Department of Industrial Engineering and Management Sciences

Northwestern University, Evanston, Illinois 60208-3119, U.S.A.

Working Paper No. 07-03

A Mixture Model Approach to Estimating the Number of True Null Hypotheses and Adaptive Control of FDR

Ajit C. Tamhane

Jiaxiao Shi

March 2007

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.

ABSTRACT

Many methods have been proposed for estimating the number, m_0 (or the proportion, π_0), of the true null hypotheses to adaptively control a type I error rate (e.g., the false discovery rate or FDR) using a multiple test procedure. Most of these methods use a formal test (see, e.g., Storey 2002) or a graphical technique (see, e.g., Spjøtvoll and Schweder 1982) to eliminate "significantly" non-null p-values. Then m_0 is estimated from the remaining p-values by assuming that they follow the $U[0,1]$ distribution. Because these methods ignore that some "nonsignificant" p-values may come from alternative hypotheses (type II errors), they tend to overestimate m_0, and hence lead to a more conservative control of FDR. In this paper we propose to use all p-values to estimate m_0 by modelling them with a parametric mixture distribution following up on the finding by Wu, Guan and Zhao (2006) that nonparametric approaches are too conservative. Two different mixture distribution models are considered. The normal model assumes that the test statistics from the true null hypotheses are i.i.d. $N(0,1)$ and those from the alternative hypotheses are i.i.d. $N(\delta,1)$ with $\delta \neq 0$, and π_0 and $1-\pi_0$ as the mixing proportions. 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. Beta(a,b) with $a < b$. Three methods of estimation of π_0 and of the associated mixture distribution parameters are developed for each model. The methods are compared via simulation with each other and with Storey's 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. The EM algorithm (Dempster, Laird and Rubin 1977) estimator performs best overall when the assumed model holds, but it is not very robust to significant model violations. An example is given to illustrate the methods.

Keywords and Phrases: Beta distribution; Bias-correction; EM algorithm; Mixture model; False discovery rate (FDR); Normal distribution; Positive false discovery rate (pFDR); p-values

1. Introduction

Suppose that m null hypotheses, $H_{01}, \ldots, H_{0m}$, are to be tested against alternatives, $H_{11}, \ldots, H_{1m}$. Denote the test statistics by $X_1, \ldots, X_m$ and their p-values by $p_1, \ldots, p_m$. Also denote the ordered values of these quantities by $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(m)}$ and $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$. Let M_0 and M_1 be the index sets of the true and false null hypotheses, respectively, with $|M_0| = m_0$ and $|M_1| = m_1$ where M_0 and M_1 are unknown sets and $m_0 + m_1 = m$. Let $\pi_0 = m_0/m$ and $\pi_1 = m_1/m$ be the corresponding unknown proportions where $\pi_0 + \pi_1 = 1$.

In many applications m_0 is much smaller than m, which is often used as a conservative upper bound on the number of true null hypotheses (i.e., the least favorable configuration is that all null hypotheses are true) in multiple comparison procedures (MCP's) to control an appropriate type I error rate, e.g., the *familywise error rate (FWE)* in the Bonferroni procedure (Hochberg and Tamhane 1987) or the *false discovery rate (FDR)* in the Benjamini and Hochberg (1995) procedure. A more powerful MCP can be implemented if a suitable estimate $\widehat{m}_0$ of m_0 based on the observed p-values is used in place of m. Such *adaptive* procedures are especially useful in large-scale multiplicity testing problems arising in microarray data involving m of the order of several thousands.

A number of methods have been proposed for estimating m_0 or equivalently π_0; see Schweder and Spjøtvoll (1982), Hochberg and Benjamini (1990), Benjamini and Hochberg (2000), Turkheimer, Smith and Schmidt (2001), Storey (2002) and Jiang and Doerge (2005). All of these methods are based on the fact that the p-values from the null hypotheses are $U[0,1]$ random variables (r.v.'s). The p-values from the alternative hypotheses are not explicitly modelled. Any p-value that differs significantly from the null $U[0,1]$ distribution is rejected as non-null and excluded from the estimation process. Different formal and graphical tests are used for this purpose.

For example, Storey's (2002) method uses a λ-level test for a sufficiently large λ (e.g., $\lambda = 0.5$) to reject any p-value $\leq \lambda$ as non-null. Let $N_r(\lambda) = \sharp(p_i \leq \lambda)$ and $N_a(\lambda) = \sharp(p_i > \lambda)$ denote the number of rejected and accepted hypotheses, respectively, at level $\lambda \in (0,1)$. If type II errors are ignored for a sufficiently large λ then

$$E[N_a(\lambda)] \approx m_0(1-\lambda). \tag{1.1}$$

Storey's (ST) estimator is given by

$$\widehat{\pi}_0(\lambda) = \frac{N_a(\lambda)}{m(1-\lambda)}. \tag{1.2}$$

On the other hand, Schweder and Spjøtvoll (1982) proposed sequentially fitting straight lines through the origin to the plot of $N_a(p_{(i)}) = m - i$ vs. $1 - p_{(i)}$ $(1 \leq i \leq m)$, dropping the smallest $p_{(i)}$ as non-null at each step until the slope of the fitted line stops decreasing. Then from Equation (1.1) the final slope is taken as the estimate $\widehat{m}_0$. Because these estimators attribute all nonsignificant p-values to the true null hypotheses (type II errors are ignored), they tend to be positively biased which results in conservative adaptive control of any type I error rate.

The goal of the present paper is to propose two approaches to modelling non-null p-values, thus enabling their inclusion in the estimation of m_0; the parameters of these models also need to be estimated. We expect that estimators of m_0 derived from *all* p-values will be more accurate. We also study the effects of the proposed improved methods of estimation of m_0 on the adaptive control of FDR.

Hsueh, Chen, and Kodell (2003) (HCK) were the first to take into account non-null p-values. They assumed the following simple hypothesis testing setup. Suppose that all m hypotheses pertain to the means of the normal distributions with $H_{0i} : \mu_i = 0$ vs. $H_{1i} : \mu_i > 0$. (HCK considered a two-sided alternative, but that is not germane to the present discussion.) The test statistic $X_i \sim N(\delta_i, 1)$, where δ_i is the standardized μ_i. Let $\delta_i = 0$ for $i \in M_0$ and $\delta_i = \delta > 0$ for $i \in M_1$[1] where δ is assumed to be *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 (or equivalently p_i's) that are greater than (or equivalently less than) selected thresholds can be derived using this setup. Then HCK applied the least squares method to estimate the unknown parameter m_0 in this expression taking the corresponding observed number of X_i's as the dependent variable.

[1]Note that since M_1 is unknown, we need to assume a common non-null distribution in which case the effect of non-null p-values depends on M_1 only through its cardinality, $m_1 = m - m_0$, which is to be estimated.

The normal model is the topic of Section 2. However, we relax the assumption of known δ. We first extend the HCK estimation method to the *unknown* δ case. Next we note that the HCK method makes use of only the number of X_i's that are greater than selected thresholds; it does not make use of the magnitudes of the X_i's. Therefore we propose two alternative methods of estimation which use 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 least squares method of HCK, but uses the sample mean (instead of the number) of the X_i's that are greater than selected thresholds. We refer to it as the test statistics (TS) method. The second method is the EM algorithm of Dempster, Laird and Rubin (1977) which uses 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 violation.

In the second approach discussed in Section 3 the non-null p-values are modelled 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.

It should be pointed out that both the normal and beta 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 the fixed δ assumption is tested by generating different δ_i's for $i \in M_1$ from a normal distribution. Robustness of the normal model assumption is tested by generating p_i's for $i \in M_1$ using the beta model and converting them to the X_i's using the inverse normal transformation. Similarly, the robustness of the beta model is tested by generating X_i's using the normal model and transforming them to the p-values.

Adaptive control of the FDR using different estimators of m_0 is discussed in Section 4.

The ST, HCK, TS and EM estimators are compared in a large simulation study in Section 5. The performance measures used in the simulation study are the biases and mean square errors of the estimators of π_0 and FDR. An example illustrating application of the proposed methods is given in Section 6. Conclusions are summarized in Section 7. Proofs of some technical results are given in the Appendix.

