Empirical Bayes FCR Controlling Confidence Intervals

Zhigen Zhao

346 Speakman Hall, Department of Statistics, Fox School of Business
Temple University, 1810 N. 13th Street, Philadelphia, 19122
USA, Supported in part by NSF Grant DMS-03-06497

J. T. Gene Hwang

Department of Mathematics and Department of Statistical Science, Malott Hall, Cornell University, Ithaca, NY, 14850, USA; Department of Statistics, Cheng Kung University, Tainan, Taiwan

June 29, 2010

Summary.
Benjamini and Yekutieli (2005) suggested that it is important to account for multiplicity correction for confidence intervals when only some of these intervals are reported after selection. They introduced the concept of *false coverage rate (FCR)* for confidence intervals which is parallel to the concept of FDR in the multiple hypothesis testing problem and developed confidence intervals for selected parameters which control FCR.
In this paper, we consider the FCR criterion to be too stringent, requiring that it is controlled (in the frequentist's sense) for all the possible unknown parameter values. This is especially true if the number of unknown parameters is large, for example tens of thousands or more as often being involved in modern applications including the microarray analysis. We propose and study a less stringent criterion, the control of the *empirical Bayes* FCR for confidence intervals. This refers to the control of average FCR with respect to a class of distributions of parameters. Under such a criterion, we study some empirical Bayes confidence intervals, which, by some analytic and numerical calculations, are demonstrated to have the FCR controlled at level q for a class of prior distributions. Furthermore, the proposed empirical Bayes intervals are always shorter, in average length, than (and could be one third or half as long as) the intervals of Benjamini and Yekutieli (2005). We apply these procedures to the data of Choe *et al.* (2005) and obtain similar results.

Keywords: Multiplicity; Simultaneous Intervals.

1. Introduction

In statistical analysis, confidence intervals are one of the most important tools. Unlike a hypothesis testing or a p-value, a confidence interval could provide a range of the true parameter θ_i while taking into consideration of the variability in estimating the parameter. The traditional evaluation of a confidence interval is based on the probability of covering the true parameter and the expected length.

Benjamini and Yekutieli (2005), however, proposes a very interesting criterion: the *False Coverage Rate (FCR)*. To explain the concept, we use the microarray data analysis as an example, although a similar question arises in many scientific studies. In a microarray experiment, a scientist selects many genes, typically the most differentially expressed genes perhaps by using a procedure controlling the false discovery rate (FDR). If the scientist is interested in reporting confidence intervals for the parameters corresponding to these selected genes, what should he do? This is the question raised

in Benjamini and Yekutieli (2005). Their proposed criterion is to examine the FCR, which is the average rate of false coverage, namely, not covering true parameters, among the selected intervals. They demonstrated that if one ignores the selection and uses the traditional (frequentist's) $1-q$ confidence interval, the FCR may be much higher than q and not controlled. They then constructed their confidence intervals, called B-Y intervals in this paper, that have a controlled FCR.

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 unknown parameters θ_i's. This requirement seems too stringent. For microarray experiments and other modern applications, there are a huge number of parameters, often tens of thousands or more; and it is customary that scientists' reasoning revolves around the probability of θ_i's, the differential expression levels (being equal to zero for example). It seems reasonable to consider the average FCR, averaging over such a distribution of θ_i's. In this paper, such average FCR is called the *Bayes FCR* while the distribution of θ_i's is called the Bayes prior distribution. In practice, the prior distribution can be speculated but never totally known. Hence a class of priors is considered instead. In this paper, we aim at constructing the *empirical* Bayes FCR controlling confidence intervals, *which are defined to be the intervals that guarantee that the Bayes FCR_π is 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 Bayes FCR controlling intervals only for one prior distribution.

Although we use the terminology of the Bayesian or empirical Bayesian, the criterion of Bayes FCR could be appropriate for frequentists too since it can be interpreted as the average FCR with respect to a weight function, the prior distribution. It is also essentially the frequentist FCR when θ_i are random as in the random effect models.

In section 2, we introduce all terminologies and our model. 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. Also some asymptotic theorems are derived to deal with any confidence intervals. In section 3, we apply these theorems in section 2 to a class of prior distributions. We establish that under certain conditions, the empirical Bayes FCR can be controlled asymptotically as the number of parameters p (number of genes) goes to infinity if the empirical Bayes confidence intervals in the sense of Morris (1983) are used. The asymptotic property holds regardless of the selection rule. We have also shown by simulations that certain empirical Bayes intervals control the *empirical* Bayes FCR when p is finite. Moreover, the empirical Bayes intervals are always shorter in average length and could be one third as long when compared to the B-Y's intervals. When applied to the data set of Choe *et al.* (2005) (after some bias correction of the data), we show that a modified version of the confidence intervals in Hwang, Qiu, and Zhao (2009) controls the actual FCR near 5% while the average length is much shorter than that of B-Y's intervals.

