J. R. Statist. Soc. B (2004)
66, *Part* 1, *pp.* 187–205

Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach

John D. Storey,
University of Washington, Seattle, USA

and Jonathan E. Taylor and David Siegmund
Stanford University, Stanford, USA

[Received June 2002. Final revision June 2003]

Summary. The false discovery rate (FDR) is a multiple hypothesis testing quantity that describes the expected proportion of false positive results among all rejected null hypotheses. Benjamini and Hochberg introduced this quantity and proved that a particular step-up p-value method controls the FDR. Storey introduced a point estimate of the FDR for fixed significance regions. The former approach conservatively controls the FDR at a fixed predetermined level, and the latter provides a conservatively biased estimate of the FDR for a fixed predetermined significance region. In this work, we show in both finite sample and asymptotic settings that the goals of the two approaches are essentially equivalent. In particular, the FDR point estimates can be used to define valid FDR controlling procedures. In the asymptotic setting, we also show that the point estimates can be used to estimate the FDR conservatively over all significance regions simultaneously, which is equivalent to controlling the FDR at all levels simultaneously. The main tool that we use is to translate existing FDR methods into procedures involving empirical processes. This simplifies finite sample proofs, provides a framework for asymptotic results and proves that these procedures are valid even under certain forms of dependence.

Keywords: Multiple comparisons; Positive false discovery rate; P-values; Q-values; Simultaneous inference

1. Introduction

Classically, the goal of multiple hypothesis testing has been to guard against making one or more type I errors among a family of hypothesis tests. In an innovative paper, Benjamini and Hochberg (1995) introduced a new multiple-hypothesis testing error measure with a different goal in mind—to control the proportion of type I errors among all rejected null hypotheses. This is useful in exploratory analyses, where we are more concerned with having mostly true findings among several, rather than guarding against one or more false positive results.

Table 1 describes the various outcomes when applying some significance rule to perform m hypothesis tests. In this work, we assume that the m hypothesis tests have corresponding p-values $p_1, \ldots, p_m$. The significance rule that we consider rejects null hypotheses with corresponding p-values that are less than or equal to some threshold, which can be fixed or data dependent. We are particularly interested in V, the number of type I errors (false positive results), and

Address for correspondence: John D. Storey, Department of Biostatistics, University of Washington, Box 357232, Seattle, WA 98195-7232, USA.
E-mail: jstorey@u.washington.edu

 1369–7412/04/66187

Table 1. Possible outcomes from m hypothesis tests based on a significance rule

	Accept null hypothesis	*Reject null hypothesis*	*Total*
Null true	U	V	m_0
Alternative true	T	S	m_1
	W	R	m

R, the total number of rejected null hypotheses. The *familywise error rate* (FWER) is defined to be

$$\text{FWER} = \Pr(V \geqslant 1),$$

and the *false discovery rate* (FDR) is defined to be (Benjamini and Hochberg, 1995)

$$\text{FDR} = E\left[\frac{V}{R \vee 1}\right] = E\left[\frac{V}{R}\middle| R > 0\right]\Pr(R > 0), \tag{1}$$

where $R \vee 1 = \max(R, 1)$. The effect of '$R \vee 1$' in the denominator of the first expectation is to set $V/R = 0$ when $R = 0$. As explained in Benjamini and Hochberg (1995), the FDR offers a less strict multiple-testing criterion than the FWER, and the FDR may be more appropriate for some applications.

Two approaches to providing conservative FDR procedures are the following. One is to fix the acceptable FDR level beforehand, and to find a data-dependent thresholding rule so that the FDR of this rule is less than or equal to the prechosen level. This is the approach that was taken by Benjamini and Hochberg (1995), for example. Another is to fix the thresholding rule, and to form an estimate of the FDR whose expectation is greater than or equal to the true FDR over that significance region. This is the approach that was taken by Storey (2002), for example. In either case, we want to be conservative regardless of the value of m_0, which is usually unknown. If this holds then 'strong control' is provided. 'Weak control' is provided when the procedure is conservative only when $m_0 = m$; in general this is not of interest in the FDR case (Benjamini and Hochberg, 1995). Throughout this paper, 'control' implicitly means strong control unless explicitly stated otherwise.

Benjamini and Hochberg (1995) proved by induction that the following procedure (referred to here as the BH procedure) controls the FDR at level α when the p-values following the null distribution are independent and uniformly distributed.

Step 1: let $p_{(1)} \leqslant \ldots \leqslant p_{(m)}$ be the ordered, observed p-values.
Step 2: calculate $\hat{k} = \max\{1 \leqslant k \leqslant m : p_{(k)} \leqslant \alpha\, k/m\}$.
Step 3: if $\hat{k}$ exists, then reject null hypotheses corresponding to $p_{(1)} \leqslant \ldots \leqslant p_{(\hat{k})}$. Otherwise, reject nothing.

Other sequential p-value FDR controlling methods have been suggested (e.g. Benjamini and Liu (1999)). The BH procedure was originally introduced by Simes (1986) to control weakly the FWER when all p-values are independent, but it provides strong control of the FDR as well.

Let $\text{FDR}(t)$ denote the FDR when rejecting all null hypotheses with $p_i \leqslant t$ for $i = 1, \ldots, m$. For $t \in [0, 1]$, we define the following empirical processes based on the notation in Table 1:

$$\left.\begin{aligned} V(t) &= \#\{\text{null } p_i : p_i \leqslant t\}, \\ S(t) &= \#\{\text{alternative } p_i : p_i \leqslant t\}, \\ R(t) = V(t) + S(t) &= \#\{p_i : p_i \leqslant t\}. \end{aligned}\right\} \tag{2}$$

Additional processes can be defined for the other variables in Table 1. In terms of these empirical processes, we have

$$\mathrm{FDR}(t) = E\left[\frac{V(t)}{R(t) \vee 1}\right].$$

For fixed t, Storey (2002) provided a family of conservatively biased point estimates of FDR(t). In this paper we denote these estimates by $\widehat{\mathrm{FDR}}_\lambda(t)$, where $\lambda \in [0, 1)$ is a tuning parameter that is explained later. The BH procedure induces a data-dependent threshold $p_{(\hat{k})}$ rather than some fixed t. In Section 2 we show that the BH procedure is simply a random thresholding procedure which is found by choosing t so that the natural empirical estimator $t/\{R(t)/m\}$ of FDR(t) (which in fact is the special case $\widehat{\mathrm{FDR}}_{\lambda=0}(t)$) is bounded by α. In theorem 2, we provide an alternative proof of strong control of the BH procedure with a simple martingale argument applied to the empirical processes $V(t)$ and $R(t)$. Theorem 3 shows that the analogous thresholding procedure using the estimates $\widehat{\mathrm{FDR}}_\lambda(t)$ for general $\lambda \in [0, 1)$ maintains strong control of the FDR at a predetermined level. Thus, we essentially provide a new family of FDR controlling procedures and show that the goals of the two approaches can be met with this one family of estimates. Another motivation for studying this general class of FDR controlling procedures is that $\widehat{\mathrm{FDR}}_{\lambda=0}(t)$ is the most conservatively biased estimate of all $\lambda \in [0, 1)$ (see Storey (2002) and the proof of theorem 1).

Our main results in the finite sample setting, theorems 2 and 3, formulate these procedures in terms of empirical processes. Besides allowing simple proofs of the finite sample results, the empirical process approach sets the stage for our asymptotic results in theorems 4 and 5, where we provide conditions under which asymptotic control of the FDR can be achieved. The proofs are straightforward, relying only on convergence of the underlying empirical processes, and not large deviation inequalities as in Genovese and Wasserman (2002a). This also allows for the presence of certain forms of dependence.

In the asymptotic setting, we provide a conservative estimate of the error rate over all significance regions *simultaneously*, which is essentially equivalent to controlling the FDR at all levels simultaneously. In theorem 6 we show that the estimates of FDR(t) defined in Storey (2002) provide an asymptotic form of simultaneous control if the corresponding empirical processes converge. Thus, asymptotically, the goals of the two approaches are equivalent and there is no need to distinguish between them. A slightly different view of this is to assign a simultaneously conservative multiple-testing measure of significance to each test. In theorem 7, we show the simultaneous conservative consistency of Storey's (2002) q-value estimates, which assign measures of significance to each test in terms of FDRs.

