

Parametric Mixture Models for Estimating the Proportion of True Null Hypotheses and Adaptive Control of FDR

Ajit C. Tamhane¹ and Jiaxiao Shi²

Northwestern University

Abstract: Estimation of the proportion or the number of true null hypotheses is an important problem in multiple testing, especially when the number of hypotheses is large. Wu, Guan and Zhao (2006) found that nonparametric approaches are too conservative. We study two parametric mixture models (normal and beta) for the distributions of the test statistics or their p -values to address this problem. The components of the mixture are the null and alternative distributions with mixing proportions π_0 and $1 - \pi_0$, respectively, where π_0 is the unknown proportion to be estimated. The normal model assumes that the test statistics from the true null hypotheses are i.i.d. $N(0, 1)$ while those from the alternative hypotheses are i.i.d. $N(\delta, 1)$ with $\delta \neq 0$. The beta model assumes that the p -values from the null hypotheses are i.i.d. $U[0, 1]$ and those from the alternative hypotheses are i.i.d. $\text{Beta}(a, b)$ with $a < 1 < b$. All parameters are assumed to be unknown. Three methods of estimation of π_0 are developed for each model. The methods are compared via simulation with each other and with Storey’s (2002) nonparametric method in terms of the bias and mean square error of the estimators of π_0 and the achieved FDR. Robustness of the estimators to the model violations is also studied by generating data from other models. For the normal model, the parametric methods perform better compared to Storey’s method with the EM method (Dempster, Laird and Rubin 1977) performing best overall when the assumed model holds; however, it is not very robust to significant model violations. For the beta model, the parametric methods do not perform as well because of the difficulties of estimation of parameters, and Storey’s nonparametric method turns out to be the winner in many cases. Therefore the beta model is not recommended for use in practice. An example is given to illustrate the methods.

Contents

1	Introduction	305
2	Normal Model	307
2.1	Hsueh, Chen, and Kodell (HCK) Method	307
2.2	Test Statistics (TS) Method	308
2.3	EM Method	310
3	Beta Model	310
3.1	Hsueh, Chen, and Kodell (HCK) Method	311
3.2	Test Statistics (TS) Method	311

¹Department of IEMS, Northwestern University, Evanston, IL 60208

²Department of Statistics, Northwestern University, Evanston, IL 60208

AMS 2000 subject classifications: Primary 62F10; secondary 62F12

Keywords and phrases: Beta model; Bias-correction; EM method; Mixture model; False discovery rate (FDR); Least squares method; Maximum likelihood method; Normal model; p -values.

1	3.3 EM Method	312	1
2	4 Adaptive Control of FDR	313	2
3	5 Simulation Results	314	3
4	5.1 Simulation Results for Normal Model	315	4
5	5.2 Simulation Results for Beta Model	319	5
6	6 Example	320	6
7	7 Concluding Remarks	322	7
8	Acknowledgment	323	8
9	Appendix A	323	9
10	References	325	10

11
12
13 **1. Introduction**

14
15 Suppose that m null hypotheses, H_{01}, \dots, H_{0m} , are to be tested against alternatives,
16 H_{11}, \dots, H_{1m} . Let X_1, \dots, X_m be the test statistics and p_1, \dots, p_m their p -values.
17 Throughout we assume that the X_i 's and hence the p_i 's are mutually independent.
18 Suppose that some unknown number m_0 of the hypotheses are true and $m_1 =$
19 $m - m_0$ are false. We wish to estimate m_0 or equivalently the proportion $\pi_0 = m_0/m$
20 of the true hypotheses based on the X_i 's or equivalently the p_i 's. The estimate \hat{m}_0 is
21 useful for devising more powerful adaptive multiple comparison procedures (MCPs)
22 to control an appropriate type I error rate, e.g., the familywise error rate (FWE)
23 (Hochberg and Tamhane 1987) in the Bonferroni procedure or the false discovery
24 rate (FDR) in the Benjamini and Hochberg (1995) procedure. These procedures
25 normally use the total number m as a conservative upper bound on the number of
26 true hypotheses. Adaptive procedures based on \hat{m}_0 are especially useful in large-
27 scale multiplicity testing problems arising in microarray data involving m of the
28 order of several thousands.

29 A number of methods have been proposed for estimating m_0 starting with
30 Schweder and Spjøtvoll (1982); see, e.g., Hochberg and Benjamini (1990), Benjamini
31 and Hochberg (2000), Turkheimer, Smith and Schmidt (2001), Storey (2002), Storey
32 et al. (2004) and Langaas et al. (2005). Many of these methods reject the p -values
33 that differ significantly from the null $U[0, 1]$ distribution as non-null and exclude
34 them from the estimation process. Different formal or graphical tests are used for
35 this purpose. For example, consider Storey's (2002) method with a fixed λ -level
36 test for a sufficiently large λ (e.g., $\lambda = 0.5$) to reject any p -value $\leq \lambda$ as non-null.
37 (It should be noted that λ is not fixed but is a tuning parameter whose value is
38 determined from the data to minimize the mean square error of the estimate of π_0
39 using bootstrap.) Let $N_r(\lambda) = \#\{p_i \leq \lambda\}$ denote the number of rejected hypotheses
40 and $N_a(\lambda) = \#\{p_i > \lambda\}$ the number of accepted hypotheses at level $\lambda \in (0, 1)$. If
41 type II errors are ignored for a sufficiently large λ then

42 (1.1)
$$E[N_a(\lambda)] \approx m_0(1 - \lambda).$$

43 Storey's (ST) estimator is given by

44 (1.2)
$$\hat{\pi}_0(\lambda) = \frac{N_a(\lambda)}{m(1 - \lambda)} \text{ or } \hat{m}_0(\lambda) = \frac{N_a(\lambda)}{1 - \lambda}.$$

45 Schweder and Spjøtvoll's (1982) method visually fits a straight line through the
46 origin to the plot of $N_a(p_{(i)}) = m - i$ vs. $1 - p_{(i)}$ ($1 \leq i \leq m$) for large values
47 of the $p_{(i)}$. The slope of the fitted line is taken as an estimate of m_0 according to
48
49
50
51

Equation (1.1). Because these estimators attribute all nonsignificant p -values to the true null hypotheses (type II errors are ignored) and do not explicitly model the non-null p -values, they tend to be positively biased which results in conservative adaptive control of any type I error rate.

To get a handle on type II errors, so that both the null and non-null p -values can be utilized to estimate π_0 , the mixture model approach has been proposed by several authors. The mixture model differs from the setup given in the first paragraph in that the number of true hypotheses is a random variable (r.v.) and m_0 is its expected value. Specifically, let Z_i be a Bernoulli r.v. which equals 1 with probability π_0 if H_{0i} is true and 0 with probability $\pi_1 = 1 - \pi_0$ if H_{0i} is false. Assume that the Z_i ($1 \leq i \leq m$) are i.i.d. Then the number of false hypotheses, $M_0 = \sum_{i=1}^m Z_i$, is a binomial r.v. with parameters m and π_0 , and $E(M_0) = m_0 = m\pi_0$.

A parametric mixture model was considered by Hsueh, Chen, and Kodell (2003) (HCK). They assumed the following simple hypothesis testing setup. Suppose that all m hypotheses pertain to the means of the normal distributions with $H_{0i} : \mu_i = 0$ versus $H_{1i} : \mu_i > 0$. (HCK considered a two-sided alternative, but that is not germane to their method.) Conditional on Z_i , the test statistic $X_i \sim N(\delta_i, 1)$, where δ_i is the standardized μ_i with $\delta_i = 0$ if $Z_i = 1$ and $\delta_i = \delta > 0$ if $Z_i = 0$ where HCK assumed that δ is *known*. We refer to this model as the *normal model*, which was also used by Black (2004) to study the bias of Storey's (2002) estimator. An expression for the expected number of X_i 's that are greater than any specified threshold can be derived using this setup. By plotting the corresponding observed number of X_i 's against the threshold, m_0 could be estimated as the slope of the straight line through the origin using least squares (LS) regression.

The normal model is the topic of Section 2. We first extend the HCK estimation method to the unknown δ case, which is a nonlinear least squares (NLS) regression problem. Next we note that the HCK method makes use of only the *number* of X_i 's that are greater than a specified threshold; it does not make use of the *magnitudes* of the X_i 's. Therefore we propose two alternative methods of estimation which utilize the magnitudes of the X_i 's in an attempt to obtain a better estimate of δ and thereby a better estimate of m_0 . The first of these alternative methods is similar to the LS method of HCK, but uses the sample mean (instead of the number) of the X_i 's that are greater than a specified threshold. We refer to it as the test statistics (TS) method. The second method is the EM method of Dempster, Laird and Rubin (1977) which finds the maximum likelihood estimators (MLEs) of the mixture distribution of the X_i 's.

This normal mixture model approach in conjunction with the EM algorithm was proposed by Guan, Wu and Zhao (2004) and most recently by Iyer and Sarkar (2007). So, although the use of the EM algorithm for estimation in the context of the present problem is not new, we perform a comprehensive comparison of it with the other two new methods, and find that it performs best when the assumed model is correct, but is not robust to model violations.

In the second approach discussed in Section 3, the non-null p -values are modeled by a beta distribution with unknown parameters a and b (denoted by $\text{Beta}(a, b)$). We refer to this model as the *beta model*. Here we restrict to estimation methods based on p -values since the X_i 's can have different null distributions. All three estimators (HCK, TS and EM) are also derived for the beta model.

