

Genome analysis

On correcting the overestimation of the permutation-based false discovery rate estimator

Shuo Jiao and Shunpu Zhang*

Department of Statistics, University of Nebraska Lincoln, Lincoln, NE 68526, USA

Received on February 14, 2008; revised on June 11, 2008; accepted on June 12, 2008

Advance Access publication June 23, 2008

Associate Editor: Alex Bateman

ABSTRACT

Motivation: Recent attempts to account for multiple testing in the analysis of microarray data have focused on controlling the false discovery rate (FDR), which is defined as the expected percentage of the number of false positive genes among the claimed significant genes. As a consequence, the accuracy of the FDR estimators will be important for correctly controlling FDR. Xie *et al.* found that the standard permutation method of estimating FDR is biased and proposed to delete the predicted differentially expressed (DE) genes in the estimation of FDR for one-sample comparison. However, we notice that the formula of the FDR used in their paper is incorrect. This makes the comparison results reported in their paper unconvincing. Other problems with their method include the biased estimation of FDR caused by over- or under-deletion of DE genes in the estimation of FDR and by the implicit use of an unreasonable estimator of the true proportion of equivalently expressed (EE) genes. Due to the great importance of accurate FDR estimation in microarray data analysis, it is necessary to point out such problems and propose improved methods.

Results: Our results confirm that the standard permutation method overestimates the FDR. With the correct FDR formula, we show the method of Xie *et al.* always gives biased estimation of FDR: it overestimates when the number of claimed significant genes is small, and underestimates when the number of claimed significant genes is large. To overcome these problems, we propose two modifications. The simulation results show that our estimator gives more accurate estimation.

Contact: szhang3@unl.edu

1 INTRODUCTION

The use of microarray technology makes it possible to monitor the expression levels of thousands of genes simultaneously. A common goal of analyzing the genome-wide expression data generated from this technology is to detect differentially expressed (DE) genes. Now, as the cost of microarray experiments keeps decreasing, replicated microarray experiments are feasible.

Numerous methods (parametric and non-parametric) have been introduced to detect DE genes. Some of the most well-known parametric approaches include the regression approach of

Thomas *et al.* (2001), the empirical Bayes (EB) methods of Newton *et al.* (2001) and Kendziorzki *et al.* (2003) and the linear models and EB methods of Smyth (2004). Among the non-parametric methods, some well known names include the EB method of Efron *et al.* (2001), the significance analysis of microarray (SAM) of Tusher *et al.* (2001) and the mixture model method (MMM) of Pan *et al.* (2003).

False discovery rate (FDR) introduced by Benjamini and Hochberg (1995) is now commonly used as the choice of the Type I error rate in microarray studies. It is defined as the expected percentage of false positive (FP) genes among the claimed significant genes. It was proved that in many cases controlling FDR is more appropriate compared to controlling family-wise error rate (FWER) since the FDR approaches typically reject more null hypotheses than the FWER approaches (Benjamini and Yekutieli, 2001; Yekutieli and Benjamini, 1999). Several FDR controlling methods are implemented in the R multtest package (Pollard *et al.*, 2004).

However, the true FDR is unknown in practice. Hence, the estimated FDR will serve as the criterion to compare different methods when controlling the error rates. The comparison results are reasonable only if the estimated FDR approximates the true FDR well. The most common method of estimating the FDR is to use the permutation method. However, it has been reported in the literature that the permutation-based FDR estimator tends to overestimate the true FDR. A number of papers has discussed the correction of the overestimation problem of the permutation method (Guo and Pan, 2005; Pan, 2003; Zhao and Pan, 2003; Zhang, 2006).

Xie *et al.* (2005) also noticed the overestimation problem of standard permutation method. Their paper showed that the over-estimation of FDR is caused by the fact that the distribution of null statistics generated from the permutation method is more dispersed than the true null distribution of the test statistics. To solve the problem, they proposed to exclude the predicted DE genes from the estimation of FDR. However, we find that their proposed method has serious under- or over-estimation problem depending on the number of genes declared significant. In addition, we found that they used an incorrect formula of FDR, and hence the comparison results reported in their paper are not correct and the conclusions they drew might be misleading. More seriously, we found that Xie *et al.* (2005) implicitly used an estimator of the proportion of equivalently expressed (EE) genes (π_0) which can only provide good estimate of π_0 when the number of genes declared significant is equal or

*To whom correspondence should be addressed.

close to the true number of DE genes in the microarray data and is otherwise biased.

2 METHODS

2.1 The test statistics and the null statistics

As in Xie *et al.* (2005), only one-sample comparison will be considered in this article. Suppose that Y_{ij} is the expression level of gene i in array j ($i = 1, 2, \dots, n; j = 1, \dots, k$). The goal is to test the following hypothesis: $H_0: E(Y_{ij}) = 0$ against $H_1: E(Y_{ij}) \neq 0$. We use the same three test statistics as in Xie *et al.* (2005) for the purpose of comparison:

1. The mean statistic: $M_i = \bar{Y}_i$,
2. The t -statistic: $T_i = \frac{\bar{Y}_i}{V_i/\sqrt{k}}$,
3. The SAM statistic: $S_i = \frac{\bar{Y}_i}{(V_i + V_0)/\sqrt{k}}$,

where $\bar{Y}_i = \sum_{j=1}^k Y_{ij}/k$, $V_i^2 = \sum_{j=1}^k (Y_{ij} - \bar{Y}_i)^2/(k-1)$, and V_0 is the fudge factor used to stabilize the variance.