The point of view that was taken in Storey (2002) allows an additional error measure to be considered, the positive false discovery rate pFDR, whose motivation can be found in Storey (2001):

$$\mathrm{pFDR} = E\left[\frac{V}{R}\middle| R > 0\right].$$

In an asymptotic setting, pFDR and the FDR are equivalent, and consequently any asymptotic results about the FDR can essentially be directly translated into results for pFDR.

Other asymptotic properties of the FDR have previously been studied (Genovese and Wasserman, 2002a, b; Finner and Roters, 2001, 2002). We are not the first to study FDR procedures when the p-values are dependent (Yekutieli and Benjamini, 1999; Benjamini and Yekutieli, 2001; Storey and Tibshirani, 2001), but our empirical process approach puts the asymptotic and dependence issues into one, coherent, framework. It should be mentioned, though, that our framework does not enable us to prove finite sample results in the dependence setting without further assumptions on the underlying distributions, e.g. the positive regression dependence property that is discussed in Benjamini and Yekutieli (2001).

It has previously been pointed out that the BH procedure can be made less conservative by incorporating an estimate of the proportion of true null hypotheses (Benjamini and Hochberg, 1995, 2000; Storey, 2002). Benjamini and Hochberg (2000) proposed a data-adaptive procedure for estimating this proportion and modifying the BH procedure. However, there has been no proof that this procedure provides strong control of the FDR. Extending Storey's (2002) more tractable method, we can prove strong control of our proposed procedure, which also incorporates an estimate of the proportion of true null hypotheses.

The lay-out of the paper is as follows. In Section 2 we formally describe our proposed procedures and main results. Some numerical results are presented in Section 3. Section 4 outlines proofs of the finite sample results, and Section 5 of the asymptotic results. Finally, Section 6 provides a numerical method for automatically picking a tuning parameter used in the procedures proposed.

2. Procedures proposed and main results

In this section we formally describe the procedures proposed and main results of this work. We first review the point estimate approach for a fixed significance region; then we introduce an FDR controlling procedure in terms of our FDR point estimate, and we describe the simultaneous conservative consistency of the estimates.

2.1. $\widehat{\mathrm{FDR}}_\lambda(t)$ is a conservative point estimate of FDR(t)

Recall that, for a non-random significance threshold t, $\mathrm{FDR}(t)$ denotes the FDR when rejecting all null hypotheses with p-values less than or equal to t. The estimate of $\mathrm{FDR}(t)$ proposed in Storey (2002) is

$$\widehat{\mathrm{FDR}}_\lambda(t) = \frac{\hat{\pi}_0(\lambda)t}{\{R(t) \vee 1\}/m}. \tag{3}$$

The term $\hat{\pi}_0(\lambda)$ is an estimate of $\pi_0 \equiv m_0/m$, the proportion of true null hypotheses. This estimate depends on the tuning parameter λ and is defined as

$$\hat{\pi}_0(\lambda) = \frac{W(\lambda)}{(1-\lambda)m}, \tag{4}$$

where $W(\lambda) = m - R(\lambda)$ (see Table 1). The reasoning behind $\hat{\pi}_0(\lambda)$ is the following. As long as each test has a reasonable power, then most of the p-values near 1 should be null. Therefore, for a well-chosen λ, we expect about $\pi_0(1-\lambda)$ of the p-values to lie in the interval $(\lambda, 1]$, because the null p-values are uniformly distributed. Therefore, $W(\lambda)/m \approx \pi_0(1-\lambda)$, where $E[\hat{\pi}_0(\lambda)] \geqslant \pi_0$ when the p-values corresponding to the true null hypotheses are uniformly distributed (or stochastically greater). There is an inherent bias–variance trade-off in the choice of λ. In most cases, when λ grows smaller, the bias of $\hat{\pi}_0(\lambda)$ grows larger, but the variance becomes smaller. Therefore, λ can be chosen to try to balance this trade-off. The interested reader is referred to Storey (2002) for a detailed motivation of this and related estimates.

Our first result is an extension of the conservative bias result of Storey (2002). The only assumption that we make is that the p-values corresponding to the true null hypotheses are independent and uniformly distributed, whereas Storey (2002) assumed an independent and identically distributed mixture model of the p-values and stochastic ordering. Our assumptions are the same as those used to show strong control of the BH procedure in Benjamini and Hochberg (1995).

Theorem 1. Suppose that the p-values corresponding to the true null hypotheses are independent and uniformly distributed. Then, for fixed $\lambda \in [0, 1)$, $E[\widehat{\mathrm{FDR}}_\lambda(t)] \geqslant \mathrm{FDR}(t)$.

2.2. Using $\widehat{FDR}_\lambda(t)$ to control the false discovery rate strongly

We now use $\widehat{\mathrm{FDR}}_\lambda(t)$ to derive a new class of FDR controlling procedures, of which the BH procedure is a special case. Since $\widehat{\mathrm{FDR}}_\lambda(t)$ is a conservative point estimate of $\mathrm{FDR}(t)$, it is tempting to try to use $\widehat{\mathrm{FDR}}_\lambda(t)$ to provide strong control, i.e., if we want to control the FDR at level α, we take the largest t such that $\widehat{\mathrm{FDR}}_\lambda(t) \leqslant \alpha$ as the significance threshold. We show that this heuristic procedure controls the FDR.

Sequential p-value methods estimate an appropriate significance threshold based on the p-values and the prechosen level α. Thus, we define the following function that chooses the cut point based on some function F defined on $[0, 1]$:

$$t_\alpha(F) = \sup\{0 \leqslant t \leqslant 1 : F(t) \leqslant \alpha\}. \tag{5}$$

In words, $t_\alpha(F)$ finds the largest t such that $F(t) \leqslant \alpha$. We shall be particularly concerned with the thresholding rule $t_\alpha(\widehat{\mathrm{FDR}}_\lambda) = \sup\{0 \leqslant t \leqslant 1 : \widehat{\mathrm{FDR}}_\lambda(t) \leqslant \alpha\}$. A slightly modified version is considered for finite sample results. Since $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$ is a random variable, we write for ease of notation

$$\mathrm{FDR}\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\} \triangleq E\left[\frac{V\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}}{R\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\} \vee 1}\right].$$

The thresholding rule $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$ has a useful interpretation in the context of the BH procedure. Under the assumption of independent and uniformly distributed null p-values, the BH procedure controls the FDR at exactly level $\pi_0\alpha$ (see Finner and Roters (2001), Benjamini and Yekutieli (2001) and theorem 2 below). Thus, it follows that, if m is replaced with $\pi_0 m$ in the BH procedure, then the FDR is controlled exactly at level α. This eliminates the conservative bias of the BH procedure and therefore increases the average proportion of false null hypotheses that are rejected at each α (which could be stated as an increase in the overall 'average power'). A strategy to increase the power is then to replace m in the BH procedure with our estimate $\hat{\pi}_0(\lambda)m$. This makes the procedure more accurate in the sense that the true level of FDR control is asymptotically closer to α than the BH procedure (theorem 4). The thresholding rule $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$ operationally makes this exact change to the BH procedure. Specifically, consider the following two lemmas.

Lemma 1. The p-value step-up method $t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})$ is equivalent to the BH procedure.

Proof. Noting that $\hat{\pi}_0(\lambda = 0) = 1$, this follows by the proof of the next proposition. □

Lemma 2. In general, the p-value step-up method $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$ is equivalent to the BH procedure with m replaced by $\hat{\pi}_0(\lambda)m$.

Proof. We must show that $p_{(\hat{k}_\lambda)} \leqslant t_\alpha(\widehat{\mathrm{FDR}}_\lambda) < p_{(\hat{k}_\lambda+1)}$ where $\hat{k}_\lambda$ is the $\hat{k}$ in the BH procedure with m replaced by $\hat{\pi}_0(\lambda)m$. Ignoring the cases where $R(t) = 0$, in which case both procedures reject no p-values, the inequalities follow immediately once it is noted that the BH procedure for selecting $\hat{k}_\lambda$ is simply $\hat{k}_\lambda = \max\{k : \widehat{\mathrm{FDR}}_\lambda(p_{(k)}) \leqslant \alpha\}$. □

For the finite sample case with independent null p-values, we propose a new FDR controlling step-up method with two minor modifications to $\widehat{\mathrm{FDR}}_\lambda$. If $\lambda > 0$, we want to guarantee that $\hat{\pi}_0(\lambda) > 0$; recall that $\hat{\pi}_0(\lambda = 0) = 1$. Thus, for $\lambda > 0$ we replace $\hat{\pi}_0(\lambda)$ in $\widehat{\mathrm{FDR}}_\lambda$ with

$$\hat{\pi}_0^*(\lambda) = \frac{W(\lambda) + 1}{(1 - \lambda)m}.$$

We must also limit the significance threshold to the region $[0, \lambda]$. Therefore, the estimate of $\mathrm{FDR}(t)$ that we use for the finite sample case is

$$\widehat{\mathrm{FDR}}_\lambda^*(t) = \begin{cases} \dfrac{\hat{\pi}_0^*(\lambda)t}{\{R(t) \vee 1\}/m}, & \text{if } t \leqslant \lambda, \\ 1, & \text{if } t > \lambda. \end{cases} \tag{6}$$

These modifications allow us to prove in the finite sample case that the $t_\alpha(\widehat{\mathrm{FDR}}_\lambda^*)$ procedure controls the FDR. This last modification has little effect on the procedure, as explained later in this section. Both modifications are unnecessary to prove asymptotic control, as is shown in theorem 4. We summarize the procedure as follows.