We stress that both the normal and beta models are simply "working" models intended to get a handle on type II errors. We do not pretend that these models are strictly true. Therefore robustness of the estimators to the model assumptions is an important issue. In the simulation comparisons for the normal model, robustness of

1 the fixed δ assumption is tested by generating different δ_i 's for the false hypotheses 1
 2 from a normal distribution. Robustness of the normal model assumption is tested by 2
 3 generating p_i 's for the false hypotheses using the beta model and transforming them 3
 4 to the X_i 's using the inverse normal transformation. Similarly, the robustness of the 4
 5 beta model is tested by generating X_i 's using the normal model and transforming 5
 6 them to p_i 's. 6

7 Adaptive control of FDR using different estimators of m_0 is the topic of Sec- 7
 8 tion 4. The ST, HCK, TS and EM estimators are compared in a large simulation 8
 9 study in Section 5. The performance measures used in the simulation study are the 9
 10 biases and mean square errors of the estimators of π_0 and FDR. An example illus- 10
 11 trating application of the proposed methods is given in Section 6. Conclusions are 11
 12 summarized in Section 7. Proofs of some technical results are given in the Appendix. 12

13

14 **2. Normal Model** 14

15

16 The normal mixture model can be expressed as 16

17

18 (2.1)
$$f(x_i) = \pi_0\phi(x_i) + \pi_1\phi(x_i - \delta),$$
 18

19

20 where $f(x_i)$ is the p.d.f. of X_i and $\phi(\cdot)$ is the p.d.f. of the standard normal distri- 20
 21 bution. Although δ will need to be estimated, we are not too concerned about its 21
 22 estimation accuracy since, after all, it is a parameter of a working model. 22

23

24 **2.1. Hsueh, Chen, and Kodell (HCK) Method** 24

25

26 Let 26

27

28 (2.2)
$$\beta(\delta, \lambda) = P_{H_{1i}}\{p_i > \lambda\} = P_{H_{1i}}\{X_i < z_\lambda\} = \Phi(z_\lambda - \delta)$$
 28

29

30 denote the type II error probability of a test performed at level λ where $\Phi(\cdot)$ is the 30
 31 standard normal c.d.f. and $z_\lambda = \Phi^{-1}(1 - \lambda)$. Then $E[N_r(\lambda)] = m_0\lambda + (m - m_0)[1 -$ 31
 32 $\beta(\delta, \lambda)]$, and hence 32

33

34 (2.3)
$$E[N_r(\lambda)] - m\Phi(-z_\lambda + \delta) = m_0[\lambda - \Phi(-z_\lambda + \delta)].$$
 34

35

36 For $\lambda = p_{(i)}$, $i = 1, 2, \dots, m$, the term inside the square brackets in the R.H.S. of 36
 37 the above equation is 37

38 (2.4)
$$x_i = p_{(i)} - \Phi(-z_{p_{(i)}} + \delta)$$
 38

39

40 and the L.H.S. can be estimated by 40

41

42 (2.5)
$$y_i = i - m\Phi(-z_{p_{(i)}} + \delta).$$
 42

43

44 If δ is assumed to be known then we can calculate (x_i, y_i) , $i = 1, 2, \dots, m$, and using 44
 45 (2.3) fit an LS straight line through the origin by minimizing $\sum_{i=1}^m (y_i - m_0x_i)^2$ with 45
 46 respect to (w.r.t.) m_0 . The LS estimator of m_0 is given by 46

47

48 (2.6)
$$\hat{m}_0 = \frac{\sum_{i=1}^m x_i y_i}{\sum_{i=1}^m x_i^2}.$$
 48

49

50 We first extend the HCK estimator to the unknown δ case by incorporating 50
 51 estimation of δ as part of the NLS problem of minimizing $\sum_{i=1}^m (y_i - m_0x_i)^2$ w.r.t. 51

m_0 and δ . The iterative algorithm for this purpose is given below. The initial values for this algorithm as well as the algorithms for the TS and EM estimators were determined by solving the following two moment equations for m_0 and δ :

$$(2.7) \quad \sum_{i=1}^m X_i = (m - m_0)\delta \text{ and } \sum_{i=1}^m X_i^2 = m_0 + (m - m_0)(\delta^2 + 1).$$

HCK Algorithm

Step 0: Compute initial estimates \hat{m}_0 and $\hat{\delta}$ by solving (2.7). Let $\hat{\pi}_0 = \hat{m}_0/m$.

Step 1: Set $\delta = \hat{\delta}$ and compute (x_i, y_i) , $i = 1, 2, \dots, m$, using (2.4) and (2.5).

Step 2: Compute \hat{m}_0 using (2.6) and $\hat{\pi}_0 = \hat{m}_0/m$.

Step 3: Find $\hat{\delta}$ to minimize $\sum_{i=1}^m (y_i - m_0 x_i)^2$.

Step 4: Return to Step 1 until convergence.

Remark: One could use weighted least squares to take into account the heteroscedasticity of the y_i 's. We tried this, but the resulting NLS problem was computationally much more intensive without a collateral gain in the efficiency of the estimators.

2.2. Test Statistics (TS) Method

As noted in Section 1, we hope to improve upon the HCK estimator by utilizing the information in the magnitudes of the X_i 's. Toward this end we first propose an estimator analogous to the HCK estimator except that it uses the sample mean (rather than the number) of the X_i 's that are significant at a specified level λ .

Define

$$S_a(\lambda) = \{i : p_i > \lambda\} = \{i : X_i < z_\lambda\} \text{ and } S_r(\lambda) = \{i : p_i \leq \lambda\} = \{i : X_i \geq z_\lambda\}.$$

Then $N_a(\lambda) = |S_a(\lambda)|$ and $N_r(\lambda) = |S_r(\lambda)|$. Finally define

$$\bar{X}_a(\lambda) = \frac{1}{N_a(\lambda)} \sum_{i \in S_a(\lambda)} X_i \text{ and } \bar{X}_r(\lambda) = \frac{1}{N_r(\lambda)} \sum_{i \in S_r(\lambda)} X_i.$$

To derive the expected values of these sample means the following lemma is useful.

Lemma 1. *Define*

$$c_{0a}(\lambda) = E_{H_{0i}}(X_i | X_i < z_\lambda), c_{0r}(\lambda) = E_{H_{0i}}(X_i | X_i \geq z_\lambda),$$

and

$$c_{1a}(\delta, \lambda) = E_{H_{1i}}(X_i | X_i < z_\lambda), c_{1r}(\delta, \lambda) = E_{H_{1i}}(X_i | X_i \geq z_\lambda).$$

Then

$$c_{0a}(\lambda) = -\frac{\phi(z_\lambda)}{1 - \lambda}, c_{0r}(\lambda) = \frac{\phi(z_\lambda)}{\lambda}$$

and

$$c_{1a}(\delta, \lambda) = \delta - \frac{\phi(z_\lambda - \delta)}{\Phi(z_\lambda - \delta)}, c_{1r}(\delta, \lambda) = \delta + \frac{\phi(\delta - z_\lambda)}{\Phi(\delta - z_\lambda)}. \quad \blacksquare$$

1 *Proof.* The proof follows from the following expressions for the conditional expect- 1
 2 ations of $X \sim N(\mu, 1)$: 2

$$3 \quad E(X|X \leq x) = \mu - \frac{\phi(x - \mu)}{\Phi(x - \mu)} \text{ and } E(X|X > x) = \mu + \frac{\phi(\mu - x)}{\Phi(\mu - x)}. \quad 4$$

5
 6
 7 The desired expected values of $\bar{X}_a(\lambda)$ and $\bar{X}_r(\lambda)$ are then given by the following 7
 8 lemma. 8

9 **Lemma 2.** *Let* 9

$$10 \quad (2.8) \quad g(\pi_0, \delta, \lambda) = P\{Z_i = 1|X_i < z_\lambda\} = \frac{\pi_0(1 - \lambda)}{\pi_0(1 - \lambda) + \pi_1\Phi(z_\lambda - \delta)} \quad 11$$

12
 13 *and* 14

$$15 \quad (2.9) \quad h(\pi_0, \delta, \lambda) = P\{Z_i = 1|X_i \geq z_\lambda\} = \frac{\pi_0\lambda}{\pi_0\lambda + \pi_1\Phi(-z_\lambda + \delta)}. \quad 16$$

17
 18 *Then* 19

$$20 \quad (2.10) \quad E[\bar{X}_a(\lambda)] = g(\pi_0, \delta, \lambda)c_{0a}(\lambda) + [1 - g(\pi_0, \delta, \lambda)]c_{1a}(\delta, \lambda) \quad 21$$

22
 23 *and* 24

$$25 \quad (2.11) \quad E[\bar{X}_r(\lambda)] = h(\pi_0, \delta, \lambda)c_{0r}(\lambda) + [1 - h(\pi_0, \delta, \lambda)]c_{1r}(\delta, \lambda), \quad 26$$

27 *where* $c_{0a}(\lambda)$, $c_{0r}(\lambda)$, $c_{1a}(\delta, \lambda)$ *and* $c_{1r}(\delta, \lambda)$ *are as given in Lemma 1.* 28

29 *Proof.* Given in the Appendix. 30

31 To develop an estimation method analogous to the HCK method note that from 31
 32 (2.3) and (2.11) we get 32

$$33 \quad (2.12) \quad E[N_r(\lambda)]E[\bar{X}_r(\lambda)] - m\Phi(-z_\lambda + \delta)c_{1r}(\delta, \lambda) = m_0[\lambda c_{0r}(\lambda) - \Phi(-z_\lambda + \delta)c_{1r}(\delta, \lambda)]. \quad 34$$

35
 36 For $\lambda = p_{(i)}$, $i = 1, 2, \dots, m$, the term inside the square brackets in the R.H.S. of 36
 37 the above equation is 37

$$38 \quad (2.13) \quad x_i = p_{(i)}c_{0r}(p_{(i)}) - \Phi(-z_{p_{(i)}} + \delta)c_{1r}(\delta, p_{(i)}) \quad 39$$

40
 41 *and* the L.H.S. can be estimated by 41

$$42 \quad (2.14) \quad y_i = i\bar{X}_r(p_{(i)}) - m\Phi(-z_{p_{(i)}} + \delta)c_{1r}(\delta, p_{(i)}) \quad 43$$

$$44 \quad = \sum_{j=m-i+1}^m X_{(j)} - m\Phi(-z_{p_{(i)}} + \delta)c_{1r}(\delta, p_{(i)}). \quad 45$$

46
 47
 48 Then from (2.12) we see that a regression line of y_i versus x_i can be fitted through 48
 49 the origin with slope m_0 by minimizing $\sum_{i=1}^m (y_i - m_0x_i)^2$ w.r.t. m_0 and δ . The 49
 50 algorithm to solve this NLS regression problem is exactly analogous to the HCK 50
 51 algorithm. 51

2.3. EM Method

Whereas the HCK and TS methods compute the LS estimators of π_0 and δ (for two different regression models), the EM method computes their MLEs. For these MLEs to exist, it is necessary that π_0 be bounded away from 0 and 1. The steps in the EM algorithm are as follows.

