Journal of the Royal Statistical Society
SERIES B Statistical Methodology

J. R. Statist. Soc. B (2012)
74, *Part* 5, *pp.* 871–891

Empirical Bayes false coverage rate controlling confidence intervals

Zhigen Zhao

Temple University, Philadelphia, USA

and J. T. Gene Hwang

Cornell University, Ithaca, USA, and National Chung Cheng University, Min-Hsiung, Taiwan

[Received June 2010. Final revision January 2012]

Summary. Benjamini and Yekutieli suggested that it is important to account for multiplicity correction for confidence intervals when only some of the selected intervals are reported. They introduced the concept of the *false coverage rate* (FCR) for confidence intervals which is parallel to the concept of the false discovery rate in the multiple-hypothesis testing problem and they developed confidence intervals for selected parameters which control the FCR. Their approach requires the FCR to be controlled in the frequentist's sense, i.e. controlled for all the possible unknown parameters. In modern applications, the number of parameters could be large, as large as tens of thousands or even more, as in microarray experiments. We propose a less conservative criterion, the Bayes FCR, and study confidence intervals controlling it for a class of distributions. The Bayes FCR refers to the average FCR with respect to a distribution of parameters. Under such a criterion, we propose some confidence intervals, which, by some analytic and numerical calculations, are demonstrated to have the Bayes FCR controlled at level q for a class of prior distributions, including mixtures of normal distributions and zero, where the mixing probability is unknown. The confidence intervals are shrinkage-type procedures which are more efficient for the θ_is that have a sparsity structure, which is a common feature of microarray data. More importantly, the centre of the proposed shrinkage intervals reduces much of the bias due to selection. Consequently, the proposed empirical Bayes intervals are always shorter in average length than the intervals of Benjamini and Yekutieli and can be only 50% or 60% as long in some cases. We apply these procedures to the data of Choe and colleagues and obtain similar results.

Keywords: Multiplicity; Simultaneous intervals

1. Introduction

In statistical analysis, confidence intervals are one of the most important inferential tools. Unlike hypothesis testing by using p-values, confidence intervals could provide ranges for the parameters while taking into consideration the variability in estimating them. The traditional evaluation of confidence intervals is based on the (simultaneous) probability of covering the true parameter and the average length.

In this paper, we focus on the situation when a high dimensional parameter $(\theta_1, \ldots, \theta_p)$ is involved, as often is the case in modern applications. See Efron (2010). In such a case, the scientific interest, at times, lies in making statistical inference regarding the θ_is that are preselected,

Address for correspondence: J. T. Gene Hwang, Department of Mathematics, National Chung Cheng University, 168 University Road, Min-Hsiung, Chia-Yi 621, Taiwan.
E-mail: hwang@math.cornell.edu

 1369–7412/12/74871

or declared to be statistically significant. If the scientist is interested in reporting, based on the same data used for selection, confidence intervals for the parameters corresponding to these selected θ_is, what should he do? This is the question that was raised in Benjamini and Yekutieli (2005). Their proposed criterion is to examine the false coverage rate (FCR), which is the average rate of false coverage, namely, not covering the true parameters, among the selected intervals. They demonstrated that, if one ignores the selection and uses the traditional (frequentist's) confidence intervals, each having one-dimensional coverage probability $1-\alpha$, the FCR may be much higher than q and is not controlled. The intervals with simultaneous coverage probability $1-q$ can control the FCR to be bounded by q. However, these intervals are very long. They then constructed their confidence intervals, called BY intervals in this paper, that have a controlled FCR and a shorter average length than the simultaneous confidence intervals.

In their approach, the FCR is defined in the frequentist's sense and is required to be less than or equal to q for every set of fixed parameters θ_i. For microarray experiments and other modern applications, there are a huge number of parameters, often tens of thousands or more, and it is convenient to describe the θ_is in terms of probability distributions, such as the percentage π_0 of non-differentially expressed genes. Indeed, there is much research aiming at or relating to the estimation of π_0. See, for example, Ruppert *et al.* (2007), Nettleton *et al.* (2006), Storey (2002), Jin and Cai (2007) and the papers cited therein. When a distribution is postulated, it seems reasonable to consider the average FCR, averaging with respect to such a distribution of θ_is. In this paper, such an average FCR is called the *Bayes FCR* whereas the distribution of θ_is is called the Bayes prior distribution. This is similar to the definition of the Bayes FDR as in Chen and Sarkar (2006) and Sarkar *et al.* (2008). More to the point, the Bayes FCR was also defined and studied in Yekutieli (2008) who studied a single prior. In practice, the prior distribution can be speculated about but is never totally known. Hence a class of priors is considered in this paper. Whether such a class of priors is plausible can even be checked by using the data to plot graphs similar to Q–Q-plots as done in Qiu and Hwang (2007). In this paper, we aim at constructing the *empirical* Bayes FCR controlling confidence intervals, which are defined to be the intervals that guarantee the Bayes FCR_π with respect to π to be less than q for any π in a class of priors. Hence the empirical Bayes FCR controlling intervals aim at a class of prior distributions whereas the Bayes FCR controlling intervals work only for one prior distribution. The former intervals are much more difficult to construct than the latter.

Our present approach also leads to shrinkage confidence intervals that are particularly appropriate for θ_is with a sparsity structure which occurs frequently in microarray data. More importantly, the centres of the shrinkage confidence intervals are shrinkage estimators or *empirical* Bayes estimators which correct much of the selection bias. We demonstrate this in Section 2. The BY intervals and the Bonferroni intervals are centred at the usual non-shrinkage estimators which have selection bias. This explains why the resulting shrinkage intervals are shorter in length while maintaining the control of the Bayes FCR.

Although we use the terminology of Bayes or empirical Bayes approaches, the Bayes FCR is exactly the frequentist's FCR when θ_is are the random effects as in the random- or mixed effect models. These random-effect models are becoming important even in the area of microarrays. See the references cited in the first paragraph of Section 3.

In Section 3, we introduce all terminologies and our model. For a single prior, we establish a theorem demonstrating that, regardless of the selection rule, Bayes intervals have a Bayes FCR controlled at q, as long as the posterior non-coverage probabilities of the Bayes intervals are controlled at the same level. This result is also established in Yekutieli (2008). However, to deal with a class of priors, we need to deal with other intervals. For any interval, some asymptotic theorems are derived to evaluate the Bayes FCR. In Section 4, we apply the theorems in

Section 3 to a class of prior distributions. Considering a class of priors is crucial even for the random-effect models which assume unknown variances. We establish that, under certain conditions, the Bayes FCR can be uniformly controlled asymptotically as the number of parameters p (the number of genes) goes to ∞ if the empirical Bayes confidence intervals in the sense of Morris (1983) are used. The asymptotic property holds regardless of which selection rules are used as long as they satisfy some minor requirements. We have also shown by simulations that certain empirical Bayes intervals control the Bayes FCR for any normal priors when p is finite. Moreover, the empirical Bayes intervals are always shorter in average length and could be half as long when compared with the BY intervals.

Section 5 deals with another important class of priors which are mixtures of a normal distribution with the zero point where the probability π_0 of being 0 is positive and unknown. This class of priors is particularly useful for the microarray data where the probability of non-differentially expressed genes π_0 could be high and cannot be ignored. We aim to construct intervals with controlled Bayes FCR for any π_0 and any normal prior. Here, the ordinary empirical Bayes intervals that are constructed in Section 4 no longer have a controlled Bayes FCR for some π_0. A novel theorem (theorem 7) is established which eventually leads to intervals which are shown numerically to have controlled Bayes FCR for any normal mixture prior with zero. These intervals have uniformly smaller average length than the BY intervals, and the biggest reduction in length over the BY intervals could be 40% approximately. Although the theorem works for any given selection rule, the recommended intervals do depend on which selection rule is being used.

Finally, in Section 6, various procedures are applied to data sets of Choe *et al.* (2005) (after some bias correction of the data). A striking feature of the data set is that the true parameters of the data set are known. We can therefore calculate the actual FCR and the average length for each procedure. The result demonstrates the superiority of our proposed intervals as observed in the simulation. When applied to the data set of Choe *et al.* (2005) (after some bias correction of the data), we show that our proposed intervals have actual FCR less than 5% whereas the average length is much shorter than that of BY intervals and other alternatives.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

`http://www.blackwellpublishing.com/rss`

2. Empirical Bayes false coverage rate reduces the selection bias

In this section, we briefly demonstrate in a simple setting that the empirical Bayes FCR can reduce the selection bias. The idea is hidden in the asymptotic theoretical arguments of Hwang (1993) and Qiu and Hwang (2007). However, the idea can be demonstrated easily by using graphs (Fig. 1). Here, we assume the setting where

$$X_i \overset{\text{IID}}{\sim} N(\theta_i, \sigma_i^2), \quad \text{and} \quad \theta_i \overset{\text{IID}}{\sim} N(\mu, \tau^2), \qquad i = 1, 2, \ldots, p. \tag{1}$$

The simulation below assumes that $\sigma_i^2 = \tau^2 = 1$ and $\mu = 10$. Let $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(p)}$ be the order statistics. Let $\theta_{(i)}$ be the θ corresponding to the population that has produced $X_{(i)}$. Note that $\theta_{(i)}$ depends on X_i. In fact, $\theta_{(i)} = \theta_j$ if $X_j = X_{(i)}$ and is called the parameter of the selected population. See, for example, Qiu and Hwang (2007) and the reference therein. Yekutieli (2008) studied settings that were more general than this simple example. The naive estimator for $\theta_{(i)}$ is $X_{(i)}$, which ignores the selection and is known to have bias, called the selection bias. Fig. 1(a) plots the expectations $E(X_{(i)})$ *versus* $E(\theta_{(i)})$ by using 'diamond' symbols. The discrep-