Step 1: let α be the prechosen level at which to control the FDR.
Step 2: for any fixed significance region $[0, t]$, estimate $\mathrm{FDR}(t)$ by either $\widehat{\mathrm{FDR}}_\lambda(t)$ given in equation (3) or $\widehat{\mathrm{FDR}}_\lambda^*(t)$ given in equation (6).
Step 3: for small m where the null p-values are independent, reject all null hypotheses corresponding to $p_i \leqslant t_\alpha(\widehat{\mathrm{FDR}}_\lambda^*)$ for $\lambda > 0$, and $p_i \leqslant t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})$ for $\lambda = 0$.
Step 4: for large m where the p-values meet the 'weak dependence' conditions of theorem 4, reject all null hypotheses corresponding to $p_i \leqslant t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$.

The following two results concerning the finite sample properties of our proposed procedure are proven in Section 4. Theorem 2 has been shown before (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001), but our new proof uses martingale methods. The proof of theorem 2 easily leads to a proof of theorem 3.

Theorem 2 (Benjamini and Hochberg, 1995). If the p-values corresponding to the true null hypotheses are independent, then

$$\mathrm{FDR}\{t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})\} = \pi_0\alpha \leqslant \alpha.$$

That is, the BH procedure controls the FDR exactly at level $\pi_0\alpha$ and conservatively at level α. Now consider the procedure with $\lambda > 0$.

Theorem 3. If the p-values corresponding to the true null hypotheses are independent, then, for $\lambda > 0$,

$$\mathrm{FDR}\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda^*)\} \leqslant (1 - \lambda^{\pi_0 m})\alpha \leqslant \alpha.$$

Therefore, the thresholding procedure $t_\alpha(\widehat{\mathrm{FDR}}^*_\lambda)$ strongly controls the FDR at level α.

Remark 1. Reasonably chosen λ will tend to be larger than $t_\alpha(\widehat{\mathrm{FDR}}^*_\lambda)$, so in practice there should be little difference between $t_\alpha(\widehat{\mathrm{FDR}}^*_\lambda)$ and $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$. Consider that

$$\mathrm{FDR}(t) \approx E[V(t)]/E[R(t)] \leqslant \alpha$$

implies

$$t \leqslant \frac{(1-\pi_0)\alpha}{\pi_0(1-\alpha)}$$

since $E[R(t)]/m \leqslant \pi_0 t + (1-\pi_0)$. Even with a small $\pi_0 = 0.75$ and a large $\alpha = 0.20$, it approximately follows that, if $\mathrm{FDR}(t) \leqslant \alpha$, then $t \ll 0.1$. Therefore, using λ over the range $\lambda = 0, 0.1, 0.2, \ldots, 0.9$, for example, implies that $t_\alpha(\widehat{\mathrm{FDR}}^*_\lambda)$ will essentially be equivalent to $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$.

FDRs are most useful in cases in which many hypotheses are tested. Moreover, in areas of potential application (astrophysics, brain imaging and gene expression data), m is typically of the order of several thousand. For large m, the assumption of independence can be replaced with 'weak dependence' (defined below). For the remainder of this section, we consider the large m case.

The asymptotic theorems require the almost sure pointwise convergence of the empirical distributions of the null p-values and alternative p-values. Recall the empirical processes $V(t)$ and $S(t)$ that are defined in equation (2). Therefore, $V(t)/m_0$ and $S(t)/m_1$ are the empirical distribution functions of the null and alternative hypotheses respectively. We make the following assumptions for our asymptotic results:

$$\lim_{m\to\infty}\left\{\frac{V(t)}{m_0}\right\} = G_0(t) \quad \text{and} \quad \lim_{m\to\infty}\left\{\frac{S(t)}{m_1}\right\} = G_1(t) \qquad \text{almost surely for each } t \in (0,1], \tag{7}$$

where G_0 and G_1 are continuous functions;

$$0 < G_0(t) \leqslant t \text{ for each } t \in (0,1]; \tag{8}$$

$$\lim_{m\to\infty}(m_0/m) \equiv \pi_0 \text{ exists.} \tag{9}$$

The asymptotic results also hold if the assumption that G_0 and G_1 are continuous is replaced by the pointwise convergence of left limits of the empirical distribution functions. We define *weak dependence* to be any type of dependence where equation (7) can hold. Dependence in finite blocks, ergodic dependence and certain mixing distributions are all candidates to meet the weak dependence criterion.

Our first result shows that $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$ asymptotically provides strong control of the FDR under the above assumptions. Define

$$\widehat{\mathrm{FDR}}^\infty_\lambda(t) \triangleq \left\{\frac{1-G_0(\lambda)}{1-\lambda}\pi_0 + \frac{1-G_1(\lambda)}{1-\lambda}\pi_1\right\} G_0(t) \Big/ \{\pi_0\, G_0(t) + \pi_1\, G_1(t)\},$$

which is the pointwise limit of $\widehat{\mathrm{FDR}}_\lambda(t)$ under the assumptions of equations (7)–(9).

Theorem 4. Suppose that the convergence assumptions of equations (7)–(9) hold. If there is a $t \in (0,1]$ such that $\widehat{\mathrm{FDR}}^\infty_\lambda(t) < \alpha$, then

$$\limsup_{m\to\infty} [\mathrm{FDR}\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}] \leqslant \alpha.$$

A generalization of theorem 1 of Genovese and Wasserman (2002a) is possible by using our approach. We specifically show that, if $\widehat{\mathrm{FDR}}_\lambda$ converges almost surely pointwise to some limit $\widehat{\mathrm{FDR}}_\lambda^\infty$, then the random thresholding rule $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$ converges to the deterministic rule $t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)$.

Theorem 5. Suppose that the convergence assumptions of equations (7)–(9) hold. Then

$$\lim_{m\to\infty} \{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\} = t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)$$

almost surely if $\widehat{\mathrm{FDR}}_\lambda^\infty(\cdot)$ has a non-zero derivative at $0 < t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty) < 1$.

We now state a conservative consistency result of $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$, but we first define

$$\mathrm{FDR}^\infty(t) \triangleq \frac{\pi_0\, G_0(t)}{\pi_0\, G_0(t) + \pi_1\, G_1(t)},$$

which is the pointwise limit of $\mathrm{FDR}(t)$ under the convergence assumptions of equations (7)–(9).

Corollary 1. Under the convergence assumptions of equations (7)–(9), we have that $\lim_{m\to\infty}\{t_\alpha(\mathrm{FDR})\} = t_\alpha(\mathrm{FDR}^\infty)$ and $\lim_{m\to\infty}\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda) - t_\alpha(\mathrm{FDR})\} \leqslant 0$ almost surely.

This final result shows that, for $\lambda > 0$, $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$ is asymptotically less conservative than the BH procedure and therefore has greater asymptotic power.

Corollary 2. Suppose that the convergence assumptions of equations (7)–(9) hold, and $G_0(t) < G_1(t)$ for $t \in (0,1)$. Then for any $0 < \lambda < 1$

$$\lim_{m\to\infty} \{t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})\} < \lim_{m\to\infty} \{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\} < \lim_{m\to\infty} \{t_\alpha(\mathrm{FDR})\} \qquad \text{almost surely.}$$

Therefore, the procedure with $0 < \lambda < 1$ will asymptotically reject a higher proportion of false null hypotheses for the same chosen α.

The proofs of corollaries 1 and 2 are straightforward, so we omit them. It is conceivable that, under the assumptions of corollary 2, the least possible conservative thresholding procedure is a limit of the family of $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$.

