A NOTE ON ESTIMATING THE FALSE DISCOVERY RATE 
UNDER MIXTURE MODEL

BY FANG LIU AND SANAT K. SARKAR

Temple University

In this note, we focus on estimating the false discovery rate (FDR) of a multiple 
testing method with a common, non-random rejection threshold under a mixture 
model. We develop a new class of estimates of the FDR and prove that it is 
less conservatively biased than what is traditionally used. Numerical evidence is 
presented to show that the mean squared error (MSE) is also often smaller for 
the present class of estimates, especially in small-scale multiple testings. A similar 
class of estimates of the positive false discovery rate (pFDR) less conservatively 
biased than what is usually used is then proposed. When modified using our 
estimate of the pFDR and applied to a gene-expression data, Storey’s q-value 
method identifies a few more significant genes than his original q-value method 
at certain thresholds. The BH method of controlling the FDR is modified using 
our estimate of the FDR. The modified BH method is more powerful and controls 
the FDR in situations where the p-values have the same dependence structure as 
required by the BH method and, for lack of information about the proportion \( \pi_0 \) 
of true null hypotheses, it is reasonable to assume that \( \pi_0 \) is uniformly distributed 
over (0,1).

KEY WORDS: Binomial distribution; UMVU estimate; microarray analysis; Multiple 
testing; Storey’s estimation based approach.

1 Introduction

The false discovery rate (FDR) introduced by Benjamini and Hochberg (1995) is one of 
the most standard measures of error in multiple testing that is being currently used in 
a wide variety of applications. With \( R_n \) and \( V_n \) representing the numbers of rejections 
and false rejections of null hypotheses, respectively, while testing \( n \) null hypotheses, it 
is defined as the expected proportion of false rejections among all rejections, that is,

\[
FDR_n = E \left\{ \frac{V_n}{R_n} I(R_n > 0) \right\},
\]

The research is supported by NSF Grant DMS-0603868.
where \( I \) is the indicator function.

The idea of estimating the FDR was first considered by Storey (2002) who advocated using an estimation based approach to controlling it, which, as stated in Storey, Taylor, and Siegmund (2004), 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. Given the observed p-value \( p_i \) associated with the \( i \)th null hypothesis \( H_i \), Storey (2002) considered a single-step test rejecting \( H_i \) if \( p_i \leq \gamma \), for some common, non-random threshold \( \gamma \), and developed his approach through conservatively (upward) biased point estimates of its FDR under the following two-component mixture model:

**Mixture Model.** Let \( P_i \) denote the random p-value corresponding to \( p_i \) and \( H_i = 0 \) or 1 according as the corresponding null hypothesis is true or false. Let \((P_i, H_i), i = 1, \ldots, n,\) be independently and identically distributed with \( \Pr (P_i \leq u \mid H_i) = (1 - H_i)u + H_i F_1(u), \ u \in (0, 1), \) for some continuous cdf \( F_1(u) \), and \( \Pr(H_i = 0) = \pi_0 = 1 - \Pr(H_i = 1) \).

He proved that the FDR of the above single-step test under this mixture model is given by

\[
\text{FDR}_n(\gamma) = \frac{\pi_0 \gamma}{F(\gamma)} \Pr \{ R_n(\gamma) > 0 \}, \tag{1.1}
\]

where

\[
R_n(\gamma) = \sum_{i=1}^{n} I(P_i \leq \gamma) \quad \text{and} \quad F(\gamma) = \Pr (P_i \leq \gamma) = \pi_0 \gamma + (1 - \pi_0) F_1(\gamma), \tag{1.2}
\]

and then developed his approach to controlling the FDR based on the following class of conservatively biased point estimates of it:

\[
\hat{\text{FDR}}_{n, \lambda}(\gamma) = \frac{\hat{\pi}_0(\lambda) \gamma}{\hat{F}_n(\gamma)}, \quad \lambda \in [0, 1), \tag{1.3}
\]

where

\[
\hat{F}_n(\gamma) = \frac{1}{n} \max\{ R_n(\gamma), 1 \} \quad \text{and} \quad \hat{\pi}_0(\lambda) = \frac{n - R_n(\lambda)}{n(1 - \lambda)}. \tag{1.4}
\]

In particular, Storey (2002) modified the FDR by considering it conditional on at least one rejection, which he called the positive false discovery rate (pFDR), i.e,

\[
\text{pFDR}_n = E \left\{ \frac{V_n}{R_n} \mid R_n > 0 \right\} = \frac{\text{FDR}}{\Pr \{ R_n > 0 \}},
\]

and noted that for the single-step test with the threshold \( \gamma \) under the mixture model it is actually equal to

\[
\frac{\pi_0 \gamma}{F(\gamma)} = \Pr \{ H_1 = 0 \mid P_1 \leq \gamma \}, \tag{1.5}
\]
a ‘Bayesian p-value’. That motivated him to introduce a pFDR analog of the p-value, called the q-value, as a test-specific measure of significance in multiple testing. For a test with the observed p-value $p$, it is defined as

$$q(p) = \inf_{\gamma \geq p} \text{pFDR}(\gamma),$$

the minimum pFDR that can be achieved for a rejection region $(0, p]$ among all rejection regions containing it. To apply his so called q-value method, Storey (2002) proposed to estimate $q(p)$ using the following class of estimates of the pFDR $n(\gamma)$ that is derived from that for the FDR $n(\gamma)$ in (1.3):

$$\hat{\text{pFDR}}_{n,\lambda}(\gamma) = \frac{\hat{\text{FDR}}_{n,\lambda}(\gamma)}{1 - (1 - \gamma)^n}, \quad \lambda \in [0, 1). \tag{1.6}$$