In this article, we will focus on the permutation-based method for estimating the FDR. The key issue in the permutation-based method is the generation of the so-called null statistics (the values of the test statistic when the genes are EE). For convenience, we shall use Z_i as a general notation to denote the test statistic and use z_i to denote its corresponding null statistic. In the standard permutation method, one set of null statistics is calculated by applying the test statistic to one set of permuted data. The set of permuted data is obtained by randomly assigning the '+' or '-' signs on each Y_{i1}, \dots, Y_{ik} (SAM). Suppose the number of permutations is B , applying the test statistic to the b -th set of permuted data will create the b -th set of null statistics $z_i^{(b)}$, where $b = 1, \dots, B$, and $i = 1, \dots, n$.

2.2 Method for FDR estimation

Given the test statistics Z_i and a fixed cutoff value d , define $TS(d) = \#\{i: |Z_i| > d\}$ as the total number of significant genes; $FP(d) = \#\{i: |Z_i| > d, i \in EE\}$ as the number of FP genes, where EE is the set of all EE genes; π_0 as the proportion of EE genes; and $\hat{\pi}_0$ as its estimator. According to Storey and Tibshirani (2003), the FDR can be approximated as

$$FDR(d) = E\left(\frac{FP(d)}{TS(d)}\right) \approx \frac{E(FP(d))}{E(TS(d))}. \quad (1)$$

A practical version of FDR is the false discovery proportions (FDP) defined by

$$FDP(d) = \frac{FP(d)}{TS(d)}. \quad (2)$$

To estimate FDR, the standard method is to use the permuted null statistics. Define

$$\widehat{FP}(d) = \sum_{b=1}^B \#\{i: |z_i^{(b)}| > d\}/B. \quad (3)$$

Notice that $\widehat{FP}(d)$ is actually an estimate of $FP(d)/\pi_0$. Storey and Tibshirani (2003) suggested to estimate the FDR by

$$\widehat{FDR}(d) = \frac{\hat{\pi}_0 \widehat{FP}(d)}{TS(d)}. \quad (4)$$

However, as shown in Xie *et al.* (2005), although the null statistics of EE genes have the true null distribution of test statistics, the null statistics of DE genes are more dispersed than those of EE genes. As a result, the empirical distribution of the null statistics from all genes is not a good approximation to the true null distribution. To overcome this problem, Xie *et al.* (2005) proposed a new FDR estimator. Their idea is as follows: since the over-estimation problem of standard permutation method is caused by the DE genes, using only EE genes to construct the null distribution will avoid this problem. Nevertheless, in practice which genes are EE genes is unknown. Therefore, they proposed to use the predicted EE genes to estimate the FDR.

Their FDR estimation procedure works as follows: suppose Z_i is the test statistic and S_i is the SAM statistic, for any given $d > 0$, any gene i with $|S_i| > d$ is said to be significant. $TS(d)$ is defined the same as before. Define a set of non-significant genes $D(d) = \{i: |S_i| \leq d'\}$, where S_i is the SAM statistic and d' is chosen so that the number of genes not in set $D(d)$ is the same as $TS(d)$. In other words, $D(d) = \Omega - TS(d)$, where Ω is the set of all genes. $\widehat{FP}(d)$ is then estimated by constructing B sets of null statistics as before. The only difference is that only genes in $D(d)$ are going to be used this time.

Let

$$\widehat{FP}(d)_0 = \sum_{b=1}^B \#\{i \in D(d): |z_i^{(b)}| > d\}/B. \quad (5)$$

Then, the FDR is estimated by

$$\widehat{FDR}(d)_0 = \frac{\widehat{FP}(d)_0}{TS(d)}. \quad (6)$$

Note that $\widehat{FP}(d)_0$ in (6) is the average number of significant genes found from the genes in $D(d)$. We can re-write (6) in the form of (4) as

$$\widehat{FDR}(d)_0 = \hat{\pi}_0 \widehat{FP}(d)/TS(d), \quad (7)$$

where $\widehat{FP}(d) = \frac{n}{n - TS(d)} \widehat{FP}(d)_0$ can be viewed as the average number of significant genes if all n genes are EE and $\hat{\pi}_0 = 1 - TS(d)/n$ is the estimated proportion of EE genes in the microarray data.

In Xie *et al.* (2005), the above method was proved to be able to correct the FDR overestimation problem of the permutation method effectively. However, our study has found that (6) has four major problems:

- (1) In Xie *et al.* (2005), the true FDR formula (2) is incorrectly defined as

$$FDR(d) = \frac{\pi_0 FP(d)}{TS(d)}. \quad (8)$$

This mistake will affect the evaluation of their proposed FDR estimator.