2. Normal Model

The normal model assumes that the test statistics $X_i \sim N(0,1)$ for $i \in M_0$ and $X_i \sim N(\delta,1)$ for $i \in M_1$ where $\delta > 0$ is unknown. This is equivalent to the mixture model:

$$f(x_i) = \pi_0\phi(x_i) + \pi_1\phi(x_i - \delta), \tag{2.1}$$

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

2.1 Hsueh, Chen, and Kodell (HCK) Method of Estimation

Let

$$\beta(\delta,\lambda) = P_{H_{1i}}\{p_i > \lambda\} = P_{H_{1i}}\{X_i \le z_\lambda\} = \Phi\left(z_\lambda - \delta\right) \tag{2.2}$$

denote the type II error probability of a test performed at level λ where $\Phi(\cdot)$ is the standard normal c.d.f. and $z_\lambda = \Phi^{-1}(1-\lambda)$. We then have

$$E[N_a(\lambda)] = m_0(1-\lambda) + (m-m_0)\Phi\left(z_\lambda - \delta\right) \text{ and } E[N_r(\lambda)] = m_0\lambda + (m-m_0)\Phi\left(-z_\lambda + \delta\right). \tag{2.3}$$

HCK used the following method to estimate m_0: Rewrite the equation for $E[N_r(\lambda)]$ as

$$E[N_r(\lambda)] - m\Phi\left(-z_\lambda + \delta\right) = m_0[\lambda - \Phi\left(-z_\lambda + \delta\right)]. \tag{2.4}$$

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

$$x_i = p_{(i)} - \Phi\left(-z_{p_{(i)}} + \delta\right) \tag{2.5}$$

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

$$y_i = i - m\Phi\left(-z_{p_{(i)}} + \delta\right). \tag{2.6}$$

If δ is known then we can calculate $(x_i, y_i),\ i = 1, 2, \ldots, m$, and using (2.4) fit a straight line of y_i vs. x_i through the origin with slope m_0. The least squares estimator of m_0 is given by

$$\widehat{m}_0 = \frac{\sum_{i=1}^{m} x_i y_i}{\sum_{i=1}^{m} x_i^2}. \tag{2.7}$$

We propose to extend the HCK estimator to the unknown δ case by incorporating estimation of δ as part of the following nonlinear least squares problem:

$$\underset{m_0,\delta}{\text{Minimize}} \left[\sum_{i=1}^{m}(y_i - m_0 x_i)^2 = \sum_{i=1}^{m}\left\{i - m\Phi\left(-z_{p_{(i)}} + \delta\right) - m_0\left[p_{(i)} - \Phi\left(-z_{p_{(i)}} + \delta\right)\right]\right\}^2\right]. \tag{2.8}$$

We used the iterative algorithm given below to solve this problem and obtain the estimates $\widehat{m}_0$ and $\widehat{\delta}$. The initial starting values for this algorithm as well as the algorithms for TS and EM estimators were determined by solving the following two moment equations for m_0 and δ:

$$\sum_{i=1}^{m} X_i = m_1\delta \text{ and } \sum_{i=1}^{m} X_i^2 = m_0 + m_1(\delta^2 + 1). \tag{2.9}$$

HCK Algorithm

Step 0: Initialize $\widehat{m}_0$ and $\widehat{\delta}$ by solving (2.9). Let $\widehat{\pi}_0 = \widehat{m}_0/m$.

Step 1: Set $\delta = \widehat{\delta}$ and compute $(x_i, y_i),\ i = 1, 2, \ldots, m$, using (2.5) and (2.6).

Step 2: Compute $\widehat{m}_0$ using (2.7) and $\widehat{\pi}_0 = \widehat{m}_0/m$.

Step 3: Find $\widehat{\delta}$ to minimize (2.8).

Step 4: Return to Step 1 until convergence.

Remark: One could use a generalized least squares estimation method to take into account heteroscedasticity of the y_i's. However, the resulting nonlinear least squares problem is computationally much more intensive without a collateral gain in the efficiency of the estimates.

2.2 Test Statistics (TS) Method of Estimation

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 (or equivalently not significant) at a specified level λ.

Define

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

Then $N_a(\lambda) = |S_a(\lambda)|$ and $N_r(\lambda) = |S_r(\lambda)|$ are the corresponding numbers of "accepted" and "rejected" hypotheses at level λ. Finally define

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

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

Lemma 1 *Define*

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

and

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

Then we have

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

and

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

Proof: We will derive general expressions for $E(X|X \leq c)$ and $E(X|X > c)$ where $X \sim N(\mu, 1)$. First we have,

$$\begin{aligned} E(X|X \leq c) &= \frac{1}{\Phi(c-\mu)} \int_{-\infty}^{c} x\phi(x-\mu)dx \\ &= \frac{1}{\Phi(c-\mu)} \int_{-\infty}^{c} [\mu + (x-\mu)]\phi(x-\mu)dx \\ &= \mu + \frac{1}{\Phi(c-\mu)} \int_{-\infty}^{c} (x-\mu)\phi(x-\mu)dx. \end{aligned}$$

Make a change of variables $y = \phi(x-\mu)$. Then $dy = -(x-\mu)\phi(x-\mu)dx$. Hence,

$$E(X|X \le c) = \mu - \frac{\phi(c-\mu)}{\Phi(c-\mu)}.$$

Similarly, we can show that

$$E(X|X > c) = \mu + \frac{\phi(\mu-c)}{\Phi(\mu-c)}.$$

The expressions E_{0a}, E_{1a} etc. follow by substitution. ■

The desired expected values are then given by the following lemma.

Lemma 2 *Let*

$$g(\pi_0,\delta,\lambda) = P\left\{i \in M_0 | X_i < z_\lambda\right\} = \frac{\pi_0(1-\lambda)}{\pi_0(1-\lambda) + \pi_1\Phi\left(z_\lambda - \delta\right)} \tag{2.10}$$

and

$$h(\pi_0,\delta,\lambda) = P\left\{i \in M_0 | X_i \ge z_\lambda\right\} = \frac{\pi_0\lambda}{\pi_0\lambda + \pi_1\Phi\left(-z_\lambda + \delta\right)}. \tag{2.11}$$

Then

$$E[\overline{X}_a(\lambda)] = g(\pi_0,\delta,\lambda)E_{0a}(\lambda) + [1 - g(\pi_0,\delta,\lambda)]E_{1a}(\delta,\lambda) \tag{2.12}$$

and

$$E[\overline{X}_r(\lambda)] = h(\pi_0,\delta,\lambda)E_{0r}(\lambda) + [1 - h(\pi_0,\delta,\lambda)]E_{1r}(\delta,\lambda), \tag{2.13}$$

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

Proof: We have

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

In the penultimate step above, we have used the fact that conditionally on $X_i \le z_\lambda$, the probability that $i \in M_0$ is $g(\pi_0,\delta,\lambda)$ and the probability that $i \in M_1$ is $1 - g(\pi_0,\delta,\lambda)$.

Furthermore, the conditional expectation of X_i in the first case is $E_{0a}(\lambda)$ and in the second case it is $E_{1a}(\delta, \lambda)$. The expression for $E[\overline{X}_r(\lambda)]$ follows similarly. $\square$

To develop an estimation method analogous to the HCK method note that from (2.3) and (2.13) we get

$$E[N_r(\lambda)]E[\overline{X}_r(\lambda)] - m\Phi\left(-z_{p_{(i)}} + \delta\right) E_{1r}(\delta, \lambda) = m_0\left[\lambda E_{0r}(\lambda) - \Phi\left(-z_{p_{(i)}} + \delta\right) E_{1r}(\delta, \lambda)\right]. \tag{2.14}$$

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

$$x_i = p_{(i)} E_{0r}(p_{(i)}) - \Phi\left(-z_{p_{(i)}} + \delta\right) E_{1r}(\delta, p_{(i)}) \tag{2.15}$$

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

$$y_i = i\overline{X}_r(p_{(i)}) - m\Phi\left(-z_{p_{(i)}} + \delta\right) E_{1r}(\delta, p_{(i)}) = \sum_{j=m-i+1}^{m} X_{(j)} - m\Phi\left(-z_{p_{(i)}} + \delta\right) E_{1r}(\delta, p_{(i)}). \tag{2.16}$$

Then from (2.14) we see that a regression line of y_i vs. x_i can be fitted through the origin with slope m_0. This leads to the following nonlinear least squares problem:

$$\begin{aligned}\underset{m_0,\delta}{\text{Minimize}} \Bigg[\sum_{i=1}^{m}(y_i - m_0 x_i)^2 = \sum_{i=1}^{m}\Big[i\overline{X}_r(p_{(i)}) - m\Phi\left(-z_{p_{(i)}} + \delta\right) E_{1r}(\delta, p_{(i)}) \\ -m_0\{p_{(i)}E_{0r}(p_{(i)}) - \Phi\left(-z_{p_{(i)}} + \delta\right) E_{1r}(\delta, p_{(i)})\}\Big]^2\Bigg].\end{aligned} \tag{2.17}$$

We can use an iterative algorithm analogous to that used for the HCK estimator to perform this minimization.