2. General Theorem on Bayes Intervals

We begin by giving the definition of *False Coverage Rate(FCR)* of confidence intervals, a term coined in Benjamini and Yekutieli (2005), which will be referred as B-Y (2005) throughout the paper. Consider one-dimensional parameters θ_i, $i = 1, \cdots, p$. 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 each gene corresponds to an ANOVA model (or a linear model). In such a case, X_i is the ANOVA 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 and θ_i is replaced by a vector (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. Application of ANOVA model including random parameters are many even in the microarray literature. See for example, Cui and Churchill (2003), which is cited in more than 400 papers, Cui *et al.* (2005); and Hwang and Liu (2009), and Qiu and Hwang (2007). Many papers in microarray literature also assume a random (Bayesian) model. See, for example, Newton *et al.* (2001), Kendziorski *et al.* (2003), Lönnstedt *et al.* (2005), and Tai and Speed (2006). None of these papers, however, address the FCR problem.

Let CI_i, based on X_i, be an interval for θ_i. Assume that $\mathcal{R}$ is a set of index i such that θ_i has been selected based on X_i's. 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 defined in B-Y (2005) is

$$FCR = E\frac{V}{R}I(R > 0),$$

where the expectation integrates out X, under the assumption that θ_i's are fixed. B-Y (2005) suggests to control FCR to be less than or equal to q, a small number, for every θ_i's. However, in modern technology like microarray, the number of parameters is very large. Therefore it is customary for biologists to describe and think about θ_i's in terms of its distribution. Therefore it seems natural to consider θ_i's as random variables having some distribution π. Hence it seems reasonable to consider the average FCR by integrating out the θ_i's with respect to their distribution π and define the Bayes FCR as

$$FCR_\pi = E_\pi E\frac{V}{R}I(R > 0).$$

Note that, similar to the finding in B-Y (2005) for the frequentist FCR, regarding t-intervals, the classical 95% z-intervals have Bayes FCR much larger than 5% and fail drastically to control it at 5% level as demonstrated in Figures 2 by the black dotted line. This is due to the fact that these parameters have been preselected - they are declared to be significantly different from zero when applying Benjamini and Hochberg (1995) 's procedure with FDR 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 p.d.f of $X = (X_1, \cdots, X_p)$ is $f_\theta(X)$ and the p.d.f. of $\theta = (\theta_1, \cdots, \theta_p)$ is $\pi(\theta)$.

THEOREM 2.1. *For any selection rule,*

$$FCR_\pi = \int_{R>0} E(Q|X)m(X)dX$$

where $E(Q|X) = \frac{1}{R}\sum_{i\in\mathcal{R}} P(\theta_i \notin CI_i|X)$ *and* $m(X) = \int f_\theta(X)\pi(\theta)d\theta$.

The proof of this theorem and all the other theorems below are given in the Appendix unless it is obvious from the context. Given Theorem 2.1, the following theorem is obvious.

THEOREM 2.2. *If* $P(\theta_i \notin CI_i|X) \le q$, $\forall i$, *then* $FCR_\pi \le qP(R > 0) \le q$, *for any selection rule based on* X *leading to* R. *(The proof is omitted.)*

In both Theorem 2.1 and 2.2 above, there is no independent assumption needed among X_i or among θ_i. This theorem provides us a straightforward way to construct confidence intervals with a controlled Bayes FCR when the prior distribution π is known. Let's consider the following example.

Example:
Assume that the independent $X_i \sim N(\theta_i, \sigma_i^2)$ $(i = 1, 2, \cdots, p)$ where σ_i^2's are known constants, and θ_i's are i.i.d. $N(\mu, \tau^2)$ distributed. Define the confidence interval CI_i^B as

$$CI_i^B = [M_iX_i + (1 - M_i)\mu] \pm z\sqrt{M_i}\sigma_i, \tag{1}$$

where $M_i = \frac{\tau^2}{\tau^2+\sigma_i^2}$ and $P(|Z| > z) = q$ for a standard normal random variable Z. Then CI_i^B has posterior coverage probability $1-q$. By Theorem 2.2, the Bayes FCR of (1) is no greater than q.

Theorem 2.2 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 3. The theorems below help to study the asymptotic properties of any confidence intervals.

THEOREM 2.3. *Assume that* $\max_{1\le i\le p} P(\theta_i \notin CI_i|X) = \alpha(p, X)$ *and*

$$\lim_{p\to\infty} P(\alpha(p, X) \le q + \epsilon) \to 1, \forall \epsilon > 0. \tag{2}$$

Then

$$\limsup_{p\to\infty} FCR_\pi \le q.$$

When condition (2) holds, we shall say that $\alpha(p, X)$ is asymptotically (as $p \to \infty$) less or equal to q in probability. Under such a condition, FCR_π is asymptotically controlled at the level q. The theorem aims at dealing with the most severe term $\max_{1\le i\le p} P(\theta_i \notin CI_i|X)$; 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 following.

THEOREM 2.4. *Assume that* $\frac{R}{p} \to \eta > 0$, *and*

$$\frac{1}{p}\sum_i |P(\theta_i \notin CI_i|X) - q| \to 0, \text{ almost surely}, \tag{3}$$

where q *is any number independent of* i. *Then* $\lim_{p\to\infty} FCR_\pi \to q$.