Under certain finite sample cases, there is positive probability that $\hat{\pi}_0(\lambda) > 1$ for $\lambda > 0$, in which case the procedure could reject fewer hypotheses than the BH procedure. However, as can easily be seen in the proofs of the above results, if we replace $\hat{\pi}_0(\lambda)$ with $\min\{\hat{\pi}_0(\lambda), 1\}$, then all the asymptotic results still hold. The procedure with this minor adjustment therefore always calls at least as many p-values significant as the BH procedure. Also note that the automatic method for choosing λ in Section 6 prevents any λ from being chosen where $\hat{\pi}_0(\lambda) > 1$, so this is a moot point in practice.

2.3. $\widehat{FDR}_\lambda(t)$ is simultaneously conservatively consistent for FDR(t)

As discussed in Section 1, it can be difficult to choose a significance threshold t or FDR controlling level α before any data are seen, so it is useful to show that $\widehat{\mathrm{FDR}}_\lambda(t) \geqslant \mathrm{FDR}(t)$ for all t simultaneously. Under the convergence assumptions of equations (7)–(9), we can show a version of the above property in the following theorem.

Theorem 6. Suppose that the convergence assumptions of equations (7)–(9) hold. Then, for each $\delta > 0$,

$$\lim_{m\to\infty} \inf_{t\geqslant\delta} \{\widehat{\mathrm{FDR}}_\lambda(t) - \mathrm{FDR}(t)\} \geqslant 0 \quad \text{and} \quad \lim_{m\to\infty} \inf_{t\geqslant\delta} \left\{\widehat{\mathrm{FDR}}_\lambda(t) - \frac{V(t)}{R(t)\vee 1}\right\} \geqslant 0$$

with probability 1.

This result holds not only for $\mathrm{FDR}(t)$ but also for the realized proportion of false discoveries $V(t)/\{R(t) \vee 1\}$.

As will be seen from the proofs in Section 5, the above uniform convergence follows directly from the arguments that are used in the proof of the Glivenko–Cantelli theorem, based on the pointwise convergence of the corresponding empirical distribution functions to continuous limits. Therefore, we have now shown that the point estimates can be used to estimate conservatively the FDR over all significance regions simultaneously. This can be equivalently viewed as controlling the FDR at all levels simultaneously, making the goals of the two approaches asymptotically equivalent.

2.4. Simultaneous conservative consistency of the estimated *q*-values

In single-hypothesis testing, it is common to report the p-value as a measure of significance of the test rather than whether or not the null hypothesis was rejected at a predetermined level. If theorem 6 holds, then in theory we can avoid having to control the FDR at a fixed level and just examine $\widehat{\mathrm{FDR}}_\lambda(p_i)$ for $i = 1, \ldots, m$. To assign a measure of significance in terms of FDR to a particular p-value p_i, there are benefits to reporting the minimum FDR at which this p-value can be called significant rather than $\widehat{\mathrm{FDR}}_\lambda(p_i)$. This can be accomplished through the 'q-value' which is proposed in Storey (2001).

To be precise, suppose that

(a) $H_i = 0$ or $H_i = 1$ according to whether the ith null hypothesis is true or not,
(b) $H_i \sim^{\mathrm{IID}}$ Bernoulli(π_1) so that $\Pr(H_i = 0) = \pi_0$ and $\Pr(H_i = 1) = \pi_1$ and
(c) $P_i | H_i \sim^{\mathrm{IID}} (1 - H_i)G_0 + H_i G_1$.

Storey (2001) showed that

$$E\left[\frac{V(t)}{R(t)}\middle| R(t) > 0\right] = \Pr(H_i = 0 | P_i \leqslant t),$$

where $\Pr(H_i = 0 | P_i \leqslant t)$ is the same for each i because of the independent and identically distributed data assumptions. The expectation on the left-hand side is called the positive false discovery rate, pFDR. In the spirit of the definition of the p-value, Storey (2001) defined

$$q\text{-value}(p_i) = \min_{t\geqslant p_i}\{\mathrm{pFDR}(t)\},$$

i.e. the q-value of a statistic is the minimum pFDR at which that statistic can be called significant. Under assumptions (a)–(c), it follows that

$$q\text{-value}(p_i) = \min_{t\geqslant p_i}\{\Pr(H_i = 0 | P_i \leqslant t)\},$$

which is a Bayesian analogue of the p-value—or rather a 'Bayesian posterior type I error rate', a concept suggested as early as Morton (1955). Further properties and motivation can be found in Storey (2001). Also, an application of the connection between $\Pr(H_i = 0 | P_i \leqslant t)$ and posterior probabilities can be found in Efron *et al.* (2001).

Storey (2002) estimated pFDR(t) by $\widehat{\mathrm{pFDR}}_\lambda(t) = \widehat{\mathrm{FDR}}_\lambda(t)/\{1-(1-t)^m\}$ and estimated q-value(p_i) by

$$\hat{q}_\lambda(p_i) = \inf_{t \geqslant p_i} \{\widehat{\mathrm{pFDR}}_\lambda(t)\}. \tag{10}$$

Motivated by the posterior probability estimate

$$\widehat{\Pr}_\lambda(H_i = 0 | P_i \leqslant t) = \frac{\hat{\pi}_0(\lambda)t}{R(t)/m},$$

another clearly motivated estimate of the q-value is

$$\hat{q}_\lambda(p_i) = \inf_{t \geqslant p_i} \{\widehat{\Pr}_\lambda(H_i = 0 | P_i \leqslant t)\}, \tag{11}$$

which equals $\inf_{t \geqslant p_i}\{\widehat{\mathrm{FDR}}_\lambda(t)\}$. Under the weak dependence criteria, we can prove that $\hat{q}_\lambda(p_i)$ is asymptotically conservative, regardless of definition (10) or (11). Moreover, we can show that $\hat{q}_\lambda(\cdot)$ is *simultaneously* conservatively consistent over all arguments in the following sense.

Theorem 7. Suppose that the convergence assumptions of equations (7)–(9) hold. Then, for each $\delta > 0$,

$$\lim_{m \to \infty} \inf_{t \geqslant \delta} \{\hat{q}_\lambda(t) - q\text{-value}(t)\} \geqslant 0 \qquad \text{almost surely.}$$

Adjusted p-values, which are defined in terms of a particular step-up or step-down method, are usually justified by strong control of the method at a *single* predetermined level α (Shaffer, 1995). Here, we have provided a more general result in the sense that we have shown control at *all* $\alpha \geqslant \delta$, which is the condition that is necessary for adjusted p-values or estimated q-values to be simultaneously conservative.

3. Numerical studies

We now present numerical results in both an independent p-value example and a dependent p-value example. In each of these examples, the *average power* is defined to be the power averaged over each true alternative hypothesis.

3.1. Independence example

We first present a simple numerical study to compare the average power of our proposed FDR controlling procedure with that of the Benjamini and Hochberg (1995) procedure. We performed $m = 1000$ one-sided hypothesis tests with null distribution $N(0,1)$ and alternative distribution $N(2,1)$. We let $m_0 = 100, \ldots, 900$ and generated 1000 sets of 1000 normal random variables for each m_0-value. The BH procedure and proposed finite sample procedure $t_\alpha(\widehat{\mathrm{FDR}}^*_\lambda)$ were performed at levels $\alpha = 0.01$ and $\alpha = 0.05$. For simplicity, we set $\lambda = 0.5$ for both cases in our procedure. Fig. 1 shows the average power of the procedure proposed PP *versus* BH. It can be seen that the increase in power that we achieve is greater the smaller m_0 is. This makes sense because the difference in our proposed procedure PP over BH is that it estimates π_0, where $\pi_0 = m_0/m$.

Secondly, we performed 3000 hypothesis tests of the above distributions with $m_0 = 2400$. The $\hat{q}_\lambda(p_i)$ were calculated at each p-value p_i, as well as the true q-value(p_i). These calculations are displayed in Fig. 2. It can be seen that $\hat{q}_\lambda(\cdot) \geqslant q$-value$(\cdot)$ over all p-values simultaneously, which is exactly the result of theorem 7.