- (2) In Xie *et al.* (2005), the SAM statistic was used to define the set $D(d)$, which is used in (5) to estimate the number of FP even if the test statistic is the mean or t -statistics. This is unreasonable. If one has chosen the mean or t -statistic as the test statistic, why would he/she use a different statistic to estimate the number of FP? The only explanation is that the mean statistic and the t -statistic do not provide results as good as the SAM statistic does. Note that the mean statistic and the t -statistic can be viewed as two extreme cases of the SAM statistic with the fudge factor equal to ∞ and 0, respectively. It is well known that the performance of the testing procedure based on the mean statistic and the t -statistic is generally inferior to that based on the SAM statistic.
- (3) It can be seen from (7) that Xie *et al.* (2005) implicitly uses $\hat{\pi}_0 = 1 - TS(d)/n$ as an estimate of π_0 . Noticing that $TS(d)$ is the number of claimed significant genes, such $\hat{\pi}_0$ can range from 0 to 1 for $TS(d)$ from n to 0. As a consequence, one will always under- or over-estimate π_0 unless $TS(d) =$ the true number of DE genes.
- (4) The over- or under-estimation of FDR due to under- or over- deletion of genes, which will be discussed in Section 2.3.

2.3 Our proposed method for FDR estimation

Considering the unreasonable estimates $\hat{\pi}_0$ of Xie *et al.* (2005) may provide, we suggest estimating π_0 by the method introduced in Storey and Tibshirani (2003), which is implemented in SAM. In their paper, they calculated P -values for each gene. Denote the P -values by p_1, p_2, \dots, p_n . Then, π_0 is estimated by $\hat{\pi}_0^{sam} = \#\{p_i > \lambda\}/(n(1 - \lambda))$, where λ is a tuning parameter. As we can see, $\hat{\pi}_0^{sam}$ is a constant no matter how $TS(d)$ changes. In addition, after a test statistic Z_i is determined, we use the same test statistic for both identifying the DE genes and defining the set $D(d)$. In other words, $D(d) = \{i: |Z_i| \leq d\}$. With $\hat{\pi}_0^{sam}$ and this new $D(d)$, we propose the following FDR estimator

$$\widehat{FDR}(d)_1 = \hat{\pi}_0^{sam} \widehat{FP}(d)/TS(d), \quad (9)$$

where $\widehat{FP}(d) = \frac{n}{n - TS(d)} \widehat{FP}(d)_0$.

The estimator $\widehat{FDR}(d)_1$ corrects Xie *et al.*'s method by using a more reasonable estimator of π_0 . However, another question comes to light: Is removing all the predicted DE genes a proper way of estimating the FDR? As we know, what we really want is to remove all the DE genes and use all the EE genes to construct the null statistics. However, in those predicted DE genes, there are some genes which are actually EE genes, but are falsely identified as positive (FP genes). It is obvious that the FP genes are the EE genes with the greatest test statistics in absolute values. Therefore, excluding such genes will cause underestimation of the tail of the null distribution. In Section 3.2, we will show that removing all the predicted DE genes gives significantly different FP estimates from those obtained by removing the true DE genes (which is not feasible in practice but good for comparison).

Since removing all predicted DE genes will cause underestimation of the FDR, an intuitive solution would be to add the FP genes back into the pool of the genes for the estimation of the FDR. For this purpose, we propose the following two-step procedure to estimate the FDR, in which the first step is to remove all the predicted DE genes and the second step is trying to re-include the possible FP genes to construct the null statistics:

- (1) Suppose Z_i is the test statistic, for any given $d > 0$, any gene i with $|Z_i| > d$ is said to be significant. Let $TS(d) = \#\{i: |Z_i| > d\}$, $D(d) = \{i: |Z_i| \leq d\}$, $\widehat{FP}(d)_0 = \sum_{b=1}^B \#\{i \in D(d): |z_i^{(b)}| > d\} / B$, and $\widehat{FDR}(d)_1 = \frac{n}{n-TS(d)} \hat{\pi}_0^{sam} \widehat{FP}(d)_0 / TS(d)$.
- (2) Using $\widehat{FDR}(d)_1$ from Step 1, let $D(d') = \{i: |Z_i| \leq d'\}$, d' is chosen such that the number of genes not in $D(d')$ is $TS(d') = TS(d)(1 - \widehat{FDR}(d)_1)$. Then following the same procedure as Step 1, we get $\widehat{FP}(d')_0 = \sum_{b=1}^B \#\{i \in D(d'): |z_i^{(b)}| > d'\} / B$, and

$$\widehat{FDR}(d)_2 = \hat{\pi}_0^{sam} \widehat{FP}(d')_0 / TS(d), \quad (10)$$

where $\widehat{FP}(d) = \frac{n}{n-TS(d)} \widehat{FP}(d')_0$.

The idea behind our proposed method is as follows: when the number of predicted DE genes is greater than the true number of significant genes, there will be a substantial number of FP genes in them. Since removing all predicted DE genes will cause biased estimation of the FDR, we only remove the genes which we consider are most likely to be true DE genes.

3 RESULTS

3.1 Problems caused by using Xie *et al.*'s estimate of π_0

In Xie *et al.* (2005), π_0 is estimated by $\hat{\pi}_0 = 1 - TS(d)/n$. As stated before, we would expect to see over- or under-estimation of FDR by this method because of the over- or under-estimation of π_0 by $\hat{\pi}_0$.

