducted to test

$$H_{i0} : \theta_{i1} = \theta_{i2} \text{ versus } H_{i1} : \theta_{i1} \neq \theta_{i2},$$

where H_{i0} means that gene i is not differentially expressed. For two realizations $\xi_{il} = x_{il}$, $l = 1, 2$ the two-sided p-value of this test is defined as

$$p_i = \frac{\sum_{\{c,d:c+d=x_{i1}+x_{i2}, f_{1,2}(c,d) \leq f_{1,2}(x_{i1},x_{i2})\}} f_{1,2}(c,d)}{\sum_{\{c,d:c+d=x_{i1}+x_{i2}\}} f_{1,2}(c,d)},$$

where $f_{1,2}(\cdot,\cdot)$ is the joint probability mass function of (ξ_{i1}, ξ_{i2}).

5.3. Summary from simulation studies

Due to the practical range of π_0 in various studies, we report in Table 2 the performances of the competing estimators of π_0 for $\pi_0 = 0.9$ with $\pi_{1s} \in \{0, 0.1, ..., 0.9, 1\}$ and for $\pi_0 = 1$. Note that $\hat{\pi}_0^{St}(0.5)$, $\tilde{\pi}_0^{PC}$ are excluded since they are both 1 for all runs (which means $\hat{S}\left(\hat{\pi}_0^{St}(0.5)\right) = \hat{S}\left(\tilde{\pi}_0^{PC}\right) = 0$) for these π_0, π_{1s} and for all underlying distributions used in our simulations to generate the discrete counts. The same is true for $\hat{\pi}_0^{PC}$ when the underlying distribution is NBP. From Table 2, it is clear that $\hat{\pi}_0^{CD}(0.5)$ is stable and the least conservative among all competing estimators for the reported values of π_0, π_{1s} except $(\pi_0, \pi_{1s}) = (0.9, 0)$ and $\pi_1 = 1$.

Our additional (but unreported) simulation studies together with those provided in Table 2 reveal the following:

(a) Overall $\hat{\pi}_0^{CD}(0.5)$ outperforms its competitors in terms of stability and accuracy, across almost all configurations π_a, $a = 1, ..., 111$. When π_0 is small, it outperforms its competitors by a larger margin but can underestimate π_0.
(b) For π_0 fixed, all estimators become more conservative as π_{1s} increases. When π_{1s} dominates π_0, all estimators tend to estimate π_{1s} and are very inaccurate. As π_0 gets closer to 1, all competing estimators perform better.
(c) Empirically $\hat{\pi}_0^{CD}(0.5)$ gives the least conservative estimate of π_0 roughly for $\pi_0 \in [0.75, 1)$ even when there are small effect signals. $\hat{\pi}_0^{CD}(\lambda, \varepsilon)$ is slightly less stable than $\hat{\pi}_0^{PC}$ in terms of sample standard deviations.

