Electronic Journal of Statistics
Vol. 0 (0000)
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

Decision Approach and Empirical Bayes FCR-Controlling Interval for Mixed Prior Model

Zhigen Zhao

e-mail: zhaozhg@temple.edu

Abstract: In this paper, I apply the decision theory and empirical Bayesian approach to construct confidence intervals for selected populations when true parameters follow a mixture prior distribution. A loss function with two tuning parameters k_1 and k_2 is coined to address the mixture prior. One specific choice of k_2 can lead to the procedure in Qiu and Hwang (2007); the other choice of k_2 provides an interval construction which controls the Bayes FCR. Both the analytical and extensive numerical simulation studies demonstrate that the new empirical Bayesian FCR controlling approach enjoys great length reduction. At the end, I apply different methods to a microarray data set. It turns out that the average length of the new approach is only 57% of that of Qiu and Hwang's procedure which controls the simultaneous non-coverage probability and 66% of that of Benjamini and Yekutieli (2005)'s procedure which controls the frequentist's FCR.

AMS 2000 subject classifications: Decision Bayes, Loss Function, Simultaneous Intervals..

1. Introduction

Simultaneous interval estimation for a large number of selected parameters is challenging especially when the number of observations for each parameter is very small. The difficulties are due to the selection bias (see [11] and [8]) and the multiplicity. The traditional approach, which treats all the parameters as fixed, seems to have little power when the dimension tends to be very large, for instance, several thousands in microarray. However, the empirical Bayesian approach is known to be able to *borrow strength* across the populations. Thus, it is very likely that this method will provide us with some satisfactory procedures.

In the past, researchers attempted to provide point estimators of the parameters for selected populations (see, for example, [4] and [8]). However, only a few interval procedures have been constructed for those selected populations. One exciting work is [11], which offers a way to construct intervals that controls the *empirical* Bayesian simultaneous coverage coefficient for selected populations. Other than the normal-normal model, they treated the so-called *normal-mixture* model where the true parameters are i.i.d. samples from a mixture of a normal random variable and a single point *zero*. Because they control the *empirical*

*This is an original survey paper

Bayesian simultaneous coverage probability, which is less stringent than the frequentist simultaneous coverage probability, their intervals are much shorter than the intervals constructed using Bonferroni's method.

An alternative criterion has been proposed in the paper [2]. They adapted the concept of FDR from multiple testing and coined a concept False Coverage Rate (FCR) for simultaneous intervals. This criterion is much less conservative than the simultaneous coverage coefficient. They constructed confidence intervals for multiple selected parameters which can control the FCR at a specified q-level, typically $q = 5\%$. They centered their intervals upon the estimator X_i's which are biased for selected populations and addressed the multiplicity by lengthening the intervals. Consequently, their intervals are extremely long and can be substantially improved compared to the intervals we shall propose.

Later, [12] introduced the Bayes FCR, *empirical* Bayesian FCR. They further connected Bayes confidence intervals which aim to control Bayesian non-coverage coefficients with the Bayesian FCR controlling intervals. They applied this general theorem to the normal-normal model where observations X_i's follow a normal distribution with means θ_i's and the parameters θ_i's follow a normal prior. They used the empirical Bayesian approach to derive explicit intervals which can control the empirical Bayes FCR. Their construction reduced the average length of [2]'s procedure dramatically. Incorporating the *empirical* Bayesian method, instead of lengthening the intervals, [12] addressed the multiplicity by reducing the bias of the center of their intervals.

Here, we use the decision approach and empirical Bayes to construct intervals for selected populations under the *normal-mixture* model as illustrated in [11]. Application of decision approach to interval/set estimation has a long history (see [5], [3], and [6]). Recently, [9] have constructed the double shrinkage *empirical* Bayesian confidence interval for a one dimensional parameter when assuming the variances to be unequal and unknown. However, the loss function they have used needs adjustment for the *normal- mixture* model we consider here (detailed argument is in section 2.2). Thus a new loss function with two tuning parameters k_1 and k_2 is proposed. One specific choice of k_2 results in [11]'s procedure which offers the controlling of *empirical* Bayesian simultaneous coverage probability. The other choice of k_2 provides us with a way to construct the empirical Bayesian FCR-controlling intervals based on the normal-mixture model.