To show this, 5 ($=k$) replicates of 4000 ($=n$) genes are generated, among which 400 are DE genes and the others are EE genes. The expression levels Y_{ij} for EE genes are generated from $N(0, 4)$ and Y_{ij} for DE gene are generated from $N(\mu_i, 4)$, while $\mu_i \sim N(0, 16)$. The SAM, mean and t -statistics are used as the test statistics. Our purpose is to compare the FDR estimator of Xie *et al.* (2005) ($\widehat{FDR}(d)_0$) from (7) and one of our proposed estimator ($\widehat{FDR}(d)_1$) from (9). The values of the standard FDR estimator from (4) and the true FDR values are also plotted as references.

Overestimation of FDR when $TS(d)$ is smaller than the true number of DE genes.

In this scenario, $TS(d)$ is set to be from 100 to 200, which is much less than the true number of DE genes ($=400$). In Figure 1, as we expected, $\widehat{FDR}(d)_0$ always overestimates the true FDR while $\widehat{FDR}(d)_1$ provides much less biased estimates. In some cases, $\widehat{FDR}(d)_1$ still gives overestimation. This overestimation is caused by the fact that $\hat{\pi}_0^{sam}$ also always overestimates the true π_0 , but to a much lesser degree.

Underestimation of FDR when $TS(d)$ is greater than the true number of DE genes.

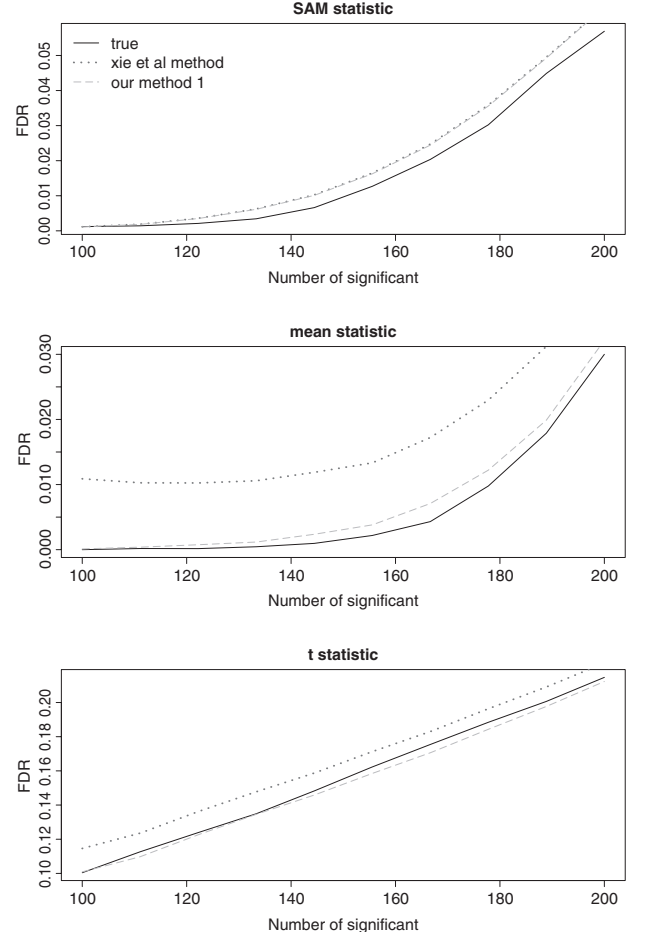


Fig. 1. FDR curves of different estimation methods using the SAM, mean and t -statistics. There are 400 DE genes among 4000 genes. The number of claimed significant gene ranges from 100 to 200. $\hat{\pi}_0^{sam}$ is used as the estimate of π_0 . Our method 1 is the estimator $\widehat{FDR}(d)_1$ from (9).

The same simulation set-up is used as above except now $TS(d)$ is set to be from 500 to 600, which is greater than the the true number of DE genes ($=400$).

As shown in Figure 2, for the t and SAM statistics, Xie *et al.*'s method underestimates the true FDR while our proposed method gives much more accurate estimates. However, for the mean statistic, our method does not give any improvement over Xie *et al.*'s method. The reason is that the SAM statistic was used to predict DE genes in Xie *et al.* (2005) while our method $\widehat{FDR}(d)_1$ uses the same mean statistic in both predicting the DE genes and estimating the FDR. The better performance of Xie *et al.*'s method in this case is due to the use of the SAM statistic in predicting DE genes, rather than the method itself. As it can be seen from the top plot of Figure 2, our estimator $\widehat{FDR}(d)_1$ performs significantly better than Xie *et al.*'s method when the SAM statistic is used.

3.2 Underestimation caused by removing the predicted DE genes

In this section, we show that removing all predicted DE genes will lead to an underestimation of the true FP number. We generate