Table 2. Partial results from all simulations

Distribution	π_0	π_{1s}	$\widehat{E}(\widehat{\pi}_0^{CD})$	$\widehat{E}(\widehat{\pi}_0^{PC})$	$\widehat{S}(\widehat{\pi}_0^{CD})$	$\widehat{S}(\widehat{\pi}_0^{PC})$
	0.9	0	**0.8977**	0.9215	0.0091	0.0051
BIN	0.9	0.1	**0.9086**	0.9296	0.0094	0.0052
	0.9	0.2	**0.9185**	0.9374	0.0092	0.0052
	0.9	0.3	**0.9290**	0.9455	0.0094	0.0054
	0.9	0.4	**0.9395**	0.9533	0.0092	0.0052
	0.9	0.5	**0.9497**	0.9612	0.0094	0.0052
	0.9	0.6	**0.9602**	0.9692	0.0095	0.0053
	0.9	0.7	**0.9704**	0.9770	0.0094	0.0053
	0.9	0.8	**0.9807**	0.9849	0.0097	0.0055
	0.9	0.9	**0.9901**	0.9928	0.0083	0.0049
	0.9	1	**0.9971**	0.9984	0.0047	0.0026
	1	0	**0.9978**	0.9988	0.0041	0.0023
	0.9	0	**0.9076**	0.9251	0.0091	0.0051
FNH	0.9	0.1	**0.9160**	0.9311	0.0094	0.0052
	0.9	0.2	**0.9238**	0.9368	0.0094	0.0052
	0.9	0.3	**0.9323**	0.9432	0.0092	0.0050
	0.9	0.4	**0.9406**	0.9495	0.0095	0.0053
	0.9	0.5	**0.9490**	0.9560	0.0097	0.0054
	0.9	0.6	**0.9586**	0.9654	0.0097	0.0053
	0.9	0.7	**0.9684**	0.9746	0.0096	0.0054
	0.9	0.8	**0.9771**	0.9821	0.0100	0.0055
	0.9	0.9	**0.9861**	0.9893	0.0091	0.0052
	0.9	1	**0.9930**	0.9955	0.0071	0.0042
	1	0	**0.9960**	0.9979	0.0057	0.0031
	0.9	0	**0.9487**	1	0.0115	0
NBP	0.9	0.1	**0.9544**	1	0.0118	0
	0.9	0.2	**0.9588**	1	0.0116	0
	0.9	0.3	**0.9638**	1	0.0112	0
	0.9	0.4	**0.9692**	1	0.0120	0
	0.9	0.5	**0.9734**	1	0.0117	0
	0.9	0.6	**0.9790**	1	0.0113	0
	0.9	0.7	**0.9839**	1	0.0106	0
	0.9	0.8	**0.9884**	1	0.0099	0
	0.9	0.9	**0.9919**	1	0.0084	0
	0.9	1	**0.9947**	1	0.0072	0
	1	0	**0.9952**	1	0.0069	0

Table 3. Results from application of the generalized estimators to *Arabidopsis thaliana* data. *CD* means using $\hat{\pi}_0^{CD}(\lambda,\varepsilon)$ and $\widetilde{FDR}_\lambda$. *St* means using $\hat{\pi}_0^{St}(\lambda) = \hat{\pi}_0^{CD}(\lambda,0)$ and $\widehat{FDR}_\lambda$. *BH* is the BH procdure, and *adaBH* the adaptive BH procedure for which α is replaced by $\alpha' = \alpha/\hat{\pi}_0^{CD}(0.5,0.9)$. $\hat{t}$ is the threshold of the procedure, and $FDR^\dagger(\hat{t})$ the FDR procedure evaluated at $\hat{t}$

Method	(λ,ε)	$\widehat{\pi}_0$	$\widehat{t}$	$FDR^\dagger(\widehat{t})$	$R(\widehat{t})$
CD	(.5,.9)	0.958801	0.00093514	0.049999348	433
St	(.5,0)	1	0.00089042	0.05	430
BH	NA	1	0.00089042	$\leq$0.05	430
adaBH	NA	0.958801	0.00093516	$\leq$0.05	433

These suggest practical preference to our estimator of π_0 when discrete p-values are induced by FETs and ENTs for three types of underlying distributions: binomial, FNH and NBP.

5.4. Application to Arabidopsis thaliana data

Di et al. (2011) fit the *Arabidopsis thaliana* data with the NBP model, estimate the parameters (μ,ϕ,β), conduct for each gene an ENT, and report 430 differentially expressed genes by applying the BH procedure to the $26,222$ two-sided p-values of these tests at FDR level $\alpha = 0.05$. We carefully looked into their source codes and found out that $2,076$ of the p-values they obtained are Not a Number (NaN) because the corresponding sizes are estimated as zero. We filtered these NaN p-values, conducted BH procedure at the same FDR level on the rest, and confirmed their finding.

To apply the generalized estimators with FDR level $\alpha = 0.05$, we use the *NBPSeq* package (Di et al., 2011) to obtain the pseudo-counts, pseudo library size, and the estimated probabilities and sizes. We then select genes (a total of $24,146$) with positive sizes and relevant quantities, conduct for each gene the ENT, and compute the two-sided p-value of each ENT. Our findings are summarized in Table 3.