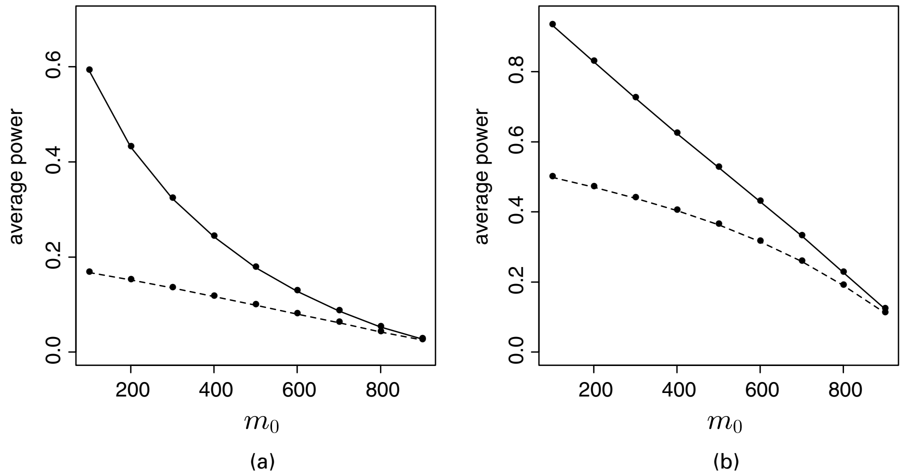

Fig. 1. Average power *versus* m_0 for the proposed procedure for small m (———) and the Benjamini and Hochberg (1995) procedure (-------): (a) FDR controlled at level $\alpha = 0.01$; (b) FDR controlled at level $\alpha = 0.05$ (there is an increase in power under the proposed procedure in both (a) and (b))

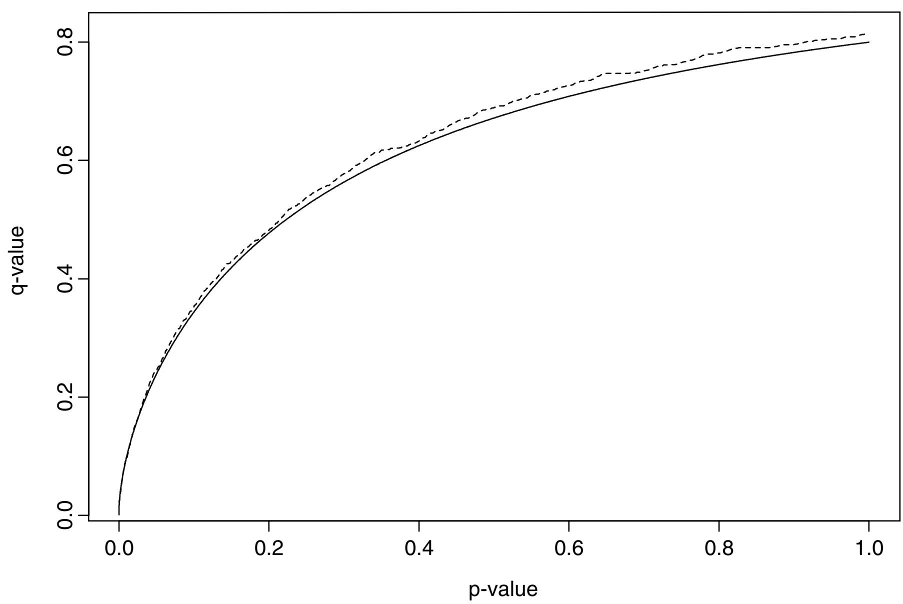

Fig. 2. $\hat{q}_\lambda(\cdot)$ and q-value$(\cdot)$ evaluated at each p-value for 3000 tests of $N(0,1)$ *versus* $N(2,1)$ with $m_0 = 2400$ (———, actual q-values; - - - - -, q-values)

3.2. Dependence example

Our second numerical study illustrates the asymptotic control of the FDR, even under a certain form of dependence. The null statistics have $N(0,1)$ marginal distributions with $m_0 = 2400$ and the alternative distributions have marginal distribution $N(2,1)$ with $m_1 = 600$. The statistics have correlation ± 0.4 in groups of 10. Specifically, for $1 \leqslant j \leqslant k \leqslant 10$,

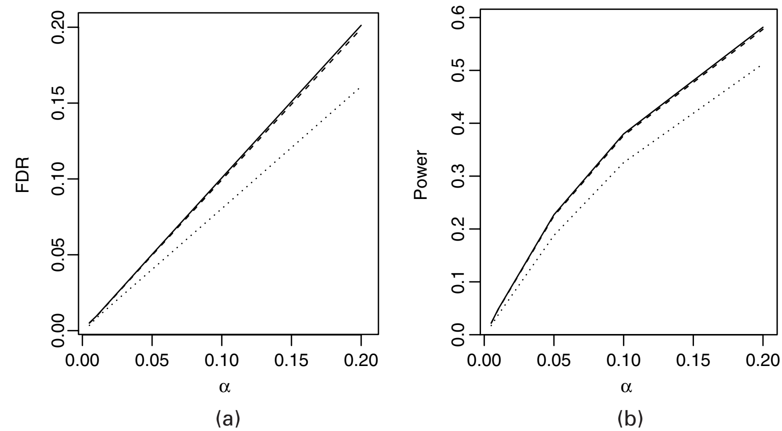

Fig. 3. (a) FDR when performing the BH (·······), PP (- - - - -) and optimal (———) procedures at level α and (b) average power attained under these three procedures: it can be seen that $t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0.5})$ (PP) attains the control and power near that of the optimal procedure whereas the more conservative $t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})$ (BH) does not

$$\Sigma_{jk} = \begin{cases} 1, & j = k, \\ 0.4, & j < k \leqslant 5, \\ -0.4, & j \leqslant 5, k > 5, \end{cases}$$

and

$$\mathrm{cov}(T_1, \ldots, T_{3000}) = \Sigma \otimes I_{300\times 300},$$

the Kronecker product of Σ and $I_{300\times 300}$. These statistics apparently do not meet the 'positive regression dependence' condition of Benjamini and Yekutieli (2001).

1000 data sets were generated, and for each one $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)$ was calculated for $\alpha = 0.005, 0.01, 0.05, 0.10, 0.20$. We used $\lambda = 0$ (BH), $\lambda = 0.5$ (PP) and the exact threshold (optimal) which can be attained by using the true $\pi_0 = 0.80$ in our proposed procedure. The true FDR and the average power were calculated for all these procedures. Fig. 3 shows that all three procedures control the FDR at level α, as the theory asserts. Moreover, $t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0.5})$ attains the control and power near that of the optimal procedure, showing the improvement of our proposed methodology over Benjamini and Hochberg's (1995) methodology. The numbers from this study are listed in Table 2.

4. Finite sample proofs

We first prove the conservative bias of $\widehat{\mathrm{FDR}}_\lambda(t)$ under the assumption that the p-values corresponding to the true null hypotheses are independent and uniformly distributed.

4.1. Proof of theorem 1

First assume that $t < \lambda$. Then

$$\begin{aligned} E[\widehat{\mathrm{FDR}}_\lambda(t) - \mathrm{FDR}(t)] &\geqslant E\left[\frac{[\{m_0 - V(\lambda)\}/(1-\lambda)]t - V(t)}{R(t) \vee 1}\right] \\ &= E\left[E\left[\frac{[\{m_0 - V(\lambda)\}/(1-\lambda)]t - V(t)}{R(t) \vee 1}\middle| V(t), S(t)\right]\right] \\ &= E\left[\frac{\{m_0 t - V(t)\}/(1-t)}{R(t) \vee 1}\right]. \end{aligned}$$

Table 2. Numerical study of $t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})$ (BH), $t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0.5})$ (PP) and the optimal procedure under dependence†

α	*FDR for the following methods:*			*Average power for the following methods:*		
	BH	*PP*	*Optimal*	*BH*	*PP*	*Optimal*
0.005	0.00343	0.00492	0.00516	0.0172	0.0218	0.0221
	(5×10^{-4})	(6×10^{-4})	(7×10^{-4})	(4×10^{-4})	(4×10^{-4})	(4×10^{-4})
0.01	0.00828	0.00934	0.00952	0.0376	0.0477	0.0483
	(7×10^{-4})	(6×10^{-4})	(6×10^{-4})	(6×10^{-4})	(6×10^{-4})	(6×10^{-4})
0.05	0.0403	0.0497	0.0503	0.188	0.225	0.227
	(6×10^{-4})	(6×10^{-4})	(6×10^{-4})	(9×10^{-4})	(9×10^{-4})	(9×10^{-4})
0.10	0.0804	0.0994	0.101	0.326	0.377	0.380
	(7×10^{-4})	(7×10^{-4})	(7×10^{-4})	(9×10^{-4})	(9×10^{-4})	(9×10^{-4})
0.20	0.161	0.199	0.201	0.512	0.578	0.582
	(8×10^{-4})	(8×10^{-4})	(8×10^{-4})	(8×10^{-4})	(9×10^{-4})	(8×10^{-4})

†The Monte Carlo standard error is listed below each number.