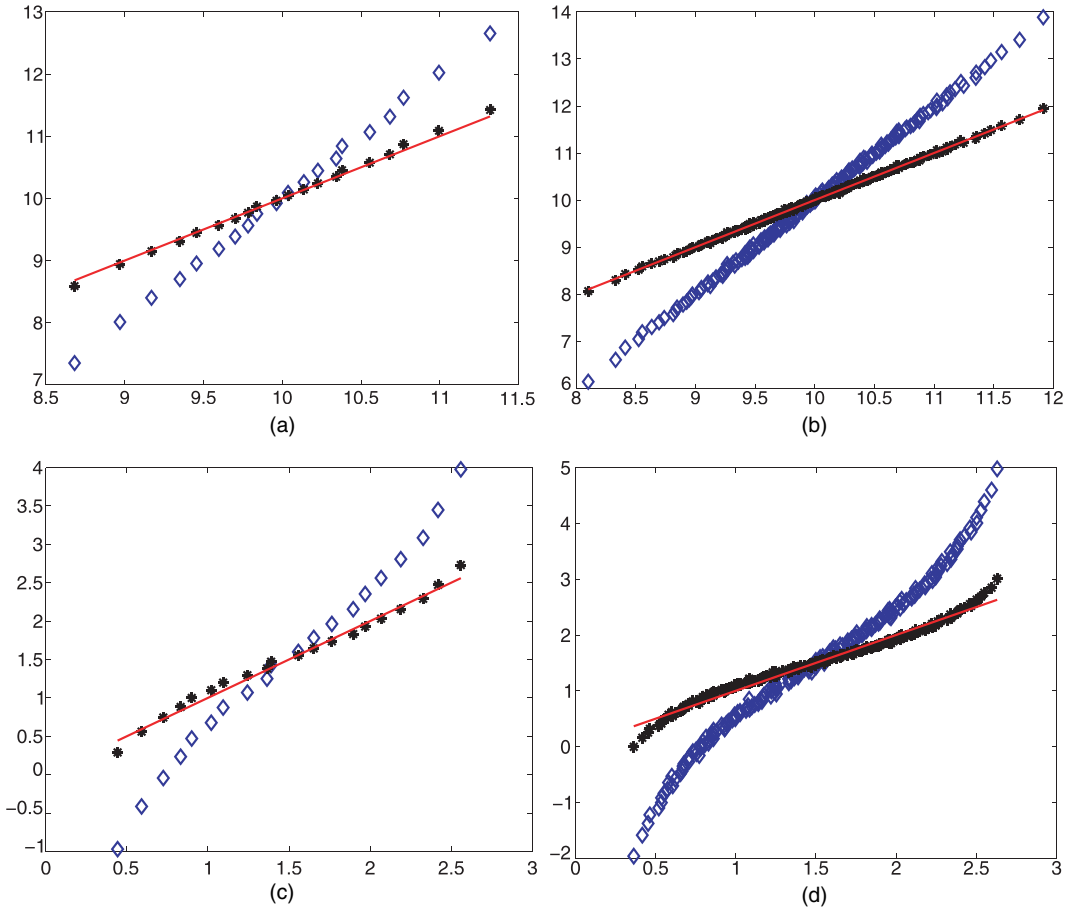

Fig. 1. (a) Expected values of the naive estimator (◇) and the Lindley–James–Stein estimator modified for the selected parameters (∗) plotted against the expectation of the selected parameters for $p = 20$ (all the asterisks are almost on the 45° line (———), showing that the Lindley–James–Stein estimator has virtually no bias, whereas the naive estimator has large bias); (b) $p = 200$ instead of $p = 20$ (both (a) and (b) assume normal prior (1) with $\sigma_i^2 = \tau^2 = 1$ and $\mu = 10$); (c) for $p = 20$ and (d) for $p = 200$ assume uniform priors over $[0, 3]$ for θs, showing a similar conclusion that the Lindley–James–Stein estimator reduces much of the bias of the naive estimator

ancy between the blue diamonds and the red line (which is the 45° line) shows that there is bias especially for small $E(\theta_{(i)})$ or large $E(\theta_{(i)})$. The empirical Bayes estimator for estimating $\theta_{(i)}$ modifies the Lindley–James–Stein estimator, which is the centre of equation (6) below, by replacing X_i with $X_{(i)}$. See Hwang (1993). (Precisely, it is $\hat{M}_i X_{(i)} + (1 - \hat{M}_i)\hat{\mu}$ where $\hat{M}_i$ and $\hat{\mu}$ are defined in the paragraph in Section 4 containing equation (5) with $\sigma_i = \sigma$.) For $p = 20$, the expectations of empirical Bayes estimators are plotted by using asterisks in Fig. 1(a), which are virtually all on the red line, showing that there is little bias even for $p = 20$. Fig. 1(b) demonstrates the same phenomenon for $p = 200$. These two graphs assume normal priors. For a uniform prior over $[0, 3]$ for θs, Figs 1(c) and 1(d) respectively for p equal to 20 and 200 demonstrate that the shrinkage estimators reduce much of the selection bias although not as perfectly as in Figs 1(a) and 1(b). These graphs show that the shrinkage estimators are better choices than the naive estimators for constructing good intervals. In this paper, we construct intervals that are centred at these better choices.

3. General theorem on Bayes intervals

We begin by giving the definition of the FCR of confidence intervals, which was a term coined in Benjamini and Yekutieli (2005). Consider one-dimensional parameters θ_i, $i=1,\ldots,p$. Assume that X_i is an estimator of θ_i. Here (X_i,θ_i) is a canonical form representation of the problem where θ_i is interpreted as a key parameter and X_i its (unbiased) estimator. The form applies to other more sophisticated models where, for each gene, one assumes an analysis-of-variance model (or a linear model) which relates the expression levels to various experimental conditions. In such a case, X_i is the analysis-of-variance estimator (or the least squares estimator) of θ_i. The result of this section holds also for the case where X_i is a vector having a distribution with parameter (θ_i,η_i) where η_i is a nuisance vector parameter as long as a prior is put on both θ_i and η_i. Applications of analysis-of-variance models including random parameters are common even in the microarray literature. See, for example, Lönnstedt and Speed (2002), Smyth (2004), Cui *et al.* (2005), Qin and Hwang (2007), Hwang and Liu (2010), Kendziorski *et al.* (2003) and Tai and Speed (2006). In all these references, an independent and identically distributed (IID) normal prior distribution of θ_i or its mixture with zero is assumed. In particular, Smyth (2004) and Lönnstedt and Speed (2002) are all well cited in biology journals. Their procedures are used, for example, in Gregory *et al.* (2008) citing the former and in Subkhankulova and Livesey (2006) citing both. Many other references in the microarray literature also assume a random non-Gaussian (Bayesian) model. See, for example, Newton *et al.* (2001) and Kendziorski *et al.* (2003), which includes both a normal model and a non-normal model.

Let CI_i, based on all the X_is, be an interval for θ_i. Assume that $\mathcal{R}$ is a set of index i such that θ_i has been selected on the basis of X_is. Let $\mathcal{V}$ consist of $i\in\mathcal{R}$ such that CI_i does not cover θ_i. Let R and V denote the numbers of elements in $\mathcal{R}$ and $\mathcal{V}$ respectively. The FCR that was defined in Benjamini and Yekutieli (2005) is $\mathrm{FCR}=E(Q)$, where $Q=V/R$ when $R>0$ and $Q=0$ when $R=0$. The expectation is calculated by integrating out X_i, under the assumption that the θ_i are fixed. The procedure in Benjamini and Yekutieli (2005) controls the FCR to be less than or equal to q, a small number, for every θ_i. However in modern technology, as in a microarray experiment, the number of parameters is very large. When a prior on θ_is is postulated, it seems natural to consider the average FCR by integrating out the θ_is with respect to their distribution π and to define the Bayes FCR as

$$\mathrm{FCR}_\pi = E_\pi\{E(Q)\}.$$

The t-intervals were shown in Benjamini and Yekutieli (2005) to have the frequentist's FCR larger than 5%. Similarly, the classical 95% z-intervals have Bayes FCR that is much larger than 5% as demonstrated in Fig. 2 by the black broken curve. This is because these parameters have been preselected—they are declared to be significantly different from 0 after applying Benjamini and Hochberg's (1995) procedure with false discovery rate set to be 5%.

In this section, we focus on the Bayes FCR. The definition of FCR_π seems unrelated to the non-coverage probability; however, the following theorems demonstrate that they are closely related. Assume that the probability density function of $X=(X_1,\ldots,X_p)$ is $f_\theta(X)$ and the probability density function of $\theta=(\theta_1,\ldots,\theta_p)$ is $\pi(\theta)$.

Theorem 1. For any selection rule,

$$\mathrm{FCR}_\pi = \int_{R>0} E(Q|X)\,m(X)\,\mathrm{d}X$$

where $E(Q|X)=(1/R)\Sigma_{i\in\mathcal{R}}P(\theta_i\notin \mathrm{CI}_i|X)$ and $m(X)=\int f_\theta(X)\,\pi(\theta)\,\mathrm{d}\theta$.

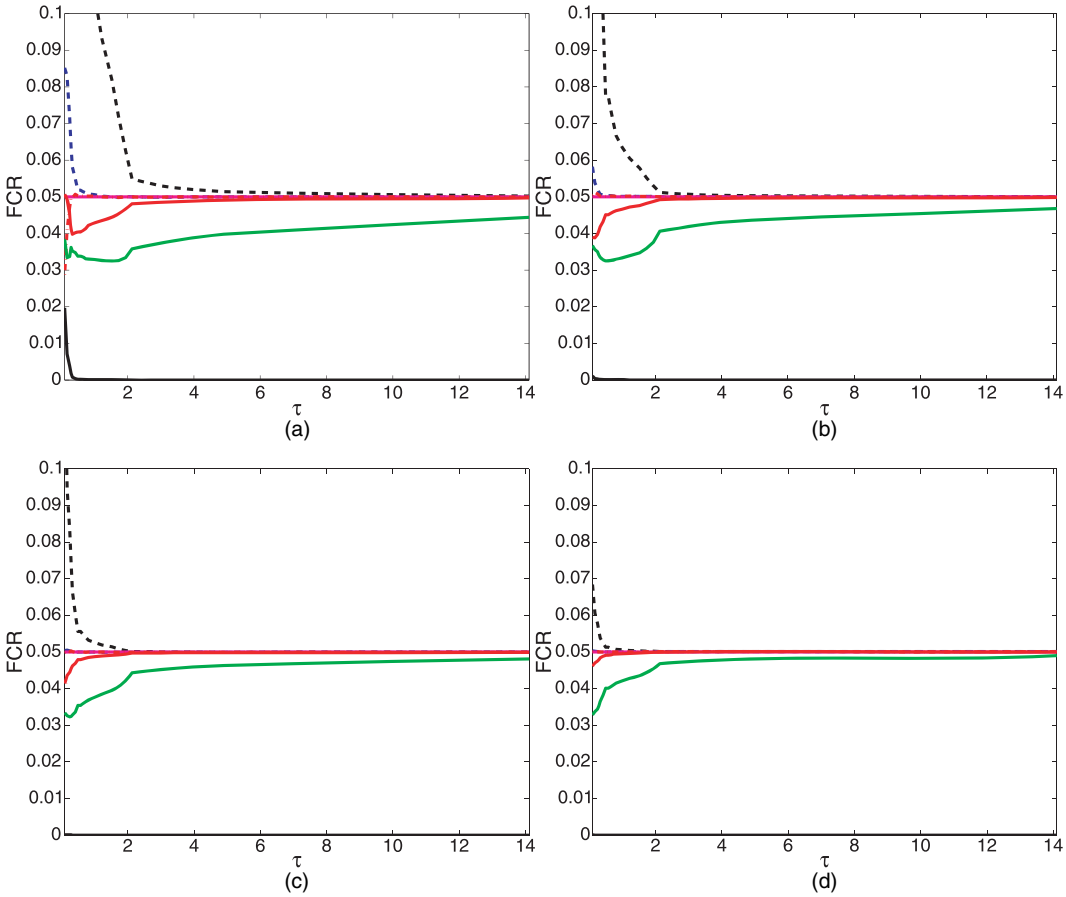

Fig. 2. Bayes FCRs of various intervals plotted against τ for $p = 2000$ and $q = 0.05$ under the normal–normal model when assuming unequal but known variances: the variances are sampled independently from inverse gamma distributions of various combinations of parameters a and b; for the definition of inverse gamma distributions, see Berger (1985), page 561, item 10, where (α,β) are denoted (a, b) here; a is chosen to be 2.1 here and the bs are chosen to be (a) 0.1, (b) 0.3, (c) 1 and (d) 3; the parameters are selected by using Benjamini and Hochberg's (1995) false discovery rate procedure at the 5% level; the naive z-intervals (– – –) and expression (6) (– – –) fail to control the FCR at the q-level (———); Benjamini and Yekutieli's procedure (———), the Bonferroni correction (———) and our empirical Bayes confidence intervals (8) (———) control the Bayes FCR at the q-level for any $N(\mu,\tau^2)$ prior where $\tau > 0$; the Bayes intervals (– – –) have also controlled the Bayes FCR but are unrealistic in assuming a known prior distribution; the same conclusion is arrived at on the basis of eight other graphs corresponding to various combinations of $a = 2.5$ and $a = 4$ and the bs listed above

The proof of this theorem and all the other theorems below is given in Appendix A unless it is obvious from the context. Given theorem 1, the following theorem is obvious. (The proof is omitted.)