Thus, the estimated q-value to be used in the q-value method, as Storey (2002) himself has suggested, is

$$\hat{q}(p) = \inf_{\gamma \geq p} \hat{\text{pFDR}}_{n,\lambda}(\gamma), \quad \lambda \in [0, 1).$$

Storey has made a connection between his estimation based approach and the approach taken by Benjamini and Hochberg (1995) to control the FDR. The approach of Benjamini and Hochberg (1995) 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 pre-chosen level. With $p_{(1)} \leq \cdots \leq p_{(n)}$ denoting the ordered p-values, Benjamini and Hochberg (1995) proposed the threshold $p_{(k)}$, where $k = \max \{1 \leq i \leq n : p_{(i)} \leq i\alpha/n\}$, if the maximum exists, otherwise, it is zero. This is the so-called BH method that is quite popular and controls the FDR in a wide variety of dependence situations [Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001) and Sarkar (2002)]. Storey (2002) pointed out that thresholding the p-values using

$$\gamma_\alpha(\hat{\text{FDR}}_{n,\lambda}) = \sup \left\{ 0 \leq \gamma \leq 1 : \hat{\text{FDR}}_{n,\lambda}(\gamma) \leq \alpha \right\} \tag{1.7}$$

can provide a wide class of BH type multiple testing methods that can potentially control the FDR. In fact, when $\lambda = 0$, it is equivalent to the BH method [see, also Storey, Taylor and Siegmund (2004)], and when $\lambda \in (0, 1)$, it can provide a valid FDR controlling method, of course only under independence of the p-values, if the estimate of $\pi_0$ is suitably modified [Storey, Taylor, and Siegmund (2004)].

Clearly, the conservative bias in estimates of FDR is a desirable property, as Storey argued, since controlling such an estimate would provide a conservative ‘control’ of the true FDR. Of course, less conservativeness is better in the sense of providing a less conservative, hence potentially more powerful, approach to controlling the FDR.
In this note, we show that there is indeed a class of estimates of the FDR\(_n(\gamma)\) under the mixture model that is less conservatively biased than the one in (1.3). We obtain this class by deriving the uniformly minimum variance unbiased estimate (UMVUE) of the FDR\(_n(\gamma)\) under the mixture model based on \(R_n(\gamma)\) with \(\pi_0\) assumed known, and then by plugging the same estimates of \(\pi_0\) as in (1.4) into this UMVUE. We prove that the resulting class of estimates of the FDR\(_n(\gamma)\) is less conservatively biased than (1.3). Our simulations indicate that the conservative bias is often noticeably reduced when our estimate of the FDR\(_n(\gamma)\) is used instead of Storey’s and \(n\) is not very large. The mean squared error (MSE) is also numerically seen to be often smaller for our class of estimates.

We use this new class of estimates of the FDR\(_n(\gamma)\) to modify Storey’s q-value method and the BH method. When applied to a gene-expression data consisting of 3170 genes, which is not that small, the modified q-value method picks up a few more significant genes than the original q-value method at certain thresholds, thus demonstrating the fact the modified q-value method can often make a difference compared to Storey’s original q-value method. With regard to our modification of the BH method, it is clearly more powerful, as it is now equivalent to thresholding a less conservatively biased estimate of the FDR\(_n(\gamma)\), and, as we prove, it is also a valid FDR controlling method in situations where, for lack of any prior information about \(\pi_0\), it is reasonable to assume that \(\pi_0\) is uniformly distributed over the (0,1) interval, and the \(p\)-values have the same type of dependence structure as in the case of the BH method.

We organize this note as follows. In Section 2, we develop the proposed class of estimates of the FDR\(_n(\gamma)\) and present the results of our simulation study examining the performances of these estimates relative to Storey’s in terms of the conservative bias and MSE. The corresponding class of estimates of the pFDR\(_n(\gamma)\) is also presented in this section. In Section 3, we present the modified q-value method and its application to a gene-expression data. The modified BH method with a proof of its FDR control are provided in Section 4.

## 2 Estimating the FDR\(_n(\gamma)\) and pFDR\(_n(\gamma)\)

In this section, we present our classes of estimates of the FDR\(_n(\gamma)\) and pFDR\(_n(\gamma)\) and the related results. First, we give the developments of these classes. Then, we report the results of a simulation study that we conducted to numerically investigate the extent of improvements we get by using our estimate \(\tilde{\text{FDR}}^*_{n,\lambda}(\gamma)\) over Storey’s estimate \(\text{FDR}_{n,\lambda}(\gamma)\) of the FDR\(_n(\gamma)\) in terms of the conservative bias and MSE.
2.1 The new estimates

We first prove the following identity that forms the basis of our class of estimates of the FDR \( n(\gamma) \) under the mixture model:

**Theorem 2.1.** Under the mixture model, the FDR of a single-step test rejecting each \( H_i \) if \( p_i \leq \gamma \), for some fixed \( \gamma \in [0, 1] \), is given by

\[
\text{FDR}_n(\gamma) = (n + 1)\pi_0 \gamma E\left( \frac{1}{\max\{R_n(\gamma), \frac{1}{n}\} + 1} \right). \tag{2.1}
\]