TS Algorithm

Step 0: Initialize $\widehat{m}_0$ and $\widehat{\delta}$ by solving (2.9). Let $\widehat{\pi}_0 = \widehat{m}_0/m$.

Step 1: Set $\delta = \widehat{\delta}$ and compute $(x_i, y_i),\ i = 1, 2, \ldots, m$, using (2.15) and (2.16).

Step 2: Compute $\widehat{m}_0$ using (2.7) and $\widehat{\pi}_0 = \widehat{m}_0/m$.

Step 3: Find $\widehat{\delta}$ to minimize (2.17).

Step 4: Return to Step 1 until convergence.

2.3 EM Method of Estimation

The steps in the EM algorithm are as follows.

Step 0: Initialize $\widehat{m}_0$ and $\widehat{\delta}$ by solving (2.9). Let $\widehat{\pi}_0 = \widehat{m}_0/m$.

Step 1: Calculate the posterior probabilities:

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

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

Step 2: Calculate new estimates:

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

Step 3: Return to Step 1 until convergence.

2.4 Bias-Corrected ST Estimator

It is possible to correct the ST estimator (1.2) for bias as follows. Use the estimate $\widehat{\delta}$ obtained from any of the above three methods and calculate the estimated type II error as $\beta(\widehat{\delta}, \lambda)$ for specified λ. Then the bias-corrected ST estimator is given by

$$\widehat{\pi}_0(\lambda) = \frac{N_a(\lambda)/m - \beta(\widehat{\delta}, \lambda)}{1 - \lambda - \beta(\widehat{\delta}, \lambda)} = \frac{N_a(\lambda)/m - \Phi\left(z_\lambda - \widehat{\delta}\right)}{1 - \lambda - \Phi\left(z_\lambda - \widehat{\delta}\right)}. \tag{2.18}$$

However, since the above three methods also yield estimators of π_0, we did not consider this additional estimator.

3. Beta Model

In many applications the normal model may be inappropriate because the test statistics may not be normally distributed or different test statistics may be used for testing different hypotheses or only the p-values of the test statistics may be available. In these cases we propose to model the non-null p-values by a Beta(a, b) distribution given by

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

with unknown parameters a and b. Since the distribution of non-null p-values must be right-skewed and generally decreasing in shape (see Hung, O'Neill, Bauer and Kohne 1997), we restrict to $a < 1$ and $b > 1$ (see Johnson and Kotz 1970, Figure 1a, pp. 42-43). 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:

$$f(p_i) = \pi_0 \times 1 + \pi_1 f(p_i|a,b). \tag{3.1}$$

The parameters a and b must be estimated along with π_0.

3.1 Hsueh, Chen, and Kodell (HCK) Method of Estimation

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

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

where $I_\lambda(a, b)$ is the incomplete beta function. Then it is easy to show that the nonlinear least squares problem (2.8) becomes

$$\underset{m_0,a,b}{\text{Minimize}} \left[\sum_{i=1}^m (y_i - m_0 x_i)^2 = \sum_{i=1}^m \left\{ i - mI_{p_{(i)}}(a,b) - m_0[p_{(i)} - I_{p_{(i)}}(a,b)] \right\}^2 \right] \tag{3.3}$$

subject to $a < 1 < b$, where

$$x_i = p_{(i)} - I_{p_{(i)}}(a,b) \text{ and } y_i = i - mI_{p_{(i)}}(a,b). \tag{3.4}$$

The HCK algorithm for the normal model can be modified to minimize $\sum_{i=1}^m (y_i - m_0 x_i)^2$ by replacing the minimization with respect to δ by minimization with respect to (a, b). Gauss-Newton method was used to perform this minimization (Gill et al. 1981). The initial starting values for this algorithm as well as the algorithms for TS and EM estimators described below were determined by solving the following three moment equations for m_0 and (a, b):

$$\begin{aligned}\sum_{i=1}^{m} p_i &= \frac{1}{2}m_0 + \frac{a}{a+b}m_1, \\ \sum_{i=1}^{m} p_i^2 &= \frac{1}{3}m_0 + \frac{a(a+1)}{(a+b)(a+b+1)}m_1, \\ \sum_{i=1}^{m} p_i^3 &= \frac{1}{4}m_0 + \frac{a(a+1)(a+2)}{(a+b)(a+b+1)(a+b+2)}m_1.\end{aligned} \tag{3.5}$$

3.2 Test Statistics (TS) Method of Estimation

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

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

Analogous to Lemma 1, we have

Lemma 3 *Define*

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

and

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

Then we have

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

and

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

Proof: The formulae for $E_{0a}(\lambda)$ and $E_{1a}(\lambda)$ are straightforward. Next consider

$$\begin{aligned}E_{1a}(a, b, \lambda) &= \frac{1}{1 - I_\lambda(a, b)} \int_\lambda^1 p f(p|a, b) dp \\ &= \frac{1}{1 - I_\lambda(a, b)} \int_\lambda^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^a (1-p)^{b-1} dp \\ &= \frac{1}{1 - I_\lambda(a, b)} \cdot \frac{a}{a+b} \int_\lambda^1 \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} p^a (1-p)^{b-1} dp \\ &= \frac{1 - I_\lambda(a+1, b)}{1 - I_\lambda(a, b)} \cdot \frac{a}{a+b}.\end{aligned}$$

The expression for $E_{1r}(a, b, \lambda)$ follows similarly. $\square$

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

Lemma 4 *Let*

$$g(\pi_0, a, b, \lambda) = P\{i \in M_0 | p_i > \lambda\} = \frac{\pi_0(1-\lambda)}{\pi_0(1-\lambda) + \pi_1[1 - I_\lambda(a, b)]}$$

and

$$h(\pi_0, a, b, \lambda) = P\{i \in M_0 | p_i \leq \lambda\} = \frac{\pi_0 \lambda}{\pi_0 \lambda + \pi_1 I_\lambda(a, b)}.$$

Then

$$E[\overline{p}_a(\lambda)] = g(\pi_0, a, b, \lambda) E_{0a}(\lambda) + [1 - g(\pi_0, a, b, \lambda)] E_{1a}(a, b, \lambda) \tag{3.6}$$

and

$$E[\overline{p}_r(\lambda)] = h(\pi_0, a, b, \lambda) E_{0r}(\lambda) + [1 - h(\pi_0, a, b, \lambda)] E_{1r}(a, b, \lambda), \tag{3.7}$$

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

The equations for the TS estimator are derived as follows. Analogous to (2.14), we obtain

$$E[N_r(\lambda)] E[\overline{p}_r(\lambda)] - m I_\lambda(a, b) E_{1r}(a, b, \lambda) = m_0[\lambda E_{0r}(\lambda) - I_\lambda(a, b) E_{1r}(a, b, \lambda)].$$

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

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

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

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

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

3.3 EM Method of Estimation

The steps in the EM algorithm are as follows.

Step 0: Initialize $\widehat{m}_0$ and $(\widehat{a}, \widehat{b})$ by solving (3.6). Let $\widehat{\pi}_0 = \widehat{m}_0/m$.

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

$$\widehat{\pi}_0(p_i) = \frac{\widehat{\pi}_0}{\widehat{\pi}_0 + \widehat{\pi}_1 f(p_i|\widehat{a}, \widehat{b})}$$

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

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

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

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

where $\psi(\cdot)$ is the digamma function. Also calculate

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

Step 3: Return to Step 1 until convergence.

4. Adaptive Control of FDR

The estimate $\widehat{m}_0$ can be used in any multiple testing procedure for adaptively controlling a suitable error rate. Here we focus on the control of FDR. Let R be the total number of rejections when testing m hypotheses and let V be the number of true hypotheses that are rejected. The proportion of rejected hypotheses that are false is V/R, where $0/0$ is defined as 0. Benjamini and Hochberg (1995) introduced the definition

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

Storey (2002) argued that FDR is meaningful only when there is at least one rejection; therefore he suggested an alternative error rate which he called positive FDR (pFDR), defined as

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

Thus,

$$\text{FDR} = \text{pFDR} \cdot P(R > 0).$$

Benjamini and Hochberg (1995) gave a step-up (SU) procedure that controls FDR $\leq \alpha$.