Theorem 2. If $P(\theta_i \notin \mathrm{CI}_i|X) \leqslant q$, $\forall i$, then $\mathrm{FCR}_\pi \leqslant q\, P(R > 0) \leqslant q$, for any selection rule based on X.

Both theorem 1 and theorem 2 above hold even if (X_i, θ_i) are dependent. These theorems provide us with a straightforward way to construct confidence intervals with a controlled Bayes FCR when the prior distribution π is known. Let us consider the following example (example 1).

Assume model (1) where σ_i^2 are known quantities. The highest posterior density confidence interval is

$$\mathrm{CI}_i^{\mathrm{B}} = M_i X_i + (1 - M_i)\mu \pm z\sigma_i M_i^{1/2}, \tag{2}$$

where $M_i = \tau^2/(\tau^2 + \sigma_i^2)$ and $P(|Z| > z) = q$ for a standard normal random variable Z. Then $\mathrm{CI}_i^{\mathrm{B}}$ has posterior coverage probability $1 - q$. By theorem 2, the Bayes FCR of interval (2) is no greater than q.

Theorem 2 could be very useful because Bayes intervals that have high coverage probabilities can automatically control the Bayes FCR. However, in practice, the Bayes prior distribution is typically unknown. Hence we need to consider other intervals, such as those in Section 4. The theorems below help to study the asymptotic properties of Bayes FCRs of any confidence intervals.

Theorem 3. Assume that $\max_{1 \leqslant i \leqslant p}\{P(\theta_i \notin \mathrm{CI}_i|X)\} = \alpha(p, X)$ and

$$\lim_{p \to \infty} P\{\alpha(p, X) \leqslant q + \varepsilon\} \to 1, \qquad \forall \varepsilon > 0. \tag{3}$$

Then

$$\limsup_{p \to \infty} \mathrm{FCR}_\pi \leqslant q.$$

When condition (3) holds, we shall say that $\alpha(p, X)$ is asymptotically (as $p \to \infty$) less than or equal to q in probability. Under such a condition, FCR_π is asymptotically controlled at the level q. Theorem 3 aims at dealing with the most severe term $\max_{1 \leqslant i \leqslant p}\{P(\theta_i \notin \mathrm{CI}_i|X)\}$; therefore, it even applies to the extreme case when only one observation is selected. A weaker sufficient condition can be obtained when R increases as p increases as in the following theorem.

Theorem 4. Assume that $R/p \to \eta > 0$, and

$$\frac{1}{p}\sum_i |P(\theta_i \notin \mathrm{CI}_i|X) - q| \to 0, \qquad \text{almost surely,} \tag{4}$$

where q is any number independent of i. Then $\lim_{p \to \infty} \mathrm{FCR}_\pi \to q$.

Similarly, if instead of result (4) we have

$$\limsup_{p \to \infty} \frac{1}{p}\sum_i (P(\theta_i \notin \mathrm{CI}_i|X) - q)_+ \leqslant 0, \qquad \text{almost surely,}$$

where, for a number a, $(a)_+$ denotes $\max(a, 0)$, then $\limsup_{p \to \infty} \mathrm{FCR}_\pi \leqslant q$.

4. Empirical Bayes approach

In Section 3, we showed that the Bayes confidence intervals can control the Bayes FCR. The result works for a single prior, which is unrealistic in real applications. More realistically, we now deal with a class of priors indexed by some unknown hyperparameters which will be estimated as in the empirical Bayes approach. The hope is that it would result in intervals that will control FCR_π for the class of priors and are called the empirical Bayes FCR controlling intervals. Although the idea seems intuitive, to construct intervals controlling FCR_π for a class of priors is technically quite difficult. Even without selection, the level of difficulty is already similar to that of constructing empirical Bayes confidence intervals for a class of priors. See Morris (1983) and Casella and Hwang (1983). It goes without saying that the selection significantly increases the level of difficulty.

Now, assume that $\theta_i \sim N(\mu, \tau^2)$ where μ and τ^2 are unknown. Using the method of moments, we estimate μ by $\hat{\mu} = \bar{X}$, and τ^2 by

$$\hat{\tau}^2 = \left(\frac{\sum_{i=1}^{p} (X_i^2 - \sigma_i^2)}{p} - \hat{\mu}^2 \right)_+. \tag{5}$$

Also, we estimate M_i by $\hat{M}_i = \hat{\tau}^2/(\hat{\tau}^2 + \sigma_i^2)$.

Substituting all the hyperparameters in the interval (2) by their estimators above results in the so-called empirical Bayes interval

$$\mathrm{CI}_i^{\mathrm{EB}} = \{\hat{M}_i X_i + (1 - \hat{M}_i)\hat{\mu}\} \pm z_q \sigma_i \hat{M}_i^{1/2}, \tag{6}$$

where z_q is chosen such that

$$P(|Z| < z_q) = 1 - q. \tag{7}$$

Since all the estimators are obtained through the method of moments, we would expect that they should converge to the Bayes interval as $p \to \infty$. Hence asymptotically interval (6) would behave like the Bayes procedure (2), having the asymptotic Bayes FCR controlled at the level q for any $N(\mu, \tau^2)$ prior with $\tau > 0$. This indeed can be proved as in the two theorems below for any $N(\mu, \tau^2)$ prior π with $\tau > 0$.

Theorem 5. For any $\varepsilon > 0$, if $\Sigma_{i=1}^{p} \sigma_i^4 = o\{p^2/\log(p)^{1+\varepsilon}\}$, then $\limsup_{p\to\infty} \mathrm{FCR}_\pi \leqslant q, \forall \pi$.

Alternatively, an application of theorem 4 provides us with the asymptotic property under a less restrictive condition when the number of selection R increases as p increases.

Theorem 6. If $R/p \to \eta > 0$, and $\Sigma_{i=1}^{p} \sigma_i^4 = o(p^2)$, then $\lim_{p\to\infty} \mathrm{FCR}_\pi = q, \forall \pi$.

Both conditions on the order of $\Sigma_i \sigma_i^4$ are mild: much weaker than the result from the law of large numbers. More specifically, when σ_i^2s are generated as samples from a population with finite second moment, these two conditions are satisfied. Both theorem 5 and theorem 6 under their assumptions imply that the empirical Bayes interval (6) asymptotically controls the FCR_π at the q-level for any non-degenerate normal prior π. However, when p is finite, FCR_π can be higher than q. See Fig. 2 for $q = 5\%$. Judging from theorem 2, this is likely because interval (6) has coverage probability lower than 95%. This suggests that we should try the confidence intervals of Hwang *et al.* (2009) adapting to the case of known variances. Namely we consider the intervals

$$\mathrm{CI}_i^{\mathrm{HQZ}} = \hat{M}_i^* X_i + (1 - \hat{M}_i^*)\hat{\mu} \pm [\hat{M}_i^* \{z_q^2 - \log(\hat{M}_i^*)\}]^{1/2} \sigma_i, \tag{8}$$

where $\hat{M}_i^* = \hat{\tau}_*^2/(\hat{\tau}_*^2 + \sigma_i^2)$, $\hat{\tau}_*^2 = \max(\hat{\tau}^2, \tau_0^2)$ and

$$\tau_0^2 = \frac{2z_q^2 \sum_i \sigma_i^2 + z_q \left\{ 4z_q^2 \left(\sum_i \sigma_i^2 \right)^2 + 2 \sum_i \sigma_i^4 (p^2 - 2pz_q^2) \right\}^{1/2}}{p^2 - 2pz_q^2}.$$

Note that $\mathrm{CI}_i^{\mathrm{EB}}$, according to theorems 5 and 6, has FCR_π controlled asymptotically. The same can be said about $\mathrm{CI}_i^{\mathrm{HQZ}}$ under the assumptions of these theorems since it contains $\mathrm{CI}_i^{\mathrm{EB}}$. More importantly, intervals (8) have good finite coverage probability and their FCR_π is about 5% or less as demonstrated in Fig. 2. Also the intervals of Bonferroni and Benjamini and Yekutieli (2005) have FCR_π controlled at 5%. However, it can be analytically proved that the BY intervals are longer than the z-intervals which are longer than interval (8) if $z_q > 1$, which is quite