In section 2, we introduce the model setting and the decision Bayes rule based on our new loss function. In section 3, we will connect the decision Bayesian rule with [11]'s procedure first and then derive a procedure which can control the Bayes FCR. In section 4, empirical Bayesian approach is constructed and evaluated both numerically and analytically. In section 5, we apply the confidence intervals constructed in section 4 to a real microarray data set and compare it with those of [2] and [11]. It turns out that our procedure out-performs theirs. The average length of our interval is only 57% of that of [11]'s procedure which controls the simultaneous coverage probability and 66% of that of [2]'s procedure which controls the frequentist's FCR. Obviously, one major reason that the proposed procedure has sharper intervals is because we take a less stringent criterion: controlling of the *empirical* Bayes FCR. However, this seems a more realistic criterion.

2. Normal-Mixture Model for the Means

2.1. Model Assumption

In history, the so-called *normal-normal* model where $X_i \sim N(\theta_i, \sigma^2)$, $\theta_i \sim N(0, \tau^2)$ has been very widely studied. [12] has constructed the *empirical* Bayesian confidence interval which controlls the *empirical* Bayesian FCR. However, in [11], they argued that it is more appropriate to consider the *normal-mixture* model where $X_i|\theta_i \sim N(\theta_i, \sigma^2)$, and

$$\pi(\theta_i) \begin{cases} = 0 & \text{with probability } \pi_0, \\ \sim N(0, \tau^2) & \text{with probability } \pi_1 = 1 - \pi_0. \end{cases} \tag{1}$$

We use an indicator function I_i to describe whether θ_i is 0, i.e. $I_i = 0$ if $\theta_i = 0$ and $I_i = 1$ if $\theta_i \sim N(0, \tau^2)$. Initially, we assume that hyper parameters τ^2 and π_0 are known and derive the corresponding decision Bayesian procedure. In section 4, we estimate them through data by using consistent estimators and derive an empirical Bayesian procedure.

2.2. Bayes Interval

Historically, there have been many attempts to apply the decision Bayes approach to construct confidence sets/intervals. [5] considered a linear loss function for confidence set CI of the parameter θ as $L(\theta, CI) = kVolume(CI) - I_{CI}(\theta)$. Also [3] uses the same loss where the tuning parameter k was determined so that the usual $1-\alpha$ confidence set is minimax. [6] used $L_i(\theta_i, CI_i) = kLen(CI_i) - I_{CI_i}(\theta_i)$ as the loss function for the interval estimator CI_i of the parameter θ_i. [9] modified the loss function above as $L(\theta_i, CI_i) = \frac{k}{\sigma_i} Len(CI_i) - I_{CI_i}(\theta_i)$ and constructed the confidence interval that shrinks both the estimated means and variances σ_i^2. However, all these loss functions are not appropriate for the normal-mixture model (1). In fact, for any given confidence interval, one can construct a new interval, which is the union of the existing procedure and *zero*. This new approach boosts the coverage probability while causing no change in the length. Consequently, the conditional expected loss of the new construction is always less than or equal to that of the original approach. As a result, the decision Bayes suggests that *zero* should always be included. However, such intervals have no power if applied to conduct tests for $\theta_i = 0$ since it will always accept the null hypothesis.

In order to avoid this phenomenon, we add extra terms which influence the loss function only when the point *zero* is included and thus define the loss function as,

$$L(\theta_i, CI_i) = k_1 Len(CI_i) I_i - I_{CI_i}(\theta_i) I_i + I_{CI_i}(0)(k_2 - (1 - I_i)), 0 \le k_2 \le 1. \tag{2}$$

The first two terms balance the length and the true coverage. The tuning parameter k_1 will be determined later in this section. The last two terms affect the loss function only when the corresponding interval does include *zero*. Under this case, if θ_i is indeed *zero*, then $k_2 - (1 - I_i) = k_2 - 1 \leq 0$, and including *zero* is beneficial as it should be. On the other hand, if θ_i is not *zero*, then $k_2 - I_i = k_2$ is positive and becomes a penalty term. Unlike the loss functions we have mentioned earlier in this section, this new loss forbid us to include *zero* in every interval. It turns out that the tuning parameter k_2 guides us to decide when *zero* should be included.