Storey (2002) proposed a single-step (SS) procedure (in contrast to Benjamini and Hochberg's (1995) step-up (SU) procedure) that rejects H_{0i} if $p_i \leq \gamma$ for some common fixed threshold γ. He showed that if all tests are independent then

$$\text{pFDR}(\gamma) = \frac{\pi_0 \gamma}{P\{p_i \leq \gamma\}}.$$

He gave the following nonparametric estimators for pFDR and FDR:

$$\widehat{\text{pFDR}}_{\lambda}(\gamma) = \frac{N_a(\lambda)\gamma}{(N_r(\gamma) \vee 1)(1-\lambda)[1-(1-\gamma)^m]}$$

and

$$\widehat{\text{FDR}}_{\lambda}(\gamma) = \frac{N_a(\lambda)\gamma}{(N_r(\gamma) \vee 1)(1-\lambda)}.$$

In the above, $x \vee y = \max(x, y)$ and $[1-(1-\gamma)^m]$ is a lower bound on $P(R > 0)$ and hence $\gamma/[1-(1-\gamma)^m]$ is an upper bound on the probability that a true null hypothesis is rejected at level γ conditional on there being at least one rejection.

We propose the following parametric estimator:

$$\widehat{\text{FDR}}(\gamma) = \frac{\widehat{\pi}_0 \gamma [1-(1-\gamma)^m]}{\widehat{\pi}_0 \gamma + \widehat{\pi}_1 [1-\beta(\cdot, \gamma)]}, \tag{4.1}$$

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 by (3.2) for the beta model. To control FDR at level α, we find the largest γ that satisfies

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

and test each hypothesis at that γ-level. We may confine attention to $\alpha \leq \pi_0$ since if $\alpha > \pi_0$ then one can choose $\gamma = 1$, and reject all hypotheses while still controlling FDR $= \pi_0 < \alpha$. Existence and uniqueness of γ for $\alpha \in (0, \pi_0]$ is proved in two lemmas in the Appendix — Lemma 5 for the normal model and Lemma 6 for the beta model.

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 $m \to \infty$) equivalent to Storey's SS procedure that rejects H_{0i} if $p_i \leq \gamma$ where γ is the solution to the equation

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

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

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

By writing $F(\gamma) = 1-\beta(\cdot,\gamma)$, we see that Equations (4.2) and (4.3) are identical if the factor $[1-(1-\gamma)^m]$ in (4.2) is ignored and π_0 is replaced by $\widehat{\pi}_0$ in (4.3). Iyer and Sarkar (2007) used the solution γ from (4.3) in Storey's SS procedure with $F(\gamma) = \Phi(\delta - z_\gamma)$ and δ and π_0 replaced by their estimates $\widehat{\delta}$ and $\widehat{\pi}_0$ obtained from the EM algorithm, which results in an adaptive FDR-controlling procedure virtually 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 nominal $\alpha = 0.10$ using the SS procedure. The bias and mean square error of the estimators were used as performance measures. The results for the normal model are presented in Section 5.1 and those for the beta model are presented in Section 5.2. Throughout we used $m = 1000$ and the number of replications were also set equal to 1000. We varied π_0 from 0.1 to 0.9 in steps of 0.1. The values $\pi_0 = 0$ and 1 were excluded because $\widehat{\pi}_0$ exhibits uneven results in some cases. Also, when $\pi_0 = 0$, there are no type I errors, so there is no FDR to estimate.

The bias of each $\widehat{\pi}_0$ estimator was estimated as the difference between the average of the $\widehat{\pi}_0$ values over 1000 replicates and the true value of π_0 (in the case of FDR, the bias was estimated as the difference between the average of the $\widehat{\text{FDR}}$ values over 1000 replicates and the nominal value of FDR $= \alpha = 0.10$). The MSE was computed as the sum of the square of the bias and the variance of the $\widehat{\pi}_0$ (or $\widehat{\text{FDR}}$) values over 1000 replicates. The detailed numerical results are given in Shi (2006); here we only present graphical plots to save space.

5.1 Simulation Results for the Normal Model

Simulations were conducted in three parts. In the first part, the true model for the non-null hypotheses was set to be the same as the assumed model by generating the X_i's for $i \in M_1$ from a $N(\delta, \sigma^2)$ distribution with a fixed $\delta = 2$ and $\sigma = 1$ (where the set M_1 of size $m_1 = \pi_1 m$ was fixed by fixing $\pi_1 = 1 - \pi_0$). In the other two parts of simulations, robustness of the assumed model was tested by generating the X_i's for $i \in M_1$ from different distributions than the assumed one. In the second part, the X_i's for $i \in M_1$ were generated from $N(\delta_i, \sigma^2)$ distributions where the δ_i's were themselves drawn from a $N(\delta_0, \sigma_0^2)$ distribution with $\delta_0 = 2$ and $\sigma_0 = 0.25$ corresponding to an approximate range of $[1, 3]$ for the δ_i (additional results for $\delta_0 = 2$ and $\sigma_0 = 0.5$ corresponding to an approximate range of $[0, 4]$ for the δ_i for $i \in M_1$ are given in Shi 2006). In the third part, the p_i's for $i \in M_1$ were generated from a Beta(a, b) distribution with $a = 0.5$ and $b = 2$, and the X_i's were computed using the inverse normal transformation $X_i = \Phi^{-1}(1 - p_i)$.

Results for Fixed δ: The bias and the square root of the mean square error ($\sqrt{\text{MSE}}$) of $\widehat{\pi}_0$ for ST, HCK, TS and EM estimators are plotted in Figure 1. Note from Equation (2.3) (also see Equation (3) of Black 2004) that the bias of the ST estimator is given by

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

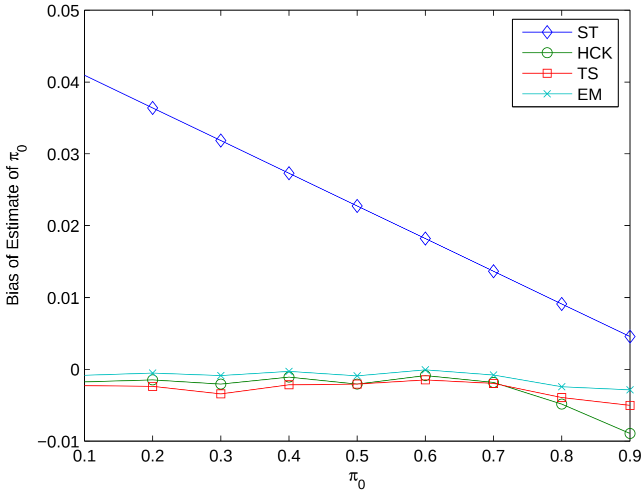
Also, using the fact that $N_a(\lambda)$ has a binomial distribution with number of trials m, and success probability,

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

we have

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

These formulae were used to compute the bias and MSE of the ST estimator instead of estimating them by simulation. Note that the bias of the ST estimator decreases linearly in



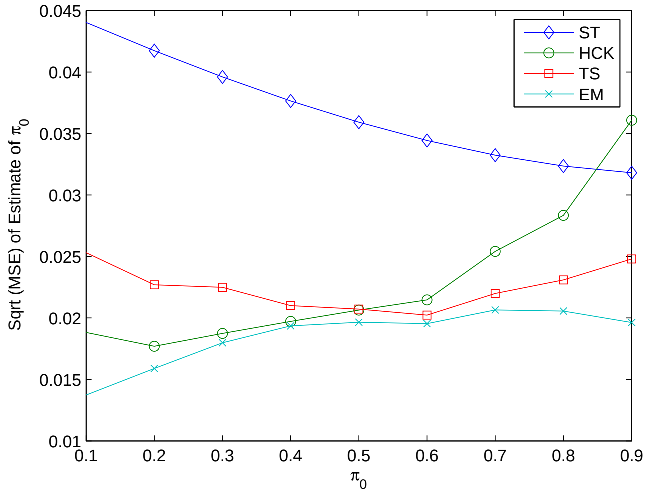


Figure 1: Bias and $\sqrt{\text{MSE}}$ of $\widehat{\pi}_0$ for ST, HCK, TS and EM Estimators (Normal Model, Fixed δ)

π_0. The $\sqrt{\text{MSE}}$ plot for the ST estimator is also approximately linear because the bias is the dominant term in MSE; the variance of $\widehat{\pi}_0(\lambda)$ is relatively small. This observation holds whenever the alternative is fixed for all false null hypotheses.

The TS estimator does not offer an improvement over the HCK estimator, as we had expected, and in fact performs worse in terms of MSE for $\pi_0 \leq 0.5$. We suspect that this result is due to the bias introduced when the term $E[N_r(\lambda)]E[\overline{X}_r(\lambda)]$ in Equation (2.14) is estimated by $i\overline{X}_r(p_{(i)})$ for $\lambda = p_{(i)}$ because of the fact that the product of the expected values does not equal to the expected value of the product of two dependent r.v.'s. The EM estimator has consistently the lowest bias and also the lowest MSE.