In going from the second to the third line, we use the fact that

$$m_0 - V(\lambda)|V(t) \sim \text{binomial}\{m_0 - V(t), (1-\lambda)/(1-t)\}.$$

Next, assume that $t \geqslant \lambda$ so that

$$V(\lambda)|V(t) \sim \text{binomial}\{V(t), \lambda/t\}.$$

Redoing the computation as above shows that

$$E[\widehat{\mathrm{FDR}}_\lambda(t) - \mathrm{FDR}(t)] \geqslant E\left[\frac{\{m_0 t - V(t)\}/(1-\lambda)}{R(t)\vee 1}\right].$$

The final result follows by showing that $E[\{m_0 t - V(t)\}/\{R(t)\vee 1\}] \geqslant 0$. However,

$$\begin{aligned} E\left[\frac{m_0 t - V(t)}{R(t)\vee 1}\right] &\geqslant E\left[\frac{\{m_0 t - V(t)\}/(1-t)}{R(t)\vee 1}\mathbf{1}_{\{V(t)\geqslant 1\}}\right] \\ &= E\left[E\left[\frac{\{m_0 t - V(t)\}/(1-t)}{V(t)+S(t)}\mathbf{1}_{\{V(t)\geqslant 1\}}\middle| S(t)\right]\right]. \end{aligned}$$

The conclusion now follows from Jensen's inequality. □

We now prove theorems 2 and 3 by using martingale methods. A similar argument has been used to prove the ballot theorem (see Chow *et al.* (1971), page 26), where the argument is attributed to G. Simons). See also Durrett (1996), page 267, where an unnecessary additional condition is imposed. We use the fact that the random variable $t_\alpha(\widehat{\mathrm{FDR}}^*_\lambda)$ is a stopping time with respect to a certain filtration, which allows the use of the optional stopping theorem. Recall the empirical processes $V(t)$, $S(t)$ and $R(t)$ defined in expression (2). We view 'time' (i.e. the thresholds t) as running backwards in these empirical processes. The proof that $V(t)/t$ is a martingale with time running backwards is easily shown, so we omit the proof here.

Lemma 3. If the p-values of the m_0 true null hypotheses are independent, then $V(t)/t$ for $0 \leqslant t < 1$ is a martingale with time running backwards with respect to the filtration $\mathcal{F}_t = \sigma(\mathbf{1}_{\{p_i \leqslant s\}}, t \leqslant s \leqslant 1, i = 1, \ldots, m)$, i.e., for $s \leqslant t$, $E[V(s)/s|\mathcal{F}_t] = V(t)/t$.

The following elementary lemma, whose proof is also omitted, incorporates the stopping times (thresholding rules) into our martingale framework.

Lemma 4. The random variable $t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})$ is a stopping time with respect to $\mathcal{F}_t$, with time running backwards. Further, for $\lambda > 0$, $t_\alpha(\widehat{\mathrm{FDR}}^*_\lambda)$ is a stopping time with respect to $\mathcal{F}^\lambda_t =^\triangle \mathcal{F}_{t\wedge\lambda}$.

We are now ready to prove theorems 2 and 3. Recall that $\pi_0 = m_0/m$.

4.2. Proof of theorem 2

Noting that the process $m\,t/R(t)$ has only upward jumps and has a final value of 1, we see that $R\{t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})\} = t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})m/\alpha$. Therefore, the ratio whose expectation we must calculate can be expressed as

$$\frac{V\{t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})\}}{R\{t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})\}} = \frac{\alpha}{m}\,\frac{V\{t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})\}}{t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})}.$$

Noting that $V(t)/t$ stopped at $t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})$ is bounded by m/α, the optional stopping theorem then implies

$$\mathrm{FDR}\{t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})\} = \frac{\alpha}{m}E\left[\frac{V\{t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})\}}{t_\alpha(\widehat{\mathrm{FDR}}_{\lambda=0})}\right] = \frac{\alpha}{m}E[V(1)] = \frac{m_0}{m}\alpha.$$

4.3. Proof of theorem 3

Abbreviate $t_\alpha(\widehat{\mathrm{FDR}}^*_\lambda)$ by t^λ_α. If $\widehat{\mathrm{FDR}}^*_\lambda(\lambda) \geqslant \alpha$ then it can be seen that $R(t^\lambda_\alpha) = t^\lambda_\alpha m\,\hat{\pi}^*_0(\lambda)/\alpha$ similarly to above. Moreover, when $\widehat{\mathrm{FDR}}^*_\lambda(\lambda) \geqslant \alpha$, then $V(t)/t$ stopped at t^λ_α is bounded by $m(m-1)/\{(1-\lambda)\alpha\}$. Thus,

$$\mathrm{FDR}(t^\lambda_\alpha) = E\left[\frac{V(t^\lambda_\alpha)}{R(t^\lambda_\alpha)}; \widehat{\mathrm{FDR}}^*_\lambda(\lambda) \geqslant \alpha\right] + E\left[\frac{V(t^\lambda_\alpha)}{R(t^\lambda_\alpha)}; \widehat{\mathrm{FDR}}^*_\lambda(\lambda) < \alpha\right].$$

Now

$$\begin{aligned}
E\left[\frac{V(t^\lambda_\alpha)}{R(t^\lambda_\alpha)}; \widehat{\mathrm{FDR}}^*_\lambda(\lambda) \geqslant \alpha\right] &= E\left[\alpha\frac{1-\lambda}{W(\lambda)+1}\,\frac{V(t^\lambda_\alpha)}{t^\lambda_\alpha}; \widehat{\mathrm{FDR}}^*_\lambda(\lambda) \geqslant \alpha\right] \\
&= E\left[\alpha\frac{1-\lambda}{W(\lambda)+1}E\left[\left.\frac{V(t^\lambda_\alpha)}{t^\lambda_\alpha}\right| \mathcal{F}_\lambda\right]; \widehat{\mathrm{FDR}}^*_\lambda(\lambda) \geqslant \alpha\right] \\
&= E\left[\alpha\frac{1-\lambda}{W(\lambda)+1}\,\frac{V(\lambda)}{\lambda}; \widehat{\mathrm{FDR}}^*_\lambda(\lambda) \geqslant \alpha\right],
\end{aligned}$$

where the last step follows by the optional stopping theorem. It also easily follows that

$$E\left[\frac{V(t^\lambda_\alpha)}{R(t^\lambda_\alpha)}; \widehat{\mathrm{FDR}}^*_\lambda(\lambda) < \alpha\right] \leqslant E\left[\alpha\frac{1-\lambda}{W(\lambda)+1}\,\frac{V(\lambda)}{\lambda}; \widehat{\mathrm{FDR}}^*_\lambda(\lambda) < \alpha\right].$$

The upper bound follows by

$$\begin{aligned}\mathrm{FDR}(t_\alpha^\lambda) &\leqslant E\left[\frac{1-\lambda}{W(\lambda)+1}\,\frac{V(\lambda)}{\lambda}\alpha\right]\\ &\leqslant E\left[\frac{1-\lambda}{m_0-V(\lambda)+1}\,\frac{V(\lambda)}{\lambda}\alpha\right]\\ &=(1-\lambda^{m_0})\alpha\leqslant\alpha.\end{aligned}$$

5. Large sample proofs

In this section, we prove several theorems that assume the convergence assumptions of equations (7)–(9).