Furthermore, the flexibility of choosing k_2 offers us interval constructions under different settings controlling different criterion. For example, when assuming the normal-normal model, the loss function (2) reduces to [6]'s if we set $k_2 = 0$. In section 3, we apply two different choices of k_2, one of which will reproduce [11]'s procedure which controls the *empirical* Bayesian coverage probability, while the other will provide a construction that can control the Bayesian FCR at q-level.

Now, we have all the pieces to construct the decision Bayes rule, i.e. we want to construct a Bayes interval CI_i^B such that it minimizes $E(L(\theta_i, CI_i|X))$ for any observation X when assuming the normal-mixture model (1) and the loss function (2).

Theorem 2.1 *Let* $\pi_i^0(X) = P(\theta_i = 0|X) = P(I_i = 0|X)$ *and* $\pi_i^1(X) = 1 - \pi_i^0(X)$. *Then*

$$EL(\theta_i, CI_i|X) = \pi_i^1(X) \int_{CI_i} (k_1 - \pi(\theta_i|X, I_i = 1)) d\theta_i + I_{CI_i}(0|X)(k_2 - \pi_i^0(X)). \tag{3}$$

The Bayes interval is

$$CI_i = \begin{cases} \{\theta_i : k_1 < \pi(\theta_i|X_i, I_i = 1)\} \setminus \{0\} & \text{if } k_2 > \pi_i^0(X), \\ \{\theta_i : k_1 < \pi(\theta_i|X_i, I_i = 1)\} \cup \{0\} & \text{if } k_2 \leq \pi_i^0(X). \end{cases} \tag{4}$$

Intuitively, for any given observation X_i, if the conditional probability $\pi_i^0(X)$ is small, it is unlikely that $\theta_i = 0$ and *zero* should be excluded. On the other hand, larger $\pi_i^0(X)$ indicates that *zero* should be included. Theorem 2.1 shows that the decision Bayes interval uses k_2 as the threshold.

Under model (1), $\pi(\theta_i|X, I_i = 1) \sim N(MX_i, M\sigma^2)$ where $M = \frac{\tau^2}{\tau^2+\sigma^2}$, therefore

$$\begin{aligned} & \{\theta_i : k_1 < \pi(\theta_i|X_i, I_i = 1)\} \\ = \quad & \{\theta_i : (\theta_i - MX_i)^2 < -M\sigma^2(2 \log k_1 \sqrt{2\pi} + \log M\sigma^2)\}. \end{aligned}$$

As in the Section 3 of [9], one wants to obtain a traditional normal interval when the non-informative prior is applied, i.e., if setting $\tau \to \infty$, $M \to 1$, one wants the corresponding interval $\{\theta_i : \frac{(\theta_i - X_i)^2}{\sigma^2} < -(2 \log k_1 \sqrt{2\pi} + \log \sigma^2)\}$ to coincide with normal interval $(X_i - z_{q/2}\sigma, X_i + z_{q/2}\sigma)$ where $z_{q/2}$ is the critical value such that $P(|Z| > z_{q/2}) = q$ when Z is a standard normal random variable. Therefore, the constant k_1 should be chosen such that $z_{q/2}^2 = -(2\log k_1\sqrt{2\pi}+\log\sigma^2)$. Plug this constant k_1 back to Bayes interval (4). Then the decision Bayes interval becomes

$$CI_i^B = \begin{cases} \{\theta_i : (\theta_i - MX_i)^2 < M\sigma^2(z_{q/2}^2 - \log M)\} \setminus \{0\} & \text{if } k_2 > \pi_i^0(X), \\ \{\theta_i : (\theta_i - MX_i)^2 < M\sigma^2(z_{q/2}^2 - \log M)\} \cup \{0\} & \text{if } k_2 \le \pi_i^0(X). \end{cases} \tag{5}$$

Unlike the interval $MX_i \pm \sqrt{M}\sigma z_{q/2}$, which is directly derived from the posterior distribution, (5) has an extra positive term $M\sigma^2(-\log M)$ which is necessary to reduce the non-coverage probability or Bayesian FCR when the hyper parameters are estimated through the data in section 4. In the next section, we will choose the value of the parameter k_2 under two different problem settings and derive the decision Bayes interval accordingly.