Three more genes are found to be differentially expressed using $\widetilde{FDR}_\lambda$ with $\hat{\pi}_0^{CD}(0.5,0.9)$. The threshold $t_{0.05}\left(\widetilde{FDR}_\lambda\right) = 0.0093514$ is slightly greater than those of $\widehat{FDR}_\lambda$ and the BH procedure, leading to more rejections; it is practically the same as that of the adaptive BH procedure since this procedure uses $\hat{\pi}_0^{CD}$. That $\widetilde{FDR}_\lambda\left(t_{0.05}\left(\widetilde{FDR}_\lambda\right)\right) = 0.0499999348$ is the consequence of the step size used in numeric search and it does not affect the validity of Theorem 2. This confirms the improvement $\hat{\pi}_0^{CD}$ and $\widetilde{FDR}_\lambda$ result in.

6. Discussion

We have proposed generalized estimators of proportion of true nulls and of the FDR of one-step MTPs. The new estimators outperform existing ones for discrete p-values whose null distributions dominate the uniform and reduce to Storey's estimators for continuous p-values. We have provided the first formal and most general justification of the stopping time property of the threshold of Storey-type FDR estimators for p-values not necessarily continuous or independent, and establised the asymptotic conservativeness of the generalized FDR estimator. Our strategy to adjust estimators of π_0 can be easily adapted to other types of discrete p-values.

However, we see that the two-sided p-values of the FET and ENT are not derived from the corresponding uniformly most powerful tests. This warns us the possible loss in power when decisions are based on such p-values, the decreased accuracy when estimating π_0 using these p-values, and advocates to estimate π_0 from the test statistics directly. Unfortunately, the non-location shiftness of the density functions of the involved test statistics and heterogeneity of discrete p-value distributions disable adaptation of methods such as those in Jin (2008) and Meinshausen and Rice (2006).

Even though Table 2 seems to show the poor performance of $\hat{\pi}_0^{CD}$ merely in terms of its upward bias, improvements on $\hat{\pi}_0^{CD}$ are possible when λ is adaptively chosen from the data instead of always being 0.5. We leave the investigation of adaptive choice of λ, ε to future research but point out that techniques in Liang and Nettleton (2012) are not applicable to heterogenous null p-value distributions. In Table 3, we only have 3 more discoveries by incorporating $\hat{\pi}_0^{CD}(0.5, 0.9)$. The reason for such a minor improvement is as follows. In addition to using more conservative tuning parameters and the possibility that π_0 is itself very close to 1, potential improvements from $\hat{\pi}_0^{CD}$ are counterbalanced by the enlarged supports of null discrete p-value distributions because most of the counts in the data are not so small. The advantage of $\hat{\pi}_0^{CD}$ will be well manifested when the discrete data are of small magnitudes and π_0 is not very close to 1.

Driven by the accumulation of massive discrete data together with accompanying statistical tests to be conducted, the need for better estimators of the proportion of true nulls and of the FDR of an MTP for this type of data is urgent. We hope that this paper could further caution the usage of methods beyond its justified scope of application, and we expect that the methods presented will justify their utility in multiple testing with discrete data and stimulate more research efforts in this area.

Acknowledgements

The authors would like to thank Darla G. French from Purdue University for providing the NGS data for American ash trees used in simulation II. Xiongzhi Chen would like to thank Douglas Baumann and Sanvesh Srivastava for helpful discussion on properties of NGS data and R codes. This research was funded in part by a National Science Foundation Plant Genome Research Program grant (no. IOS-1025976) to RWD.

A. Proofs

A.1. Proof of Lemma 1

Clearly

$$R(t) = \begin{cases} 0 & \text{if} \quad 0 \le t < p_{(1)}, \\ T_j & \text{if} \quad p_{(j)} \le t < p_{(j+1)}, j = 1, .., n-1, \\ m & \text{if} \quad p_{(n)} \le t \le p_{(n+1)}, \end{cases}$$

and (6) holds. Obviously the points of discontinuities of $L(\cdot)$ are the original distinct p-values. This justifies the first part of the assertion.