The bias and $\sqrt{\text{MSE}}$ of $\widehat{\text{FDR}}$ for ST, HCK, TS and EM estimators are plotted in Figure 2. We see that the ST estimator leads to a large negative bias which means that, on the average, $\widehat{\text{FDR}}$ 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 MSE's of the four estimators. The EM estimator still is the best choice with low bias and MSE throughout the entire range of π_0 values.

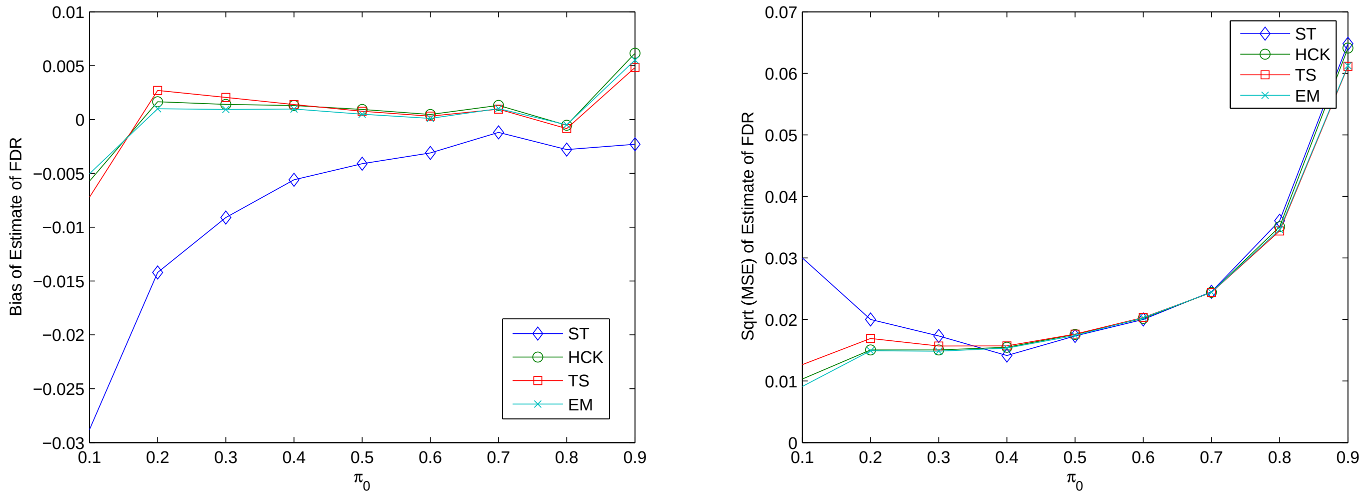


Figure 2: Bias and $\sqrt{\text{MSE}}$ of $\widehat{\text{FDR}}$ for ST, HCK, TS and EM Estimators (Normal Model, Fixed δ)

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

Robustness Results for Data Generated by Beta Model: The bias and $\sqrt{\text{MSE}}$ of $\widehat{\pi}_0$ and of $\widehat{\text{FDR}}$ for ST, HCK, TS and EM estimators are plotted in Figures 5 and 6, respectively. Looking at Figure 5 first, we see that the biases and MSE's of all four estimators are an order of magnitude higher compared to the normal model data which reflects a lack of robustness. The ST estimator has a high positive bias throughout. It is interesting to note that the EM estimator is no longer uniformly best for estimating π_0. In fact, the HCK estimator has a lower bias and MSE for $0.2 \leq \pi_0 \leq 0.7$. As far as control of FDR is concerned, there are not large differences between the estimators. In conclusion, the HCK estimator performs best for the middle range of π_0 values.

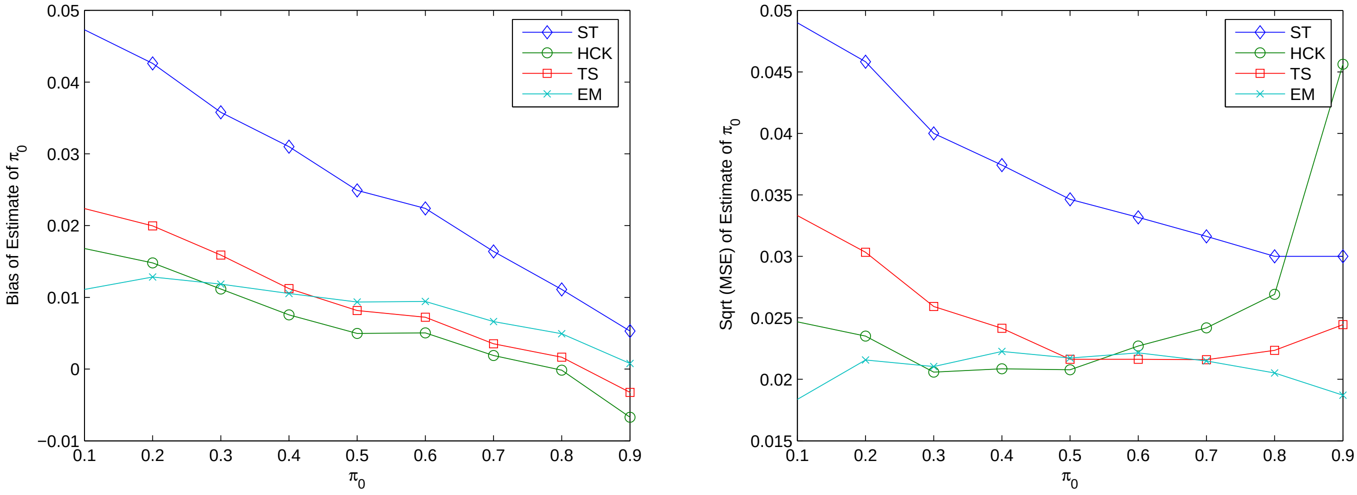


Figure 3: Bias and $\sqrt{\text{MSE}}$ of $\widehat{\pi}_0$ for ST, HCK, TS and EM Estimators (Normal Model, Random δ)

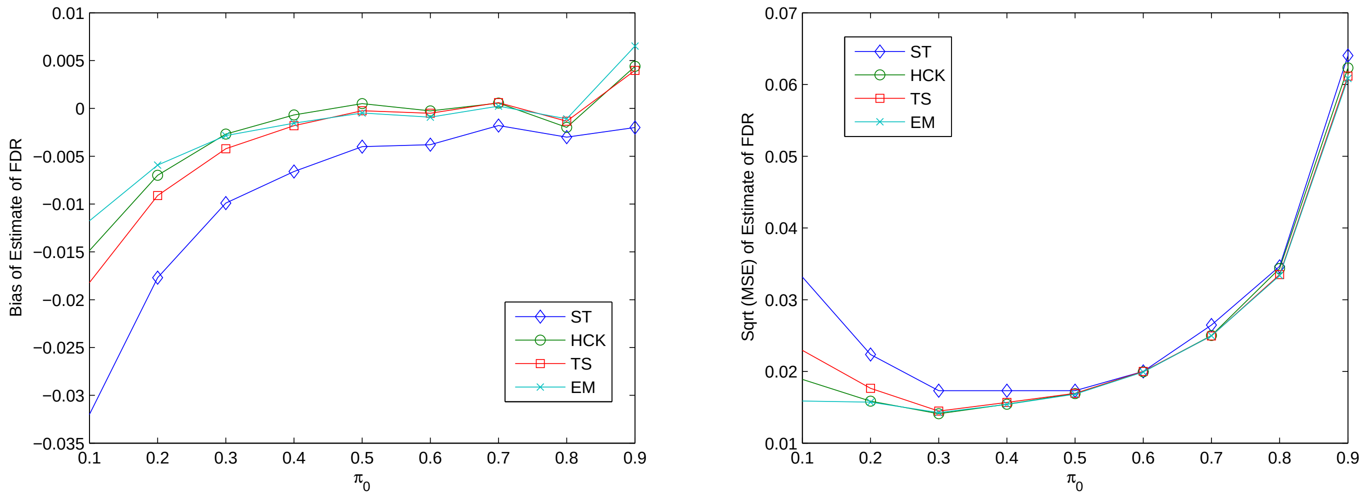


Figure 4: Bias and $\sqrt{\text{MSE}}$ of $\widehat{\text{FDR}}$ for ST, HCK, TS and EM Estimators (Normal Model, Random δ)