EM Algorithm

Step 0: Compute initial estimates \hat{m}_0 and $\hat{\delta}$ by solving (2.7). Let $\hat{\pi}_0 = \hat{m}_0/m$.

Step 1: Calculate the posterior probabilities:

$$\hat{\pi}_0(X_i) = \frac{\hat{\pi}_0 \phi(X_i)}{\hat{\pi}_0 \phi(X_i) + \hat{\pi}_1 \phi(X_i - \hat{\delta})}$$

and $\hat{\pi}_1(X_i) = 1 - \hat{\pi}_0(X_i)$, $i = 1, 2, \dots, m$.

Step 2: Calculate new estimates:

$$\hat{\pi}_0 = \frac{\sum_{i=1}^m \hat{\pi}_0(X_i)}{m} \quad \text{and} \quad \hat{\delta} = \frac{\sum_{i=1}^m \hat{\pi}_1(X_i) X_i}{\sum_{i=1}^m \hat{\pi}_1(X_i)}.$$

Step 3: Return to Step 1 until convergence.

3. Beta Model

In many applications the normal model may be inappropriate because the test statistics may not be normally distributed or different types of test statistics (e.g., normal, t , chi-square, Wilcoxon, log-rank) may be used to test different hypotheses or only the p -values of the test statistics may be available. In these cases we use the p -values to estimate π_0 .

We propose to model the non-null p -values by a Beta(a, b) distribution given by

$$g(p|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1-p)^{b-1}$$

with unknown parameters a and b with $a < 1$ and $b > 1$. This restriction is imposed in order to ensure that $g(p|a, b)$ is decreasing in p . It is well-known that the non-null distribution of the p -values must be right-skewed and generally decreasing in shape (see Hung, O'Neill, Bauer and Kohne 1997). Langaas et al. (2005) imposed the same restriction in their nonparametric estimate of the non-null distribution of p -values.

Of course, the null distribution of a p -value is Beta(1, 1), i.e., the $U[0, 1]$ distribution. As in the case of the normal model, the beta model can be represented as a mixture model for the distribution of the p_i :

$$(3.1) \quad f(p_i) = \pi_0 \times 1 + \pi_1 g(p_i|a, b).$$

The parameters a and b must be estimated along with π_0 . This problem is analogous to that encountered for the normal model with the difference that in addition to π_0 , we have to estimate two parameters, a and b , instead of a single parameter δ . We first extend the HCK method for the normal model discussed in Section 2.1 to this beta model.

3.1. Hsueh, Chen, and Kodell (HCK) Method

Denote the type II error probability of a test performed at level λ by

$$(3.2) \quad \beta(a, b, \lambda) = P_{H_{1i}}\{p_i > \lambda\} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_{\lambda}^1 p^{a-1}(1-p)^{b-1} dp = 1 - I_{\lambda}(a, b),$$

where $I_{\lambda}(a, b)$ is the incomplete beta function. Put

$$(3.3) \quad x_i = p_{(i)} - I_{p_{(i)}}(a, b) \text{ and } y_i = i - mI_{p_{(i)}}(a, b).$$

Then the HCK method amounts to solving the NLS problem of minimizing $\sum_{i=1}^m (y_i - m_0 x_i)^2$ w.r.t. m_0 and (a, b) (subject to $a < 1 < b$). Gauss-Newton method (Gill et al. 1981) was used to perform minimization w.r.t. (a, b) . The initial starting values for this algorithm as well as the algorithms for the TS and EM estimators described below were determined by solving the following three moment equations for m_0 and (a, b) :

$$(3.4) \quad \begin{aligned} \sum_{i=1}^m p_i &= \frac{1}{2}m_0 + \frac{a}{a+b}m_1, \\ \sum_{i=1}^m p_i^2 &= \frac{1}{3}m_0 + \frac{a(a+1)}{(a+b)(a+b+1)}m_1, \\ \sum_{i=1}^m p_i^3 &= \frac{1}{4}m_0 + \frac{a(a+1)(a+2)}{(a+b)(a+b+1)(a+b+2)}m_1. \end{aligned}$$

3.2. Test Statistics (TS) Method

Here the TS estimator is based on the average of the “accepted” or “rejected” p -values defined as

$$\bar{p}_a(\lambda) = \frac{1}{N_a(\lambda)} \sum_{i \in S_a(\lambda)} p_i \text{ and } \bar{p}_r(\lambda) = \frac{1}{N_r(\lambda)} \sum_{i \in S_r(\lambda)} p_i.$$

Analogous to Lemma 1, we have the following lemma.

Lemma 3. *Define*

$$d_{0a}(\lambda) = E_{H_{0i}}(p_i | p_i > \lambda), d_{0r}(\lambda) = E_{H_{0i}}(p_i | p_i \leq \lambda),$$

and

$$d_{1a}(a, b, \lambda) = E_{H_{1i}}(p_i | p_i > \lambda), d_{1r}(a, b, \lambda) = E_{H_{1i}}(p_i | p_i \leq \lambda).$$

Then we have

$$d_{0a}(\lambda) = \frac{\lambda+1}{2}, d_{0r}(\lambda) = \frac{\lambda}{2}$$

and

$$d_{1a}(a, b, \lambda) = \frac{1 - I_{\lambda}(a+1, b)}{1 - I_{\lambda}(a, b)} \cdot \frac{a}{a+b}, d_{1r}(a, b, \lambda) = \frac{I_{\lambda}(a+1, b)}{I_{\lambda}(a, b)} \cdot \frac{a}{a+b}.$$

1 *Proof.* Straightforward. 1

2
3 The next lemma gives $E[\bar{p}_a(\lambda)]$ and $E[\bar{p}_r(\lambda)]$; its proof is exactly analogous to
4 that of Lemma 2. 4

5 **Lemma 4.** *Let* 5

$$6 \quad g(\pi_0, a, b, \lambda) = P\{Z_i = 1 | p_i > \lambda\} = \frac{\pi_0(1 - \lambda)}{\pi_0(1 - \lambda) + \pi_1[1 - I_\lambda(a, b)]} \quad 6$$

7
8
9 *and* 9

$$10 \quad h(\pi_0, a, b, \lambda) = P\{Z_i = 1 | p_i \leq \lambda\} = \frac{\pi_0\lambda}{\pi_0\lambda + \pi_1 I_\lambda(a, b)}. \quad 10$$

11
12
13 *Then* 13

$$14 \quad (3.5) \quad E[\bar{p}_a(\lambda)] = g(\pi_0, a, b, \lambda)d_{0a}(\lambda) + [1 - g(\pi_0, a, b, \lambda)]d_{1a}(a, b, \lambda) \quad 14$$

15
16
17 *and* 17

$$18 \quad (3.6) \quad E[\bar{p}_r(\lambda)] = h(\pi_0, a, b, \lambda)d_{0r}(\lambda) + [1 - h(\pi_0, a, b, \lambda)]d_{1r}(a, b, \lambda), \quad 18$$

19
20 *where* $d_{0a}(\lambda)$, $d_{0r}(\lambda)$, $d_{1a}(a, b, \lambda)$ *and* $d_{1r}(a, b, \lambda)$ *are as given in Lemma 3.* 20

21 The equations for the TS estimator are derived as follows. Analogous to (2.12),
22 we obtain 22

$$23 \quad E[N_r(\lambda)]E[\bar{p}_r(\lambda)] - mI_\lambda(a, b)d_{1r}(a, b, \lambda) = m_0[\lambda d_{0r}(\lambda) - I_\lambda(a, b)d_{1r}(a, b, \lambda)]. \quad 24$$

25
26 For $\lambda = p_{(i)}$, $i = 1, 2, \dots, m$, the term in the square brackets in the R.H.S. of the
27 above equation equals 27

$$28 \quad x_i = \frac{p_{(i)}^2}{2} - \frac{a}{a+b} I_{p_{(i)}}(a+1, b) \quad 28$$

29
30
31 *and the L.H.S. can be estimated by* 31

$$32 \quad y_i = \sum_{j=1}^i p_{(j)} - \frac{a}{a+b} I_{p_{(i)}}(a+1, b). \quad 32$$

33
34
35
36 The TS algorithm for the normal model can be modified to minimize $\sum_{i=1}^m (y_i -$
37 $m_0 x_i)^2$ by replacing the minimization with respect to δ by minimization with re-
38 spect to (a, b) . 38

39 40 41 **3.3. EM Method** 41

42
43 The steps in the EM algorithm, which gives the MLEs of π_0 and (a, b) , are as fol-
44 lows. As in the case of the normal model, for these MLEs to exist, it is necessary
45 that π_0 be bounded away from 0 and 1. 45

46
47 **Step 0:** Initialize \hat{m}_0 and (\hat{a}, \hat{b}) by solving (3.5). Let $\hat{\pi}_0 = \hat{m}_0/m$. 47

48 **Step 1 (E-Step):** Calculate the posterior probabilities: 48

$$49 \quad \hat{\pi}_0(p_i) = \frac{\hat{\pi}_0}{\hat{\pi}_0 + \hat{\pi}_1 g(p_i | \hat{a}, \hat{b})} \quad 49$$

1 and $\widehat{\pi}_1(p_i) = 1 - \widehat{\pi}_0(p_i)$, $i = 1, 2, \dots, m$. 1

2 **Step 2 (M-Step):** Calculate \widehat{a} and \widehat{b} as solutions of the equations (see Equations 2
 3 (21.1) and (21.2) in Johnson and Kotz 1970): 3

$$4 \quad \psi(a) - \psi(a + b) = \frac{\sum_{i=1}^m \widehat{\pi}_1(p_i) \ln p_i}{\sum_{i=1}^m \widehat{\pi}_1(p_i)}, \quad 4$$

$$5 \quad \psi(b) - \psi(a + b) = \frac{\sum_{i=1}^m \widehat{\pi}_1(p_i) \ln(1 - p_i)}{\sum_{i=1}^m \widehat{\pi}_1(p_i)}, \quad 5$$