Now we show that $L(\cdot)$ can only have downward jumps at points of discontinuity. Define

$$\varphi(t, \eta) = \frac{t + \eta}{R(t + \eta)} - \frac{t}{R(t)} = \frac{\eta R(t) + t\{R(t) - R(t + \eta)\}}{R(t + \eta) R(t)}.$$

From the fact

$$R\left(p_{(j)}\right) - R\left(p_{(j)}-\right) = n_j > 0$$

but $R\left(p_{(j)}+\right) - R\left(p_{(j)}\right) = 0$ for each $1 \le j \le n$, it follows that $\varphi\left(p_{(j)}, 0+\right) = \lim_{\eta \downarrow 0} \varphi(t, \eta) = 0$ but

$$\varphi\left(p_{(j)}, 0-\right) = \lim_{\eta \uparrow 0} \varphi\left(p_{(j)}, \eta\right) = \frac{p_{(j)} n_j}{R\left(p_{(j)}\right)\left\{R\left(p_{(j)}\right) - n_j\right\}} > 0.$$

Thus $L\left(p_{(j)}-\right) - L\left(p_{(j)}\right) = \varphi\left(p_{(j)}, 0-\right) > 0$ and the proof is completed.

A.2. Proof of Theorem 2

Define

$$\mathbf{X}_t^{(m)}(\omega) = \left(1_{\{p_1 \le t\}}(\omega), ..., 1_{\{p_m \le t\}}(\omega)\right), \ \omega \in \Omega.$$

Then $\mathcal{F}_t = \sigma\left(\mathbf{X}_s^{(m)}(\omega), 1 \ge s \ge t\right)$, $t \in [0, 1]$ and $\{\mathcal{F}_t : 0 \le t \le 1\}$ is a decreasing sequence of sub-sigma-algebras of $\mathcal{A}$. Write $\widetilde{FDR}_\lambda(t) = \widetilde{FDR}_\lambda(t, \omega)$ and

$t_\alpha\left(\widetilde{FDR}_\lambda\right) = \tilde{t}_\alpha$. When $\tilde{t}_\alpha = 1$ or 0 corresponding to $\alpha = 1$ or $\alpha = 0$, it is automatically a stopping time. So we only consider $\alpha \in (0,1)$ and $0 < \tilde{t}_\alpha < 1$. By definition,

$$\left\{\omega \in \Omega : \tilde{t}_\alpha \leq s\right\} = \bigcap_{\{t:s<t\leq 1\}} A_t = \tilde{A}_s,$$

where $A_t = \left\{\omega \in \Omega : \widetilde{FDR}_\lambda(t,\omega) > \alpha\right\}$, $1 \geq t > s$. We need to show $\tilde{A}_s \in \mathcal{F}_{s\wedge\lambda}$.

Since either $\mathcal{F}_{s\wedge\lambda} = \mathcal{F}_s \supseteq \mathcal{F}_\lambda$ when $s \leq \lambda$ or $\mathcal{F}_{s\wedge\lambda} = \mathcal{F}_\lambda \supseteq \mathcal{F}_s$ when $s \geq \lambda$, the stopping time property holds once we prove $\tilde{A}_s \in \mathcal{F}_s$. With the decomposition

$$A_t = \bigcup\nolimits_{r\in\mathbb{Q}} (A_{t,r} \cap B_r)$$

and

$$\tilde{A}_s = \bigcap_{\{t:s<t\leq 1\}} \bigcup_{r\in\mathbb{Q}} (A_{t,r} \cap B_r) = \bigcup_{r\in\mathbb{Q}} \bigcap_{\{t:s<t\leq 1\}} (A_{t,r} \cap B_r),$$

where $A_{t,r} = \left\{\omega \in \Omega : \dfrac{t}{m^{-1}\{R(t) \vee 1\}} \geq r\right\}$ and $B_r = \left\{\omega \in \Omega : \dfrac{\alpha}{\hat{\pi}_0^{CD}(\lambda,\epsilon)} < r\right\}$, it suffices to show

$$\bigcap_{\{t:s<t\leq 1\}} (A_{t,r} \cap B_r) = \bigcap_{\{t:s<t\leq 1\}} A_{t,r} \cap B_r \in \mathcal{F}_s,$$

or equivalently

$$A^*_{s,r} = \bigcap_{\{t:s<t\leq 1\}} A_{t,r} \in \mathcal{F}_s,$$

since $B_r \in \mathcal{F}_\lambda$ holds already.