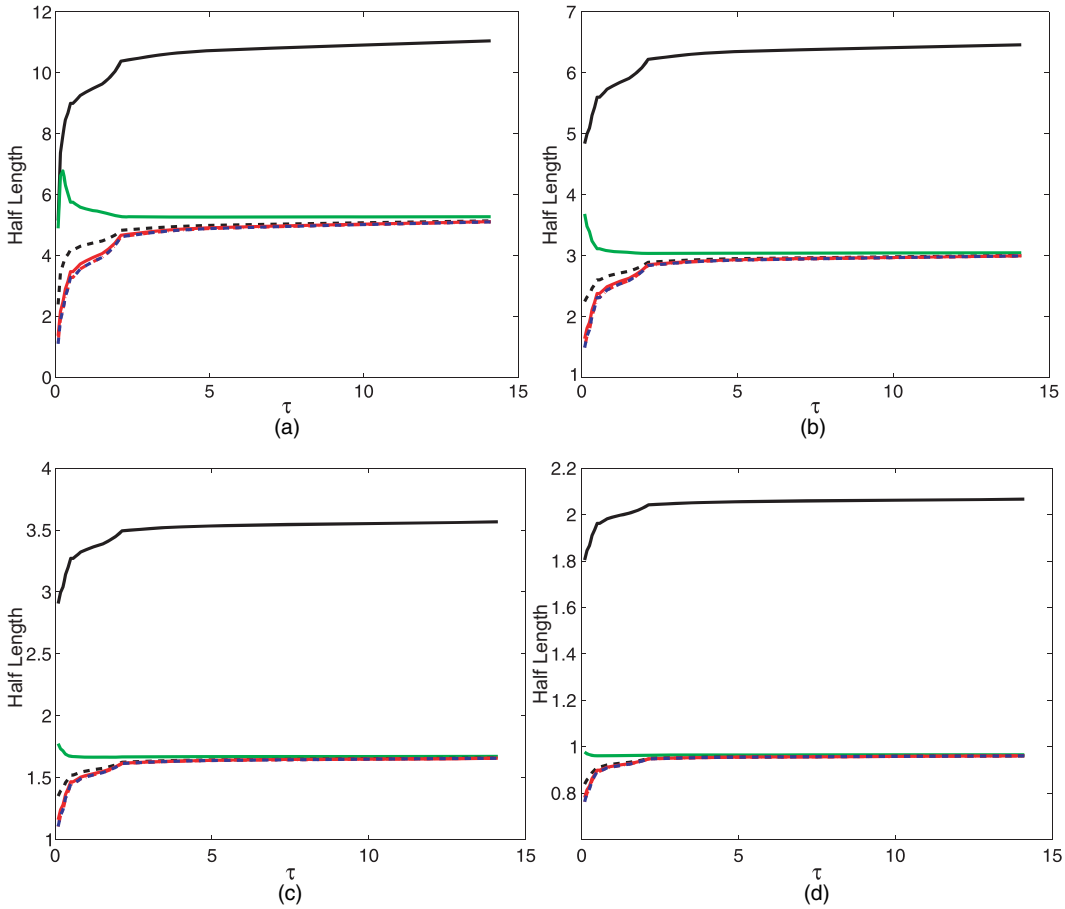

Fig. 3. Average half-length of intervals plotted for various combinations of parameters a and b which are given in Fig. 2 (the average lengths of the Bonferroni and BY intervals are uniformly larger than expression (8) and, in the most extreme case, three times as large; the proposed intervals have average lengths similar to those of the Bayes intervals; ——, Bonferroni; ——, Benjamini and Yekutieli (2005); – – –, z-interval; ——, empirical Bayes (8); –·–·–, Bayes; – – –, expression (6)): (a) $b = 0.1$; (b) $b = 0.3$; (c) $b = 1$; (d) $b = 3$

a minor restriction. Similar comments apply to the unknown σ_i case when $t > 1$. Moreover, in Fig. 3, we see that the average lengths of the Bonferroni and BY intervals are much longer than that of $\mathrm{CI}^{\mathrm{HQZ}}$, being three times longer in the most extreme cases. The intervals $\mathrm{CI}^{\mathrm{HQZ}}$ have average length almost identical to the Bayes intervals, which have the minimum average lengths. However, unlike $\mathrm{CI}^{\mathrm{HQZ}}$, the Bayes intervals assume knowledge of τ^2, which is unrealistic in real application. We plotted Figs 2 and 3 for other cases (see the caption of Fig. 2) and the results are similar.

5. Confidence intervals for mixture prior model

In the previous sections, we assumed a normal prior distribution. In many applications including the microarray data analysis, it is often more appropriate to assume a mixture model. Hence we consider the model

$$X_i|\theta_i \sim N(\theta_i, \sigma_i^2) \quad \text{and} \quad \theta_i \begin{cases} = 0, & \text{with probability } \pi_0; \\ \sim N(\mu, \tau^2), & \text{with probability } \pi_1 = 1 - \pi_0. \end{cases} \tag{9}$$

In microarray experiments, π_0 can be as large as or larger than 0.9. Let $\mathcal{R}$ denote the set of indices which are selected for the interval construction. Our goal is to construct the confidence intervals CI_i for each parameter θ_i, such that the Bayes FCR is controlled for any hyperparameters (π_0, τ^2).

Theorem 7. Let $\mathrm{CI}_i(q')$ be a confidence interval for each θ_i such that $P(\theta_i \notin \mathrm{CI}_i|X, \theta_i \neq 0) \leqslant q'$ for a positive number q'. Then the Bayes FCR is bounded above by $\int \mathrm{Int}(q', x)\, m(x)\, \mathrm{d}x$, where

$$\mathrm{Int}(q', x) = \{q' + \sum_{i \in \mathcal{R}, 0 \notin \mathrm{CI}_i} \mathrm{fdr}_i(x)(1 - q') - \sum_{i \in \mathcal{R}, 0 \in \mathrm{CI}_i} \mathrm{fdr}_i(x) q'\}\, I(R > 0). \tag{10}$$

Here $I(R > 0)$ is the indicator function equal to 1 or 0 depending on whether $R > 0$ and $\mathrm{fdr}_i(x) = P(\theta_i = 0|X = x)$ is the local false discovery rate defined in Efron (2005, 2007, 2008, 2010).

For a given observation X, assume that the hyperparameter (π_0, τ^2) is known. Then

$$\mathrm{fdr}_i(x) = \frac{(\pi_0/\sigma_i)\,\phi(x_i/\sigma_i)}{(\pi_0/\sigma_i)\,\phi(x_i/\sigma_i) + \{\pi_1/(\sigma_i^2 + \tau^2)^{1/2}\}\,\phi\{x_i/(\sigma_i^2 + \tau^2)^{1/2}\}}, \tag{11}$$

where $\phi(x)$ is the density function of the standard normal distribution.

Let $\mathrm{CI}_i(q') = M_i X_i \pm z_{q'} \sigma_i M_i^{1/2}$ where $z_{q'}$ is defined in equation (7) with q being replaced by q', and

$$q_R = \arg\max_{0 \leqslant q' \leqslant q} \{q' : \mathrm{Int}(q', X) \leqslant q\}. \tag{12}$$

Then, according to theorem 7, the Bayes FCR of the intervals $\mathrm{CI}_i(q_R)$, which is $\mathrm{CI}_i(q')$ with q' being replaced by q_R, is controlled at the level q.

For empirical Bayes intervals or other intervals considered before this point, q_R will be too small, leading to long intervals. To construct sharp intervals, we need to force the interval to include zero when data indicate that θ_i is near 0 as in Qiu and Hwang (2007).

Instead, we consider the mixture confidence interval

$$\mathrm{CI}_i^{\mathrm{Mixed}}(q_R) = \begin{cases} \mathrm{CI}_i(q_R) & \mathrm{fdr}_i(x) < 0.20, \\ \mathrm{CI}_i(q_R) \cup \{0\}, & \mathrm{fdr}_i(x) > 0.20 \end{cases} \tag{13}$$

and q_R is chosen according to equations (10) and (12) with $\mathrm{CI}_i(q')$ replaced by $\mathrm{CI}_i^{\mathrm{Mixed}}(q')$. As $\pi_0 \to 1$, $\mathrm{fdr}_i(X) \to 0$. Then all intervals $\mathrm{CI}_i^{\mathrm{Mixed}}$ include zero and $q_R = q$. The mixture intervals appear to be very short.

Efron (2005, 2007) also suggested the use of 0.2 as a reasonable choice to declare genes to be differentially expressed if $\mathrm{fdr}_i < 0.2$.

5.1. Empirical Bayes approach

In the section above, the hyperparameters are assumed to be known. In practice, they obviously should be estimated from the data. We shall take an approach that is similar to that of Hwang and Liu (2010). Under model (9), direct calculation shows that

$$\begin{aligned} m_1 &\equiv E(X_i) = \pi_1 \mu, \\ m_2 &\equiv E(X_i^2) - \sigma_i^2 = \pi_1(\tau^2 + \mu^2). \end{aligned} \tag{14}$$

Let $\hat{m}_1 = (1/p)\Sigma_i X_i$ and $\hat{m}_2 = (1/p)\Sigma_i (X_i^2 - \sigma_i^2)$, which are obviously unbiased moment estimators of m_1 and m_2. Solving expression (14) for μ and τ^2 gives

$$\begin{aligned} \mu &= m_1/\pi_1, \\ \tau^2 &= m_2/\pi_1 - \mu^2. \end{aligned} \tag{15}$$

The log-likelihood function of X_i is

$$l(x) = \sum_i \log\left[\frac{\pi_0}{(2\pi^{1/2})\sigma_i}\exp\left(-\frac{x_i^2}{2\sigma_i^2}\right) + \frac{\pi_1}{\{2\pi(\tau^2+\sigma_i^2)\}^{1/2}}\exp\left\{-\frac{(x_i-\mu)^2}{2(\tau^2+\sigma_i^2)}\right\}\right]. \tag{16}$$

Instead of directly using the maximum likelihood estimator for (π_1, μ, τ^2), we replace μ and τ^2 in equation (16) by expression (15) where m_1 and m_2 are estimated by $\hat{m}_1$ and $\hat{m}_2$. Then

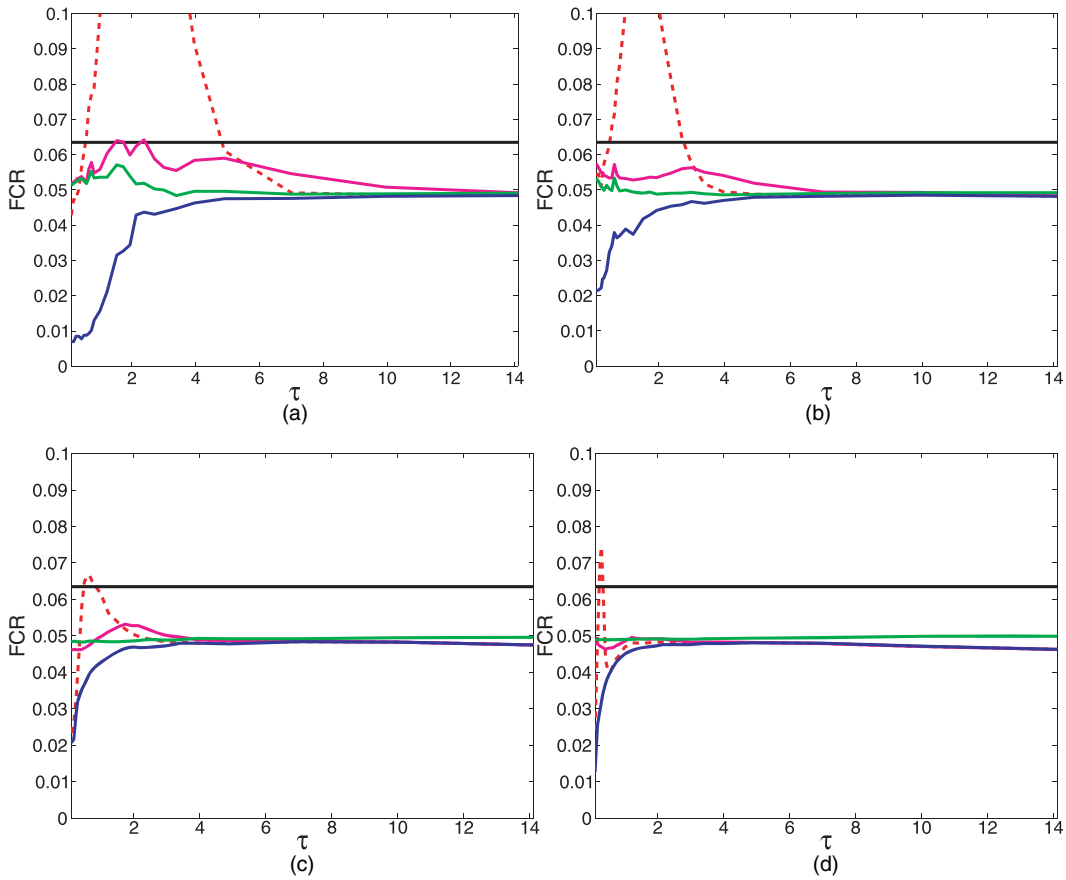

Fig. 4. Bayes FCRs of various intervals plotted against τ for $p = 5000$ and $q = 0.05$ under the normal–mixture model with $\pi_0 = 0.95$ when assuming unequal but known variances (the variances are sampled independently from the inverse gamma distribution for $a = 2.1$ and b being chosen to be (a) 0.1, (b) 0.3, (c) 1 and (d) 3; the parameters are selected by using Benjamini and Hochberg's (1995) false discovery rate procedure at the 5% level; similar graphs are plotted for $\pi_0 = 0.5$, 0.7, 0.8, 0.9, 0.99; all these graphs show that all the intervals studied here, including the proposed intervals (18), have Bayes FCR controlled at the 5% nominal level except the intervals (17)): - - -, expression (17); ——, q^*-level (where $q^* = q + 1.96 \times$ simulation standard deviation); ——, expression (18); ——, Benjamini and Yekutieli; ——, Bayes

we choose $\pi = \hat{\pi}_1$ to maximize the one-dimensional version of equation (16), which is easy to compute. We can now estimate μ and τ^2 as $\hat{\mu} = \hat{m}_1/\hat{\pi}_1$ and $\hat{\tau}^2 = \hat{m}_2/\hat{\pi}_1 - \hat{\mu}^2$. This estimator was first proposed in Hwang and Liu (2010) and was shown to work very well in their testing hypothesis context. We truncate π_1 at both ends at 0.001 and 0.999.