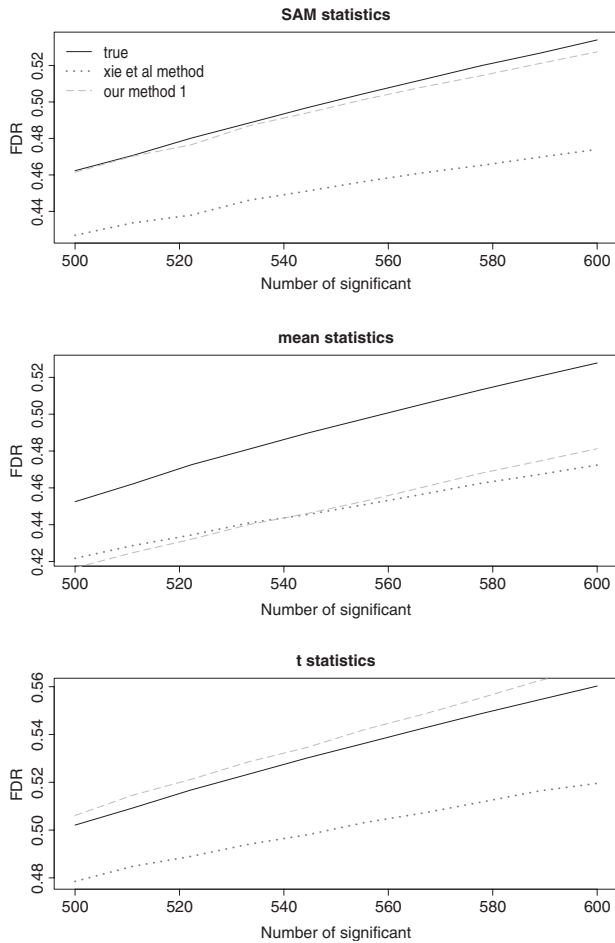


Fig. 2. FDR curves of different estimation methods using the SAM mean and t -statistics. There are 400 DE genes among 4000 genes. The number of claimed significant gene ranges from 500 to 600. Our method 1 is the estimator $\widehat{FDR}(d)_1$ from (9).

$n=4000$ genes with $k=5$, while 150 of them are DE genes. The expression levels for EE and DE genes are generated in the same way as in Section 3.1. The number of claimed significant genes is set to be 150, which is the number of true DE genes. Table 1 lists the true FP number, the estimated FP number with 150 predicted DE genes removed (\widehat{FP}_p), and the estimated FP number with 150 true DE genes removed (\widehat{FP}_t). The results reported are the averages from 50 replicates.

From Table 1, we can see \widehat{FP}_p is always less than \widehat{FP}_t . This shows removing predicted DE genes gives a smaller estimate of FP number than that of removing the true DE genes.

3.3 Performance of our methods

To evaluate the performance of our methods, the same simulation set-ups are used as those in Section 3.2. We want to see whether our proposed estimator $\widehat{FDR}(d)_2$ from (10) can overcome the problems or at least has some advantages over other estimators.

We compare four different FDR estimation methods: the standard estimator $\widehat{FDR}(d)$ from (4), Xie *et al.* (2005) estimator $\widehat{FDR}(d)_0$

Table 1. Comparison of estimated FP numbers and the true FP numbers using the SAM, mean and t -statistics

Statistic	True FP	\widehat{FP}_p	\widehat{FP}_t
SAM	64.38	61.62	65.30
mean	58.96	53.96	60.81
t	79.78	77.21	81.19

from (7), and two estimators we proposed: $\widehat{FDR}(d)_1$ from (9), $\widehat{FDR}(d)_2$ from (10).

Figure 3 shows that the estimator of Xie *et al.* (2005) always significantly underestimates the true FDR's. The estimator $\widehat{FDR}(d)_1$ also underestimates FDR due to over-deletion, but is much better than Xie *et al.*'s estimator for the SAM statistic. For the mean and t -statistics, Xie *et al.*'s estimator outperforms $\widehat{FDR}(d)_1$ sometimes due to the same reason discussed previously—the use of the SAM statistic in obtaining the predicted DE genes. In contrast, $\widehat{FDR}(d)_2$ does not have this problem. However, for the SAM statistic and the t -statistic, $\widehat{FDR}(d)_2$ slightly overestimates the true FDR. This overestimation is not caused by the estimator $\widehat{FDR}(d)_2$, but by the overestimation of π_0 caused by $\hat{\pi}_0^{sam}$. To see this, we replaced $\hat{\pi}_0^{sam}$ in (9) for $\widehat{FDR}(d)_1$ and in (10) for $\widehat{FDR}(d)_2$ with the true $\pi_0=3850/4000$. Figure 4 shows the comparison between the true FDR and the estimated FDR from (9) and (10) with the true value of π_0 . We can see that $\widehat{FDR}(d)_2$ now gives smaller estimates of FDR for all three test statistics compared to Figure 3. Another fact worth noticing in Figure 3 and 4 is that when the number of claimed significant genes is small, $\widehat{FDR}(d)_2$ does not show much advantage. The reason is that, in such a case, most of the significant genes are true DE genes and the number of FP genes is much smaller than the number of true DE genes. Hence, removing the FP genes is not going to have significant impact on the estimation of the FDR.

3.4 Comparisons under other simulation set-ups

We also want to see how the ratio of induced (I) and repressed (R) genes influences the performance of the FDR estimators. Here, $k=5, n=4000$ and there are 150 DE genes. The expression level Y_{ij} for EE genes are generated from $N(0, 4)$. For DE genes, n' of them are generated from $N(4, 4)$, and the rest of them are generated from $N(-4, 4)$; where $n'=150, 100, 50, 0$. We set the number of claimed significant genes as 300. The results reported in Table 2 are the averages from 50 replications.

The results confirm that our methods are stable to the change of ratios of the induced and repressed genes.

We have also conducted another simulation which tries to mimic the real data. Similar simulation set-up as above is used except the expression level Y_{ij} for EE genes are generated from $N(0, \sigma_i^2)$ while $\sigma_i^2 \sim Gamma(4, 2)$ and Y_{ij} for DE gene are generated from $N(\mu_i, \sigma_i^2)$ while $\mu_i \sim N(0, 16), \sigma_i^2 \sim Gamma(4, 2)$.