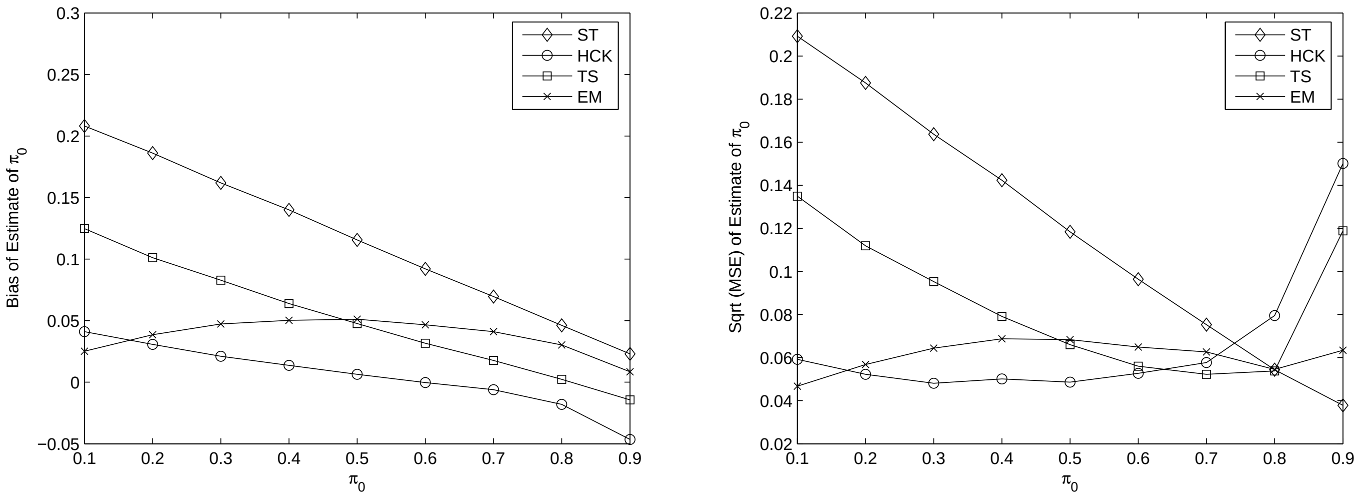


Figure 5: Bias and $\sqrt{\text{MSE}}$ of $\widehat{\pi}_0$ for ST, HCK, TS and EM Estimators (Normal Model, Data Generated by Beta Model)

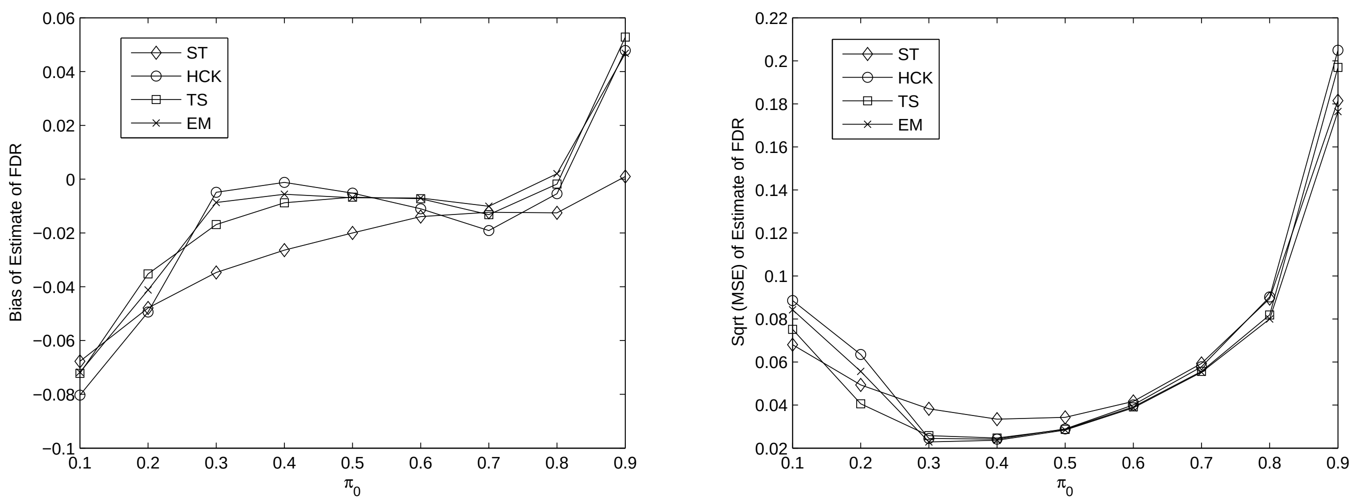


Figure 6: Bias and $\sqrt{\text{MSE}}$ of $\widehat{\text{FDR}}$ for ST, HCK, TS and EM Estimators (Normal Model, Data Generated by Beta Model)

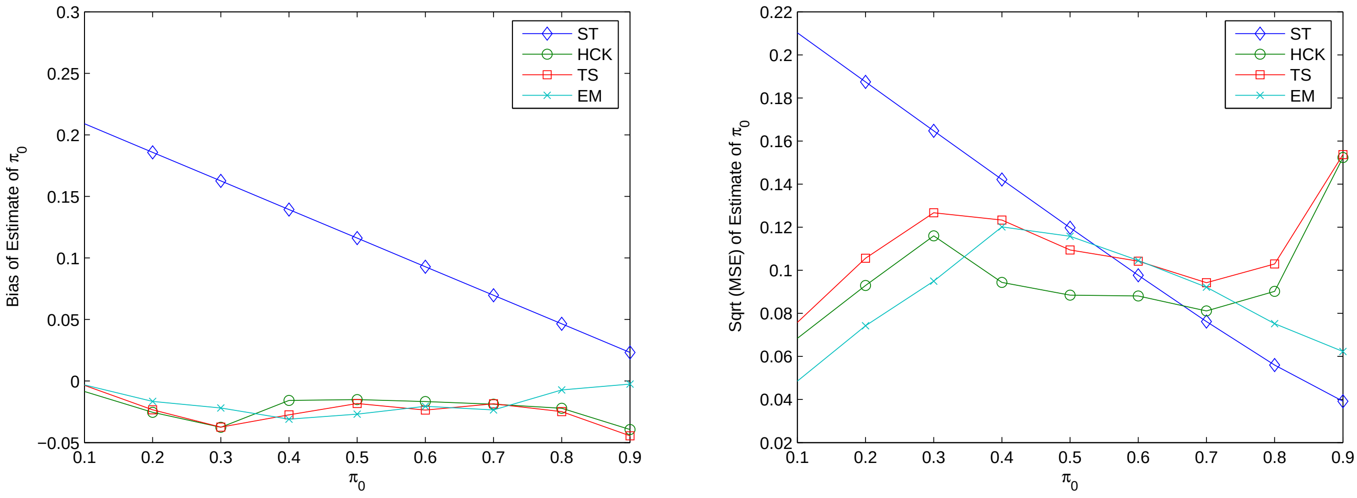


Figure 7: Bias and $\sqrt{\text{MSE}}$ of $\widehat{\pi}_0$ for ST, HCK, TS and EM Estimators (Beta Model)

5.2 Simulation Results for the 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 the formulae (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. The results are plotted in Figures 7 and 8. From these figures we see that the EM estimator performs best overall. Comparing the results here with the corresponding results for the normal model with the fixed δ case, we see that the biases and MSE's of all estimators are an order of magnitude higher in the present case. Thus, when the assumed model holds, the normal model estimators appear to perform much better.

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 Figures 9 and 10. From these figures we see that the EM estimator performs rather poorly.