6 where $\psi(\cdot)$ is the digamma function (i.e., the derivative of the natural logarithm of 6
 7 the gamma function). Also calculate 7

$$8 \quad \widehat{\pi}_0 = \frac{\sum_{i=1}^m \widehat{\pi}_0(p_i)}{m}. \quad 8$$

9 **Step 3:** Return to Step 1 until convergence. 9

10 **4. Adaptive Control of FDR** 10

11 We now discuss the use of the estimate \widehat{m}_0 for adaptively controlling the FDR. 11
 12 The control is assumed to be strong control (Hochberg and Tamhane 1987), i.e., 12
 13 $\text{FDR} \leq \alpha$ for some specified $\alpha < 1$ for all possible combinations of true and false 13
 14 null hypotheses and the respective parameter values. Let R be the total number of 14
 15 rejected hypotheses and let V be the number of true hypotheses that are rejected. 15
 16 Benjamini and Hochberg (1995) introduced the definition 16

$$17 \quad \text{FDR} = E \left[\frac{V}{R} \right] = E \left[\frac{V}{R} \mid R > 0 \right] P(R > 0), \quad 17$$

18 where $0/0$ is defined as 0. Benjamini and Hochberg (1995) gave a step-up (SU) 18
 19 procedure that controls $\text{FDR} \leq \alpha$. 19

20 Storey (2002) considered a single-step (SS) procedure (which he referred to as 20
 21 the fixed rejection region method) that rejects H_{0i} if $p_i \leq \gamma$ for some common 21
 22 fixed threshold γ . His focus was on estimating the FDR. He proposed the following 22
 23 nonparametric estimator: 23

$$24 \quad (4.1) \quad \widehat{\text{FDR}}_\lambda(\gamma) = \frac{\widehat{\pi}_0(\lambda)\gamma}{\{N_r(\lambda) \vee 1\}/m}, \quad 24$$

25 where $\widehat{\pi}_0(\lambda)$ is given by (1.2). The solution $\widehat{\gamma}$ to the equation $\widehat{\text{FDR}}_\lambda(\gamma) = \alpha$ can 25
 26 be used in an MCP that tests each hypothesis at the $\widehat{\gamma}$ -level. Storey, Taylor and 26
 27 Siegmund (2004, Theorem 3) have shown that this heuristic procedure (which uses 27
 28 a slightly modified estimator of π_0) controls the FDR. The heuristic procedures 28
 29 proposed below along the same lines (which use parametric estimators of the FDR) 29
 30 have not been rigorously shown to control the FDR. 30

31 We propose the following parametric estimator of the FDR: 31

$$32 \quad (4.2) \quad \widehat{\text{FDR}}(\gamma) = \frac{\widehat{\pi}_0\gamma}{\widehat{\pi}_0\gamma + \widehat{\pi}_1[1 - \beta(\cdot, \gamma)]}, \quad 32$$

33 where $\beta(\cdot, \gamma)$ is either $\beta(\widehat{\delta}, \gamma)$ given by (2.2) for the normal model or $\beta(\widehat{a}, \widehat{b}, \gamma)$ given 33
 34 by (3.2) for the beta model. To adaptively control the FDR at level α , we use 34
 35 35

the same heuristic procedure as above except that $\hat{\gamma}$ is obtained by setting this parametric estimator equal to α .

We may confine attention to $\alpha \leq \pi_0$ since if $\alpha > \pi_0$ then one can choose $\hat{\gamma} = 1$, and reject all hypotheses while still controlling the FDR = $\pi_0 < \alpha$. Existence and uniqueness of $\hat{\gamma}$ for $\alpha \in (0, \pi_0]$ is proved in the following two lemmas for the normal and beta models, respectively.

Lemma 5. *For the normal model, the solution $\hat{\gamma}$ to the equation $\widehat{FDR}(\gamma) = \alpha$, where $\widehat{FDR}(\gamma)$ and $\beta(\hat{\delta}, \gamma)$ are given by (4.2) and (2.2), respectively, exists and is unique for $\alpha \in (0, \pi_0]$.*

Proof. Given in the Appendix.

Lemma 6. *For the beta model, assuming $0 < \hat{a} < 1 < \hat{b}$, the solution $\hat{\gamma}$ to the equation $\widehat{FDR}(\gamma) = \alpha$, where $\widehat{FDR}(\gamma)$ and $\beta(\hat{\delta}, \gamma)$ are given by (4.2) and (3.2), respectively, exists and is unique for $\alpha \in (0, \pi_0]$.*

Proof. Given in the Appendix.

To develop an adaptive FDR-controlling procedure for the normal mixture model, Iyer and Sarkar (2007) took a somewhat different approach via the following asymptotic result of Genovese and Wasserman (2002): Assume that the p_i are independent $U[0, 1]$ when the H_{0i} are true and have a common distribution F when the H_{0i} are false. Then the nominal α -level Benjamini and Hochberg SU procedure is asymptotically (as $m \rightarrow \infty$) equivalent to Storey's SS procedure that rejects H_{0i} if $p_i \leq \hat{\gamma}$ where $\hat{\gamma}$ is the solution to the equation

$$F(\gamma) = \rho\gamma \text{ and } \rho = \frac{1 - \alpha\pi_0}{\alpha(1 - \pi_0)}.$$

Furthermore, since the SU procedure actually controls the FDR conservatively at $\pi_0\alpha$ level, exact control at level α is achieved by replacing α in the expression for ρ by α/π_0 , which results in the following equation for γ :

$$(4.3) \quad F(\gamma) = \rho\gamma \text{ and } \rho = \frac{\pi_0(1 - \alpha)}{\alpha(1 - \pi_0)}.$$

By writing $F(\gamma) = 1 - \beta(\cdot, \gamma)$, we see that $\widehat{FDR}(\gamma) = \alpha$ and (4.3) are identical if π_0 is replaced by $\hat{\pi}_0$ in (4.3). Iyer and Sarkar (2007) used the solution $\hat{\gamma}$ from (4.3) in Storey's SS procedure with $F(\gamma) = \Phi(\delta - z_\gamma)$, and δ and π_0 replaced by their estimates $\hat{\delta}$ and $\hat{\pi}_0$ obtained from the EM method, which results in an adaptive FDR-controlling procedure, which is identical to the one proposed before.

5. Simulation Results

We compared different estimators by conducting an extensive simulation study. The ST estimator was used with $\lambda = 0.5$ throughout. The estimators were compared in terms of their accuracy of estimation of π_0 and control of FDR at a nominal $\alpha = 0.10$ using the SS procedure. The bias and mean square error (MSE) of the estimators were used as the performance measures. The results for the normal model are presented in Section 5.1 and for the beta model in Section 5.2. Throughout we used $m = 1000$ and the number of replications was also set equal to 1000. We varied

1 π_0 from 0.1 to 0.9 in steps of 0.1. The values $\pi_0 = 0$ and 1 were excluded because 1
 2 $\widehat{\pi}_0$ exhibits erratic results in these extreme cases; also $\text{FDR} = 0$ when $\pi_0 = 0$. 2

3 In each simulation run, first the random indexes of the true and false hypotheses 3
 4 were generated by generating Bernoulli r.v.'s Z_i . Then the respective X_i 's or the 4
 5 p_i 's were generated using the appropriate null or alternative distributions. The bias 5
 6 of each $\widehat{\pi}_0$ estimator was estimated as the difference between the average of the 6
 7 $\widehat{\pi}_0$ values over 1000 replicates and the true value of π_0 . In the case of FDR, the 7
 8 bias was estimated as the difference between the average of the $\widehat{\text{FDR}}$ values over 8
 9 1000 replicates and the nominal $\alpha = 0.10$. The MSE was computed as the sum of 9
 10 the square of the bias and the variance of the $\widehat{\pi}_0$ (or $\widehat{\text{FDR}}$) values averaged over 10
 11 1000 replicates. The detailed numerical results are given in Shi (2006); here we only 11
 12 present graphical plots to save space. 12
 13

14 **5.1. Simulation Results for Normal Model** 14

15 Simulations were conducted in three parts. In the first part, the true model for the 15
 16 non-null hypotheses was set to be the same as the assumed model by generating the 16
 17 X_i 's for the false hypotheses from a $N(\delta, \sigma^2)$ distribution with a fixed $\delta = 2$ and 17
 18 $\sigma = 1$. In the other two parts of simulations, robustness of the assumed model was 18
 19 tested by generating the X_i 's for the false hypotheses from different distributions 19
 20 than the assumed one. In the second part, the X_i 's for the false hypotheses were 20
 21 generated from $N(\delta_i, \sigma^2)$ distributions where the δ_i 's were themselves drawn from a 21
 22 $N(\delta_0, \sigma_0^2)$ distribution with $\delta_0 = 2$ and $\sigma_0 = 0.25$ corresponding to an approximate 22
 23 range of $[1, 3]$ for the δ_i . In the third part, the p_i 's for the false hypotheses were 23
 24 generated from a $\text{Beta}(a, b)$ distribution with $a = 0.5$ and $b = 2$, and the X_i 's were 24
 25 computed using the inverse normal transformation $X_i = \Phi^{-1}(1 - p_i)$. 25
 26

27 **Results for Fixed δ :** The bias and the square root of the mean square error 27
 28 ($\sqrt{\text{MSE}}$) of $\widehat{\pi}_0$ for ST, HCK, TS and EM estimators are plotted in Fig. 1. Note 28
 29 from Equation (2.3) that the bias of the ST estimator is given by 29
 30

31 (5.1)
$$\text{Bias}[\widehat{\pi}_0(\lambda)] = \frac{1 - \pi_0}{1 - \lambda} \Phi(z_\lambda - \delta).$$
 31
 32

33 Also, using the fact that $N_a(\lambda)$ has a binomial distribution with number of trials 33
 34 m and success probability, 34

35
$$p = P\{p_i > \lambda\} = \pi_0(1 - \lambda) + (1 - \pi_0)\Phi(z_\lambda - \delta),$$
 35
 36

37 and using Equation (1.2) for $\widehat{\pi}_0(\lambda)$, we have 37
 38

39 (5.2)
$$\text{Var}[\widehat{\pi}_0(\lambda)] = \frac{p(1 - p)}{m(1 - \lambda)^2}.$$
 39
 40

41 These formulae were used to compute the bias and MSE of the ST estimator instead 41
 42 of estimating them by simulation. Note that the bias of the ST estimator decreases 42
 43 linearly in π_0 . The $\sqrt{\text{MSE}}$ plot for the ST estimator is also approximately linear 43
 44 because the bias is the dominant term in MSE. This is true whenever the alternative 44
 45 is fixed for all false null hypotheses. 45
 46

47 The TS estimator does not offer an improvement over the HCK estimator, as we 47
 48 had hoped, and in fact performs slightly worse in terms of MSE for $\pi_0 \leq 0.5$. We sus- 48
 49 pect that this result is due to the bias introduced when the term $E[N_r(\lambda)]E[\overline{X}_r(\lambda)]$ 49
 50

in Equation (2.12) is estimated by $i\bar{X}_r(p_{(i)})$ for $\lambda = p_{(i)}$ because of the fact that the product of the expected values does not equal the expected value of the product of two dependent r.v.'s. The EM estimator has consistently the lowest bias and the lowest MSE.