From Figure 5, we can see that the results are similar as before for the SAM and t -statistics: the standard method always overestimates and method of Xie *et al.* (2005) always underestimates. $\widehat{FDR}(d)_1$ performs better than Xie *et al.*'s method and $\widehat{FDR}(d)_2$ always performs the best.

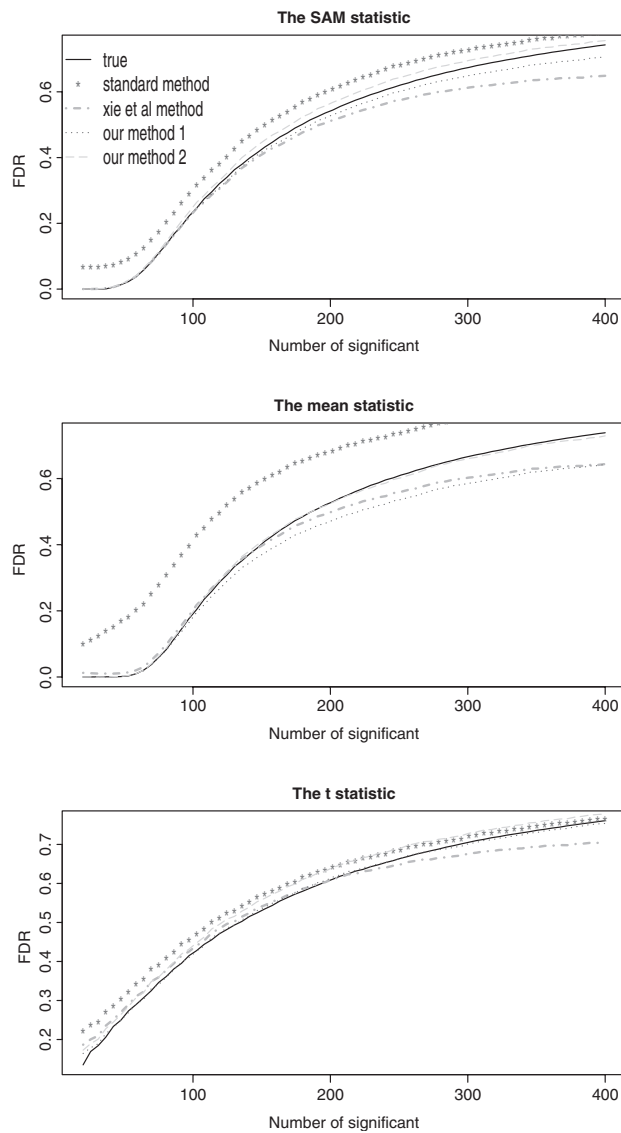


Fig. 3. FDR curves of different estimation methods using the SAM, mean and t -statistics. There are 150 DE genes among 4000 genes. The number of claimed significant gene ranges from 20 to 400. $\hat{\pi}_0^{sam}$ is used as estimate of π_0 . Our methods 1 and 2 are the estimators $\widehat{FDR}(d)_1$ from (9) and $\widehat{FDR}(d)_2$ from (10), respectively.

3.5 Biological data

In [Zhong *et al.* \(2004\)](#), duplications and deletions in an evolved strain (DD2459) were identified by a whole-genome *Escherichia coli* MG1655 spotted DNA microarray experiment with three replicates. Thirty-eight genes have been confirmed to be true duplicated/deleted genes by rtPCR. To compare our proposed estimator $\widehat{FDR}(d)_2$ with Xie *et al.*'s estimator $\widehat{FDR}(d)_0$, we used this data to construct a table summarizing the upper bound of true FDR (the proportion of detected DE genes which are not in the confirmed 38 DE genes), FDR estimates given by $\widehat{FDR}(d)_2$ and $\widehat{FDR}(d)_0$ for different number of total significant genes ($TS(d)$). Because the confirmed 38 true DE genes are mostly genes with largest mean in absolute value, we can see from [Table 3](#) that the mean statistic gives the smallest FDR upper