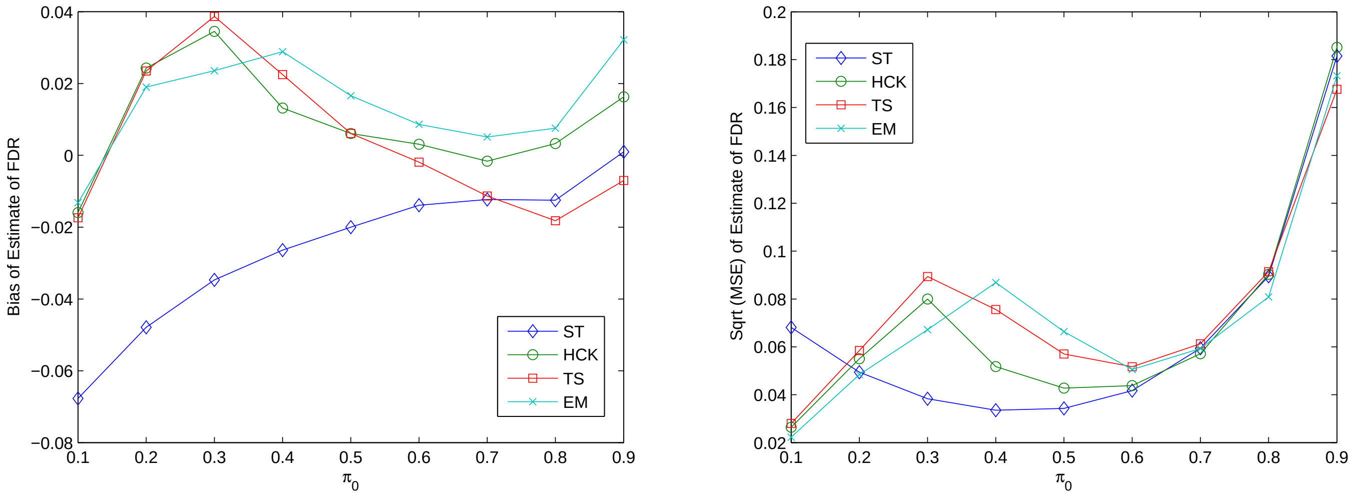


Figure 8: Bias and $\sqrt{\text{MSE}}$ of $\widehat{\text{FDR}}$ for ST, HCK, TS and EM Estimators (Beta Model)

Thus lack of robustness of the EM estimator to the model assumptions is demonstrated again. The TS estimator generally has the lowest bias for estimating π_0 and its achieved FDR is closest to the nominal α; other methods are more conservative.

Comparing the robustness results here with the corresponding robustness results for the normal model estimators when the data are generated by the beta model, we again see the phenomenon noted earlier that the performance of the estimators in the present case is much poorer. This suggests preference for the normal model estimators over the beta model estimators.

6. Example

We consider the National Assessment of Educational Progress (NAEP) data analyzed by Benjamini and Hochberg (2000). The data pertain to the change 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 Bonferroni and Hochberg (1988) FWE controlling procedure identified only 4 significant results (those with p-values $\leq p_{(4)} = 0.0002$) Application of the FDR controlling non-adaptive BH procedure resulted in 11 significant results. By applying

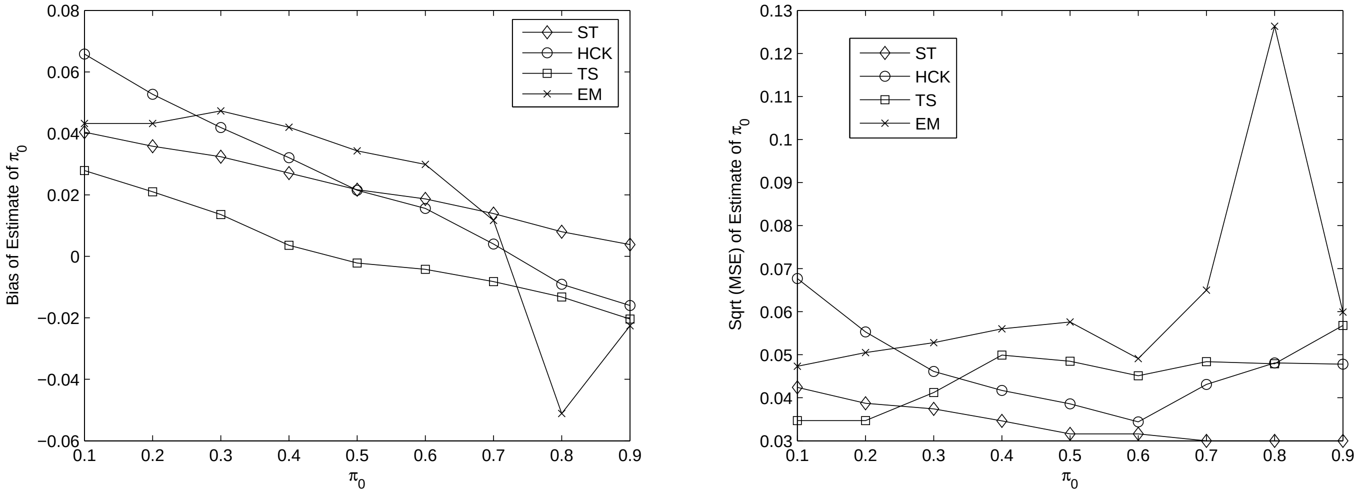


Figure 9: Bias and $\sqrt{\mathrm{MSE}}$ of $\widehat{\pi}_0$ for ST, HCK, TS and EM Estimators (Beta Model, Data Generated by Normal Model)

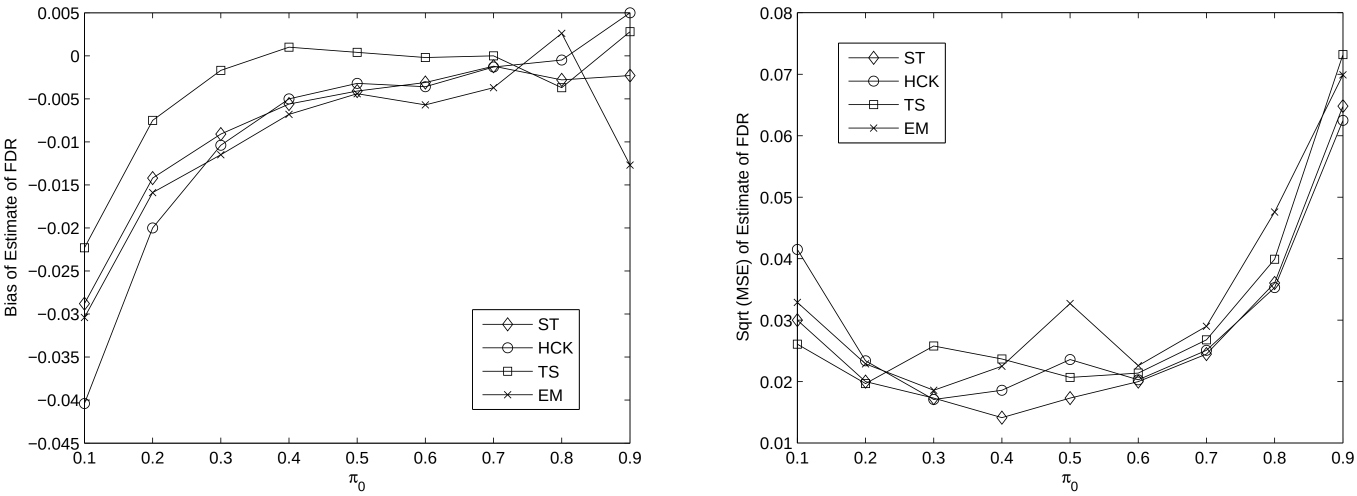


Figure 10: Bias and $\sqrt{\mathrm{MSE}}$ of $\widehat{\mathrm{FDR}}$ for ST, HCK, TS and EM Estimators (Beta Model, Data Generated by Normal Model)

their method they estimated $\widehat{m}_0 = 7$ ($\widehat{\pi}_0 = 0.2059$); using this value in the adaptive BH 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 γ-values obtained by solving the inequality (4.2) for $\alpha = 0.05$ are inversely ordered.

Table 1: Estimates of π_0, δ (for the Normal Model), a and b (for the Beta Model), Value of γ and Number of Rejected Hypotheses for 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
γ	0.3163	0.0948	0.2946	1.0000	0.3093	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 rejections

The p-values $\leq \gamma$ 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 method.

Before fitting the beta mixture model, it is useful to plot a histogram of the p-values. This histogram shown in Figure 11 has a decreasing shape, which (ignoring a few large p-values possibly from true null hypotheses) corresponds to $a < 1$ and $b > 1$. HCK and EM methods yield $\widehat{\pi}_0 < \alpha = 0.05$, hence $\gamma = 1$ which means that all 34 hypotheses are rejected. The TS method yields $\widehat{\pi}_0 = 0.1307$ and $\gamma = 0.3093$, which are close to the estimates produced by the HCK and EM methods for the normal model and it rejects 27 hypotheses.