**Proof.** Let \( V_n(\gamma) = \sum_{i=1}^{n} I(p_i \leq \gamma, H_i = 0) \), the total number of false rejections. Then, we have

\[
\text{FDR}_n(\gamma) = E\left\{ \frac{V_n(\gamma)}{R_n(\gamma)} I(R_n(\gamma) > 0) \right\}
= E\left\{ \sum_{i=1}^{n} \sum_{r=1}^{n} \frac{1}{r} I(p_i \leq \gamma, H_i = 0, R_n(\gamma) = r) \right\}
= \sum_{i=1}^{n} \Pr(p_i \leq \gamma, H_i = 0) \sum_{r=1}^{n} \frac{1}{r} \Pr(R_{n-1}^{-i}(\gamma) = r - 1)
= \pi_0 \gamma E\left( \sum_{i=1}^{n} \frac{1}{R_{n-1}^{-i}(\gamma) + 1} \right), \tag{2.2}
\]

where \( R_{n-1}^{-i}(\gamma) = \sum_{j \neq i=1}^{n} I(p_j \leq \gamma) \).

The summation inside the expectation in the last line of (2.2) can be expressed in terms of \( R_n(\gamma) \). When \( R_n(\gamma) > 0 \), since \( R_{n-1}^{-i}(\gamma) = R_n(\gamma) - I(p_i \leq \gamma) \), this summation is equal to

\[
\sum_{i=1}^{n} \frac{1}{R_{n-1}^{-i}(\gamma) + 1} = \sum_{i: p_i \leq \gamma} \frac{1}{R_{n-1}^{-i}(\gamma) + 1} + \sum_{i: p_i > \gamma} \frac{1}{R_{n-1}^{-i}(\gamma) + 1}
= \frac{R_n(\gamma)}{R_n(\gamma)} + \frac{n - R_n(\gamma)}{R_n(\gamma) + 1} = \frac{n + 1}{R_n(\gamma) + 1}.
\]

When \( R_n(\gamma) = 0 \), since \( R_{n-1}^{-i}(\gamma) = 0 \), this summation is equal to \( n \). In other words, the right-hand side in (2.2) is equal to

\[
(n + 1)\pi_0 \gamma E\left( \frac{1}{\max\{R_n(\gamma), \frac{1}{n}\} + 1} \right), \tag{2.3}
\]

which is the right-hand side in (2.1). Thus, the theorem is proved.

**Remark 2.1.** Since \( R_n(\gamma) \sim \text{Binomial}(n, F(\gamma)) \) and \( R_{n-1}^{-i}(\gamma) \sim \text{Binomial}(n - 1, F(\gamma)) \), for each \( i \), alternatively one can verify the correctness of expression (2.2) for \( \text{FDR}_n(\gamma) \).
and that it is also equal to (2.3) from the following result which might of independent interest. A proof of this result is given in Appendix.

**Result 1.** Let $X_n \sim \text{Binomial } (n, \theta), 0 < \theta < 1$. Then,

$$E \left( \frac{n}{X_{n-1} + 1} \right) = E \left( \frac{n + 1}{\max\{X_n, \frac{1}{n} \} + 1} \right) = \frac{1}{\theta} \{1 - (1 - \theta)^n\}.$$

Thus,

$$\hat{\text{FDR}}_n(\gamma | \pi_0) = \frac{(n + 1) \pi_0 \gamma}{\max\{R_n(\gamma), \frac{1}{n} \} + 1}$$

is an unbiased estimate of $\text{FDR}_n(\gamma)$ with a known $\pi_0$. In fact, since $R_n(\gamma)$ is complete sufficient for $F(\gamma)$ with a known $\pi_0$ [see, for example, Casella and Berger (2001)], (2.4) is actually the best unbiased estimate in the sense of being the UMVUE with a known $\pi_0$ based on $R_n(\gamma)$. When $\pi_0$ is unknown, though one may consider using any reasonable estimate of it to incorporate into (2.4), we choose the estimates in (1.4) that Storey (2002) originally considered and propose our new class of estimates of the $\text{FDR}_n(\gamma)$ as follows:

$$\hat{\text{FDR}}_{n, \lambda}^*(\gamma) = \frac{(n + 1) \hat{\pi}_0(\lambda) \gamma}{\max\{R_n(\gamma), \frac{1}{n} \} + 1}$$

$$= \frac{(n + 1) \left[ n - R_n(\lambda) \right] \gamma}{n(1 - \lambda) \left[ \max\{R_n(\gamma), \frac{1}{n} \} + 1 \right]}, \lambda \in [0, 1). \quad (2.5)$$

We will now prove that $\hat{\text{FDR}}_{n, \lambda}^*(\gamma)$ is less conservatively biased than $\hat{\text{FDR}}_{n, \lambda}(\gamma)$.