Similarly, if instead of (3), we have

$$\limsup_{p\to\infty} \frac{1}{p}\sum_i (P(\theta_i \notin CI_i|X) - q)_+ \le 0, \text{ almost surely},$$

where for a number a, $(a)_+$ denotes $\max(a, 0)$. Then $\limsup_{p\to\infty} FCR_\pi \le q$.

3. Empirical Bayes Approach

In Section 2, we showed that the Bayes confidence intervals can control the Bayes FCR. The result works for a prior which is precisely known, unrealistic in real applications. More realistically, we now deal with a class of priors indexed by some unknown hyper-parameters which will be estimated as in the *empirical* Bayes approach. This results intervals that will control the FCR_π for the class of priors and are called the *empirical* Bayes FCR controlling intervals.

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

$$\hat{\tau}^2 = (\frac{\sum_{i=1}^p (X_i^2 - \sigma_i^2)}{p} - \hat{\mu}^2)_+. \tag{4}$$

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

Substituting all the hyper-parameters in the interval (1) by their estimators above results in the so-called empirical Bayesian interval

$$CI_i^{EB} = [\hat{M}_i X_i + (1 - \hat{M}_i)\hat{\mu}] \pm z\sqrt{\hat{M}_i}\sigma_i. \tag{5}$$

Since all the estimators are obtained through the method of moment, we would expect that they should converge to the Bayes interval as $p \to \infty$. Hence asymptotically (5) would behave like the Bayes procedure (1), having the asymptotic Bayes FCR controlled 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 3.1. *For any* $\epsilon > 0$, *if* $\sum_{i=1}^p \sigma_i^4 = o(\frac{p^2}{(\log p)^{1+\epsilon}})$, *then* $\limsup_{p\to\infty} FCR_\pi \le q, \forall\pi$.

Alternatively, an application of theorem 2.4 provides us the asymptotic property under a less restrictive condition when the number of selection R increases as p increases.

THEOREM 3.2. *If* $\frac{R}{p} \to \eta > 0$, *and* $\sum_{i=1}^p \sigma_i^4 = o(p^2)$, *then* $\lim_{p\to\infty} FCR_\pi = q, \forall\pi$.

Both conditions on the order of $\sum_i \sigma_i^4$ are mild, much weaker than the result from the law of large numbers. More specifically, when σ_i^2's are generated as samples from a population with finite second moment, these two conditions are satisfied. Both Theorems 3.1 and 3.2 under their assumptions imply that the empirical Bayesian interval (5) asymptotically controls the FCR_π at the q-level for any non-degenerate normal prior π. However, when p is finite, the FCR_π can be higher than q. See Figure 2 for $q = 5\%$. Judging from Theorem 2.2, this is likely due to the fact that (5) has coverage probability lower than 95%. This suggests that we should try the confidence intervals of Hwang, Qiu, and Zhao (2009) adapting to the case of known variances. Namely we consider the intervals

$$CI_i^{HQZ} = [\hat{M}_i^* X_i + (1 - \hat{M}_i^*)\hat{\mu}] \pm \sqrt{\hat{M}_i^* V(\hat{M}_i^*)}\sigma_i \tag{6}$$

where $V(\hat{M}_i^*) = \sqrt{z^2 - \log(\hat{M}_i^*)}$, $\hat{M}_i^* = \frac{\hat{\tau}_*^2}{\hat{\tau}_*^2 + \sigma_i^2}$, $\hat{\tau}_*^2 = \max(\hat{\tau}^2, \tau_0^2)$, and

$$\tau_0^2 = \frac{2z^2\sum_i \sigma_i^2 + z\sqrt{4z^2(\sum_i \sigma_i^2)^2 + 2\sum_i \sigma_i^4(p^2 - 2pz^2)}}{p^2 - 2pz^2}.$$

Since (6) contains (5), under the assumptions of Theorems 3.1 and 3.2, (6) also have FCR_π controlled asymptotically. More importantly, intervals (6) have good finite coverage probability and its FCR_π is about 5% or less as demonstrated in Figure 2. Also the intervals of Bonferroni and B-Y have FCR_π controlled at 5%. However, it can be analytically proved that the B-Y's intervals are longer than the Z intervals which are longer than (6) if $z > 1$. Similar comments apply to the unknown σ_i's case when $t > 1$. Moreover, in Figure 3, we see that the average lengths of Bonferroni and B-Y are much longer than that of CI^{HQZ}, being three times longer in the most extreme cases. The intervals CI^{HQZ} have average length almost identical to the Bayes intervals, which have the minimum average lengths. However, unlike CI^{HQZ}, the Bayes intervals assume the knowledge of τ^2, unrealistic in real application. We plotted Figures 2 and 3 for other cases (see the caption of Figure 2) and the results are similar.

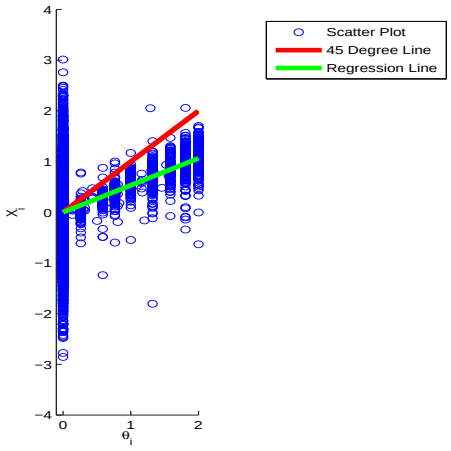

Fig. 1. The regression line of the observed differential expression levels X_i vs the true differential expression levels θ_i is below the 45° line, showing that X_i's tend to underestimate θ_i's.