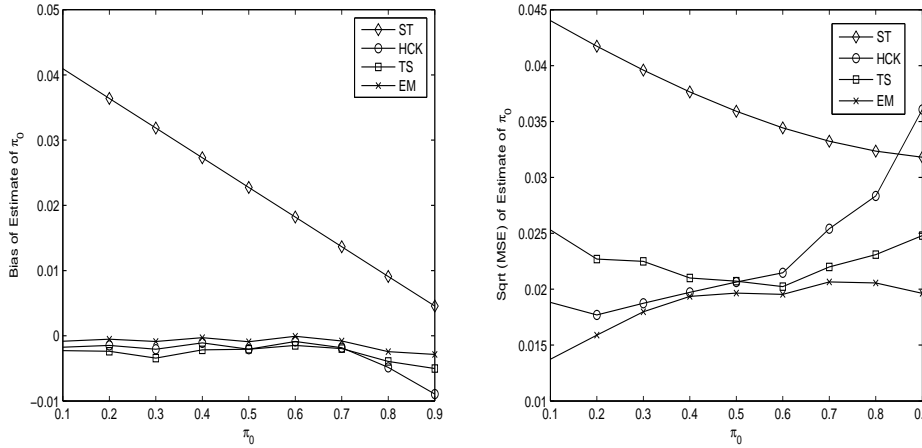


FIG 1. Bias and \sqrt{MSE} of $\hat{\pi}_0$ for ST, HCK, TS and EM Estimators for Normal Model (Data Generated by Normal Model with Fixed δ)

The bias and \sqrt{MSE} of \widehat{FDR} for ST, HCK, TS and EM estimators are plotted in Fig. 2. We see that the ST estimator leads to a large negative bias which means that, on the average, \widehat{FDR} is less than the nominal $\alpha = 0.10$ resulting in conservative control of FDR. The other three estimators yield approximately the same level of control. Surprisingly, there is very little difference in the MSEs of the four estimators. The EM estimator still is the best choice with the lowest bias and the lowest MSE throughout the entire range of π_0 values.

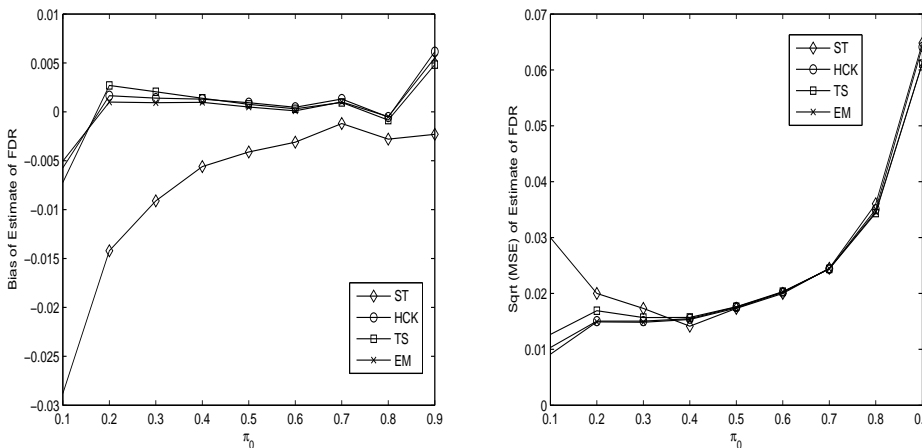
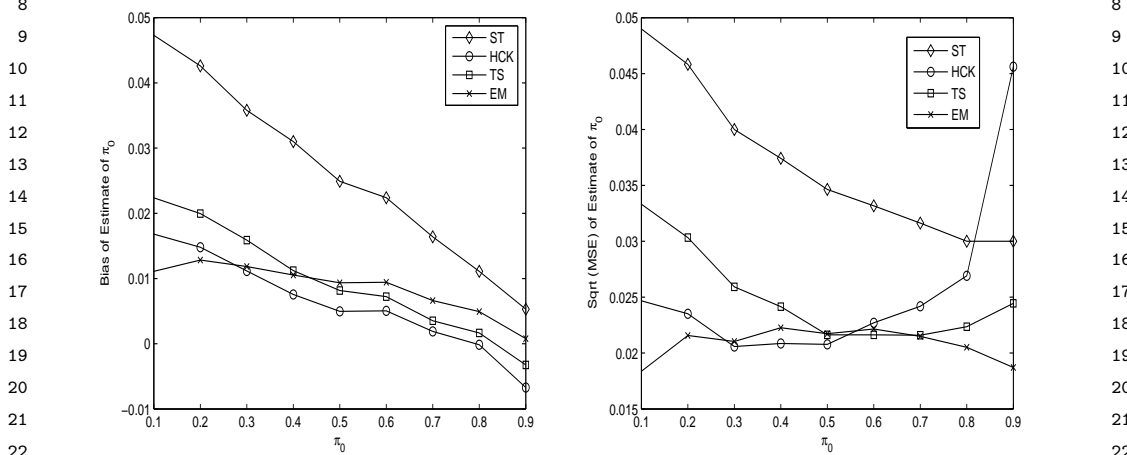
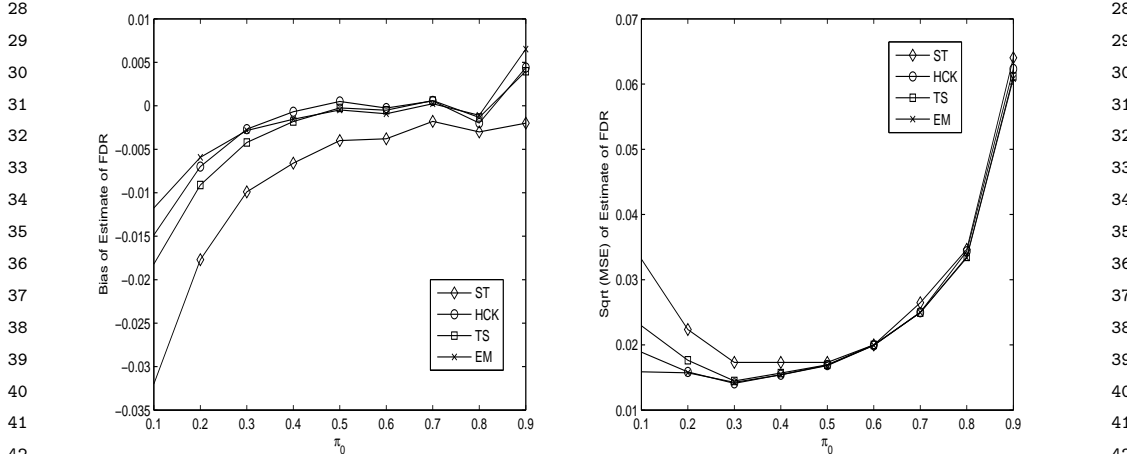


FIG 2. Bias and \sqrt{MSE} of \widehat{FDR} for ST, HCK, TS and EM Estimators for Normal Model (Data Generated by Normal Model with Fixed δ)

1 **Results for Random δ :** The bias and $\sqrt{\text{MSE}}$ of $\hat{\pi}_0$ and of $\widehat{\text{FDR}}$ for ST, HCK, TS
 2 and EM estimators are plotted in Figs. 3 and 4, respectively. By comparing these
 3 results with those for fixed $\delta = 2$, we see that, as one would expect, there is a slight
 4 degradation in the performance of every estimator because the assumed model does
 5 not hold. The comparisons between the four estimators here are similar to those
 6 for fixed δ with the estimators ranked as EM > HCK > TS > ST.
 7



23
 24 FIG 3. Bias and $\sqrt{\text{MSE}}$ of $\hat{\pi}_0$ for ST, HCK, TS and EM Estimators for Normal Model (Data
 25 Generated by Normal Model with Random δ)
 26



27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44 FIG 4. Bias and $\sqrt{\text{MSE}}$ of $\widehat{\text{FDR}}$ for ST, HCK, TS and EM Estimators for Normal Model (Data
 45 Generated by Normal Model with Random δ)
 46

47 **Robustness Results for Data Generated by Beta Model:** The bias and
 48 $\sqrt{\text{MSE}}$ of $\hat{\pi}_0$ and of $\widehat{\text{FDR}}$ for ST, HCK, TS and EM estimators are plotted in
 49 Figs. 5 and 6, respectively. Looking at Fig. 5 first, we see that the biases and MSEs
 50 of all four estimators are an order of magnitude higher compared to the normal
 51 model data which reflects lack of robustness.
 52

1 It is interesting to note that the EM estimator is no longer uniformly best for
 2 estimating π_0 . In fact, the HCK estimator has a lower bias and MSE for $0.2 \leq$
 3 $\pi_0 \leq 0.7$. The lack of robustness of the EM estimator is likely due to the strong
 4 dependence of the likelihood methods on distributional assumptions. On the other
 5 hand, for the least squares methods, the dependence on the assumed distribution
 6 is only through its first moment and hence is less strong. As far as control of
 7 FDR is concerned, there are not large differences between the proposed estimators.
 8 However, when $\pi_0 = 0.9$ the proposed estimators exceed the nominal
 9 FDR by as much as 0.05, while the ST estimator still controls FDR conservatively.
 10 In conclusion, the HCK estimator performs best for the middle range of π_0 values.
 11
 12