**Theorem 2.2.** For any fixed $\lambda \in [0, 1)$ and $\gamma \in (0, 1)$,

$$\text{FDR}_n(\gamma) \leq E \left( \hat{\text{FDR}}_{n, \lambda}^*(\gamma) \right) \leq E \left( \hat{\text{FDR}}_{n, \lambda}(\gamma) \right). \quad (2.6)$$

**Proof.** Let us suppress the subscript $n$ in $R_n$.

$$E \left( \hat{\text{FDR}}_{n, \lambda}^*(\gamma) \right) = (n + 1) \gamma E \left( \hat{\pi}_0(\lambda) E \left( \frac{1}{\max\{R(\gamma), \frac{1}{n} \} + 1} \left| R(\lambda) \right. \right) \right). \quad (2.7)$$

Note that, given $R(\lambda)$,

$$R(\gamma) \sim \text{Binomial } \left( \frac{F(\gamma)}{F(\lambda)}, R(\lambda) \right), \text{ when } \gamma \leq \lambda,$$

and

$$n - R(\gamma) \sim \text{Binomial } \left( \frac{1 - F(\gamma)}{1 - F(\lambda)}, n - R(\lambda) \right), \text{ when } \gamma \geq \lambda.$$
Since a binomial random variable is stochastically increasing in the number of trials, the distribution of $R(\gamma)$ is stochastically increasing in $R(\lambda)$. Therefore, the conditional expectation, given $R(\lambda)$, of any decreasing function of $R(\gamma)$, in particular $1/[\max\{R(\gamma), \frac{1}{n}\} + 1]$, is a decreasing function of $R(\lambda)$. As $\hat{\pi}_0(\lambda)$ is also a decreasing function of $R(\lambda)$, it is positively correlated with this conditional expectation, which implies that the outer expectation in (2.7) is greater than or equal to

$$E\{\hat{\pi}_0(\lambda)\} E\left\{E \left( \frac{1}{\max\{R(\gamma), \frac{1}{n}\} + 1} \right | R(\lambda) \right\} = E\{\hat{\pi}_0(\lambda)\} E\left( \frac{1}{\max\{R(\gamma), \frac{1}{n}\} + 1} \right),$$

since

$$E\{\pi_0(\lambda)\} = \frac{1 - F(\lambda)}{1 - \lambda} = \pi_0 + (1 - \pi_0) \frac{1 - F(\lambda)}{1 - \lambda} \geq \pi_0.$$

Thus, we have

$$E\left( \hat{\text{FDR}}_{n,\lambda}(\gamma) \right) \geq (n + 1) \gamma \pi_0 E\left( \frac{1}{\max\{R(\gamma), \frac{1}{n}\} + 1} \right) = \text{FDR}_n(\gamma),$$

the first inequality in (2.6). The second inequality follows from the fact that $\hat{\text{FDR}}_{n,\lambda}(\gamma) \leq \text{FDR}_{n,\lambda}(\gamma)$ with probability one.

As in Storey (2002), we now propose the following class of estimates of $p\text{FDR}_n(\gamma)$ based on what we propose for $\hat{\text{FDR}}_{n,\lambda}(\gamma)$ in (2.5) as follows:

$$p\text{FDR}_{n,\lambda}(\gamma) = \frac{\hat{\text{FDR}}_{n,\lambda}(\gamma)}{1 - (1 - \gamma)^n} = \frac{(n + 1) [n - R_n(\lambda)] \gamma}{n (1 - \lambda) \{\max\{R_n(\gamma), \frac{1}{n}\} + 1\} \{1 - (1 - \gamma)^n\}}, \quad \lambda \in [0, 1). \quad (2.8)$$

We have the following result on the pFDR that is similar to and follows easily from Theorem 2.2.

**Corollary to Theorem 2.2.** For any fixed $\lambda \in [0, 1)$ and $\gamma \in (0, 1)$,

$$p\text{FDR}_n(\gamma) \leq E\left( p\text{FDR}_{n,\lambda}(\gamma) \right) \leq E\left( \hat{p}\text{FDR}_{n,\lambda}(\gamma) \right). \quad (2.9)$$

**Remark 2.2.** It is important to note the difference between the approaches taken here and in other papers, like Dalmasso, Broet, and Moreau (2005) and Pounds and Cheng (2006), towards reducing the conservative bias of the estimate $\hat{\text{FDR}}_{n,\lambda}(\gamma)$ of
FDR_n(γ). We start with the best unbiased estimate of FDR_n(γ) assuming π₀ known before using a conservatively biased point estimate of π₀, while the others have started with the estimate π₀γ/\hat{F}_n(γ) when π₀ is known that is already conservatively biased before developing an estimate of π₀ better than Storey’s, thus attempting to improve FDR_n,λ(γ). It would be interesting to see what happens if one uses the type of estimates of π₀ suggested in these other papers in our UMVUE, instead of using the same estimate as Storey originally used. While we expect that this would produce better performing estimates of FDR_n(γ) than ours when compared to FDR_n,λ(γ), we have to rely on simulations to see it, as seeing it theoretically seems difficult. The only reason we have kept Storey’s estimate of π₀ while replacing the π₀ in the UMVUE is that we can theoretically see why the resulting estimate is in fact less conservatively biased than (1.3). As seen numerically in the next sub-section, such an estimate also often outperforms (1.3) in terms of the MSE, especially in small-scale multiple testings.

2.2 A numerical study

We conducted a simulation study to investigate the reductions in bias that our estimates of FDR_n(γ) and pFDR_n(γ) offer compared to Storey’s estimates. We only report the results related to the FDR in this section, as similar results can be expected for the pFDR.