4. 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, Qiu, and Zhao (2009) applied to the data their double shrinkage confidence intervals for the parameters θ_i's where θ_i is the true differential expression of the i-th gene. Zhao (2010) applied his double shrinkage point estimator to estimate all the θ_i's.

We download the data from `http://www.elwood9.net/spike` and then manipulate the data in the same way as Hwang, Qiu, and Zhao (2009). The data set, for each of the 14,010 genes, consists of six estimated differential expression levels, where the first three, denoted as Y_1, Y_2 ,and Y_3, correspond to the control group whereas Z_1, Z_2 ,and Z_3, 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_j's and Z_j's. We then use the distribution of $|T|$, where T has a t-distribution with degrees of freedom (between 2 and 4) estimated by Satterthwaite's approximation and $|X|/S$ to calculate the p-value, and we apply the procedure of Benjamini and Hochberg (1995) to select genes where the false discovery rate (FDR) is controlled at 5% level.

Our recommended intervals denoted as CI_i^{HQZ*}, which are adapted from (6) where we replace the z in $V(\cdot)$ by the t cutoff point, but not the z in τ_0^2, which was derived based on the central limit theorem for a large p. Also σ_i^2 is replaced by the exponential Lindley James-Stein estimator, $\hat{\sigma}_i^2$, proposed in Cui *et al.* (2005). Hence

$$CI_i^{HQZ*} = (6) \quad \text{with } \sigma_i^2 \text{ and } z \text{ in } V(\cdot) \text{ replaced by } \hat{\sigma}_i^2 \text{ and } t. \tag{7}$$

As in Hwang, Qiu, and Zhao (2009), which is based on Cui *et al.* (2005), designed only for equal degrees of freedom, we take degrees of freedom to be 2 for intervals in (7) when applied to the real data reported in Table (1).

We calculate the proportion of selected genes whose corresponding intervals fail to cover the true parameters for both our approach, (7), and B-Y's procedure. The proportion is the actual FCR of the data. Surprisingly, the actual FCRs of these two procedure are unreasonably high, being 65.45% and 53.34% respectively.

In 2009, one of the authors, Hwang, and his collaborators Jia-Chiun Pan and Guan-Hua Huang at Chiao Tung University, Taiwan, discovered that the data of Choe *et al.* (2005) consistently un-

derestimate the "true" parameter in the Gold Spike-in data set, indicating that there is a serious violation of model assumptions. In Figure 1, we reproduced their result and plotted X_i vs θ_i. The red solid line is the 45 degrees line. The green one is the regression line of X_i's and θ_is based on the linear model $X = a\theta$. The least squares estimator $\hat{a} = \frac{\sum_i X_i\theta_i}{\sum_i \theta_i^2}$ equals 0.5327, which indicates that X_i's tend to under estimate θ_i's severely. This explains why the FCR's in the previous paragraph are surprisingly high. In this paper, all the confidence intervals considered all aim at the expectations of X_i's which seems to be $a\theta_i$. Although in practice, it is not possible to estimate a, here with the knowledge of θ_i's, 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 this bias correction, we now recalculate the actual FCR's of the procedures and report them in Table 1. Aiming at controlling the FCR at 5%-level, our intervals have the FCR that is close to the nominal level. B-Y's procedure, using 2 as the degrees of freedom (df), has a small FCR and much longer average length than ours, almost twice as long. Further, B-Y's intervals, which uses the integer parts of Satterthwaite's df, has FCR 7.17%, larger than the nominal level and still is 20% longer than (7) on average. As anticipated, Bonferroni's correction is way too conservative, with an extreme long average length, about 3.5 times as long as the B-Y procedure (df=2) while the FCR is controlled at an extremely low level, 0.28%. Finally, CI_{SS}, the recommended intervals of Hwang, Qiu, and Zhao (2009), which replaces t^2 by t^2e^m in (7) where $m = E\log\frac{\chi^2_{df}}{df} \le 0$. This procedure reduces the average length but fails to control the FCR at 5% for this data set. Consequently, our approach (7) is recommended for applications. The corrected data most likely fail the assumptions we made to derive (7) and yet (7) still works well. There seems to be no point of trying the non-integer degrees of freedom of Scatterthwaite approximation for B-Y, since the FCR would be higher than 7.17%.

The code for calculating the *empirical* Bayes FCR-controlling confidence interval can be downloaded at `http://astro.temple.edu/~zhaozhg/publications.html`

5. Conclusion

In this paper, we propose controlling the empirical Bayes FCR as a criterion alternative to the (frequentist) FCR proposed by Benjamini and Yekutieli (2005). Especially when there are many parameters θ_i, it seems too stringent to require that the frequentist FCR be controlled for every θ_i's. However, the FCR averaging over θ_i's (or the Bayes FCR) would be more appropriate. By controlling the Bayes FCR for a class of priors, or the empirical Bayes FCR, we derive sharper confidence intervals.