When the dimension is very large, the estimator of the hyperparameter can be very accurate, leading to intervals with controlled FCR. However for moderately large p, one needs to make some adjustment which is given below.

Assume that π_0, μ and τ^2 are estimated as above and $\widehat{\mathrm{fdr}}_i(X)$s are calculated with π_0, μ and τ^2 in equation (11) being replaced by the corresponding estimators. To control the FCR at the level q, consider interval (8), where q is replaced by q' which is to be determined below:

$$\mathrm{CI}_i^{\mathrm{EB}}(q') = \begin{cases} \text{interval (8)}, & \text{if } \widehat{\mathrm{fdr}}_i(x) < 0.2, \\ \text{interval (8)} \cup \{0\}, & \text{if } \widehat{\mathrm{fdr}}_i(x) > 0.2. \end{cases} \tag{17}$$

Let

$$q_R = \arg\max_{0 \leqslant q' \leqslant q} \{q' : \widehat{\mathrm{Int}}(q', X) \leqslant q\},$$

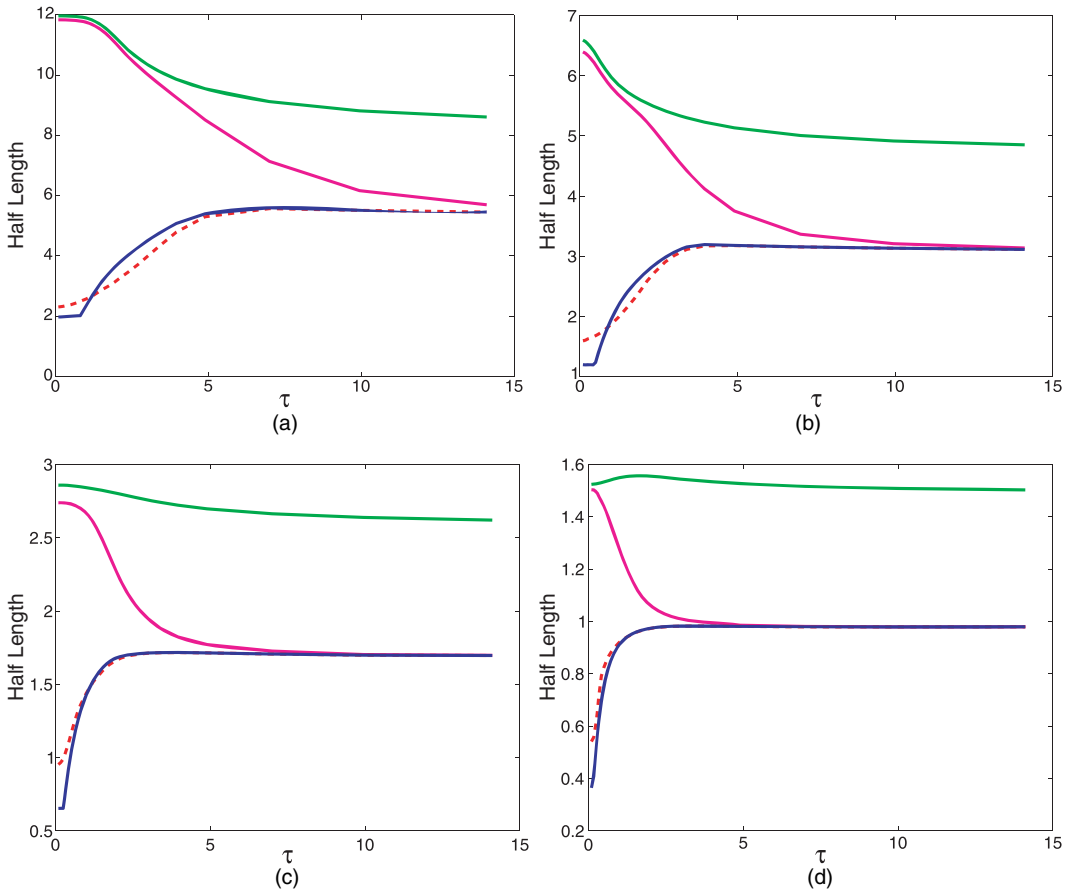

Fig. 5. Average half-length of intervals constructed plotted for $p = 0.95$ and $a = 2.1$ and b being chosen to be (a) 0.1, (b) 0.3, (c) 1 and (d) 3: the average lengths of the BY intervals are shown to be uniformly larger than the proposed intervals (18)); ——, BY; ——, proposed intervals (18); - - -, expression (17); ——, Bayes

where $\widehat{\text{Int}}$ is defined as in equation (10) with fdr_i being replaced by $\widehat{\text{fdr}}_i$. We then consider the intervals $\text{CI}_i^{\text{EB}}(q_R)$.

These intervals, by simulation, are shown to have the Bayes FCR larger than q for large π_0 and small τ^2. Note that π_1 and τ^2 in expression (9) are nearly unidentifiable when $\pi_1\tau^2$ is small. When this happens, probably, one cannot do any better than the frequentist's intervals of Benjamini and Yekutieli. Hence, when the ratio of estimated value $\hat{\pi}_1\hat{\tau}^2$ and σ_i^2 is small, we turn to use the BY intervals to ensure a satisfactory Bayes FCR. Consequently, the interval proposed is

$$\text{CI}_i^{\text{EBM}} = \begin{cases} \text{interval (17)}, & \text{if } \hat{\pi}_1\hat{\tau}^2/\sigma_i^2 > \text{cut} = \min\{0.6, (270/p)^{1/2}\}, \\ X_i \pm z_{Rq/2p}\sigma_i, & \text{if } \hat{\pi}_1\hat{\tau}^2/\sigma_i^2 > \text{cut}, \end{cases} \tag{18}$$

where $z_{Rq/2p}$ is z such that $P(|Z| < z) = Rq/p$ and the cut-off point is chosen numerically to have a controlled FCR. Note that $\pi_1\tau^2$ is always identifiable even though π_1 and τ^2 are nearly identifiable for small $\pi_1\tau^2$. Qiu and Hwang (2007) employed a similar cut-off point where 720 is used instead of 270, which gives a shorter interval in expression (18).

In Fig. 4, we graphed the simulated Bayes FCR for various procedures. The red curve corresponds to intervals (18). It is clearly seen that the Bayes FCR is controlled at 5%, the nominal level for all the hyperparameter settings. Hence, numerical evidence shows that these intervals control the empirical Bayes FCR. The red broken curve corresponds to the intervals (17) with no mixing with Benjamini and Yekutieli's (2005) procedure. It fails to control the Bayes FCR at 5% when τ^2 is small. Thus intervals (17) are not empirical Bayes FCR controlling intervals, indicating that a correction proposed in expression (18) is necessary. In Fig. 5, we plot the average half-length of all the procedures. It is clearly seen that intervals (18) have uniformly shorter average length than that of Benjamini and Yekutieli's (2005) procedure and could be 40% shorter in many cases.

We have examined other settings corresponding to various choices of π_0 and τ^2 that are not reported here. (See the caption of Fig. 4.) All the figures indicate that the intervals (18) control the empirical Bayes FCR at 5% and have uniformly shorter average length than Benjamini and Yekutieli's (2005) procedure with much reduction in length in many cases.

6. Application to the golden spike-in data of Choe *et al.* (2005)

In this section, we apply the various approaches to the golden spike-in data set of Choe *et al.* (2005). A striking feature of this data set is that the true parameters are known. Because of this, many researchers use this data set to test their statistical procedures. Hwang *et al.* (2009) applied to the data their double-shrinkage confidence intervals for the parameters θ_i where θ_i is the true differential expression of the ith gene. Zhao (2010) applied his double-shrinkage point estimator to estimate all the θ_is.

We downloaded the data from `http://www.elwood9.net/spike` and then manipulated the data in the same way as Hwang *et al.* (2009). The data set, for each of the 14010 genes, consists of six estimated differential expression levels, where the first three, denoted Y_1, Y_2 and Y_3, correspond to the control group whereas Z_1, Z_2 and Z_3 correspond to the treatment group. Let $X = \bar{Y} - \bar{Z}$ and $S^2 = \frac{1}{3}(S_Y^2 + S_Z^2)$ where $\bar{Y}$ and $\bar{Z}$ are the sample average and S_Y^2 and S_Z^2 the unbiased estimators of the variances of Y_js and Z_js. We then use $|X|/S$ and the distribution of $|T|$, where T has a t-distribution with ν degrees of freedom (between 2 and 4) estimated by Satterthwaite's approximation to calculate the p-value, and we apply the procedure of Benjamini and Hochberg (1995) to select genes where the false discovery rate is controlled at the 5% level.

Our recommended intervals are denoted as $\mathrm{CI}_i^{\mathrm{EBM}*}$, which are adapted from intervals (18) where we replace z_q by t except that the z_q in τ_0^2 remains unchanged where τ_0^2 was derived on the basis of the central limit theorem for a large p. Also σ_i^2 is replaced by the exponential Lindley–James–Stein estimator $\hat{\sigma}_i^2$ that was proposed in expression (2.3) of Cui *et al.* (2005), where X_g is replaced by νS_i^2. It is necessary to use a shrinkage estimator such as $\hat{\sigma}_i^2$. Note that here the selection is based on X_i and S_i^2. Unless shrinking on S_i^2 is done, there will be selection bias and the FCR will be too big, just like in the known σ_i^2 case where the naive z-intervals do not work well since X_is are not shrunken. Hence

$$\mathrm{CI}_i^{\mathrm{EBM}*} = \text{expression (18)}, \qquad \text{with } \sigma_i^2 \text{ and } z \text{ replaced by } \hat{\sigma}_i^2 \text{ and } t. \tag{19}$$

We calculate the proportion of selected genes whose corresponding intervals fail to cover the true parameters for both our approach (19) and Benjamini and Yekutieli's (2005) procedure. The proportion is the actual FCR of the data. Surprisingly, the actual FCRs of these two procedures are unreasonably high, being 72.77% and 72.45% respectively.