5.1. Proof of theorem 4

Let t' be the $t'>0$ such that $\alpha-\widehat{\mathrm{FDR}}_\lambda^\infty(t')=\varepsilon>0$. Therefore, we can take m sufficiently large that $|\widehat{\mathrm{FDR}}_\lambda^\infty(t')-\widehat{\mathrm{FDR}}_\lambda(t')|<\varepsilon/2$ which implies that $\widehat{\mathrm{FDR}}_\lambda(t')<\alpha$ and $t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\geqslant t'$. Therefore, $\liminf_{m\to\infty}\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}\geqslant t'$ with probability 1. By theorem 6 it follows that with probability 1

$$\liminf_{m\to\infty}\left[\widehat{\mathrm{FDR}}_\lambda\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}-\frac{V\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}}{R\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}\vee 1}\right]\geqslant\lim_{m\to\infty}\inf_{t\geqslant\delta}\left\{\widehat{\mathrm{FDR}}_\lambda(t)-\frac{V(t)}{R(t)\vee 1}\right\}\geqslant 0$$

for $\delta=t'/2$. Since $\widehat{\mathrm{FDR}}_\lambda\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}\leqslant\alpha$ it follows that

$$\limsup_{m\to\infty}\left[\frac{V\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}}{R\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}\vee 1}\right]\leqslant\alpha,$$

with probability 1. By Fatou's lemma,

$$\limsup_{m\to\infty}\left(E\left[\frac{V\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}}{R\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}\vee 1}\right]\right)\leqslant E\left[\limsup_{m\to\infty}\left[\frac{V\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}}{R\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}\vee 1}\right]\right]\leqslant\alpha.$$

5.2. Proof of theorem 5

For $t'>t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)$, we have that $\widehat{\mathrm{FDR}}_\lambda^\infty(t')-\widehat{\mathrm{FDR}}_\lambda^\infty\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)\}=\varepsilon$ for some $\varepsilon>0$. Thus, we can take m sufficiently large that $|\widehat{\mathrm{FDR}}_\lambda^\infty(t')-\widehat{\mathrm{FDR}}_\lambda(t')|<\varepsilon/2$, and thus $\widehat{\mathrm{FDR}}_\lambda(t')>\alpha$ eventually with probability 1. Hence $\limsup_{m\to\infty}\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}\leqslant t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)$ almost surely. If $\widehat{\mathrm{FDR}}_\lambda^\infty(\cdot)$ has a non-zero derivative at $t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)$, then it must be a positive derivative. Otherwise, the definition of $t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)$ would be violated. Thus, there is a neighbourhood, say of size $\delta>0$, such that, for $t'\in[t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)-\delta, t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)]$, we have

$$\widehat{\mathrm{FDR}}_\lambda^\infty(t')<\widehat{\mathrm{FDR}}_\lambda^\infty\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)\}.$$

By a similar argument to that for the previous case, we have that, for any t' in this neighbourhood, $\widehat{\mathrm{FDR}}_\lambda(t')<\alpha$ eventually with probability 1. Thus $\liminf_{m\to\infty}\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda)\}\geqslant t_\alpha(\widehat{\mathrm{FDR}}_\lambda^\infty)$ almost surely. Putting these together, we have the result.

5.3. Proof of theorem 6

By an easy modification of the Glivenko–Cantelli theorem,

$$\lim_{m\to\infty} \sup_{0\leqslant t\leqslant 1} \left| \frac{V(t)}{m} - \pi_0\, G_0(t) \right| = 0 \qquad \text{almost surely,}$$

$$\lim_{m\to\infty} \sup_{0\leqslant t\leqslant 1} \left| \frac{\{V(t)+S(t)\} \vee 1}{m} - \{\pi_0\, G_0(t) + \pi_1\, G_1(t)\} \right| = 0 \qquad \text{almost surely,}$$

where $V(t) + S(t) = R(t)$. Take any $\delta > 0$. Then

$$\begin{aligned} &\lim_{m\to\infty} \sup_{t\geqslant\delta} \left| \frac{V(t)}{\{V(t)+S(t)\}\vee 1} - \frac{\pi_0 m\, G_0(t)}{\{V(t)+S(t)\}\vee 1} \right| \\ &\quad \leqslant \lim_{m\to\infty} \left| \frac{m}{\{V(\delta)+S(\delta)\}\vee 1} \right| \sup_{t\geqslant\delta} \left| \frac{V(t)}{m} - \pi_0\, G_0(t) \right| = 0 \qquad \text{almost surely.} \end{aligned} \tag{12}$$

Since $\lim_m \inf_{t\geqslant\delta}\{\hat{\pi}_0(\lambda)t - \pi_0\, G_0(t)\} \geqslant 0$ almost surely, it follows that

$$\lim_{m\to\infty} \inf_{t\geqslant\delta} \left\{ \widehat{\mathrm{FDR}}_\lambda(t) - \frac{V(t)}{R(t)\vee 1} \right\} \geqslant 0 \qquad \text{almost surely.}$$

to show that $\lim_{m\to\infty} \inf_{t\geqslant\delta}\{\widehat{\mathrm{FDR}}_\lambda(t) - \mathrm{FDR}(t)\} \geqslant 0$, it suffices to show that

$$\lim_{m\to\infty} \sup_{t\geqslant\delta} \left| \frac{V(t)}{R(t)\vee 1} - \mathrm{FDR}(t) \right| = 0 \qquad \text{almost surely.} \tag{13}$$

Since $\pi_0\, G_0(\delta) + \pi_1\, G_1(\delta) > 0$ and these are both non-decreasing functions, it is straightforward to show that

$$\lim_{m\to\infty} \sup_{t\geqslant\delta} \left| \frac{m}{\{V(t)+S(t)\}\vee 1} - \frac{1}{\pi_0\, G_0(t) + \pi_1\, G_1(t)} \right| = 0 \qquad \text{almost surely.}$$

Using this, inequality (12) from above and the triangle inequality, we obtain

$$\begin{aligned} 0 &= \lim_{m\to\infty} \sup_{t\geqslant\delta} \left| \frac{V(t) - m\pi_0\, G_0(t)}{\{V(t)+S(t)\}\vee 1} \right| \\ &\quad + \lim_{m\to\infty} \sup_{t\geqslant\delta} \left| \frac{m\pi_0\, G_0(t)}{\{V(t)+S(t)\}\vee 1} - \frac{\pi_0\, G_0(t)}{\pi_0\, G_0(t) + \pi_1\, G_1(t)} \right| \qquad \text{almost surely.} \\ &\geqslant \lim_{m\to\infty} \sup_{t\geqslant\delta} \left| \frac{V(t)}{R(t)\vee 1} - \frac{\pi_0\, G_0(t)}{\pi_0\, G_0(t) + \pi_1\, G_1(t)} \right| \geqslant 0 \qquad \text{almost surely.} \end{aligned} \tag{14}$$

Now

$$\left| \frac{V(t)}{R(t)\vee 1} - \frac{\pi_0\, G_0(t)}{\pi_0\, G_0(t) + \pi_1\, G_1(t)} \right| \leqslant 2,$$

so it follows that

$$\begin{aligned}
0 &= E\left[\lim_{m\to\infty}\sup_{t\geqslant\delta}\left|\frac{V(t)}{R(t)\vee 1}-\frac{\pi_0\,G_0(t)}{\pi_0\,G_0(t)+\pi_1\,G_1(t)}\right|\right]\\
&= \lim_{m\to\infty}\left(E\left[\sup_{t\geqslant\delta}\left|\frac{V(t)}{R(t)\vee 1}-\frac{\pi_0\,G_0(t)}{\pi_0\,G_0(t)+\pi_1\,G_1(t)}\right|\right]\right)\\
&\geqslant \lim_{m\to\infty}\sup_{t\geqslant\delta}\left|E\left[\frac{V(t)}{R(t)\vee 1}\right]-\frac{\pi_0\,G_0(t)}{\pi_0\,G_0(t)+\pi_1\,G_1(t)}\right|\geqslant 0,
\end{aligned}$$

where $E[V(t)/\{R(t)\vee 1\}]=\mathrm{FDR}(t)$. Applying the triangle inequality to this result and equation (14) implies that equation (13) holds, which completes the proof.