Our research indicates that interval with *empirical* Bayes coverage probability usually would have the *empirical* Bayes FCR controlled at level q, regardless of the selection rule. The classical frequentist z-confidence intervals, however, fail to have the empirical Bayes FCR controlled. The empirical Bayes intervals center at the bias corrected estimator whereas other intervals do not. This is why it is so sharp, having controlled Bayes FCR for a class of priors and having much shorter length than other intervals.

Appendix: Proofs

Proof of Theorem 2.1. Let $Q = \frac{V}{R}I(R > 0)$. By the definition of FCR_π,

$$FCR_\pi = E(EQ|X) = \int_{\{R>0\}} (EQ|X)m(X)dX.$$

Since conditioning on X, R is non-random, $EQ|X = E\frac{V}{R}|X = \frac{EV|X}{R}$. By the definition of V, we have

$$EV|X = \sum_i E1_{\{\theta_i \notin CI_i, \text{ and i is selected}\}}|X = \sum P(\theta_i \notin CI_i|X)I(i \text{ is selected}).$$

This implies that $EV|X = \sum_{i\in\mathcal{R}} P(\theta_i \notin CI_i|X)$, completing the proof.

Proof of Theorem 2.3. Theorem 2.1 and the first assumption of this theorem imply that FCR_π is bounded above by $E\alpha(p,X)$, which equals $A+B$ where $A = \int_{\alpha(p,X)>q+\epsilon} \alpha(p,X)m(X)dX$ and $B = \int_{\alpha(p,X)\le q+\epsilon} \alpha(p,X)m(X)dX$. The fact that $\alpha(p,X) \le 1$ implies that A is bounded above by $P(\alpha(p,X) > q+\epsilon) \to 0$ by the assumption of this theorem. Obviously B is bounded above by $q+\epsilon$. These conclude that $\limsup_{p\to\infty} E(\alpha(p,X)) \le q+\epsilon$ for every $\epsilon > 0$ and hence the same inequality is true for $\epsilon = 0$. We now conclude the theorem.

Proof of Theorem 2.4. Since $FCR_\pi = E(\frac{\sum_{i=1}^p P(\theta_i\notin CI_i|X)I(\text{i is selected})}{R})$,

$$\begin{aligned} |FCR_\pi - q| &= |E(\frac{1}{R}\sum_{i=1}^p (P(\theta_i \notin CI_i|X) - q)I(\text{i is selected}))| \\ \le\quad & E(\frac{1}{R/p}\frac{1}{p}\sum_{i=1}^p |(P(\theta_i \notin CI_i|X) - q)|I(\text{i is selected})). \end{aligned}$$

Letting $p\to\infty$ and passing the limit inside the expectation, which is allowed by the dominated convergence theorem, we obtain

$$\lim_{p\to\infty} |FCR - q| \le \frac{1}{\eta} E \lim_{p\to\infty} \frac{1}{p}\sum |P(\theta_i \notin CI_i|X) - q|, \tag{A1}$$

which by (3) equals zero, establishing the first part. The second part can be similarly proved.

Proof of Theorem 3.1. Before we prove this theorem, we state and prove the following lemma.

LEMMA 5.1. *If $\sum_{i=1}^p \sigma_i^4 = o(\frac{p^2}{(\log p)^{1+\epsilon}})$, then $(\log p)^{\frac{\epsilon+1}{2}}(\hat\mu - \mu) \to 0$ in probability, and $(\log p)^{\frac{\epsilon+1}{2}}(\hat\tau^2 - \tau^2) \to 0$ in probability. Similarly, if $\sum_{i=1}^p \sigma_i^4 = o(p^2)$, then both $\hat\tau^2 - \tau^2$ and $\hat\mu - \mu$ converge to 0 in probability.*

Pf: Since $(\sum_{i=1}^p \sigma_i^2)^2 \le p(\sum_{i=1}^p \sigma_i^4)$,

$$\sum_{i=1}^p \sigma_i^2 \le \sqrt{p(\sum_{i=1}^p \sigma_i^4)} = o(\frac{p^{3/2}}{(\log p)^{(1+\epsilon)/2}}).$$

Since $\hat\mu = \bar X$, then $E\hat\mu = E\bar X = \mu$, and $Var(\hat\mu) = \frac{\sum_{i=1}^p \sigma_i^2}{p^2}$. For any $\delta_1 > 0, \delta_2 > 0$, Chebyshev's Inequality implies that

$$P((\log p)^{\frac{\epsilon+1}{2}}|\hat\mu - \mu| > \frac{\delta_1}{2}) < \frac{4(\log p)^{\epsilon+1}}{\delta_1^2} Var(\hat\mu) = \frac{4(\log p)^{\epsilon+1}\sum_{i=1}^p \sigma_i^2}{p^2\delta_1^2} \to 0.$$