In 2009, one of the authors, Hwang, and his collaborators, Jia-Chiun Pan and Professor Guan-Hua Huang, rediscovered a phenomenon which had been known in Bolstad *et al.* (2003), Irizarry *et al.* (2003), Wu and Irizarry (2004) and Cope *et al.* (2004). It was found that the golden spike-in data of Choe *et al.* (2005) consistently underestimate the 'true' parameter, indicating that there is a serious violation of model assumptions of unbiasedness. In Fig. 6, we reproduce their result and plot X_i against θ_i. The red line is the 45° line. The green line is the regression line of X_is on θ_is based on the linear model $X = a\theta$. The least squares estimator $\hat{a} = \Sigma_i X_i\theta_i / \Sigma_i \theta_i^2$ equals 0.5327, which indicates that X_is tend to underestimate θ_is severely. This explains why the FCRs that were depicted in the previous paragraph are surprisingly high. In this paper, all the confidence intervals considered aim at the expectations of X_is which seems to be $a\theta_i$. Although in practice it is not possible to estimate a, here with the knowledge of θ_is we estimate a by $\hat{a} = 0.5327$. Hence it seems reasonable to evaluate the procedures in terms of capturing $\hat{a}\theta_i$. After the bias correction, we then recalculate the actual FCRs of the procedure and report them in Table 1. The nominal level for the FCR is set to be 5%. Among all these intervals, expression (19) is the best. Although the data most probably fail the assumptions that are used to derive expression (19), expression (19) has the actual FCR equal to 3.82%, which is close to but smaller than 5%. See the first FCR column. From the same column, the Bonferroni intervals have the actual FCR equal to 0.16%, which is much smaller than 5% and are 10 times as long as interval (19). See the first column under the heading 'Average half-length'. Looking at the first FCR column, one may be surprised to see that the actual FCR of the BY intervals is 26.43%, which is much higher than 5%. This does not agree with their theory which asserts control of the FCR. We suspect that this is due to the failure of the t-approximation using Satterthwaite's approximated degrees of freedom. When we use the conservative choice, the 2 degrees of freedom, the actual FCR of Benjamini and Yekutieli (2005) becomes 1.11%, which is less than 5%, agreeing with their theory. See the second FCR column in Table 1. Out of curiosity, we also calculate the actual FCR of expression (19) by using 2 degrees of freedom, which turns out to be 0.64%. See the second FCR column. Even with such a conservative FCR, the average half-length is still good. Here we can see another advantage of intervals that are derived on the basis of a Bayesian model with zero mixture prior; they are well protected in the FCR even when models are not well approximated. Consequently when applying intervals (19), one need not be concerned about making further correction of the degrees of freedom.

Another major advantage of expression (19) using Satterthwaite's degrees of freedom is that it has the smallest average length, compared with all intervals studied in Table 1, some of which do not have FCR controlled. The intervals $\mathrm{CI}_{\mathrm{SS}}$ are the intervals of Hwang *et al.* (2009). They

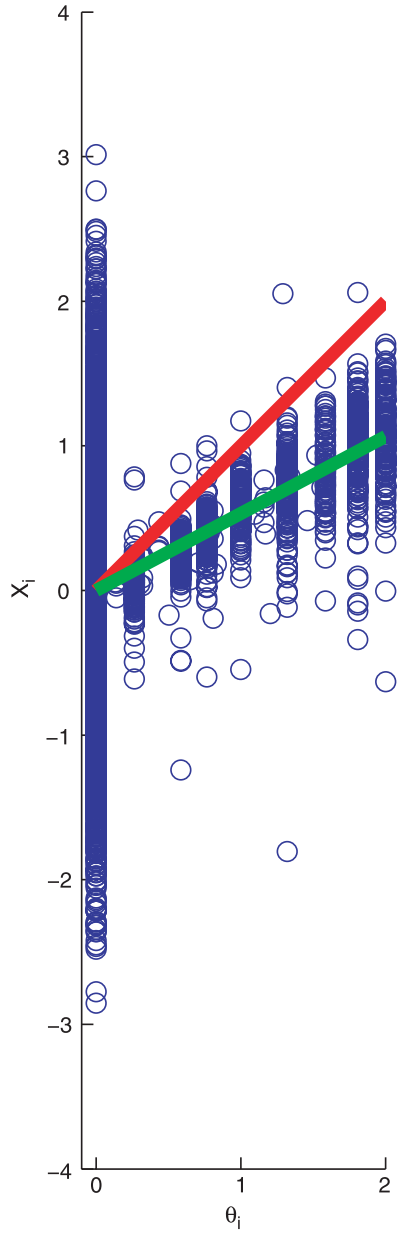

Fig. 6. The regression line (——) of the scatter plot (O) of the observed differential expression levels X_i *versus* the true differential expression levels θ_i is below the 45° line (——), showing that X_is tend to underestimate θ_is

have high FCR (9.55%) for 2 degrees of freedom. Although not calculated here, we expect the intervals to have even higher FCR if Satterthwaite's approximated degrees of freedom are used.

All these intervals aim at the parameters that are selected by using the 5% false discovery rate controlling procedure of Benjamini and Hochberg (1995), based on the p-value calculated by using the t-distribution with Satterthwaite's degrees of freedom. We expect that the comparison results will be similar for other selection procedures.

Table 1. Actual FCR and average half-lengths for various intervals†

	FCR (%)	*Average half-length*		*FCR (%)*	*Average half-length*
Interval (19), Satterthwaite degrees of freedom	3.82	0.4072	Interval (19), 2 degrees of freedom	0.64	0.5781
BY interval, Satterthwaite degrees of freedom	15.29	0.4774	BY interval, 2 degrees of freedom	1.11	1.1315
Bonferroni interval, Satterthwaite degrees of freedom	0.16	4.1439	CI_{SS} interval, 2 degrees of freedom	9.55	0.4085

†The parameters of interest are selected according to the Benjamini and Hochberg (1995) false discovery rate controlling procedures with $\alpha = 0.05$ by using a t-distribution with the degrees of freedom calculated from Satterthwaite's approximation. The best intervals are expression (19) with Satterthwaite's approximated degrees of freedom, having the shortest average length while controlling the actual FCR. BY intervals with the same approximated degrees of freedom surprisingly fail to control the FCR. BY intervals with a more conservative 2 degrees of freedom have controlled the FCR. The Bonferroni intervals have controlled the FCR, but have long average length. CI_{SS} are the intervals that were proposed in Hwang *et al.* (2009), which fail to control the actual FCR.

7. Conclusion

Benjamini and Yekutieli (2005) constructed intervals (which we called BY intervals) that control the frequentist's FCR for every possible value of key parameters θ_i. In this paper, we propose to control the average of the frequentist's FCR with respect to the weight function $\pi(\theta)$ or the Bayes FCR with respect to the prior $\pi(\theta)$. When θ_is are the random effects in a mixed analysis-of-variance model, the Bayes FCR is exactly the frequentist's FCR. By controlling a class of Bayes FCR, shorter intervals are constructed, which are applicable to random-effect models. Unlike BY intervals, the centres of the intervals proposed reduce much of the selection bias. Therefore, the resulting intervals are shorter and still have the controlled Bayes FCR for a class of priors.

Acknowledgements

We thank Professor Guan-Hua Huang at Chiao Tung University for pointing out several interesting references. We also thank Jing-huei Chen and Becky Brody for careful editing, which led to a better version of the paper.

Zhigen Zhao was supported in part by National Science Foundation grant DMS-03-06497.

Appendix A: Proofs

A.1. Proof of theorem 1

Let $Q = (V/R)I(R > 0)$. By the definition of FCR_π,

$$\mathrm{FCR}_\pi = E\{E(Q|X)\} = \int_{\{R>0\}} E(Q|X)\, m(X)\, \mathrm{d}X.$$

Since, conditioning on X, R is non-random,

$$E(Q|X) = E\left(\frac{V}{R}\middle| X\right) = \frac{E(V|X)}{R}.$$

By the definition of V, we have

$$E(V|X) = \sum_i E(I_{\{\theta_i \notin \mathrm{CI}_i, \text{ and } i \text{ is selected}\}}|X) = \sum P(\theta_i \notin \mathrm{CI}_i|X)\, I(i \text{ is selected}).$$

This implies that $E(V|X) = \Sigma_{i\in\mathcal{R}}\, P(\theta_i \notin \mathrm{CI}_i|X)$, completing the proof.

A.2. Proof of theorem 3

Theorem 1 and the first assumption of this theorem imply that FCR_π is bounded above by $E\{\alpha(p, X)\}$, which equals $A + B$ where

$$A = \int_{\alpha(p,X)>q+\varepsilon} \alpha(p, X)\, m(X)\, \mathrm{d}X$$

and

$$B = \int_{\alpha(p,X)\leqslant q+\varepsilon} \alpha(p, X)\, m(X)\, \mathrm{d}X.$$

The fact that $\alpha(p, X) \leqslant 1$ implies that A is bounded above by $P\{\alpha(p, X) > q + \varepsilon\} \to 0$ by the assumption of this theorem. Obviously B is bounded above by $q + \varepsilon$. These conclude that $\limsup_{p\to\infty} E\{\alpha(p, X)\} \leqslant q + \varepsilon$ for every $\varepsilon > 0$ and hence the same inequality is true for $\varepsilon = 0$. We now conclude the theorem.

A.3. Proof of theorem 4

Since

$$\mathrm{FCR}_\pi = E\left\{ \frac{\sum_{i=1}^{p} P(\theta_i \notin \mathrm{CI}_i|X)\, I(i \text{ is selected})}{R} \right\},$$

$$\begin{aligned} |\mathrm{FCR}_\pi - q| &= \left| E\left[\frac{1}{R} \sum_{i=1}^{p} \{ P(\theta_i \notin \mathrm{CI}_i|X) - q \}\, I(i \text{ is selected}) \right] \right| \\ &\leqslant E\left[\frac{1}{R/p} \frac{1}{p} \sum_{i=1}^{p} |\{ P(\theta_i \notin \mathrm{CI}_i|X) - q \}\, | I(i \text{ is selected}) \right]. \end{aligned}$$

Letting $p \to \infty$ and passing the limit inside the expectation, which is allowed by the bounded convergence theorem and the fact that the integrand is bounded by 2, we obtain

$$\lim_{p\to\infty} |\mathrm{FCR} - q| \leqslant \frac{1}{\eta} E\left\{ \lim_{p\to\infty} \frac{1}{p} \sum |P(\theta_i \notin \mathrm{CI}_i|X) - q| \right\}, \tag{20}$$

which by equation (4) equals 0, establishing the first part. The second part can be similarly proved.

A.4. Proof of theorem 5

Before we prove theorem 5, we state and prove the following lemma.