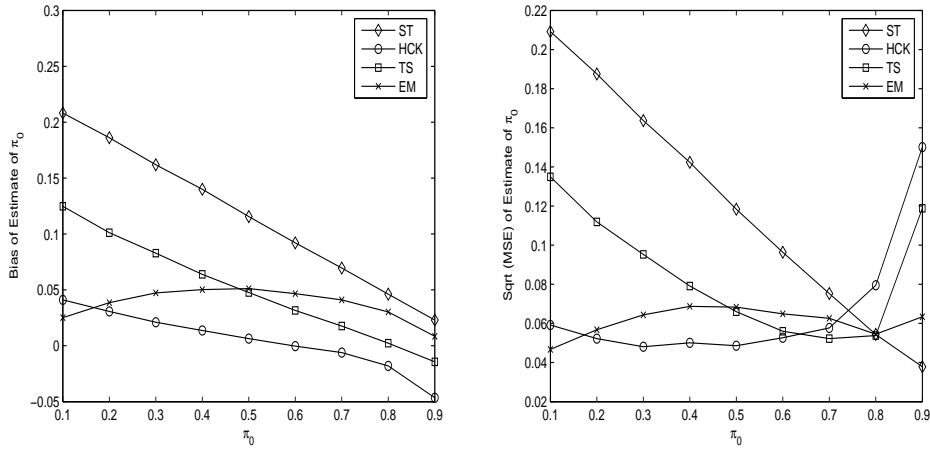


FIG 5. Bias and \sqrt{MSE} of $\hat{\pi}_0$ for ST, HCK, TS and EM Estimators for Normal Model (Data Generated by Beta Model)

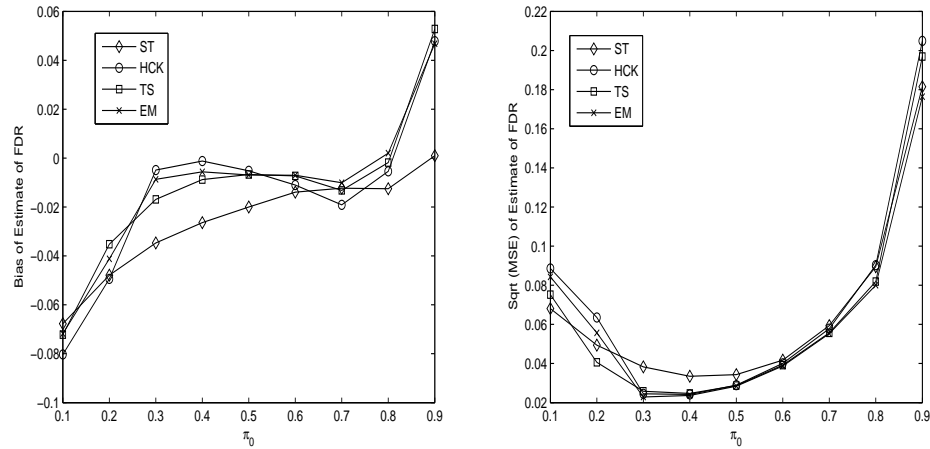


FIG 6. Bias and \sqrt{MSE} of \hat{FDR} for ST, HCK, TS and EM Estimators for Normal Model (Data Generated by Beta Model)

5.2. Simulation Results for Beta Model

Results for Beta(0.5, 2) Data: In this case the non-null p -values were generated from a Beta(a, b) distribution with $a = 0.5, b = 2.0$ and the null p -values were generated from the $U[0, 1]$ distribution. As before, the bias and variance of the ST estimator were not estimated from simulations, but were computed using Equations (5.1) and (5.2) with $\Phi(z_\lambda - \delta)$ replaced by $1 - I_\lambda(a, b)$. Note that the bias of the ST estimator decreases linearly in π_0 in this case as well and $\sqrt{\text{MSE}}$ decreases approximately linearly. From Fig. 7 we see that all estimators of π_0 , except ST, have significant negative biases particularly over the interval $[0.2, 0.5]$ and for $\pi_0 \geq 0.7$, resulting in the achieved FDR significantly exceeding the nominal value of $\alpha = 0.10$ over the corresponding ranges of π_0 as can be seen from Fig. 8. Comparing the results here with those for the normal model with the fixed δ case, we see that the biases and MSEs of all estimators are an order of magnitude higher in the present case. The reason behind this poor performance of the beta model probably lies in the difficulty of estimating the parameters a, b of the beta distribution. Only the ST estimator controls FDR conservatively and has the smallest MSE for $0.2 \leq \pi_0 \leq 0.7$. Thus the ST estimator has the best performance since it is a nonparametric estimator (and the performance would be even better if λ is not fixed, but is used a tuning parameter). In other words, the benefits of using a parametric model are far outweighed by the difficulty of estimating the parameters of the model resulting in less efficient estimators.

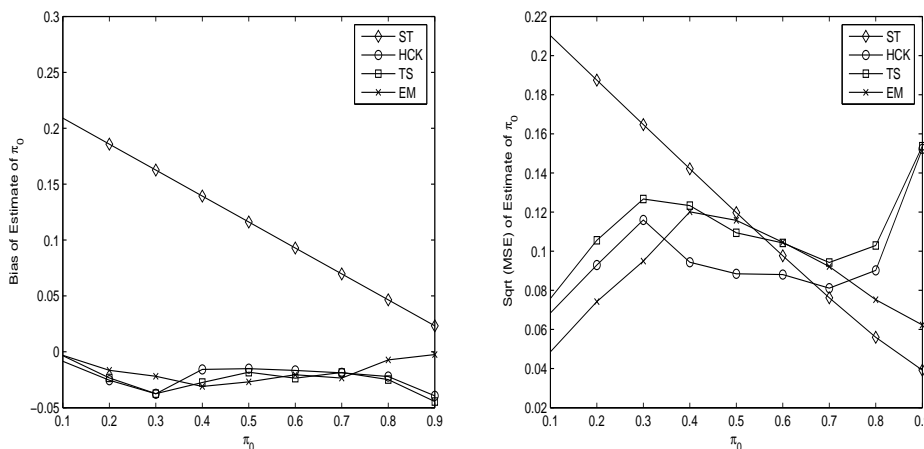


FIG 7. Bias and $\sqrt{\text{MSE}}$ of $\hat{\pi}_0$ for ST, HCK, TS and EM Estimators for Beta Model (Data Generated by Beta Model)

Robustness Results for Data Generated by Normal Model: In this case we generated the data by the normal model with $N(2, 1^2)$ as the alternative distribution. The p -values were then computed and all four methods of estimation were applied. The results are plotted in Figs. 9 and 10. From these figures we see that none of the proposed estimators exhibit consistent negative bias as they did when the data were generated according to the beta model. This is surprising since one would expect these estimators to perform more poorly when the assumed model does not hold as in the present case. We also see that the EM estimator performs worse than other estimators. Thus lack of robustness of the EM estimator to the

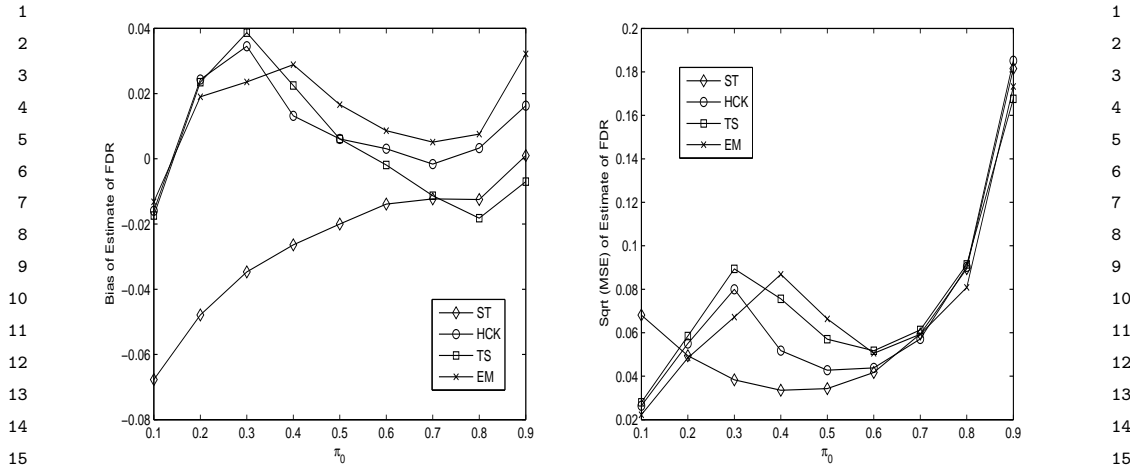


FIG 8. Bias and \sqrt{MSE} of \widehat{FDR} for ST, HCK, TS and EM Estimators for Beta Model (Data Generated by Beta Model)

model assumptions is demonstrated again, and for the same reason. The TS estimator generally has the lowest bias for estimating π_0 and its achieved FDR is closest to the nominal α ; ST estimator has the second best performance.