3. Choose k_2

3.1. Qiu and Hwang (2007)

[11] constructed the interval for K parameters $\theta_{(p-K+1)}, \cdots, \theta_{(p)}$ under the model (1) where the observations $\theta_{(j)}$ is the parameter corresponding to $X_{(j)}$ and $X_{(j)}$'s are permutation of $X_1, \cdots, X_p$, so that

$$|X_{(1)}| \le |X_{(2)}| \le \cdots \le |X_{(p)}|.$$

The parameter $\theta_{(j)}$'s are called the parameters of selected population. In particular, $\theta_{(p)}$ is the parameter of the population which happens to have produced the largest $|X_i|$ or the population selected to have the largest X_i in magnitude. Note that $|\theta_{(p)}|$ is not necessarily equal to $\max_{1\le j\le p} |\theta_{(j)}|$. We construct the interval for $\theta_{(j)}$ where $p-K+1 \le j \le p$ as

$$\begin{aligned} & CI_{(j)}^B \qquad (6) \\ = & \begin{cases} \{\theta_{(j)} : (\theta_{(j)} - MX_{(j)})^2 < M\sigma^2(z_{q/2K}^2 - \log M)\} \setminus \{0\} & \text{if } k_2 > \pi_{(j)}^0(X), \\ \{\theta_{(j)} : (\theta_{(j)} - MX_{(j)})^2 < M\sigma^2(z_{q/2K}^2 - \log M)\} \cup \{0\} & \text{if } k_2 \le \pi_{(j)}^0(X). \end{cases} \end{aligned}$$

When compared with (5), the major difference is that we use the critical value $z_{q/2K}$ to address the multiplicity.

Direct calculation shows that for each j,

$$P(\theta_{(j)} \notin CI_{(j)}^B | X) \le q/K + \pi_{(j)}^0(X)((\pi_{(j)}^0(X) < k_2) - q/K).$$

Consequently, the simultaneous non-coverage coefficient satisfies

$$P(\theta_{(j)} \notin CI_{(j)}, j = p-K+1, \cdots, p) \le q + E\sum_{j=p-K+1}^{p} \pi_{(j)}^0(X)(I(\pi_{(j)}^0(X) < k_2) - q/K). \tag{7}$$

If k_2 is chosen to be the maximum k such that the summation above is non-positive, i.e.

$$k_2 = \arg\max_k \{E \sum_{j=p-K+1}^{p} \pi^0_{(j)}(X)(I(\pi^0_{(j)}(X) < k) - q/K) \leq 0\}, \tag{8}$$

then the non-coverage coefficient $P(\theta_{(j)} \notin CI_{(j)}, j = p-K+1, \cdots, p)$ is controlled at the q-level. Using this choice of k_2, (6) is identical to [11]'s Bayes procedure, hence providing a surprising satisfaction of [11]. This always indicates that the loss function (2) is reasonable and useful.

3.2. Bayes FCR Controlling Interval

Constructing the confidence intervals for selected parameters, [2] initiated the concept of FCR as following: assume that $\mathcal{R}$ as the set of parameters that are selected for interval construction and $\mathcal{V}$ be the subset of selected parameters such that the corresponding intervals don't cover the true parameter. Define $FCR = E\frac{V}{R}$ where $V = \#\mathcal{V}$ and $R = \#\mathcal{R}$.

This definition of FCR is based on frequentist viewpoint. As argued in [12], it is way to conservative. They further defined the Bayesian FCR, denoted as FCR_π, by integrating the FCR with respect to the prior distribution $\pi(\theta)$. They have showed that for a known prior π, the Bayesian confidence interval controls FCR_π automatically. Later in that paper, they considered a family of prior distribution Π and introduced the *empirical* Bayesian FCR controlling intervals. In other words, a confidence interval construction controls the *empirical* Bayesian FCR at q-level if $FCR_\pi \leq q$ for any $\pi \in \Pi$. In what follows, we will prove that (5) controls the Bayesian FCR when asuming known hyper parameters and then derive the *empirical* Bayesian FCR controlling intervals.