We generated n independent pairs of observations (X_i, Z_i), i = 1, … , n, where Z_i is the outcome of a Bernoulli experiment with 1 − π₀ as the success probability and X_i ∼ N(µ_i, 1), with µ_i = δZ_i for a fixed δ > 0. For each X_i, we calculated the p-value p_i = 1 − Φ(X_i), where Φ is the cdf of N(0, 1), and considered testing µ_i = 0 against µ_i = δ based on the rejection region \{p_i ≤ γ\}. This was repeated 20,000 times before simulating the bias and MSE for each estimate and for each combination of values of π₀, γ, n, and δ selected from the range 0.1, 0.2, … , 0.9, in increments of 0.1, for π₀, the two values 0.001 and 0.01 for γ, the two values 100 and 500 for n, and the value 1 for δ. For λ, we considered the values 0.5, which Storey (2002) originally used, and 0.95, which is often used in practice. The formula (1.1) was used to compute the true FDR_n for the bias and MSE calculations.

The numerical findings are summarized in Tables 1 (for λ = 0.5) and 2 (for λ = 0.95). The proposed estimate of FDR_n(γ) is seen to have better performance than Storey’s both in terms of the conservative bias and MSE, with the improvement in terms of the bias appearing quite significant particularly when π₀ is large.
Table 1: Numerical comparison between $\widehat{\text{FDR}}_{n,\lambda}(\gamma)$ and $\widehat{\text{FDR}}^*_n(\gamma)$ in terms of the bias and MSE when $\lambda = 0.5$

\begin{table}[h]
\centering
\begin{tabular}{cccccccccc}
\hline
$\pi_0$ & FDR & $\widehat{\text{FDR}}_n(t)$ & $\widehat{\text{FDR}}_n(\gamma)$ & $\widehat{\text{FDR}}^*_n(\gamma)$ & $\widehat{\text{FDR}}^*_n(\gamma)$ & FDR & $\widehat{\text{FDR}}^*_n(\gamma)$ & $\widehat{\text{FDR}}^*_n(\gamma)$ & $\widehat{\text{FDR}}^*_n(\gamma)$ \\
\hline
\multicolumn{10}{c}{$n = 500$} \\
$\gamma = 0.01$ & & & & & & & & & \\
0.1 & 0.012 & 0.03511 & 0.03405 & 0.00131 & 0.00123 & 0.012 & 0.04140 & 0.03441 & 0.00268 & 0.00166 \\
0.2 & 0.026 & 0.03515 & 0.03361 & 0.00138 & 0.00126 & 0.026 & 0.04460 & 0.03407 & 0.00384 & 0.00201 \\
0.3 & 0.044 & 0.03533 & 0.03305 & 0.00150 & 0.00132 & 0.044 & 0.04981 & 0.03361 & 0.00630 & 0.00266 \\
0.4 & 0.067 & 0.03572 & 0.03233 & 0.00175 & 0.00146 & 0.067 & 0.05797 & 0.03291 & 0.01102 & 0.00388 \\
0.5 & 0.098 & 0.03665 & 0.03144 & 0.00229 & 0.00180 & 0.098 & 0.07073 & 0.03212 & 0.02029 & 0.00670 \\
0.6 & 0.140 & 0.03861 & 0.03019 & 0.00354 & 0.00257 & 0.140 & 0.09158 & 0.03066 & 0.03709 & 0.01131 \\
0.7 & 0.202 & 0.04299 & 0.02834 & 0.00692 & 0.00464 & 0.196 & 0.12113 & 0.02819 & 0.06396 & 0.02084 \\
0.8 & 0.302 & 0.05457 & 0.02551 & 0.01845 & 0.01093 & 0.283 & 0.16329 & 0.02352 & 0.10229 & 0.03732 \\
0.9 & 0.493 & 0.09397 & 0.01998 & 0.00883 & 0.00409 & 0.418 & 0.20844 & 0.01485 & 0.13695 & 0.06374 \\
0.95 & 0.672 & 0.16051 & 0.01628 & 0.29442 & 0.10346 & 0.513 & 0.22469 & 0.00946 & 0.14044 & 0.07928 \\
\hline
\multicolumn{10}{c}{$n = 100$} \\
$\gamma = 0.001$ & & & & & & & & & \\
0.1 & 0.006 & 0.02124 & 0.01735 & 0.00068 & 0.00040 & 0.005 & 0.02248 & 0.01413 & 0.00069 & 0.00033 \\
0.2 & 0.013 & 0.02322 & 0.01722 & 0.00103 & 0.00049 & 0.010 & 0.02354 & 0.01347 & 0.00079 & 0.00037 \\
0.3 & 0.023 & 0.02616 & 0.01699 & 0.00169 & 0.00066 & 0.017 & 0.02433 & 0.01264 & 0.00088 & 0.00042 \\
0.4 & 0.035 & 0.03094 & 0.01669 & 0.00301 & 0.00099 & 0.024 & 0.02478 & 0.01162 & 0.00093 & 0.00047 \\
0.5 & 0.051 & 0.04026 & 0.01736 & 0.00507 & 0.00188 & 0.032 & 0.02511 & 0.01033 & 0.00094 & 0.00053 \\
0.6 & 0.074 & 0.05334 & 0.01692 & 0.01087 & 0.00323 & 0.042 & 0.02536 & 0.00989 & 0.00089 & 0.00059 \\
0.7 & 0.108 & 0.07214 & 0.01626 & 0.01854 & 0.00603 & 0.053 & 0.02614 & 0.00709 & 0.00077 & 0.00062 \\
0.8 & 0.160 & 0.09688 & 0.01439 & 0.02824 & 0.01086 & 0.065 & 0.01808 & 0.00507 & 0.00057 & 0.00061 \\
0.9 & 0.246 & 0.11543 & 0.01022 & 0.03189 & 0.01779 & 0.079 & 0.01260 & 0.00262 & 0.00033 & 0.00053 \\
0.95 & 0.310 & 0.11072 & 0.00610 & 0.02601 & 0.02025 & 0.087 & 0.00887 & 0.00135 & 0.00022 & 0.00044 \\
\hline
\end{tabular}
\end{table}
Table 2: Numerical comparison between $\hat{FDR}_{n,\lambda}(\gamma)$ and $\hat{FDR}^*_n(\gamma)$ in terms of the bias and MSE when $\lambda = 0.95$