Lemma 1. If $\Sigma_{i=1}^{p} \sigma_i^4 = o\{p^2/\log(p)^{1+\varepsilon}\}$, then $\log(p)^{(\varepsilon+1)/2}(\hat{\mu} - \mu) \to 0$ in probability, and $\log(p)^{(\varepsilon+1)/2} \times (\hat{\tau}^2 - \tau^2) \to 0$ in probability. Similarly, if $\Sigma_{i=1}^{p} \sigma_i^4 = o(p^2)$, then both $\hat{\tau}^2 - \tau^2$ and $\hat{\mu} - \mu$ converge to 0 in probability.

Since $(\Sigma_{i=1}^{p} \sigma_i^2)^2 \leqslant p(\Sigma_{i=1}^{p} \sigma_i^4)$,

$$\sum_{i=1}^{p} \sigma_i^2 \leqslant \left\{ p\left(\sum_{i=1}^{p} \sigma_i^4 \right) \right\}^{1/2} = o\left\{ \frac{p^{3/2}}{\log(p)^{(1+\varepsilon)/2}} \right\}.$$

Since $\hat{\mu} = \bar{X}$, then $E(\hat{\mu}) = E(\bar{X}) = \mu$, and $\mathrm{var}(\hat{\mu}) = (\Sigma_{i=1}^{p} \sigma_i^2)/p^2$. For any $\delta_1 > 0$ and $\delta_2 > 0$, Chebyshev's inequality implies that

$$P\left\{\log(p)^{(\varepsilon+1)/2}|\hat{\mu}-\mu|>\frac{\delta_1}{2}\right\}<\frac{4\log(p)^{\varepsilon+1}}{\delta_1^2}\mathrm{var}(\hat{\mu})=\frac{4\log(p)^{\varepsilon+1}\sum_{i=1}^{p}\sigma_i^2}{p^2\delta_1^2}\to 0.$$

Therefore $\log(p)^{(\varepsilon+1)/2}(\hat{\mu}-\mu)\to 0$ in probability as $p\to\infty$. Later we shall need the fact

$$\log(p)^{(\varepsilon+1)/2}(\hat{\mu}^2-\mu^2)\to 0, \qquad \text{in probability.}$$

This can easily be proved by writing the left-hand side as a product of $\log(p)^{(\varepsilon+1)/2}(\hat{\mu}-\mu)$ and $\hat{\mu}+\mu$, where the first term converges to 0 in probability and the other to a constant in probability.

Now to prove the second part, by equation (5), we have

$$\begin{aligned}P\{\log(p)^{(\varepsilon+1)/2}|\hat{\tau}^2-\tau^2|>\delta_1\}&\leqslant P\left\{\log(p)^{(\varepsilon+1)/2}\left|\frac{\sum(X_i^2-\sigma_i^2)}{p}-\hat{\mu}^2-\tau^2\right|>\delta_1\right\}\\&\leqslant P\left\{\log(p)^{(\varepsilon+1)/2}\left|\frac{\sum(X_i^2-\sigma_i^2)}{p}-\mu^2-\tau^2\right|>\frac{\delta_1}{2}\right\}+P\left\{\log(p)^{(\varepsilon+1)/2}|\hat{\mu}^2-\mu^2|>\frac{\delta_1}{2}\right\}.\end{aligned}$$

Note that the second term converges to 0, and we only need to deal with the first term. Let

$$f(X)=\log(p)^{(\varepsilon+1)/2}\frac{\sum(X_i^2-\sigma_i^2-\mu^2-\tau^2)}{p}.$$

Then $E\{f(X)\}=0$ and

$$\mathrm{var}\{f(X)\}=\frac{\log(p)^{(\varepsilon+1)}}{p^2}\sum\mathrm{var}(X_i^2)$$

by independence of X_is. Direct calculation shows that

$$\mathrm{var}(X_i^2)=2(\sigma_i^2+\tau^2)(\tau^2+\sigma_i^2+2\mu^2),$$

and consequently

$$\mathrm{var}\{f(X)\}=\left\{\frac{2\sum\sigma_i^4}{p^2}+\frac{4(\tau^2+\mu^2)\sum\sigma_i^2}{p^2}+\frac{2\tau^2(\tau^2+2\mu^2)}{p}\right\}\log(p)^{\varepsilon+1}=o(1).$$

Chebyshev's inequality then implies that $\log(p)^{(\varepsilon+1)/2}(\hat{\tau}^2-\tau^2)\to 0$ in probability. The same argument applies to the second part of the lemma.

Now we are ready to prove theorem 5. Let $X_{(i)}$ be the order statistics of X_i in magnitude so that $|X_{(1)}|\leqslant|X_{(2)}|\leqslant\ldots\leqslant|X_{(p)}|$, and let $\theta_{(i)}$ and $\sigma_{(i)}^2$ be the parameters corresponding to the observation $X_{(i)}$. Write X_i as $\mu+(\sigma_i^2+\tau^2)^{1/2}Z_i$. Since $Z_1,Z_2,\ldots,Z_p$ are IID standard normal random variables, $\max|Z_i|/\{2\log(p)\}^{1/2}$ converges to some random variable in distribution. (See example 9.5.3 on page 259 of Woodroofe (1975).) Consequently, for any $\varepsilon>0$,

$$\frac{\max|Z_{(i)}|}{\{2\log(p)\}^{(1+\varepsilon)/2}}\to 0 \qquad \text{in probability.}$$

According to lemma 1, both $\log(p)^{(\varepsilon+1)/2}(\hat{\mu}-\mu)$ and $\log(p)^{(\varepsilon+1)/2}(\hat{\tau}^2-\tau^2)$ converge to 0 in probability. Now, for any positive number δ_1, let

$$A_p=\left[\log(p)^{(\varepsilon+1)/2}|\hat{\mu}-\mu|\leqslant\delta_1,\ \log(p)^{(\varepsilon+1)/2}|\hat{\tau}^2-\tau^2|\leqslant\delta_1,\ \frac{\max|Z_{(i)}|}{\{2\log(p)\}^{(1+\varepsilon)/2}}\leqslant\delta_1\right].$$

The above results imply that $P(A_p)\to 1$ as $p\to\infty$.

According to the construction of $\mathrm{CI}_i^{\mathrm{EB}}$ in equation (6), $P(\theta_i\notin\mathrm{CI}_i^{\mathrm{EB}}|X)=P[|\theta_i-\{\hat{M}_iX_i+(1-\hat{M}_i\}\hat{\mu})|>zM_i^{1/2}\sigma_i|X]$. Since $\theta_i|X_i\sim N(M_iX_i,M_i\sigma_i^2)$, we can write θ_i as $M_iX_i+M_i^{1/2}\sigma_iZ$ where Z is a standard normal random variable independent of X_i. Let $g_1=(\hat{M}_i-M_i)(X_i-\mu)/M_i^{1/2}\sigma_i$ and $g_2=(1-\hat{M}_i)(\hat{\mu}-\mu)/M_i^{1/2}\sigma_i$. The above probability $P(\theta_i\notin\mathrm{CI}_i^{\mathrm{EB}}|X)$ can be written as $P\{|Z-(g_1+g_2)|>z(\hat{M}_i/M_i)^{1/2}\}$ which is bounded above by the same expression with g_1 and g_2 replaced by their absolute values.

Assuming that A_p holds, and using the fact that $\hat{\tau}^2+\sigma_i^2 \geqslant 2\hat{\tau}\sigma_i$, we have

$$|g_1| = \left| \frac{\sigma_i(\hat{\tau}^2-\tau^2)}{\tau(\hat{\tau}^2+\sigma_i^2)} Z_i \right| \leqslant \left| \frac{\hat{\tau}^2-\tau^2}{2\hat{\tau}\tau} \right| |\max(Z_{(i)})| \leqslant C_1\delta_1.$$

The other term g_2 can be written as $\sigma_i(\tau^2+\sigma_i^2)^{1/2}/\{(\hat{\tau}^2+\sigma_i)^{1/2}\tau\} \times (\hat{\mu}-\mu)$. Since

$$\frac{\sigma_i(\tau^2+\sigma_i^2)^{1/2}}{(\hat{\tau}^2+\sigma_i^2)\tau} = \frac{\sigma_i}{\tau(\hat{\tau}^2+\sigma_i^2)^{1/2}} \left(\frac{\tau^2+\sigma_i^2}{\tau^2+\sigma_i^2} \right)^{1/2} \leqslant \frac{1}{\tau} \max\left(1, \frac{\tau^2}{\hat{\tau}^2}\right) \leqslant C_2,$$

$g_2 \leqslant C_2\delta_1$. In the above calculations, C_1 and C_2 denote constants depending on τ^2 only and not on σ_i^2s or P. Furthermore

$$\left| \left(\frac{\hat{M}_i}{M_i} \right)^{1/2} - 1 \right| \leqslant \left| \frac{\hat{M}_i}{M_i} - 1 \right| = \left| \frac{\sigma_i^2(\hat{\tau}^2-\tau^2)}{\tau^2(\hat{\tau}^2+\sigma_i^2)} \right| \leqslant \left| \frac{\hat{\tau}^2-\tau^2}{\tau^2} \right| \leqslant \frac{\delta_1}{\tau^2}.$$

Therefore, when A_p holds, for any $i=1,2,\ldots,p$,

$$P(\theta_i \notin \mathrm{CI}_i^{\mathrm{EB}}|X) \leqslant P\left\{ ||Z| - (C_1+C_2)\delta_1| > z\left(1-\frac{\delta_1}{\tau^2}\right) \right\} \to q, \qquad \text{as } \delta_1 \to 0.$$

Therefore, for any $\varepsilon > 0$, we can always find sufficiently small δ_1, such that

$$\alpha(p,X) = \max_{1\leqslant i\leqslant p} \{P(\theta_i \notin \mathrm{CI}_i|X)\} < q+\varepsilon \qquad \text{when } A_p \text{ holds.}$$

As a result $P\{\alpha(p,X)-q>\varepsilon\} \leqslant P(A_p^{\mathrm{c}}) \to 0$ as $p \to \infty$. Now theorem 3 concludes the theorem.

A.5. Proof of theorem 6

It suffices to show that condition (4) holds. According to lemma 1, for any $\delta_1>0$, $\delta_2>0$, $\lim P(A_p)=1$ where $A_p = \{|\hat{\mu}-\mu| \leqslant \delta_1, |\hat{\tau}^2-\tau^2| \leqslant \delta_2\}$. In the proof below, we could and would impose or remove the constraint A_p without affecting the asymptotic probability.