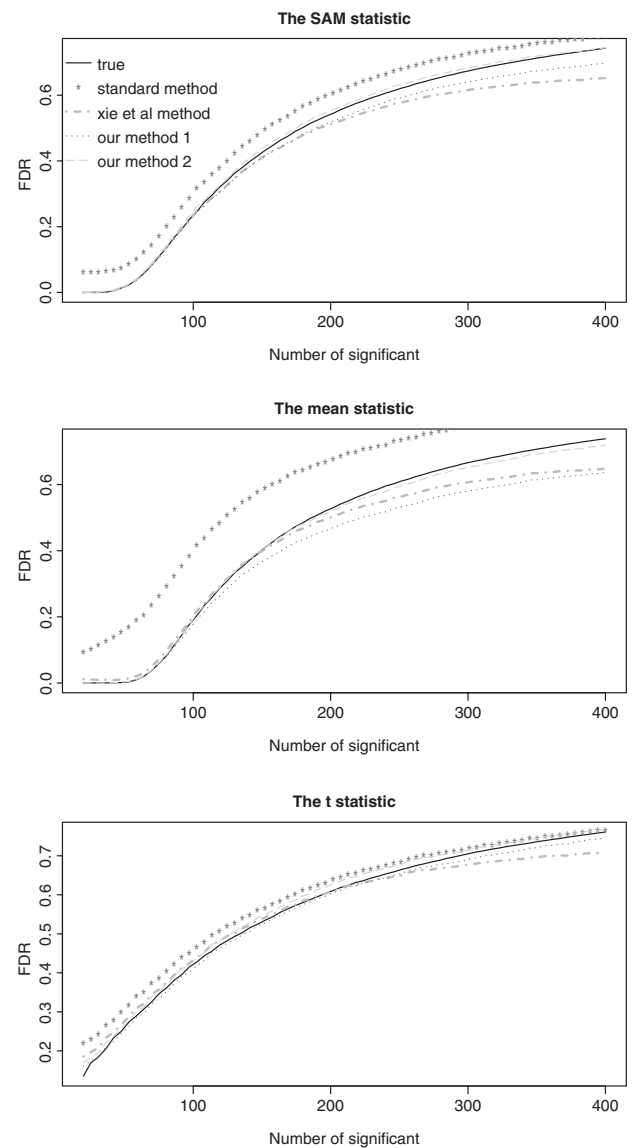


Fig. 4. FDR curves of different estimation methods using the SAM, mean and t -statistics. There are 150 DE genes among 4000 genes. The number of claimed significant gene ranges from 20 to 400. The true $\pi_0 = 3850/4000$ is used as estimate of π_0 . Our methods 1 and 2 are the estimators $\widehat{FDR}(d)_1$ from (9) and $\widehat{FDR}(d)_2$ from (10), respectively.

bound while the t -statistic does not detect any one of the 38 true DE genes. [Table 3](#) also shows that $\widehat{FDR}(d)_2$ always gives more accurate FDR estimates than $\widehat{FDR}(d)_0$.

4 DISCUSSION

In this article, we have showed that the bias-corrected FDR estimator proposed in [Xie *et al.* \(2005\)](#) uses an inappropriate estimate of π_0 and still has severe under- or over-estimation problem. We have proposed two new modifications to overcome those problems. Simulation studies and application to real data have confirmed that our estimator $\widehat{FDR}(d)_2$ gives significantly better FDR estimates than $\widehat{FDR}(d)_0$ in [Xie *et al.* \(2005\)](#).

Table 2. Comparison of the performance of FDR estimator when the ratio of induced and repressed genes changes

I/R		FDR_{true}	$\widehat{FDR}(d)$	$\widehat{FDR}(d)_0$	$\widehat{FDR}(d)_1$	$\widehat{FDR}(d)_2$
150/0	SAM	0.507	0.572	0.461	0.486	0.521
	mean	0.504	0.672	0.423	0.446	0.504
	<i>t</i>	0.558	0.564	0.513	0.539	0.560
100/50	SAM	0.508	0.566	0.463	0.489	0.520
	mean	0.504	0.665	0.416	0.439	0.498
	<i>t</i>	0.557	0.569	0.512	0.538	0.562
50/100	SAM	0.509	0.570	0.460	0.485	0.520
	mean	0.504	0.670	0.424	0.445	0.499
	<i>t</i>	0.557	0.565	0.512	0.537	0.558
0/150	SAM	0.507	0.566	0.465	0.491	0.522
	mean	0.504	0.661	0.427	0.449	0.504
	<i>t</i>	0.556	0.562	0.514	0.544	0.562

Table 3. Comparison of the performance of $\widehat{FDR}(d)_2$ and $\widehat{FDR}(d)_0$ using microarray data from Zhong *et al.* (2004)

Statistic	TS(d)	Upper bound	$\widehat{FDR}(d)_0$	$\widehat{FDR}(d)_2$
SAM	35	0.457	0.347	0.506
	40	0.500	0.304	0.443
	45	0.533	0.272	0.404
	50	0.560	0.267	0.386
mean	35	0.371	0.230	0.356
	40	0.375	0.158	0.264
	45	0.422	0.171	0.242
	50	0.480	0.177	0.231
<i>t</i>	35	1.000	0.871	1.000
	40	1.000	0.870	1.000
	45	1.000	0.817	1.000
	50	1.000	0.810	0.997

Current null statistics are constructed by randomly assigning the ‘+’ or ‘-’ signs to replicates of genes. As a consequence, the number of ‘+’ and ‘-’ signs can be different in this random assignment. Mean expression levels of EE genes will always be 0 regardless of the way of assigning the signs. However, when there is an unbalanced number of ‘+’ and ‘-’, the mean expression levels of DE genes will not be 0, which may cause the null statistics of DE genes to have different distributions from that of EE genes. Hence, it is intuitive to deduce that if we make the number of ‘+’ and ‘-’ stay balanced, this problem can be avoided. In Pan (2003) and Zhang (2006), they proposed a series of such kind of ‘balanced’ null statistics, which have the same distribution for both DE and EE genes. It would be interesting to compare the performance of our FDR estimators and estimators based on ‘balanced’ null statistics in the future research.