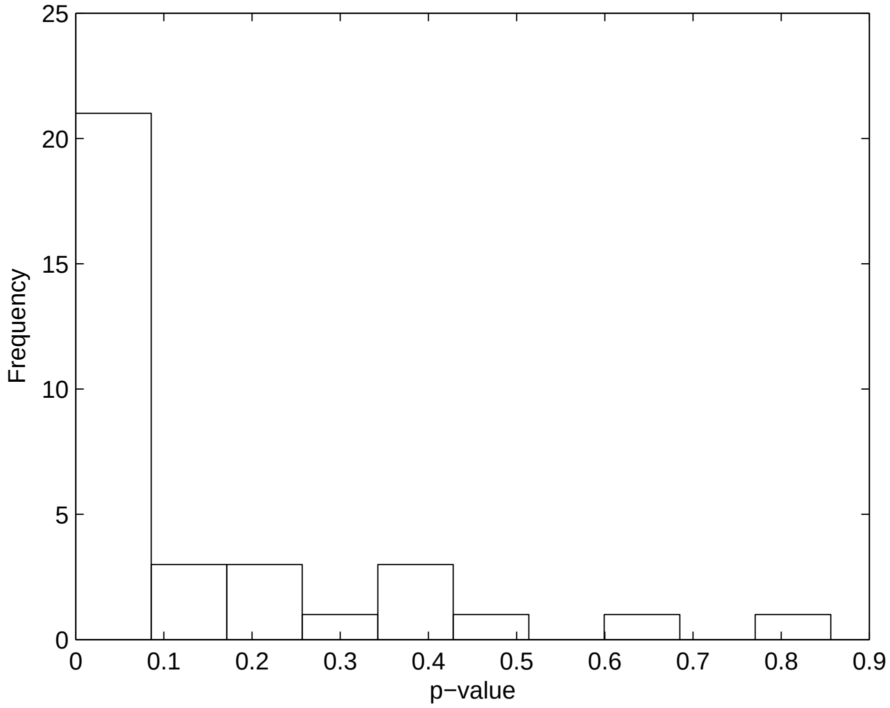


Figure 11: Histogram of the p-Values for the NAEP Data

Rejections of hypotheses with large p-values will justifiably raise some 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.

7. Concluding Remarks

In this paper we offered two different mixture models for improved estimation of the number of true null hypotheses by modelling the non-null p-values. For each model (the normal and beta) three methods of estimation have been developed: HCK, TS and EM. Generally speaking, these methods outperform (in terms of the accuracy of the estimate of π_0 and control of the FDR) the Storey estimator which ignores small p-values. Among these three improved estimators, the EM estimator generally performs the best followed by the HCK and TS estimators when the assumed model holds. 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 estimator. The TS estimator did not improve on the HCK estimator as we had expected.

Comparing the results for the estimators for the normal model and the beta model we see that all estimators perform much worse in the latter case even when the assumed model holds. Robustness performance is also poorer. The example in Section 6 also gave highly unusual results with the beta model for the HCK and EM methods, but the TS method gave reasonable results for both models. As a practical matter, we recommend use of the normal model. If only the p-values are available then they can be transformed to X_i's by applying the inverse normal transformation. The normal model methods can then be applied to the X_i's. The EM estimator should be used when the test statistics can be assumed to be normally distributed; otherwise the HCK estimator should be used.

Table 2: NAEP Trial State Assessment: Test Results for 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 $\gamma = 0.3093$ significant, i.e., the same as those under the EM column in this table.

Appendix

Lemma 5 *For the normal model, the largest $\gamma \in (0,1]$ satisfying (4.2) for $\alpha \in (0, \pi_0]$ is the solution to the equation $\widehat{FDR}(\gamma) = \alpha$. This solution exists and is unique.*

Proof: By substituting for $\beta(\cdot, \gamma)$ from (2.2) and dropping carets on $\widehat{\mathrm{FDR}}(\gamma), \widehat{\pi}_0, \widehat{\pi}_1$ and $\widehat{\delta}$ for notational convenience, the inequality (4.2) becomes

$$\mathrm{FDR}(\gamma) = \frac{\pi_0[1-(1-\gamma)^m]}{\pi_0 + \pi_1\Phi(\delta - z_\gamma)/\gamma} \le \alpha.$$

It is easy to check that $\mathrm{FDR}(0) = 0$ and $\mathrm{FDR}(1) = \pi_0$. We shall show that $\mathrm{FDR}(\gamma)$ is an increasing function of γ which will prove the lemma. Since $[1-(1-\gamma)^m]$ is increasing in γ, it only remains to show that $u(\delta, \gamma) = \Phi(\delta - z_\gamma)/\gamma$ is decreasing in γ. By implicit differentiation of the equation $\Phi(z_\gamma) = 1 - \gamma$, it follows that

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

Hence,

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

Therefore we need to show that

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

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

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

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

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

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

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

which is the Mill's ratio inequality (Johnson and Kotz 1970, p. 279). This completes the proof of the lemma. $\square$

Lemma 6 *For the beta model, assuming $0 < \widehat{a} < 1 < \widehat{b}$, the largest $\gamma \in (0, 1]$ satisfying (4.2) for $\alpha \in (0, \pi_0]$ is the solution to the equation $\widehat{FDR}(\gamma) = \alpha$. This solution exists and is unique.*

Proof: By substituting for $\beta(\cdot, \gamma)$ from (3.2) and dropping carets on $\widehat{\mathrm{FDR}}(\gamma), \widehat{\pi}_0, \widehat{\pi}_1, \widehat{a}$ and $\widehat{b}$ for notational convenience, the inequality (4.2) becomes

$$\mathrm{FDR}(\gamma) = \frac{\pi_0[1 - (1 - \gamma)^m]}{\pi_0 + \pi_1 I_\gamma(a, b))/\gamma} \leq \alpha. \tag{A.1}$$

Note that $\mathrm{FDR}(0) = 0$ and $\mathrm{FDR}(1) = \pi_0$. However, $\mathrm{FDR}(\gamma)$ is not a monotone function of γ as was the case for the normal model. As a result, the proof is a little more complicated.

Let us consider the above inequality but without the $[1-(1-\gamma)^m]$ term (i.e., $\mathrm{pFDR}(\gamma) \leq \alpha$) and rewrite it as

$$\gamma - \theta I_\gamma(a, b) \leq 0 \text{ where } \theta = \left(\frac{\alpha}{1 - \alpha}\right)\left(\frac{\pi_1}{\pi_0}\right). \tag{A.2}$$

Note that we may assume $\theta \leq 1$ since $\alpha \leq \pi_0$.

We will first show that $u(\gamma) = \gamma - \theta I_\gamma(a, b)$ is a convex function with $u(0) = 0, u'(0+) < 0$ for $\theta \in (0, 1]$ and $0 < a < 1 < b$. Therefore the largest γ, say γ^*, that satisfies (A.2) solves $u(\gamma^*) = 0$ and $u'(\gamma^*) > 0$. Since $[1 - (1 - \gamma)^m]$ is an increasing function of γ and is always ≤ 1, it will follow that the largest γ that satisfies (A.1) will be $\geq \gamma^*$.

We first compute

$$u'(\gamma) = 1 - \theta \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \gamma^{a-1}(1 - \gamma)^{b-1}. \tag{A.3}$$

Let $\eta = \theta\Gamma(a + b)/[\Gamma(a)\Gamma(b)] > 0$. We then have

$$\begin{aligned} u''(\gamma) &= -\eta\left[(a - 1)\gamma^{a-1}(1 - \gamma)^{b-1} - (b - 1)\gamma^{a-1}(1 - \gamma)^{b-2}\right] \\ &= \eta\gamma^{a-2}(1 - \gamma)^{b-2}\left[(1 - a)(1 - \gamma) + (b - 1)\gamma\right]. \end{aligned} \tag{A.4}$$

Because $a < 1 < b$, we have $u''(\gamma) > 0$ for any $\gamma \in (0, 1)$, hence $u(\gamma)$ is convex. Furthermore, it is easy to check that $u'(\gamma) \to -\infty$ when $\gamma \to 0$, and $u(0) = 0$ and $u(1) = 1 - \theta \geq 0$. Therefore there exists a unique solution to $u(\gamma) = 0$. Since $[1 - (1 - \gamma)^m]$ is a strictly increasing function of γ, it follows that the original equation $\mathrm{FDR}(\gamma) = \alpha$ has a unique solution in $\gamma \in (0, 1)$ for $\alpha \leq \pi_0$. □

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," *Journal of Statistical Planning and Inference*, **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. Schweder, T. and Spjøtvoll, E. (1982), "Plots of p-values to evaluate many tests simultaneously," *Biometrika*, **69**, 493-502.

18. Shaffer, J. P. (2005), "Multiple requirements for multiple test procedures," Paper presented at the IVth International Conference on Multiple Comparison Procedures, Shanghai, China.

19. 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.

20. Storey, J. (2002), "A direct approach to false discovery rates," *Journal of Royal Statistical Society, Ser. B*, **64**, 297-304.

21. 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.

22. Wu, B., Guan, Z. and Zhao, H. (2006), “Parametric and nonparametric FDR estimation,” *Biometrics*, **735-744**.