Theorem 3.1 *Assume that $\mathcal{R}(X)$ is the index set of observations that are selected for interval estimation. $R = \#\mathcal{R}$. Define*

$$f(p, \tau^2, \pi_0, k) = E(\sum_{i \in \mathcal{R}} \frac{\pi^0_i(X)(I(\pi^0_i(X) < k) - q)}{R} I(R > 0)),$$

and $k_2 = \max_k\{k, f(p, \tau^2, \pi_0, k) \leq 0\}$. Then intervals (5) satisfies

$$FCR_\pi \leq qP(R > 0).$$

In other words, the Bayes FCR of the intervals (5) is controlled at the q level.

Now assume that the selection rule in [11] is applied in Theorem 3.1, i.e., the population with K largest X_i are selected where $K > 1$ and hence $\mathcal{R}$ is defined accordingly. If we had used the choice k_2 in (8) which is now denoted as k_2', $f(p, \tau^2, \pi_0, k_2')$ is less than or equal to zero. Consequently, the k_2 chosen according to Theorem 3.1 is larger than or equal to k_2'. Therefore, the frequency that

(5) includes *zero* is less than that of [11]. Furthermore, according to their simultaneous confidence interval construction, the half length $M\sigma^2(z_{q/2K} - \log M)$ is much larger than the half length of the Bayes FCR controlling interval (5). The discrepancy becomes large when K is large. These two facts imply that the Bayes FCR controlling interval is less conservative than [11]. However, the construction of [11] could control the *empirical* Bayesian simultaneous coverage probability, which is a stronger criteria than the empirical Bayesian FCR.

Another advantage of this theorem is that it holds for any selection rule, including pre-determined and data-driven selection rules. For example, when observations are selected according to [1], which controls the False Discovery Rate at q-level, and k_2 is simulated accordingly, the above theorem guarantees that (5) still controls the Bayes FCR at the q-level.

A disadvantage of this approach is that the choice of k_2 depends on the expectation, which prevents us from finding k_2 explicitly. However, k_2 can be easily determined by simulation once the hyper-parameters are known or estimated by the data as shown below.

4. Empirical Bayes Approach

In this section, we estimate unknown hyper-parameters through the data and obtain an *empirical Bayes* confidence interval. Our goal is to construct the confidence intervals for selected parameters such that the Bayes FCR can always be controlled for a class of prior distributions which are determined by the hyper-parameters π_0 and τ^2. This approach is named *empirical Bayes FCR controlling intervals*, according to [12].

Recall the model 1. Then $EX_i^2 = \sigma^2 + \pi_1\tau^2$, and $EX_i^4 = 3(\sigma^4 + 2\pi_1\sigma^2\tau^2 + \pi_1\tau^4)$. By using the method of moments, one could get reliable estimators of π_0 and τ^2 when p is sufficiently large,

$$\hat{\pi}_1 = \frac{(m_2 - \sigma^2)^2}{m_4/3 + \sigma^4 - 2\sigma^2 m_2}, \hat{\tau}^2 = \frac{(m_2 - \sigma^2)}{\hat{\pi}_1}. \tag{9}$$

Plug these two estimators back to the function of f as in Theorem 3.1 and obtain the value of k_2, denoted by $\hat{k}_2$. Assume that $\hat{M}$ and $\hat{\pi}_i^0(X)$ are the estimators of M and $\pi_i^0(X)$ when π_0 and τ^2 are replaced by (9). Then we can construct the empirical Bayes interval as,

$$\begin{aligned} & CI_i^{EB} \qquad\qquad (10) \\ = & \begin{cases} \{\theta_i : (\theta_i - \hat{M}X_i)^2 < \hat{M}\sigma^2(z_{q/2}^2 - \log \hat{M})\} \setminus \{0\} & \text{if } \hat{k}_2 > \hat{\pi}_i^0(X), \\ \{\theta_i : (\theta_i - \hat{M}X_i)^2 < \hat{M}\sigma^2(z_{q/2}^2 - \log \hat{M})\} \cup \{0\} & \text{if } \hat{k}_2 \le \hat{\pi}_i^0(X). \end{cases} \end{aligned}$$

The following theorem describes the asymptotic property of the construction.