ACKNOWLEDGEMENTS

We would like to thank the Associate Editor and three referees for their constructive comments which have significantly improved the

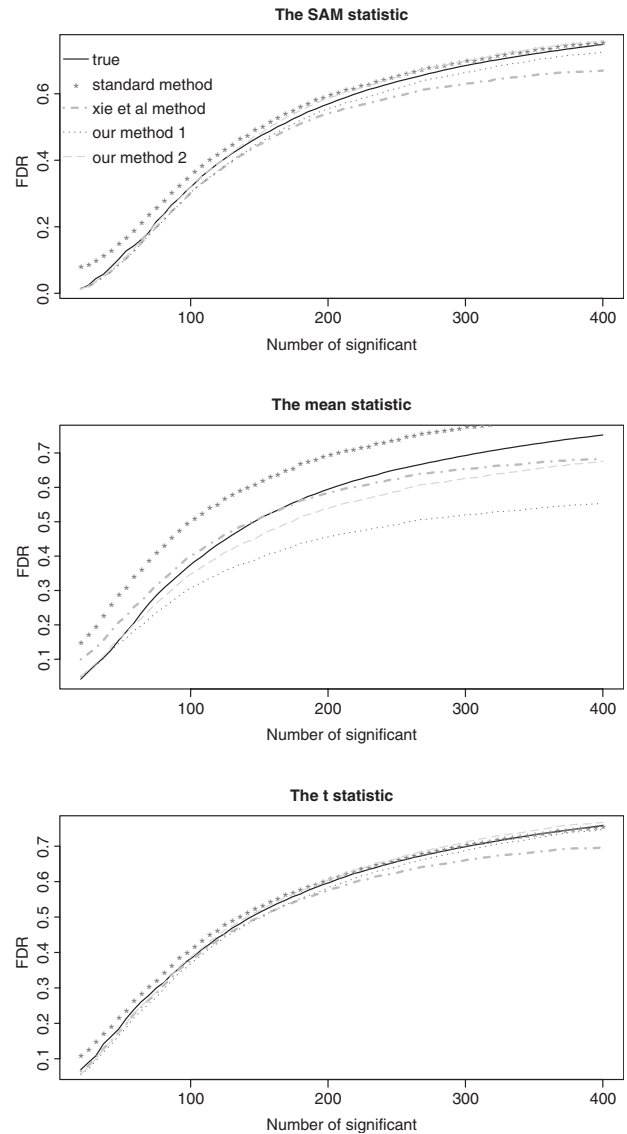


Fig. 5. FDR curves of different estimation methods using the SAM, mean, and *t*-statistics. Mimicking the real data. There are 150 DE genes among 4000 genes. The number of claimed significant gene ranges from 20 to 400. Our methods 1 and 2 are the estimators $\widehat{FDR}(d)_1$ from (9) and $\widehat{FDR}(d)_2$ from (10), respectively.

quality of the article. We also thank Dr Shaobin Zhong for providing us the microarray data.

Conflict of Interest: none declared.

REFERENCES

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.*, **57**, 289–300.
 Benjamini, Y. and Yekutieli, D. (2001) The control of the False discovery rate in multiple testing under dependency. *Ann. Stat.*, **29**, 1165–1188.
 Efron, B. *et al.* (2001) Empirical Bayes analysis of a microarray experiment. *J. Am. Stat. Assoc.*, **96**, 1151–1160.
 Guo, X. and Pan, W. (2005) Using weighted permutation scores to detect differential gene expression with microarray data. *J. Comput. Biol.*, **3**, 989–1006.

- Kendzioriski, C.M. *et al.* (2003) On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Stat. Med.*, **22**, 3899–3914.
- Kerr, M.K. *et al.* (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.*, **7**, 819–837.
- Newton, M.A. *et al.* (2001) On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *J. Comput. Biol.*, **8**, 37–52.
- Pan, W. *et al.* (2003) A mixture model approach to detecting differentially expressed genes with microarray data. *Funct. Integr. Genomics*, **3**, 117–124.
- Pan, W. (2003) On the use of permutation in the performance of a class of nonparametric methods to detect differential gene expression. *Bioinformatics*, **19**, 1333–1040.
- Pollard, K.S. *et al.* (2004) Multiple testing procedures: R multtest package and applications to genomics. *U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 164*. Available at <http://www.bepress.com/ucbbiostat/paper164>. (last accessed date December 2004)
- Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Gene. Mol. Biol.*, **3**, Article 3.
- Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
- Thomas, J.G. *et al.* (2001) An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Res.*, **11**, 1227–1236.
- Tusher, V.G. *et al.* (2001) Significant analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.
- Xie, Y. *et al.* (2005) A note on using permutation based false discovery rate estimate to compare different analysis methods for microarray data. *Bioinformatics*, **21**, 4280–4288.
- Yekutieli, D. and Benjamini, Y. (1999) Resampling based false discovery rate controlling multiple testing procedure for correlated test statistics. *J. Stat. Plann. Inference*, **82**, 171–196.
- Zhao, Y. and Pan, W. (2003) Modified nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments. *Bioinformatics*, **19**, 1046–1054.
- Zhang, S. (2006) An improved nonparametric approach for detecting differentially expressed genes with replicated microarray data. *Stat. Appl. Gene. Mol. Biol.*, **5**, Article 30.
- Zhong, S. *et al.* (2004) Evolutionary genomics of ecological specialization. *Proc. Natl Acad. Sci. USA*, **101**, 11719–11724.