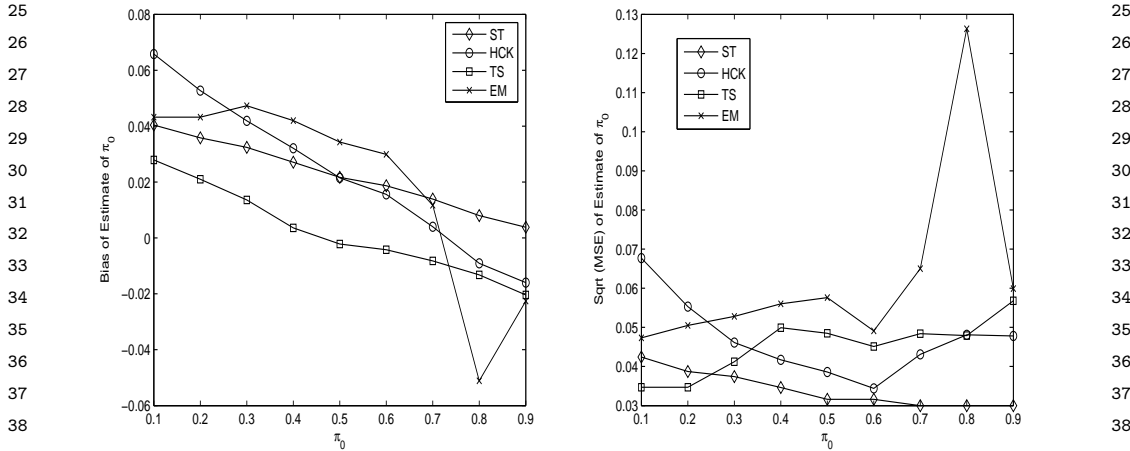


FIG 9. Bias and \sqrt{MSE} of $\widehat{\pi}_0$ for ST, HCK, TS and EM Estimators for Beta Model (Data Generated by Normal Model)

6. Example

We consider the National Assessment of Educational Progress (NAEP) data analyzed by Benjamini and Hochberg (2000). The data pertain to the changes in the average eighth-grade mathematics achievement scores for the 34 states that participated in both the 1990 and 1992 NAEP Trial State Assessment. The raw p -values for the 34 states are listed in the increasing order in Table 2. The FWE controlling

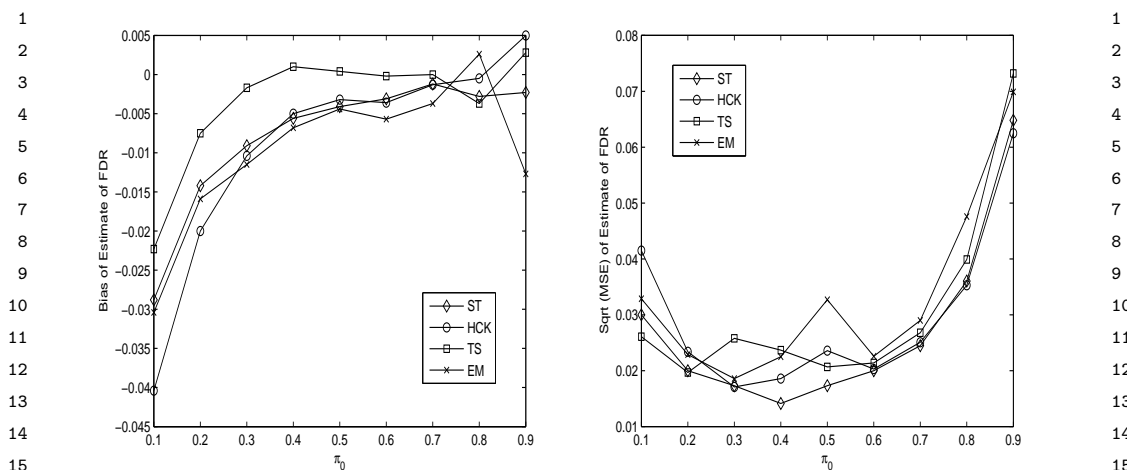


FIG 10. Bias and \sqrt{MSE} of \widehat{FDR} for ST, HCK, TS and EM Estimators for Beta Model (Data Generated by Normal Model)

Bonferroni procedure and the Hochberg (1988) procedure both identified only 4 significant results (those with p -values $\leq p_{(4)} = 0.0002$) Application of the FDR controlling non-adaptive Benjamini-Hochberg SU procedure resulted in 11 significant results. By applying their method they estimated $\widehat{m}_0 = 7$ ($\widehat{\pi}_0 = 0.2059$); using this value in the adaptive version of their procedure yielded 24 significant results.

We applied the three methods of estimation considered in this paper to these data under both the normal and beta models. The estimates $\widehat{\pi}_0$ and the associated $\widehat{\delta}$ or $(\widehat{a}, \widehat{b})$ values are given in Table 1. We see that for both models, the HCK and EM methods give smaller estimates of π_0 than does the TS method. The $\widehat{\gamma}$ -values obtained by solving the equation $\widehat{FDR}(\gamma) = \alpha$ for $\alpha = 0.05$ are inversely ordered.

TABLE 1
Estimates of the Parameters for the Normal and Beta Models, Value of $\widehat{\gamma}$ and Number of Rejected Hypotheses for the HCK, TS and EM Methods

	Normal Model			Beta Model		
	HCK	TS	EM	HCK	TS	EM
$\widehat{\pi}_0$	0.1317	0.3233	0.1407	0.0096	0.1307	0.0160
$\widehat{\gamma}$	0.3163	0.0918	0.2946	1.0000	0.3092	1.0000
$\widehat{\delta}$	1.8285	2.2657	1.9221	—	—	—
\widehat{a}	—	—	—	0.3291	0.4474	0.3210
\widehat{b}	—	—	—	2.0764	3.2842	1.9313
N_r	28	21	27	34	27	34

N_r = Number of rejected hypotheses

The p -values $\leq \widehat{\gamma}$ are declared significant. From Table 2, we see that the number of significant p -values for HCK, TS and EM for the normal model are 28, 21 and 27, respectively. Thus, HCK and EM methods give more rejections than Benjamini and Hochberg's (2000) adaptive SU procedure.

Before fitting the beta mixture model, it is useful to plot a histogram of the p -values. This histogram is shown in Fig. 11. It has a decreasing shape, and assuming that the majority of the p -values are non-null, it corresponds to $a < 1$ and $b > 1$.

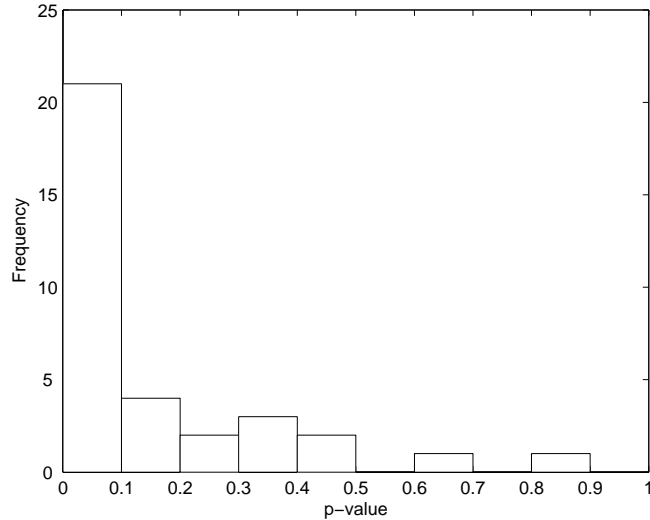


FIG 11. Histogram of the p -Values for the NAEP Data

HCK and EM methods yield $\hat{\pi}_0 < \alpha = 0.05$, hence $\hat{\gamma} = 1$ which means that all 34 hypotheses are rejected. This evidently liberal result is likely due to underestimation of π_0 using the beta model as noted in Section 5.2. The TS method yields $\hat{\pi}_0 = 0.1307$ and $\hat{\gamma} = 0.3092$, which are close to the estimates produced by the HCK and EM methods for the normal model and it rejects the same 27 hypotheses.

Rejections of hypotheses with large p -values will justifiably raise many eyebrows. This appears to be a problem with FDR-controlling procedures when there are many hypotheses that are clearly false (with p -values close to zero) which lowers the bar for rejection for other hypotheses. Shaffer (2005) has discussed this problem and has suggested imposing additional error controlling requirements in order to limit such dubious rejections. This is a topic for further research.

7. Concluding Remarks

In this paper we offered two different mixture models for estimating the number of true null hypotheses by modeling the non-null p -values. For each model (the normal and beta), three methods of estimation were developed: HCK, TS and EM. Generally speaking, these parametric estimators outperform (in terms of the accuracy of the estimate of π_0 and control of the FDR) the nonparametric ST estimator for the normal model but not for the beta model. The reason for this is that the normal model is easier to estimate and so the benefits of the parametric estimators are not significantly compromised by the errors of estimation. On the other hand, the beta model is difficult to estimate and so the benefits of the parametric estimators are lost. Therefore we do not recommend the use of the beta model in practice.

For normally distributed test statistics, the EM estimator generally performs best followed by the HCK and TS estimators. However, the EM estimator is not robust to the violation of the model assumptions. If the EM estimator for the normal model is applied to the data generated from the beta model or vice versa, its performance is often worse than that of the HCK estimator, and sometimes even that of the ST

TABLE 2
 NAEP Trial State Assessment: Test Results for the HCK, TS and EM Methods (Normal Model)

State	p -value	HCK	TS	EM	State	p -value	HCK	TS	EM
RI	0.00000	*	*	*	NY	0.05802	*	*	*
MN	0.00002	*	*	*	OH	0.06590	*	*	*
HI	0.00002	*	*	*	CA	0.07912	*	*	*
NC	0.00002	*	*	*	MD	0.08226	*	*	*
NH	0.00180	*	*	*	WV	0.10026	*	*	*
IA	0.00200	*	*	*	VA	0.14374	*	*	*
CO	0.00282	*	*	*	WI	0.15872	*	*	*
TX	0.00404	*	*	*	IN	0.19388	*	*	*
ID	0.00748	*	*	*	LA	0.20964	*	*	*
AZ	0.00904	*	*	*	MI	0.23522	*	*	*
KY	0.00964	*	*	*	DE	0.31162	*	*	*
OK	0.02036	*	*	*	ND	0.36890	*	*	*
CT	0.04104	*	*	*	NE	0.38640	*	*	*
NM	0.04650	*	*	*	NJ	0.41998	*	*	*
WY	0.04678	*	*	*	AL	0.44008	*	*	*
FL	0.05490	*	*	*	AR	0.60282	*	*	*
PA	0.05572	*	*	*	GA	0.85628	*	*	*

*Significant p -values are indicated by asterisks. For the beta model, the HCK and EM methods find all p -values significant, while the TS method finds the p -values less than $\hat{\gamma} = 0.3093$ significant, i.e., the same as those under the EM column in this table.

estimator. The TS estimator did not improve on the HCK estimator in all cases as we had hoped. Thus our final recommendation is to use the normal model with the EM method if the test statistics follow approximately normal distributions and the HCK method otherwise. If only the p -values calculated from various types of test statistics are available then the ST method is recommended; alternatively the p -values may be transformed using the inverse normal transform and then the HCK method may be applied.