Define $I_i = [p_{(i)}, p_{(i+1)})$, $i = 1, ..., n-1$. We will add $I_0 = [p_{(0)}, p_{(1)})$ and $I_n = [p_{(n)}, p_{(n+1)}]$ when $p_{(n)} < 1$. When $p_{(n)} = 1$, we take I_{n-1} to be $\left[p_{(n-1)}, p_{(n)}\right]$. Obviously there must be a unique j^* with $0 \leq j^* \leq n$ such that $s \in I_{j^*}$. Given $R(1) = m$ and $\hat{\pi}_0^{CD}(\lambda,\epsilon) \in [0,1]$, the properties of $L(\cdot)$ in Lemma 1 imply

$$A^*_{s,r} = A_{s,r} \bigcap \left(\bigcap\nolimits_{j=j^*+1}^{n+1} A_{p_{(j)},r}\right).$$

Consequently $A^*_{s,r} \in \mathcal{F}_s$ and $\tilde{A}_s \in \mathcal{F}_s$, which validates the first part of the assertion.

To show $\widetilde{FDR}_\lambda\left(\tilde{t}_\alpha\right) = \alpha$, it suffices to consider the case $\alpha > 0$. Since $\alpha < 1$ and $R(t) > 0$ by assumption, we can write $\widetilde{FDR}_\lambda(t)$ as

$$\widetilde{FDR}_\lambda(t) = \frac{\hat{\pi}_0^{CD}(\lambda,\epsilon)}{\rho^*_m(t)},$$

where $\hat{F}_m(t) = m^{-1}\sum_{i=1}^{m} \mathbf{1}_{\{p_i \le t\}}$ for $t \in [0,1]$ and

$$\rho_m^*(t) = t^{-1}\left\{\hat{F}_m(t) - \hat{F}_m(0)\right\}.$$

Therefore

$$\tilde{t}_\alpha = \sup\left\{t \in [0,1] : \rho_m^*(t) \ge \alpha^{-1}\hat{\pi}_0^{CD}(\lambda,\epsilon)\right\},$$

i.e., $\tilde{t}_\alpha$ is the last time the "slope" $\rho_m^*(t)$ is no less than $\alpha^{-1}\hat{\pi}_0^{CD}(\lambda,\epsilon)$.

Obviously, $\rho_m^*\left(\tilde{t}_\alpha\right) \ge \alpha$ and there must be a unique $0 \le j' \le n$ such that $\tilde{t}_\alpha \in I_{j'}$. Since $\hat{\pi}_0^{CD}(\lambda,\epsilon) > \alpha$, $\tilde{t}_\alpha < 1$ holds. Thus there must be some $d' > 0$ such that $I^* = [\tilde{t}_\alpha, \tilde{t}_\alpha + d') \subseteq I_{j'}$. Noting that $\rho_m^*(t) = m^{-1}L^{-1}(t)$ is continuous and decreasing on $I_{j'}$, we see that if $\rho_m^*\left(\tilde{t}_\alpha\right) > \alpha$, then we can choose $\hat{t}_\alpha \in I^*$ such that $\hat{t}_\alpha > \tilde{t}_\alpha$ and $\rho_m^*\left(\hat{t}_\alpha\right) > \alpha$. This contradicts the definition of $\tilde{t}_\alpha$. Hence $\rho_m^*\left(\tilde{t}_\alpha\right) = \alpha$ must hold, and this is equivalent to $\widetilde{FDR}_\lambda\left(\tilde{t}_\alpha\right) = \alpha$, which completes the proof.

A.3. Proof of Theorem 4

Define $K_\alpha = \left\{t \in [0,1] : \widetilde{FDR}_\lambda(t) < 1\right\}$ and $K_{\alpha,\delta} = K_\alpha \cap [\delta, 1]$. Let $FDP(t) = \dfrac{V(t)}{R(t) \vee 1}$. We verify the claims for $t \in K_{\alpha,\delta}$ since they automatically hold when the supremum is taken over $t \in [0,1] \backslash K_\alpha$. By the Glivenko-Cantelli theorem,

$$\lim_{m\to\infty} \sup_{0 \le t \le 1} \left|m^{-1}V(t) - \pi_{0,\infty}G_0(t)\right| = 0 \tag{12}$$

and

$$\lim_{m\to\infty} \sup_{0 \le t \le 1} \left|m^{-1}\{R(t) \vee 1\} - \{\pi_{0,\infty}G_0(t) + \pi_{1,\infty}G_1(t)\}\right| = 0 \tag{13}$$

hold P-a.s.. Since $\pi_{0,\infty}G_0(\delta) + \pi_{1,\infty}G_1(\delta) > 0$ and these are both non-decreasing functions, using (13) it can be easily shown that

$$\lim_{m\infty} \sup_{1 \ge t \ge \delta} \left|\frac{m}{R(t) \vee 1} - \frac{1}{\pi_{0,\infty}G_0(t) + \pi_{1,\infty}G_1(t)}\right| = 0,\ P\text{-a.s..} \tag{14}$$