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

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

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

Now to prove the second part, by (4), we have

$$P((\log p)^{\frac{\epsilon+1}{2}}|\hat{\tau}^2-\tau^2|>\delta_1) \le P((\log p)^{\frac{\epsilon+1}{2}}|\frac{\sum(X_i^2-\sigma_i^2)}{p}-\hat{\mu}^2-\tau^2|>\delta_1)$$

$$\le \quad P((\log p)^{\frac{\epsilon+1}{2}}|\frac{\sum(X_i^2-\sigma_i^2)}{p}-\mu^2-\tau^2|>\frac{\delta_1}{2}) + P((\log p)^{\frac{\epsilon+1}{2}}|\hat{\mu}^2-\mu^2|>\frac{\delta_1}{2}).$$

Note that the second term converges to zero, and we only need to deal with the first term. Let $f(X)=(\log p)^{\frac{\epsilon+1}{2}}(\frac{\sum(X_i^2-\sigma_i^2-\mu^2-\tau^2)}{p})$. Then $Ef(X)=0$ and $Varf(X)=\frac{(\log p)^{\epsilon+1}}{p^2}\sum Var(X_i^2)$ by independence of X_i's. Direct calculation shows that

$$Var(X_i^2)=2(\sigma_i^2+\tau^2)(\tau^2+\sigma_i^2+2\mu^2), \quad \text{and consequently,}$$

$$Varf(X)=(\frac{2\sum\sigma_i^4}{p^2}+\frac{4(\tau^2+\mu^2)\sum\sigma_i^2}{p^2}+\frac{\tau^2(\tau^2+2\mu^2)}{p^1})(\log p)^{\epsilon+1}=o(1).$$

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

Now we are ready to prove Theorem 3.1. Let $X_{(i)}$ be the order statistics of X_i in magnitude so that $|X_{(1)}|\le|X_{(2)}|\le\cdots\le|X_{(p)}|$, and let $\theta_{(i)}$ and $\sigma^2_{(i)}$ are the parameters corresponding to the observation $X_{(i)}$. Write $X_{(i)}$ as $\mu+\sqrt{\sigma^2_{(i)}+\tau^2}Z_{(i)}$. Since $Z_1,Z_2,\cdots,Z_p$ are i.i.d. standard normal random variables, $\frac{\max|Z_{(i)}|}{\sqrt{2\log p}}$ converges to some random variable in distribution. (See example 9.5.3 on Page 259 of Woodroofe (1975).) Consequently, for any $\epsilon>0$,

$$\frac{|Z_{(p)}|}{(\sqrt{2\log p})^{1+\epsilon}}\le\frac{\max|Z_{(i)}|}{(\sqrt{2\log p})^{1+\epsilon}}\to 0 \text{ in probability.}$$

According to the Lemma 5.1, both $(\log p)^{\frac{\epsilon+1}{2}}(\hat{\mu}-\mu)$ and $(\log p)^{\frac{\epsilon+1}{2}}(\hat{\tau}^2-\tau^2)$ converge to 0 in probability. Now for any positive number δ_1, let

$$A_p=\{(\log p)^{\frac{\epsilon+1}{2}}|\hat{\mu}-\mu|\le\delta_1,(\log p)^{\frac{\epsilon+1}{2}}|\hat{\tau}^2-\tau^2|\le\delta_1,|\frac{Z_{(p)}}{(\sqrt{2\log p})^{1+\epsilon}}|\le\delta_1\}.$$

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

According to the construction of CI_i^{EB} in (5), $P(\theta_i\notin CI_i|X)=P(|\theta_i-(\hat{M}_iX_i+(1-\hat{M}_i)\hat{\mu})|>z\sqrt{\hat{M}_i}\sigma_i|X)$. Since $\theta_i|X_i\sim N(M_iX_i,M_i\sigma_i^2)$, we can write θ_i as $M_iX_i+\sqrt{M_i}\sigma_iZ$ where Z is a standard normal random variable independent of X_i. Let $g_1=\frac{(\hat{M}_i-M_i)(X_i-\mu)}{\sqrt{M_i}\sigma_i}$ and $g_2=\frac{(1-\hat{M}_i)(\hat{\mu}-\mu)}{\sqrt{M_i}\sigma_i}$. The above probability $P(\theta_i\notin CI_i^{EB}|X)$ can be written as $P(|Z-(g_1+g_2)|>z\sqrt{\frac{\hat{M}_i}{M_i}}$ 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 $\hat{\tau}^2+\sigma_i^2\ge 2\hat{\tau}\sigma_i$, we have

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

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

$$\frac{\sigma_i\sqrt{\tau^2+\sigma_i^2}}{(\hat{\tau}^2+\sigma_i^2)\tau}=\frac{\sigma_i}{\tau\sqrt{\hat{\tau}^2+\sigma_i^2}}\sqrt{\frac{\tau^2+\sigma_i^2}{\hat{\tau}^2+\sigma_i^2}}\leq\frac{1}{\tau}\max(1,\frac{\tau^2}{\hat{\tau}^2})\leq C_2,$$

$g_2\leq C_2\delta_1$. In the above calculations, C_1 and C_2 can be found because of A_p condition and depend on τ^2 only and not on σ_i^2. Furthermore

$$|\frac{\sqrt{\hat{M}_i}}{\sqrt{M_i}}-1|\leq|\frac{\hat{M}_i}{M_i}-1|=|\frac{\sigma_i^2(\hat{\tau}^2-\tau^2)}{\tau^2(\hat{\tau}^2+\sigma_i^2)}|\leq|\frac{\hat{\tau}^2-\tau^2}{\tau^2}|\leq\frac{\delta_1}{\tau^2}.$$