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.012</td>
<td>0.00942</td>
<td>0.00007</td>
<td>0.0021</td>
<td>0.0019</td>
<td>0.012</td>
<td>0.01230</td>
<td>0.00910</td>
</tr>
<tr>
<td>0.2</td>
<td>0.026</td>
<td>0.00976</td>
<td>0.00034</td>
<td>0.00031</td>
<td>0.026</td>
<td>0.01531</td>
<td>0.00909</td>
<td>0.00255</td>
</tr>
<tr>
<td>0.3</td>
<td>0.044</td>
<td>0.01034</td>
<td>0.00056</td>
<td>0.00050</td>
<td>0.044</td>
<td>0.02037</td>
<td>0.00915</td>
<td>0.00529</td>
</tr>
<tr>
<td>0.4</td>
<td>0.067</td>
<td>0.01119</td>
<td>0.00095</td>
<td>0.00083</td>
<td>0.067</td>
<td>0.02837</td>
<td>0.00913</td>
<td>0.01079</td>
</tr>
<tr>
<td>0.5</td>
<td>0.098</td>
<td>0.01252</td>
<td>0.00168</td>
<td>0.00142</td>
<td>0.097</td>
<td>0.04001</td>
<td>0.00836</td>
<td>0.02041</td>
</tr>
<tr>
<td>0.6</td>
<td>0.140</td>
<td>0.01512</td>
<td>0.00332</td>
<td>0.00270</td>
<td>0.138</td>
<td>0.06107</td>
<td>0.00827</td>
<td>0.04145</td>
</tr>
<tr>
<td>0.7</td>
<td>0.202</td>
<td>0.02077</td>
<td>0.00754</td>
<td>0.00574</td>
<td>0.196</td>
<td>0.09230</td>
<td>0.00798</td>
<td>0.07868</td>
</tr>
<tr>
<td>0.8</td>
<td>0.302</td>
<td>0.03455</td>
<td>0.00707</td>
<td>0.00124</td>
<td>0.283</td>
<td>0.13888</td>
<td>0.00643</td>
<td>0.14050</td>
</tr>
<tr>
<td>0.9</td>
<td>0.493</td>
<td>0.07789</td>
<td>0.00580</td>
<td>0.00491</td>
<td>0.418</td>
<td>0.19080</td>
<td>0.00286</td>
<td>0.21361</td>
</tr>
<tr>
<td>0.95</td>
<td>0.672</td>
<td>0.14930</td>
<td>0.00688</td>
<td>0.32371</td>
<td>0.12495</td>
<td>0.513</td>
<td>0.21397</td>
<td>0.00183</td>
</tr>
</tbody>
</table>

3 The modified q-value method and an application

Now that we have an estimate of the pFDR less conservatively biased than what Storey (2002) used while estimating his q-value, it would be natural to use it to modify this estimated q-value. The modified estimated q-value of an observed p-value $p$ is

$$\hat{q}^*_\lambda(p) = \inf_{\gamma \geq p} \{ \hat{FDR}^*_{n,\lambda}(\gamma) \}. \quad (3.1)$$

We used the breast cancer data of Hedenfalk et al. (2001) available at http://www.nejm.org/general/content/supplemental/hedenfalk/index.html and applied both the original and modified q-value methods to see if any additional significant genes are picked up by the modified q-value method. The data consist of 3,226 genes with $n_1 = 7$ BRCA1 arrays, $n_2 = 8$ BRCA2 arrays and $n_3 = 7$ sporadic tumors, which have been analyzed in numerous other articles, often using only the BRCA1 and BRCA2 groups and sometimes using all three groups. Here, we considered both these cases, and used a permutation
Table 3: Number of significant genes when BRCA1 and BRCA2 groups are compared using a permutation \(t\)-test

<table>
<thead>
<tr>
<th>(\lambda)</th>
<th>threshold</th>
<th>Storey</th>
<th>New</th>
<th>Additional Significant Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.03</td>
<td>76</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>159</td>
<td>161</td>
<td>M-phase phosphoprotein 4 (clone 785816), v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1 (clone 273435)</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>229</td>
<td>231</td>
<td>actin, beta (clone 34357), ligase I (clone 29627)</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>314</td>
<td>314</td>
<td></td>
</tr>
<tr>
<td>Smooth</td>
<td>0.03</td>
<td>76</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Spline</td>
<td>0.05</td>
<td>162</td>
<td>162</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>233</td>
<td>233</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>319</td>
<td>319</td>
<td></td>
</tr>
<tr>
<td>Bootstrap</td>
<td>0.03</td>
<td>76</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>161</td>
<td>162</td>
<td>acyl-Coenzyme A oxidase (clone 210862)</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>231</td>
<td>232</td>
<td>thyroid hormone receptor interactor 7 (clone 781704)</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>317</td>
<td>319</td>
<td>KIAA0303 protein (clone 768031), ESTs, Weakly similar to putative p150 (clone 137417)</td>
</tr>
</tbody>
</table>

\(t\) test to compare BRCA1 and BRCA2 and permutation \(F\) test to compare the three groups. As in Storey and Tibshirani (2003), if any gene had one or more measurement (log\(_2\) expression value) exceeding 20, then that gene was eliminated. This left \(n = 3170\) genes for the permutation \(t\) test and \(n = 3169\) genes for the permutation \(F\) test.