We may write $\theta_i = M_iX_i + (1-M_i)\mu + Z(M_i\sigma_i^2)^{1/2}$, where $Z \sim N(0,1)$ and is independent of X_i. (This is because $Z|X$ is $N(0,1)$ and it has $N(0,1)$ unconditionally as well.) Consequently, Z is independent of X.

$$P(\theta_i \notin \mathrm{CI}_i|X) = P\left\{ \left| Z - \frac{(\hat{M}_i-M_i)(X_i-\mu)+(1-\hat{M}_i)(\hat{\mu}-\mu)}{M_i^{1/2}\sigma_i} \right| > z\left(\frac{\hat{M}_i}{M_i}\right)^{1/2} \right\},$$

where CI and z are the abbreviated notation for $\mathrm{CI}^{\mathrm{EB}}$ and z_q defined in expressions (6) and (7). In the above probability, Z is the only random variable and X_i and $\hat{M}_i$ are viewed as constants until after expression (22) when we need to apply the law of large numbers. We write that $X_i-\mu = Z_i(\tau^2+\sigma_i^2)^{1/2}$ where Z_is are viewed as non-random. Hence

$$P(\theta_i \notin \mathrm{CI}_i|X) = P\left\{ \left| Z - \frac{(\hat{M}_i-M_i)(\tau^2+\sigma_i^2)^{1/2}}{M_i^{1/2}\sigma_i} Z_i + \frac{(1-\hat{M}_i)(\hat{\mu}-\mu)}{M_i^{1/2}\sigma_i} \right| > z\left(\frac{\hat{M}_i}{M_i}\right)^{1/2} \right\}.$$

Under the assumption that A_p holds, similarly to in the proof of theorem 5,

$$\left| \frac{(\hat{M}_i-M_i)(\tau^2+\sigma_i^2)^{1/2}}{M_i^{1/2}\sigma_i} \right| = \left| \frac{\sigma_i(\hat{\tau}^2-\tau^2)}{\tau(\hat{\tau}^2+\sigma_i^2)} \right| < \left| \frac{\hat{\tau}^2-\tau^2}{2\tau\hat{\tau}} \right| < C_1\delta_2,$$

where, in the first inequality, we use the inequality $\tau^2+\sigma_i^2 > 2\hat{\tau}\sigma_i$. Also, under A_p,

$$\left| \frac{(1-\hat{M}_i)(\hat{\mu}-\mu)}{M_i^{1/2}\sigma_i} \right| = \left| \frac{\sigma_i(\tau^2+\sigma_i^2)^{1/2}(\hat{\mu}-\mu)}{\tau(\hat{\tau}^2+\sigma_i^2)} \right| < C_2\delta_1,$$

and $|(\hat{M}_i/M_i)^{1/2}-1| \leqslant \delta_1/\tau^2$. In the above expressions, C_1 and C_2 depend on τ only and not on i, σ_i^2 or p Consequently, $P(\theta_i \notin \mathrm{CI}_i|X) \leqslant P\{|Z-C_1\delta_2|Z_i| - C_2\delta_1| > z(1-\delta_1/\tau^2)\}$. Also we could similarly establish the lower bound

$$P(\theta_i \notin \mathrm{CI}_i|X) \geqslant P\{|Z| \geqslant z(1+\delta_1/\tau^2)\},$$

implying that

$$|P(\theta_i \notin \mathrm{CI}_i|X) - q| \leqslant \max[|q - P\{|Z| \geqslant z(1+\delta_1/\tau^2)\}|, |q - P\{|Z - C_1\delta_2|Z_i| - C_2\delta_1| > z(1-\delta_1/\tau^2)\}|].$$

Summing over i on both sides, we have

$$\frac{1}{p}\sum |P(\theta_i \notin \mathrm{CI}_i|X) - q| \leqslant \max(A, B), \tag{21}$$

where $A = |q - P\{|Z| > z(1+\delta_1/\tau^2)\}|$, and

$$B = \frac{1}{p}\sum_i \left| q - P\left\{ |Z - C_1\delta_2|Z_i| - C_2\delta_1| > z\left(1 - \frac{\delta_1}{\tau^2}\right)\right\}\right|. \tag{22}$$

Now remove the condition A_p. Obviously, $A \to 0$ as $\delta_1 \to 0$. Also, the terms in B inside the summation are functions of Z_i, IID $N(0, 1)$. The law of large numbers implies that

$$B \to E\left| q - P\left\{ \left| Z - C_1\delta_2 Z_i - C_2\delta_1 \right| > z\left(1 - \frac{\delta_1}{\tau^2}\right)\right\}\right|,$$

where the expectation is with respect to Z_i. The dominated convergence theorem then implies that the expectation converges to $|q - P(|Z| > z)| = 0$ as δ_1 and δ_2 approach 0. This concludes that expression (21) converges to 0 as $p \to \infty$. Condition (4) is established and so is the theorem.

A.6. Proof of theorem 7

Theorem 1 implies that

$$\mathrm{FCR}_\pi = \int_{R>0} \frac{1}{R}\sum_{i\in\mathcal{R}} P(\theta_i \notin \mathrm{CI}_i|X)\, m(X)\, \mathrm{d}X.$$

The posterior non-coverage probability $P(\theta_i \notin \mathrm{CI}_i|X)$ can be written as

$$\begin{aligned} P(\theta_i \notin \mathrm{CI}_i|x) &= P(\theta_i \notin \mathrm{CI}_i|x, \theta_i = 0)\, P(\theta_i = 0|x) + P(\theta_i \notin \mathrm{CI}_i|x, \theta_i \neq 0)\, P(\theta_i \neq 0|x) \\ &\leqslant \mathrm{fdr}_i(x)\, I(0 \notin \mathrm{CI}_i|x) + \{1 - \mathrm{fdr}_i(x)\}q' = q' + \mathrm{fdr}_i(x)\{I(0 \notin \mathrm{CI}_i|x) - q'\}. \end{aligned}$$

As a result,

$$\mathrm{FCR}_\pi \leqslant q'\, P(R>0) + \int_{R>0} \frac{1}{R}\sum_{i\in\mathcal{R}} \mathrm{fdr}_i(x)\{I(0 \notin \mathrm{CI}_i|x) - q'\}\, m(x)\, \mathrm{d}x.$$

The upper bound equals

$$q'\, P(R>0) + \int_{R>0} \frac{1}{R}\left\{ \sum_{i\in\mathcal{R}, 0\notin \mathrm{CI}_i} \mathrm{fdr}_i(x)(1-q') - \sum_{i\in\mathcal{R}, 0\in \mathrm{CI}_i} q'\, \mathrm{fdr}_i(x)\right\} m(X)\, \mathrm{d}X,$$

which is identical to $\int \mathrm{Int}(q', x)\, m(x)\, \mathrm{d}x$, establishing the theorem.

References

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* B, **57**, 289–300.

Benjamini, Y. and Yekutieli, D. (2005) False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Am. Statist. Ass.*, **100**, 71–93.

Berger, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. New York: Springer.

Bolstad, B. M., Irizarry, R. A., Åstrand, M. and Speed, T. P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.

Casella, G. and Hwang, J. T. (1983) Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Am. Statist. Ass.*, **78**, 688–698.

Chen, J. and Sarkar, S. (2006) A Bayesian determination of threshold for identifying differentially expressed genes in microarray experiments. *Statist. Med.*, **25**, 3174–3189.
Choe, S. E., Bouttros, M., Michelson, A. M., Chruch, G. M. and Halfon, M. (2005) Preferred analysis methods for affymetrix genechips revealed by a wholly defined control dataset. *Genome Biol.*, **6**, article R16.1–16.
Cope, L. M., Irizarry, R. A., Jaffee, H. A., Wu, Z. and Speed, T. P. (2004) A benchmark for Affymetrix GeneChip expression measures. *Bioinformatics*, **20**, 323–331.
Cui, X., Hwang, J. T., Qiu, J., Blades, N. J. and Churchill, G. A. (2005) Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics*, **6**, 59–75.
Efron, B. (2005) Local false discovery rates. Department of Statistics, Stanford University, Stanford. (Available from `http://statistics.stanford.edu/ckirby/techreports/BIO/BIO%20234.pdf`.)
Efron, B. (2007) Size, power and false discovery rates. *Ann. Statist.*, **35**, 1351–1377.
Efron, B. (2008) Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.*, **23**, 1–22.
Efron, B. (2010) *Large-scale Inference, Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge: Cambridge University Press.
Gregory, P., Bert, A., Paterson, E., Barry, S., Tsykin, A., Farshid, G., Vadas, M., Khew-Goodall, Y. and Goodall, G. (2008) The mir-200 family and mir-205 regulate epithelial to mesenchymal transition by targeting zeb1 and sip1. *Nat. Cell Biol.*, **10**, 593–601.
Hwang, J. T. (1993) Empirical Bayes estimation for the means of the selected populations. *Sankhya* A, **55**, 285–304.
Hwang, J. T. and Liu, P. (2010) Optimal tests shrinkage both means and variances applicable to microarray data analysis. *Statist. Applic. Genet. Molec. Biol.*, **9**, article 36.
Hwang, J. T., Qiu, J. and Zhao, Z. (2009) Empirical Bayes confidence intervals shrinking both means and variances. *J. R. Statist. Soc.* B, **71**, 265–285.
Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B. and Speed, T. P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, **31**, article e15.
Jin, J. and Cai, T. T. (2007) Estimating the null and the proportional of nonnull effects in large-scale multiple comparisons. *J. Am. Statist. Ass.*, **102**, 495–506.
Kendziorski, C., Newton, M., Lan, H. and Gould, M. (2003) On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Statist. Med.*, **22**, 3899–3914.
Lönnstedt, I. and Speed, T. (2002) Replicated microarray data. *Statist. Sin.*, **12**, 31–46.
Morris, C. N. (1983) Parametric empirical Bayes inference: theory and applications (with discussion). *J. Am. Statist. Ass.*, **78**, 47–65.
Nettleton, D., Hwang, J. T., Caldo, R. A. and Wise, R. P. (2006) Estimating the number of true null hypotheses from a histogram of p values. *J. Agric. Biol. Environ. Statist.*, **11**, 337–356.
Newton, M. A., Kendziorski, C. M., Richmond, C. S., Blattner, F. R. and Tsui, K. W. (2001) On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *J. Computnl Biol.*, **8**, 37–52.
Qiu, J. and Hwang, J. T. (2007) Sharp simultaneous intervals for the means of selected populations with application to microarray data analysis. *Biometrics*, **63**, 767–776.
Ruppert, D., Nettleton, D. and Hwang, J. T. (2007) Exploring the information in p-values for the analysis and planning of multiple-test experiments. *Biometrics*, **63**, 483–495.
Sarkar, S. K., Zhou, T. and Ghosh, D. (2008) A general decision theoretic formulation of procedures controlling fdr and fnr from a Bayesian perspective. *Statist. Sin.*, **18**, 925–945.
Smyth, G. K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statist. Applic. Genet. Molec. Biol.*, **3**, no. 1, article 3.
Storey, J. D. (2002) A direct approach to false discovery rates. *J. R. Statist. Soc.* B, **64**, 479–498.
Subkhankulova, T. and Livesey, F. (2006) Comparative evaluation of linear and exponential amplification techniques for expression profiling at the single-cell level. *Genome Biol.*, **7**, no. 3, article R18.
Tai, Y. C. and Speed, T. P. (2006) A multivariate empirical Bayes statistic for replicated microarray time course data. *Ann. Statist.*, **34**, 2387–2412.
Woodroofe, M. (1975) *Probability with Applications*. New York: McGraw-Hill.
Wu, Z. and Irizarry, R. A. (2004) Preprocessing of oligonucleotide array data. *Nat. Biotech.*, **22**, 656–658.
Yekutieli, D. (2008) Adjusted Bayesian inference for selected parameters. *Arxiv Preprint arXiv:0801.0499*.
Zhao, Z. (2010) Double shrinkage empirical Bayesian estimation for unknown and unequal variances. *Statist. Interface*, **3**, 533–541.