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

$$P(\theta_i\notin CI_i|X)\leq P(||Z|-(C_1+C_2)\delta_1|>z(1-\frac{\delta_1}{\tau^2}))\to q,\quad\text{ as }\delta_1\to 0.$$

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

$$\alpha(p,X)=\max_{1\leq i\leq p}P(\theta_i\notin CI_i|X)<q+\epsilon\text{ when }A_p\text{ holds.}$$

As a result $P(\alpha(p,X)-q>\epsilon)\leq P(A_p^c)\to 0$ as p goes to infinity. Now Theorem 2.3 concludes the theorem.

Proof of Theorem 3.2: It suffices to show that condition (3) holds. According to Lemma 5.1, for any $\delta_1>0,\delta_2>0$, $\lim P(A_p)=1$ where $A_p=\{|\hat{\mu}-\mu|\leq\delta_1,|\hat{\tau}^2-\tau^2|\leq\delta_2\}$. 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 due to the fact that the $Z|X$ is $N(0,1)$ and it has $N(0,1)$ unconditionally as well. Consequently, Z is independent of X.

$$P(\theta_i\notin CI_i|X)=P(|Z-\frac{(\hat{M}_i-M_i)(X_i-\mu)+(1-\hat{M}_i)(\hat{\mu}-\mu)}{\sqrt{M_i}\sigma_i}|>z\sqrt{\frac{\hat{M}_i}{M_i}}),$$

where in the above probability, Z is the only random variable and X_i and $\hat{M}_i$ are viewed as constants. Later on, 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_i s are viewed as non-random from now on until (A2). Hence

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

Under the assumption that A_p holds, similarly as in the proof of Theorem 3.1,

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

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

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

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

$$P(\theta_i \notin CI_i|X) \geq P(|Z| \geq z(1 + \frac{\delta_1}{\tau^2}), \quad \text{implying that}$$

$$|P(\theta_i \notin CI_i|X) - q| \leq \max(|q - P(|Z| \geq (1 + \frac{\delta_1}{\tau^2})z)|, |q - P(|Z - C_1\delta_2 Z_i - C_2\delta_1| > z(1 - \frac{\delta_1}{\tau^2}))|).$$

Summing over i on both sides, we have

$$\frac{1}{p}\sum |P(\theta_i \notin CI_i|X) - q| \leq max(A, B), \tag{A2}$$

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

$$B = \frac{1}{p}\sum_i |q - P(|Z - C_1\delta_2 Z_i - C_2\delta_1| > z(1 - \frac{\delta_1}{\tau^2}))|. \tag{A3}$$

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

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

where the expectation is with respect to Z_i. Dominated convergence theorem then implies that the expectation converges to $|q - P(|Z| > z)| = 0$ as δ_1 and δ_2 approach zero. This concludes that (A2) converges to zero as $p \to \infty$. Condition (3) is established and so is the theorem.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**(1), 289–300.

Benjamini, Y. and Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Amer. Statist. Assoc.*, **100**(469), 71–93. With comments and a rejoinder by the authors.

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 Biology*, **6**(2), R16.1–16.

Cui, X. and Churchill, G. (2003). Statistical tests for differential expression in cdna microarray experiments. *Genome Biology*, **4**(4), 210.

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.

Hwang, J. T. and Liu, P. (2009). Optimal tests shrinkage both means and variances applicable to microarray data analysis.

Hwang, J. T., Qiu, J., and Zhao, Z. (2009). Empirical Bayes confidence intervals shrinking both means and variances. *Journal of the Royal Statistical Society. Series B (Methodological)*, **71**(1), 265–285.

Kendziorski, C., Newton, M., Lan, H., and Gould, M. (2003). On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Statistics in Medicine*, **22**, 3899–3914.

Lönnstedt, I., Rimini, R., and Nilsson, P. (2005). Empirical bayes microarray anova and grouping cell lines by equal expression levels. *Statistical Applications in Genetics and Molecular Biology*, **4**(1).

Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *J. Amer. Statist. Assoc.*, **78**(381), 47–65. With discussion.

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. *Journal of Computational Biology*, pages 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**(3), 767–776.

Tai, Y. C. and Speed, T. P. (2006). A multivariate empirical Bayes statistic for replicated microarray time course data. *Ann. Statist.*, **34**(5), 2387–2412.

Woodroofe, M. (1975). *Probability with applications*. New York, McGraw-Hill.

Zhao, Z. (2010). Double shrinkage empirical Bayesian estimation for unknown and unequal variances. *submitted*.