Observing

$$\begin{aligned} \widetilde{FDR}_\lambda(t) - FDP(t) &= \frac{\hat{\pi}_0^{CD}(\lambda,\epsilon)\,t - \pi_{0,\infty}G_0(t)}{\{R(t) \vee 1\}/m} - \frac{m^{-1}V(t) - \pi_{0,\infty}G_0(t)}{\{R(t) \vee 1\}/m} \\ &= I_1(t) - I_2(t), \end{aligned}$$

we see, from (12), (13) and properties of $G_i(\cdot)$, $i = 0, 1$, that

$$\lim_{m\to\infty} \sup_{t\in K_{\alpha,\delta}} |I_2(t)| \leq \lim_{m\to\infty} \sup_{t\in K_{\alpha,\delta}} \frac{m}{R(t)\vee 1} \lim_{m\to\infty} \sup_{t\in K_{\alpha,\delta}} \left| \frac{V(t)}{m} - \pi_{0,\infty} G_0(t) \right| = 0.$$

Since $\lim_{m\to\infty} \hat{\pi}_0^{CD}(\lambda, \epsilon) \geq \pi_0$, P-a.s. and $G_0(t) \leq t$ for all $t \in (0, 1]$, then

$$\liminf_{m\to\infty} \inf_{1\geq t\geq\delta} \left\{ \hat{\pi}_0^{CD}(\lambda, \epsilon)\, t - \pi_0 G_0(t) \right\} \geq 0$$

and $\liminf_{m\to\infty} \inf_{t\in K_{\alpha,\delta}} I_1(t) \geq 0$. This justifies (8).

To show (9), in view of (8) and

$$\widetilde{FDR}_\lambda(t) - FDR(t) = \left\{ \widetilde{FDR}_\lambda(t) - FDP(t) \right\} + \left\{ FDP(t) - FDR(t) \right\},$$

it suffices to show

$$\lim_{m\to\infty} \sup_{1\geq t\geq\delta} |FDP(t) - FDR(t)| = 0, \ P\text{-a.s.}. \tag{15}$$

With (14), we have, P-a.s.,

$$\begin{aligned} & \lim_{m\to\infty} \sup_{1\geq t\geq\delta} \left| FDP(t) - \frac{\pi_{0,\infty} G_0(t)}{\pi_{0,\infty} G_0(t) + \pi_{1,\infty} G_1(t)} \right| \qquad (16) \\ \leq \ & \lim_{m\to\infty} \sup_{1\geq t\geq\delta} \left| \frac{V(t)/m - \pi_{0,\infty} G_0(t)}{\{R(t)\vee 1\}/m} \right| \\ & + \lim_{m\to\infty} \sup_{1\geq t\geq\delta} \left| \left[\frac{1}{\{R(t)\vee 1\}/m} - \frac{1}{\pi_{0,\infty} G_0(t) + \pi_{1,\infty} G_1(t)} \right] \pi_{0,\infty} G_0(t) \right| \\ = \ & 0. \end{aligned}$$

By (16), the dominated convergence theorem and Fatou's lemma, it follows that

$$\begin{aligned} 0 &= \lim_{m\to\infty} \sup_{1\geq t\geq\delta} E\left[\left| FDP(t) - \frac{\pi_{0,\infty} G_0(t)}{\pi_{0,\infty} G_0(t) + \pi_{1,\infty} G_1(t)} \right| \right] \\ &= E\left[\lim_{m\to\infty} \sup_{1\geq t\geq\delta} \left| FDP(t) - \frac{\pi_{0,\infty} G_0(t)}{\pi_{0,\infty} G_0(t) + \pi_{1,\infty} G_1(t)} \right| \right] \\ &\geq \lim_{m\to\infty} \sup_{1\geq t\geq\delta} \left| FDR(t) - \frac{\pi_{0,\infty} G_0(t)}{\pi_{0,\infty} G_0(t) + \pi_{1,\infty} G_1(t)} \right|. \end{aligned}$$

Thus (15) holds and so does (9), which completes the proof.