Theorem 4.1 *Assume that $0 < \pi_0 < 1$, $\tau^2 > 0$. For any $\epsilon > 0$, if there always exists $\delta, N > 0$ such that for all $p > N$,*

$$|f(p, \tau'^2, \pi_0', k') - f(p, \tau^2, \pi_0, k)| < \epsilon. \tag{11}$$

when $(\tau'^2 - \tau^2)^2 + (\pi'_0 - \pi_0)^2 + (k' - k)^2 < \delta$*, then under the model (1), the empirical Bayes interval (10) satisfies*

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

Proposition 4.1 *If we select the first* R *parameters with* $R \to \infty$ *when* $p \to \infty$*, then* f *satisfies the condition in Theorem 4.1.*

This proposition implies that when all observations are selected for interval estimation, (10) can control the empirical Bayes FCR asymptotically.

However, like all other existing constructions such as [3], [11], and [9], the interval (10) cannot provide a satisfactory answer automatically for the finite sample case.

In Figure 1, we have plotted a figure of Bayes FCR of the empirical Bayes interval versus the procedure of [2] under different settings of hyper-parameter (π_0, τ^2) when $p = 1000$ and only the top 100 observations are selected for interval estimation. [2]'s procedure can always control the FCR at the 5% level; however, their procedures are conservative in the sense that the Bayes FCR is very low when M is close to 1. In addition, the average length of their construction is large. The green line, corresponding to the construction (10), performs well when τ^2 is relatively large; however some modifications are required when τ^2 is small.

[11] has argued that π_0 is nearly unidentifiable when τ is small. This will cause the estimator (9) to be very inaccurate. If the estimator of p and τ^2 satisfy the condition that $\hat{p}\hat{\tau}^2 < \min(\sqrt{720/p}, 0.6)$, they use the Bonferroni's correction $(X_i - z_{q/2p}\sigma, X_i + z_{q/2p}\sigma)$ as their interval which controls the frequentist coverage probability. For the other cases, they used their own *empirical* Bayesian intervals. It is necessary to mix the procedure (10) with the frequentist FCR controlling intervals $(X_i - z_{Rq/2p}\sigma, X_i + z_{Rq/2p}\sigma)$, which is constructed by [2]. In what follows, we have an analytic argument that helps us to find the proposed threshold determining the mixture intervals.

Recall that $EX_i^2 = \sigma^2 + \pi_1\tau^2$ and $EX_i^4 = 3(\sigma^4 + 2\pi_1\sigma^2\tau^2 + \pi_1\tau^4)$, therefore $\tau^2 + 2\sigma^2 = \frac{EX_i^4/3 - \sigma^4}{EX_i^2 - \sigma^2}$. Use $m_2 = \sum X_i^2/p$ and $m_4 = \sum X_i^4/p$ to denote the second and fourth moments, then

$$\hat{\tau}^2 + 2\sigma^2 = \frac{m_4/3 - \sigma^4}{m_2 - \sigma^2}.$$

Since the left hand side of the above formula is always greater than or equal to $2\sigma^2$, τ^2 is not estimable when the right hand side is less than $2\sigma^2$. Therefore, we can carefully choose a proper τ_0^2, such that the probability of the right hand side is smaller than $2\sigma^2$, i.e. the probability that π_0 and τ^2 are not estimable, is controlled at the level of q. Therefore, we set the threshold value τ_0^2 to satisfy $P_{\tau^2=\tau_0^2}(\frac{m_4/3-\sigma^4}{m_2-\sigma^2} \le 2\sigma^2) \le q$.

Now consider the special case when $\pi_1 = 1$ and calculate τ_0^2. Use m'_4 and m'_2 to denote the second and fourth moments of the standard normal distribution when there are p observations. Then $m_4 = (\tau^2+\sigma^2)^2 m_4'$ and $m_2 = (\tau^2+\sigma^2)m_2'$. We can use simulation to find τ_0^2 such that

$$P_{\tau^2=\tau_0^2}((\tau^2+\sigma^2)^2\frac{m_4'}{3} - 2\sigma^2(\tau^2+\sigma^2)m_2' + \sigma^4 < 0) \le q.$$

Based on the cutoff, the final empirical Bayes FCR controlling interval with mixture is defined as

$$CI_i^{Final} = \begin{cases} X_i \pm z_{Rq/(2p)}\sigma & \text{if } m_2 - \sigma^2 < \tau_0^2, \\ CI_i^{EB}, & \text{if } m_2 - \sigma^2 > \tau_0^2. \end{cases} \tag{12}$$

In Figure 1, the red solid line corresponds to the above empirical Bayes intervals. They perform the same as [2] when τ^2 is small because of the mixed procedure. The portion of the mixture increases when π_0 increases. However, (12) performs better than theirs when τ^2 is larger. The discrepancy is significant when $M \to 1$.