5.4. Proof of theorem 7

It follows from a minor modification to the proof of theorem 6 that with probability 1

$$\lim_{m\to\infty}\inf_{t\geqslant\delta}\{\widehat{\mathrm{pFDR}}_\lambda(t)-\mathrm{pFDR}(t)\}\geqslant 0$$

for each $\delta>0$. Therefore, for each $t>0$, it is straightforward to show that with probability 1

$$\lim_{m\to\infty}\left|\inf_{s\geqslant t}\{\widehat{\mathrm{pFDR}}_\lambda(s)\}-\inf_{s\geqslant t}\{\mathrm{pFDR}(s)\}\right|\geqslant 0.$$

This is just another way of writing that with probability 1

$$\lim_{m\to\infty}\{\hat{q}_\lambda(t)-q\text{-value}(t)\}\geqslant 0,$$

for each $t>0$. Now $\hat{q}_\lambda(t)\wedge 1$ and q-value(t) are both non-decreasing cadlag functions with range in [0, 1]. Moreover, they each show pointwise convergence to continuous limits where the limit of $\hat{q}_\lambda(t)\wedge 1$ dominates that of q-value(t). It follows by the same arguments proving the Glivenko–Cantelli theorem (Billingsley, 1968) and similar steps to the proof of theorem 6 that

$$\lim_{m\to\infty}\inf_{t\geqslant\delta}\{\hat{q}_\lambda(t)\wedge 1-q\text{-value}(t)\}\geqslant 0.$$

The desired result follows because $\lim_{m\to\infty}\inf_{t\geqslant\delta}\{\hat{q}_\lambda(t)-\hat{q}_\lambda(t)\wedge 1\}\geqslant 0$. If we use the alternative definition $\hat{q}_\lambda(t)=\inf_{s\geqslant t}\{\widehat{\Pr}_\lambda(H=0|P\leqslant s)\}=\inf_{s\geqslant t}\{\widehat{\mathrm{FDR}}_\lambda(s)\}$, then the proof is analogous.

6. Automatically choosing λ

The procedures that we have considered require the choice of a tuning parameter λ in the estimate $\hat{\pi}_0(\lambda)$, which has a clear bias–variance trade-off. Since the mean-squared error (MSE) lends a reasonable balance between bias and variance, our proposed automatic choice of λ is an estimate of the value that minimizes $E[\{\hat{\pi}_0(\lambda)-\pi_0\}^2]$. The procedure that we propose is summarized in the following algorithm.

Step 1: for some range of λ, say $\mathcal{R}=\{0, 0.05, 0.10, \ldots, 0.95\}$, calculate $\hat{\pi}_0(\lambda)$ as in Section 2.
Step 2: for each $\lambda\in\mathcal{R}$, form B bootstrap versions $\hat{\pi}_0^{*b}(\lambda)$ of the estimate, $b=1,\ldots,B$, by taking bootstrap samples of the p-values.

Step 3: for each $\lambda \in \mathcal{R}$, estimate its respective MSE as

$$\widehat{\mathrm{MSE}}(\lambda) = \frac{1}{B} \sum_{b=1}^{B} [\hat{\pi}_0^{*b}(\lambda) - \min_{\lambda' \in \mathcal{R}} \{\hat{\pi}_0(\lambda')\}]^2.$$

Step 4: set $\hat{\lambda} = \arg\min_{\lambda \in \mathcal{R}} \{\widehat{\mathrm{MSE}}(\lambda)\}$. Our overall estimate of π_0 is $\hat{\pi}_0 = \hat{\pi}_0(\hat{\lambda})$.

Motivation for this can be found in Storey (2002), where a similar method for automatically choosing λ in $\widehat{\mathrm{FDR}}_\lambda(t)$ was proposed. To assess the accuracy of the procedure proposed, we performed numerical experiments that are similar to those in Storey (2002) (the data are not shown). This procedure worked equally well in the sense that the MSE that was estimated from the bootstrap procedure on average had a minimum that was close to that of the true MSE.

The theoretical results in this work are for fixed λ and do not include an automatic choice of λ. It is straightforward to show that, for a fixed and finite $\mathcal{R}$, the results of theorems 4, 6 and 7 continue to hold with $\lambda = \hat{\lambda}$. Moreover, as long as the size of $\mathcal{R}$ grows at an appropriate rate (slower than m), it should be possible to show that the asymptotic results still hold. We leave the details to the interested reader. Genovese and Wasserman (2002b) have derived conservative asymptotic properties of the $\widehat{\mathrm{FDR}}_\lambda(t)$ estimates as $\lambda \to 1$ smoothly in m as $m \to \infty$, which is a more aggressive procedure than the procedure above.

7. Discussion

We have shown that $\widehat{\mathrm{FDR}}_\lambda(t)$ can be used

(a) to estimate FDRs for fixed significance regions,
(b) to estimate significance regions for fixed FDRs and
(c) to estimate FDRs simultaneously over the naturally occurring significance regions.

We have approached the problem of dependence by using asymptotic arguments. When weak dependence exists and the number of tests is large, then our methods are valid to use. Under independence and mixture model assumptions, Genovese and Wasserman (2002b) showed that $\widehat{\mathrm{FDR}}_\lambda(t)$ and $\widehat{\mathrm{pFDR}}_\lambda(t)$ converge to a tight Gaussian process. A functional delta method is employed to derive asymptotic validity results for procedures based on data-dependent thresholds. Genovese and Wasserman (2002b) also studied the realized proportion of false discoveries $V(t)/R(t)$ as a stochastic process. There are two main differences between our approach and theirs. The first is that they viewed these procedures in terms of data-dependent thresholding rules (which is very similar to the BH viewpoint), whereas we worked from the standpoint of estimating the FDR from a fixed threshold. The second difference is that they made stronger assumptions (independence, a mixture model and large m), whereas we only require independence of the null p-values for the finite sample results, and pointwise convergence of the empirical distribution functions of the null and alternative p-values for the asymptotic results.

We have shown in both a finite sample setting and an asymptotic setting that the goals of the 'fixed FDR level' and 'fixed significance threshold' approaches can be accomplished with the same estimates. For this family of estimates, this is a unification of the two approaches. In the asymptotic setting, we have also shown that the point estimates can be used to estimate conservatively the FDR over all significance regions simultaneously, which is equivalent to controlling the FDR at all levels simultaneously. This allowed us to prove the simultaneous conservative consistency of the estimated q-values, which is the first proof (to our knowledge) that FDR-adjusted p-values can be used in a simultaneous fashion with some guarantee that they still provide control of the FDR. The main tool that we have used in this work is to translate existing FDR methods into procedures involving empirical processes. This simplifies finite sample proofs and provides a framework for asymptotic results. Of future interest will be to study the optimality properties of our procedures, as well as to develop methods that exploit any known structure for a particular application.

8. Software

The methodology that has been described in this paper is available as a suite of S-PLUS and R functions at `http://faculty.washington.edu/~storey/qvalue/`.

Acknowledgements

We thank the Associate Editor and referees for their helpful comments. It was brought to our attention after this paper had been accepted for publication that an unpublished manuscript by Benjamini, Yekutieli and Krieger proves FDR control at a single level α for another related procedure. We thank them for pointing this out.

References

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* B, **57**, 289–300.
Benjamini, Y. and Hochberg, Y. (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Statist.*, **25**, 60–83.
Benjamini, Y. and Liu, W. (1999) A step-down multiple hypothesis procedure that controls the false discovery rate under independence. *J. Statist. Planng Inf.*, **82**, 163–170.
Benjamini, Y. and Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, **29**, 1165–1188.
Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.
Chow, Y. S., Robbins, H. E. and Siegmund, D. (1971) *Great Expectations: the Theory of Optimal Stopping*. Boston: Houghton Mifflin.
Durrett, R. (1996) *Probability: Theory and Examples*, 2nd edn. Belmont: Duxbury.
Efron, B., Tibshirani, R., Storey, J. D. and Tusher, V. (2001) Empirical Bayes analysis of a microarray experiment. *J. Am. Statist. Ass.*, **96**, 1151–1160.
Finner, H. and Roters, M. (2001) On the false discovery rate and expected type I errors. *Biometr. J.*, **43**, 985–1005.
Finner, H. and Roters, M. (2002) Multiple hypotheses testing and expected number of type I errors. *Ann. Statist.*, **30**, 220–238.
Genovese, C. and Wasserman, L. (2002a) Operating characteristics and extensions of the false discovery rate procedure. *J. R. Statist. Soc.* B, **64**, 499–517.
Genovese, C. and Wasserman, L. (2002b) False discovery rates. *Technical Report*. Department of Statistics, Carnegie Mellon University, Pittsburgh.
Morton, N. E. (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.*, **7**, 277–318.
Shaffer, J. (1995) Multiple hypothesis testing. *A. Rev. Psychol.*, **46**, 561–584.
Simes, R. J. (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**, 751–754.
Storey, J. D. (2001) The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.*, to be published.
Storey, J. D. (2002) A direct approach to false discovery rates. *J. R. Statist. Soc.* B, **64**, 479–498.
Storey, J. D. and Tibshirani, R. (2001) Estimating false discovery rates under dependence, with applications to DNA microarrays. *Technical Report 2001-18*. Department of Statistics, Stanford University, Stanford.
Yekutieli, D. and Benjamini, Y. (1999) Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Statist. Planng Inf.*, **82**, 171–196.