A.4. Proof of Theorem 5

Let $t' > 0$ be such that $\alpha - \widetilde{FDR}_\lambda^\infty(t') = \varepsilon'$ for some $\varepsilon' > 0$. We can take m sufficiently large such that $\left|\widetilde{FDR}_\lambda^\infty(t') - \widetilde{FDR}_\lambda(t')\right| < \varepsilon'/2$, which implies $\widetilde{FDR}_\lambda(t') < \alpha$ and $t_\alpha\left(\widetilde{FDR}_\lambda\right) \geq t'$. Therefore $\liminf_{m\to\infty} t_\alpha\left(\widetilde{FDR}_\lambda\right) \geq t'$, P-a.s.. By (8), it follows that, P-a.s.,

$$\begin{aligned} & \liminf_{m\to\infty}\left[\widetilde{FDR}_\lambda\left\{t_\alpha\left(\widetilde{FDR}_\lambda\right)\right\} - FDP\left\{t_\alpha\left(\widetilde{FDR}_\lambda\right)\right\}\right] \\ \geq\ & \liminf_{m\to\infty}\inf_{1\geq t\geq t'/2}\left\{\widetilde{FDR}_\lambda(t) - FDP(t)\right\} \geq 0. \end{aligned}$$

Since $\widetilde{FDR}_\lambda\left\{t_\alpha\left(\widetilde{FDR}_\lambda\right)\right\} \leq \alpha$, we have $\limsup_{m\to\infty} FDP\left\{t_\alpha\left(\widetilde{FDR}_\lambda\right)\right\} \leq \alpha$, P-a.s.. By Fatou's lemma,

$$\limsup_{m\to\infty} E\left[FDP\left\{t_\alpha\left(\widetilde{FDR}_\lambda\right)\right\}\right] \leq E\left[\limsup_{m\to\infty} FDP\left\{t_\alpha\left(\widetilde{FDR}_\lambda\right)\right\}\right] \leq \alpha.$$

This completes the proof.

References

Agresti, A. (2002). *Categorical Data Analysis.* John Wiley & Sons, Inc., New Jersey.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B 57*(1), 289–300.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika 93*(3), 491–507.

Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in mutliple testing under dependency. *Ann. Statist. 29*(4), 1165–1188.

Blanchard, G. and E. Roquain (2009). Adaptive false discovery rate control under independence and dependence. *Journal of Machine Learning Research 10*, 2837–2871.

Di, Y., D. W. Schafer, J. S. Cumbie, and J. H. Chang (2011). The nbp negative binomial model for assessing differential gene expression from rna-seq. *Statistical Applications in Genetics and Molecular Biology 10*(1), 24 pages.

Jin, J. (2008). Proportion of non-zero normal means: universal oracle equivalences and uniformly consistent estimators. *J. R. Statist. Soc. B 70*(3), 461–493.

Lehmann, E. and J. P. Romano (2005). *Testing Statistical Hypotheses*. Springer, New York.

Liang, K. and D. Nettleton (2012). Adaptive and dynamic adaptive procedures for false discovery rate control and estimation. *J. R. Statist. Soc. B 74*(1), 163–182.

Meinshausen, N. and J. Rice (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist. 34*(1), 373–393.

Pena, E. A., J. D. Habiger, and W. Wu (2011). Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. *Ann. Statist. 39*(1), 556–583.

Pounds, S. and C. Cheng (2006). Robust estimation of the false discovery rate. *Bioinformatics 22*(16), 1979–1987.

Robinson, M. D. and G. K. Smyth (2008). Small-sample estimation of negative binomial dispersion, with applications to sage data. *Biostatistics 9*(2), 321–332.

Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Statist. Soc. B 64*(3), 479–498.

Storey, J. D., J. E. Taylor, and D. Siegmund (2004). Strong control, conservative point estimation in simultaneous conservative consistency of false discover rates: a unified approach. *J. R. Statist. Soc. B 66*(1), 187–205.