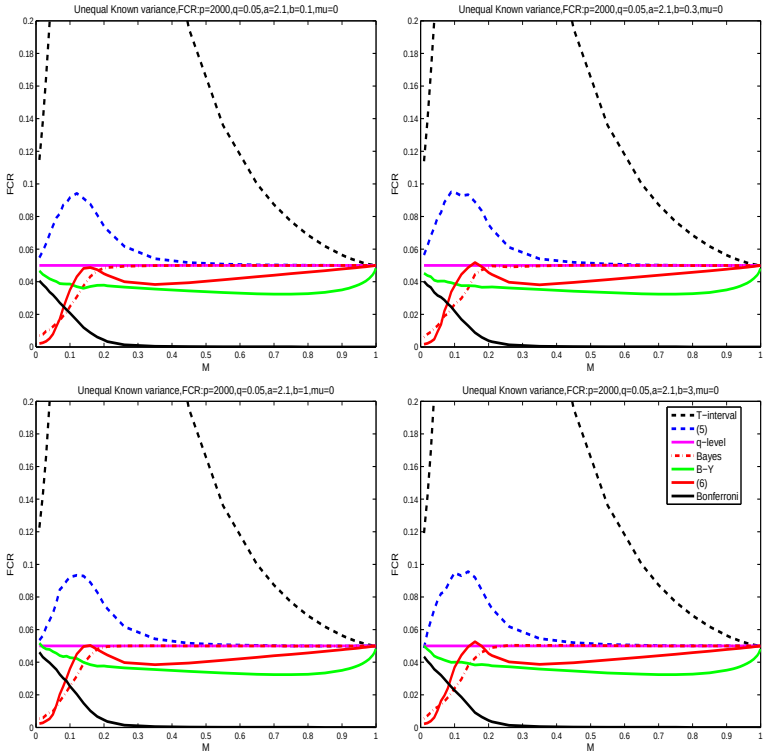

Fig. 2. The Bayes FCRs of different interval constructions are plotted against $M = \frac{\tau^2}{1+\tau^2}$ for $p = 2000$ and $q = 0.05$ under the Normal-Normal Model when assuming the unequal but known variances. The variances are sampled independently from the inverse gamma random variable for various combinations of parameters a and b. The a is chosen to be 2.1 in this figure and b varies among $0.1, 0.3, 1,$ and 3, corresponding to the four graphs above. The parameters are selected using Benjamini and Hochberg (1995)'s FDR procedure at 5% level. The naive z-interval and (5) fail to control the FCR at the q-level; B-Y's procedure, Bonferroni correction, and our empirical Bayes confidence intervals (6) control Bayes FCR at q-level for all M, hence any $N(\mu, \tau^2)$ prior where $\tau > 0$. The same conclusion is arrived for $a = 2.5$ and 4, unreported here.

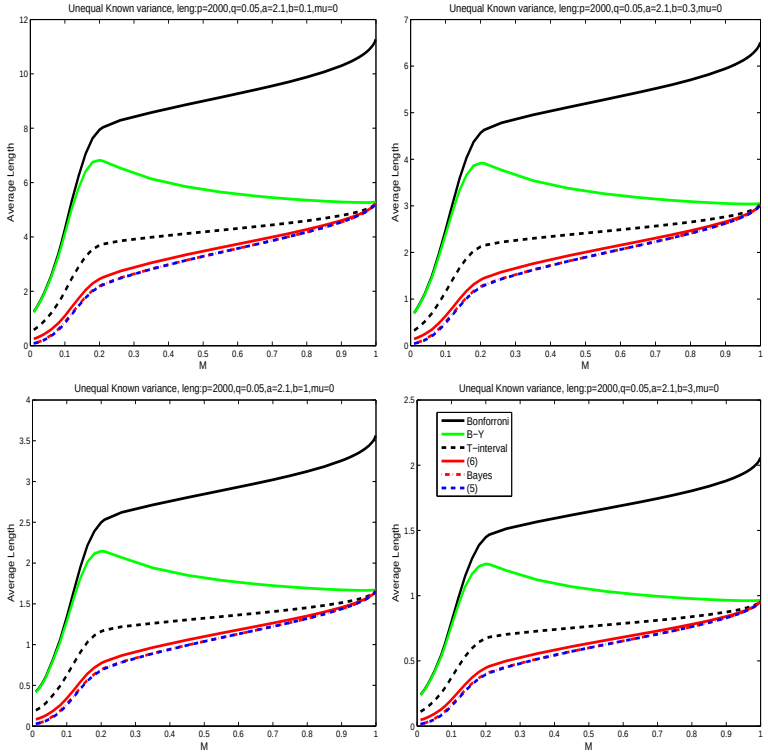

Fig. 3. The average half length of intervals that are constructed are plotted for various combinations of parameters a and b which are given in Figure 2. The average length of Bonferroni's and B-Y's intervals are uniformly larger than (6), and in the most extreme case, three times as large.

Table 1. We calculate the actual FCR and average half lengths for various intervals. For B-Y's procedure, the df is taken to be 2 or the integer part of Satterthwaite's df. BF is the Bonferroni's intervals with df=2. The recommended intervals (7), which modifies CI_{SS} of Hwang, Qiu, and Zhao (2009), has actual FCR nearly at the 5% nominal level and is short.

	(7)	B-Y (df=2)	B-Y (df=int)	BF	CI_{SS}
FCR(%)	5.1	1.1	7.17	0.16	9.87
Ave. Length	0.5449	1.1315	0.6541	4.1439	0.4093