For both the \(t\) and \(F\) tests, we calculated the \(q\)-values using Storey’s original estimate as well as our estimate, with \(\lambda\) being decided in three different ways: (i) Fixed \(\lambda\) method: \(\lambda = 0.5\) [Storey (2002)], (ii) Smoother method [Storey and Tibshirani (2003)], and (iii) Bootstrap method [Storey, Taylor, and Siegmund (2004)].

The results of the permutation \(t\) test are shown in Table 3. As expected, the new \(q\)-value method picks up at least the same number of significant genes as Storey’s method. For example, with the fixed \(\lambda = 0.5\), by thresholding the \(q\)-value at 0.03, 0.05 and 0.07 we find 76, 161 and 231 significant genes, respectively, using the modified \(q\)-value method, while these numbers are 76, 159 and 229, respectively, for the original \(q\)-value method. The two additional genes picked up by the new \(q\)-value method are M-phase phosphoprotein 4 (Symbol:ILF3, clone 785816) gene and v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1 (Symbol:YES1, clone 273435) gene. These two genes are
Table 4: Number of significant genes when all three groups are compared using a permutation $F$ test

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>threshold</th>
<th>Storey</th>
<th>New</th>
<th>Additional Significant Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.05</td>
<td>77</td>
<td>77</td>
<td>EphB4 (clone 75009), aldo-keto reductase family 1 gene (clone 33008), ataxia-telangiectasia group D-associated protein gene (clone 377275), actin (clone 34357), chromosome 14 open reading frame 2 (clone 110772), fibrogenic lymphokine (clone 366580), tyrosine kinase 2 (clone 756452), synuclein (clone 40764)</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>180</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>293</td>
<td>300</td>
<td>EphB4 (clone 75009), aldo-keto reductase family 1 gene (clone 33008), ataxia-telangiectasia group D-associated protein gene (clone 377275), actin (clone 34357), chromosome 14 open reading frame 2 (clone 110772), fibrogenic lymphokine (clone 366580), tyrosine kinase 2 (clone 756452), synuclein (clone 40764)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Smooth 0.05 124 138 KIAA0265 protein (clone 140716), adenylosuccinate lyase (clone 813280), Ras suppressor protein 1 (clone 687397), gamma-aminobutyric acid (GABA) A receptor, pi (clone 563598), ESTs, Weakly similar to trg [R.norvegicus] (clone 246749), ESTs, Weakly similar to B0495.6 [C.elegans] (clone 144926), protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65) (clone 205490), MAD (mothers against decapentaplegic, Drosophila) homolog 4 (clone 140827), cAMP responsive element binding protein 1 (clone 362332), DKFZP434D1335 protein (clone 161195), PDZ domain containing guanine nucleotide exchange factor(GEF)1 (clone 824895), lung resistance-related protein (clone 591281), cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4) (clone 291057), ESTs, Weakly similar to ORF YDL040c [S.cerevisiae] (clone 196866)</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>212</td>
<td>212</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>381</td>
<td>385</td>
<td>thyroid hormone receptor interactor 6 (clone 811108), sarcoma amplified sequence (clone 361692), transcription elongation factor A (SII), 1 (clone 257458), LIM and SH3 protein 1 (clone 44050)</td>
</tr>
<tr>
<td>0.05</td>
<td>138</td>
<td>139</td>
<td></td>
<td>ribosomal protein L38 (clone 839594)</td>
</tr>
<tr>
<td>0.07</td>
<td>218</td>
<td>218</td>
<td></td>
<td>mitogen-activated protein kinase kinase 4 (clone 726147)</td>
</tr>
<tr>
<td>0.1</td>
<td>392</td>
<td>393</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12
already found to be related with breast cancer in the literature. Vumbaca et al. (2007) found that ILF3 promotes breast cancer growth and angiogenesis in vivo, and Sommerra et al. (2005) stated that changes in the expression of YES1 have been associated with the aggressiveness of human breast cancer cells. There is a possibility that these two genes are only related to one of the BRCA1 and BRCA2 mutation, which needs to be further investigated biologically.

Similarly, for the permutation $F$ test, the number of significant genes identified by the new q-value method is more than those found by Storey’s original q-value method, as shown in Table 4. When we set the cutoff at 0.1 with $\lambda = 0.5$, we find 7 more significant genes with the new q-value method than Storey’s original q-value method.

### 4 The modified BH method

We modify the BH method in this section by redefining it in terms of the new threshold

\[
\gamma_\alpha(\hat{\text{FDR}}_{n,\lambda}^*) = \sup \left\{ 0 \leq \gamma \leq 1 : \hat{\text{FDR}}_{n,\lambda}^*(\gamma) \leq \alpha \right\}.
\]

Clearly, this is more powerful than the original BH method, since $\gamma_\alpha(\hat{\text{FDR}}_{n,\lambda=0}^*) \leq \gamma_\alpha(\hat{\text{FDR}}_{n,\lambda=0})$, and, as stated in the following theorem, it is often a valid FDR-controlling method.