Acknowledgment

This research was partially supported by the National Heart, Lung and Blood Institute Grant 1 R01 HL082725-01A1 and the National Security Agency Grant H98230-07-1-0068. The authors are extremely grateful especially to one of the two referees who pointed out some crucial problems in the earlier version of the paper.

Appendix A

Proof of Lemma 2. We have

$$\begin{aligned}
 E[\bar{X}_a(\lambda)] &= E\left\{\frac{1}{N_a} \sum_{i \in S_a(\lambda)} X_i\right\} \\
 &= E\left\{E\left[\frac{1}{n_a} \sum_{i \in s_a} X_i \mid S_a(\lambda) = s_a, N_a(\lambda) = n_a\right]\right\} \\
 &= E\left\{\frac{1}{n_a} \cdot n_a[g(\pi_0, \delta, \lambda)c_{0a}(\lambda) + [1 - g(\pi_0, \delta, \lambda)]c_{1a}(\delta, \lambda)]\right\} \\
 &= g(\pi_0, \delta, \lambda)c_{0a}(\lambda) + [1 - g(\pi_0, \delta, \lambda)]c_{1a}(\delta, \lambda).
 \end{aligned}$$

1 In the penultimate step above, we have used the fact that conditionally on $X_i \leq z_\lambda$, 1
 2 the probability that $Z_i = 1$ is $g(\pi_0, \delta, \lambda)$ and the probability that $Z_i = 0$ is 2
 3 $1 - g(\pi_0, \delta, \lambda)$. Furthermore, the conditional expectation of X_i in the first case 3
 4 is $c_{0a}(\lambda)$ and in the second case it is $c_{1a}(\delta, \lambda)$. The expression for $E[\bar{X}_r(\lambda)]$ follows 4
 5 similarly. 5
 6

7 *Proof of Lemma 5.* By substituting for $\beta(\cdot, \gamma)$ from (2.2) and dropping carets on 7
 8 $\widehat{\text{FDR}}(\gamma)$, $\widehat{\pi}_0$, $\widehat{\pi}_1$ and $\widehat{\delta}$ for notational convenience, the equation to be solved is 8
 9

$$10 \quad \text{FDR}(\gamma) = \frac{\pi_0}{\pi_0 + \pi_1 \Phi(\delta - z_\gamma)/\gamma} = \alpha. \quad 10$$

11
 12 It is easy to check that $\text{FDR}(0) = 0$ and $\text{FDR}(1) = \pi_0$. We shall show that $\text{FDR}(\gamma)$ 12
 13 is an increasing function of γ which will prove the lemma. Thus we need to show 13
 14 that $u(\delta, \gamma) = \Phi(\delta - z_\gamma)/\gamma$ is decreasing in γ . By implicit differentiation of the 14
 15 equation $\Phi(z_\gamma) = 1 - \gamma$, we get 15
 16

$$17 \quad \frac{dz_\gamma}{d\gamma} = -\frac{1}{\phi(z_\gamma)}. \quad 17$$

18
 19 Hence, 19

$$20 \quad \frac{du(\delta, \gamma)}{d\gamma} = \frac{\gamma\phi(\delta - z_\gamma) - \phi(z_\gamma)\Phi(\delta - z_\gamma)}{\gamma^2\phi(z_\gamma)}. \quad 20$$

21
 22 Therefore we need to show that 22
 23

$$24 \quad v(\delta, \gamma) = \phi(z_\gamma)\Phi(\delta - z_\gamma) - \gamma\phi(\delta - z_\gamma) > 0 \quad \forall \delta > 0. \quad 24$$

25
 26 Now $v(0, \gamma) = 0$. Therefore we must show that 26
 27

$$28 \quad \frac{dv(\delta, \gamma)}{d\delta} = \phi(\delta - z_\gamma)[\phi(z_\gamma) + \gamma(\delta - z_\gamma)] > 0, \quad 28$$

29
 30 which reduces to the condition: $w(\delta, \gamma) = \phi(z_\gamma) + \gamma(\delta - z_\gamma) > 0$. Since $w(\delta, \gamma)$ is 30
 31 increasing in δ , it suffices to show that 31

$$32 \quad w(0, \gamma) = \phi(z_\gamma) - \gamma z_\gamma > 0. \quad 32$$

33
 34 By putting $x = z_\gamma$ and hence $\gamma = \Phi(-x)$ the above inequality becomes 34
 35

$$36 \quad \frac{\Phi(-x)}{\phi(x)} < \frac{1}{x}, \quad 36$$

37
 38 which is the Mills' ratio inequality (Johnson and Kotz 1970, p. 279). This completes 38
 39 the proof of the lemma. 39
 40

41 *Proof of Lemma 6.* By substituting for $\beta(\cdot, \gamma)$ from (3.2) and dropping carets on 41
 42 $\widehat{\text{FDR}}(\gamma)$, $\widehat{\pi}_0$, $\widehat{\pi}_1$, \widehat{a} and \widehat{b} for notational convenience, the equation to be solved is 42
 43

$$44 \quad (\text{A.1}) \quad \text{FDR}(\gamma) = \frac{\pi_0}{\pi_0 + \pi_1 I_\gamma(a, b)/\gamma} = \alpha. \quad 44$$

45
 46 Note that $\text{FDR}(0) = 0$ and $\text{FDR}(1) = \pi_0$. To show that $\text{FDR}(\gamma)$ is an increasing 46
 47 function of γ we need to show that $I_\gamma(a, b)/\gamma$ decreases in γ . To see this, note that 47
 48 the derivative of $I_\gamma(a, b)/\gamma$ w.r.t. γ is proportional to $\gamma g(\gamma|a, b) - I_\gamma(a, b)$, which is 48
 49 negative since the beta p.d.f. $g(\gamma|a, b)$ is strictly decreasing in γ for $a < 1$ and $b > 1$, 49
 50 and so $\gamma g(\gamma|a, b) < I_\gamma(a, b)$. It follows therefore that the equation $\text{FDR}(\gamma) = \alpha$ has 50
 51 a unique solution in $\gamma \in (0, 1)$ for $\alpha \in (0, \pi_0]$. 51

References

- [1] BENJAMINI, Y. AND HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Ser. B*, **57**, 289–300.
- [2] BENJAMINI, Y. AND HOCHBERG, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational Statistics*, **25**, 60–83.
- [3] BLACK, M. A. (2004). A note on the adaptive control of false discovery rates. *Journal of Royal Statistical Society, Ser. B*, **66**, 297–304.
- [4] DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society, Ser. B*, **39**, 1–38.
- [5] FINNER, H. AND ROTERS, M. (2001). On the false discovery rate and expected type I error. *Biometrical Journal*, **8**, 985–1005.
- [6] GENOVESE, C. AND WASSERMAN, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *Journal of Royal Statistical Society, Ser. B*, **64**, 499–517.
- [7] GILL, P. E., MURRAY, W. AND WRIGHT, M. H. (1981). *Practical Optimization*. London; New York: Academic Press.
- [8] GUAN, Z., WU, B. AND ZHAO, H. (2004). Model-based approach to FDR estimation. *Tech. Report 2004-016*, Division of Biostatistics, Univ. of Minnesota, Minneapolis, MN.
- [9] HOCHBERG, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, **75**, 800–803.
- [10] HOCHBERG, Y. AND BENJAMINI, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in Medicine*, **9**, 811–818.
- [11] HOCHBERG, Y. AND TAMHANE, A. C. (1987). *Multiple Comparison Procedures*. New York: John Wiley and Sons.
- [12] HSUEH, H., CHEN, J. J. AND KODELL, R. L. (2003). Comparison of methods for estimating the number of true null hypotheses in multiple testing. *Journal of Biopharmaceutical Statistics*, **13**, 675–689.
- [13] HUNG, H. M., O'NEILL, R. T., BAUER, P. AND KOHNE, K. (1997). The behavior of the p-value when the alternative hypothesis is true. *Biometrics*, **53**, 11–22.
- [14] IYER, V. AND SARKAR, S. (2007). An adaptive single-step FDR procedure with applications to DNA microarray analysis. *Biometrical Journal*, **49**, 127–135.
- [15] JIANG, H. AND DOERGE, R. W. (2005). Estimating the proportion of the true null hypotheses for multiple comparisons. Preprint.
- [16] JOHNSON, N. L. AND KOTZ (1970), *Continuous Univariate Distributions-I*. New York: John Wiley & Sons, Inc.
- [17] LANGAAS, M., LINDQUIST, B. H. AND FERKINGSTAD, E. (2004). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *Journal of Royal Statistical Society, Ser. B*, **67**, 555–572.
- [18] SCHWEDER, T. AND SPJØTVOLL, E. (1982). Plots of p -values to evaluate many tests simultaneously. *Biometrika*, **69**, 493–502.
- [19] SHAFFER, J. P. (2005). Multiple requirements for multiple test procedures. Paper presented at the IVth International Conference on Multiple Comparison Procedures, Shanghai, China.
- [20] SHI, J. (2006). Improved Estimation of the Proportion of True Null Hypotheses with Applications to Adaptive Control of FDR and Drug Screening. *Doctoral Dissertation*, Department of Statistics, Northwestern University, Evanston, IL.
- [21] STOREY, J. (2002). A direct approach to false discovery rates. *Journal of Royal Statistical Society, Ser. B*, **64**, 297–304.
- [22] STOREY, J., TAYLOR, J. E. AND SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach, *Journal of Royal Statistical Society, Ser. B*, **66**, 187–205.
- [23] TURKHEIMER, F. E., SMITH, C. B. AND SCHMIDT, K. (2001). Estimation of the number of 'true' null hypotheses in multivariate analysis of neuroimaging data. *NeuroImage*, **13**, 920–930.
- [24] WU, B., GUAN, Z. AND ZHAO, H. (2006). Parametric and nonparametric FDR estimation revisited. *Biometrics*, **62**, 735–744.