**Theorem 4.1** Assume that $\Pr(P_i \leq u) \leq u$, for each $P_i$ when the corresponding null hypothesis is true, and the p-values are independent or positively dependent in the following sense:

\[
E\{\phi(P_1, \ldots, P_n) \mid P_i \leq u\} \uparrow u \in (0, 1),
\]

for each $P_i$ that corresponds to a null hypothesis and any increasing (coordinatewise) function $\phi$. Also assume that the proportion of true null hypotheses, $\pi_0$, is random and distributed as $U(0, 1)$. Then, the FDR of the method that rejects $H_i$ if $p_i \leq \gamma_\alpha(\hat{\text{FDR}}_{n,\lambda=0}^*)$ is less than or equal to $\alpha$.

Before we proceed to explain how to prove this theorem, it is important to keep in mind that the FDR here is defined as

\[
\text{FDR} = \int_0^1 \text{FDR}(\pi_0) \, d\pi_0,
\]

where FDR($\pi_0$) is the conditional FDR given a $\pi_0$.

This theorem can be proved somewhat in the same way, that is, using the following lemma, as in proving the FDR control of the BH method with a fixed $\pi_0$ under the same
type of positive dependency as assumed in this theorem, once we make the same kind of arguments as in Lemma 1 or 2 of Storey, Taylor, and Siegmund (2004) and see that thresholding the p-values at $\gamma_\alpha(\hat{\text{FDR}}_{n,\lambda=0}^*)$ is equivalent to the following stepup test: Reject $H_i$ if $p_i \leq p(l)$, where

$$l = \max \left\{ 1 \leq i \leq n : p(i) \leq \frac{(i+1)\alpha}{n+1} \right\},$$

(4.4)

provided the maximum exists, otherwise, accept all null hypotheses. Several versions of a proof of this lemma can be seen in the literature [e.g., Benjamini and Yekutieli (2001), Sarkar (2002) and Blanchard and Roquain (2008)], so we ignore proving it here.

**Lemma 4.1.** Consider a stepup test with the critical values $\alpha_1 \leq \cdots \leq \alpha_n$, that is, reject $H_i$ if $p_i \leq p(r)$, where

$$r = \max \left\{ 1 \leq i \leq n : p(i) \leq \alpha_i \right\},$$

(4.5)

provided the maximum exists, otherwise, accept all null hypotheses. Given $n_0$ true null hypotheses, the FDR of this method is less than or equal to $n_0\alpha_1$ under the positive dependency assumed in Theorem 4.1 if $\alpha_i/i$ is non-increasing in $i$.

**Proof of Theorem 4.1.** From Lemma 4.1, we see that the FDR of the method in this theorem satisfies the following:

$$\text{FDR} = \int_0^1 \text{FDR}(\pi_0) \, d\pi_0 \leq \frac{2n\alpha}{n+1} \int_0^1 \pi_0 \, d\pi_0 \leq \alpha.$$

(4.6)

Thus, the theorem is proved.

5 Concluding remarks

This note presents some interesting, new results on the FDR. For estimating the FDR of a single-step test with a common, fixed rejection threshold under the mixture model, it is shown that there is a class of estimates less conservatively biased than what Storey (2002) considered under the same setup while introducing his estimation based approach to controlling the FDR. These new estimates can improve the q-value method. Also, these estimates suggest a more powerful alternative to the BH method in commonly encountered situations where the proportion of true null hypotheses can reasonably be assumed to follow a uniform distribution over the (0,1) interval.

Storey, Taylor, and Siegmund (2004) proved some asymptotic optimum properties of $\hat{\text{FDR}}_{n,\lambda}(\gamma)$ and the related q-value. Similar results can be obtained in terms of our estimate, which we intend to do in a different communication.
6 Appendix

Proof of Result 1.

\[
E \left( \frac{n}{X_{n-1} + 1} \right) = \sum_{j=0}^{n-1} \frac{n}{j+1} \binom{n-1}{j} \theta^j (1 - \theta)^{n-1-j} = \sum_{j=0}^{n-1} \binom{n}{j+1} \theta^j (1 - \theta)^{n-1-j} = \sum_{j=1}^{n} \binom{n}{j} \theta^{j-1} (1 - \theta)^{n-j} = \frac{1}{\theta} \{1 - (1 - \theta)^n\}.
\]

(6.1)

\[
E \left( \frac{n + 1}{\max \{X_n, \frac{1}{n} \} + 1} \right) = n(1 - \theta)^n + \sum_{j=1}^{n} \frac{n+1}{j+1} \binom{n}{j} \theta^j (1 - \theta)^{n-j} = n(1 - \theta)^n + \sum_{j=1}^{n} \frac{n+1}{j+1} \theta^j (1 - \theta)^{n-j} = n(1 - \theta)^n + \sum_{j=2}^{n+1} \binom{n+1}{j} \theta^{j-1} (1 - \theta)^{n-j+1} = n(1 - \theta)^n + \frac{1}{\theta} \{1 - (1 - \theta)^{n+1} - (n+1)\theta(1 - \theta)^n\} = \frac{1}{\theta} \{1 - (1 - \theta)^n\}.
\]

(6.2)

Thus, the result is proved.

References