We have also plotted the simulated average length in figure 2 that corresponds to the same model settings in figure 1. The average length of (12) is for all M less than or equal to the average length of [2]'s procedure. The ratio of these two lengths can be as small as 56%.

In figures 3 and 4, we repeat the simulation setting but change the selection rule to [1]'s procedure which aims at finding significant observations while controlling the False Discovery Rate at a 5%-level. The intervals (12) can control the empirical Bayesian FCR at the 5%-level based on this data-driven selection. Compared with [2]'s procedure, the improvement of the average length is even more significant than that corresponding to the fixed selection rule. The ratio can be as small as 43%.

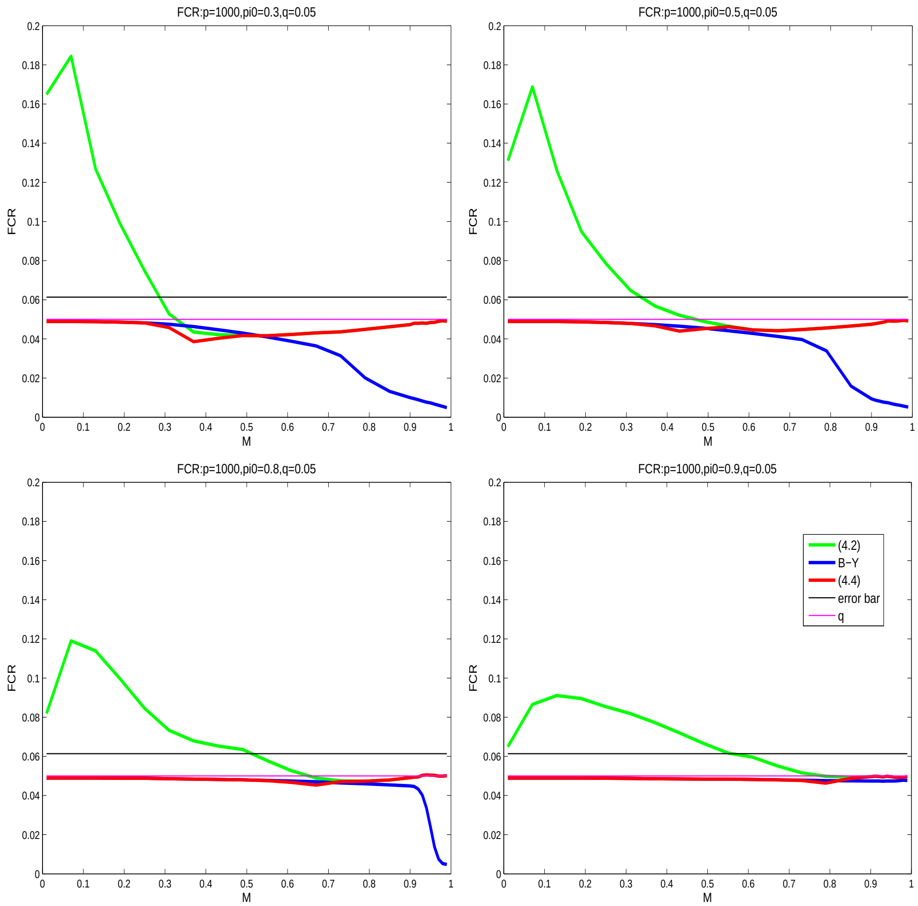

Fig 1. *These figures are the simulated Bayes FCR under different model settings against* $M = \frac{\tau^2}{1+\tau^2}$. *The dimension is set to be 1000, and top 100 observations after ordering all* X_i*'s according to their magnitude are selected for confidence interval construction. The hyper parameter* π_0 *varies among* $0.3, 0.5, 0.8$ *and* 0.9. *The Bayes FCR level that we aim at is* 5%. *When* τ^2 *is small, (10) doesn't control the Bayes FCR at 5%. However, the mixed procedure (12) does control the Bayes FCR for any hyper parameters. The portion of the mixture increases as* π_0 *increases.*

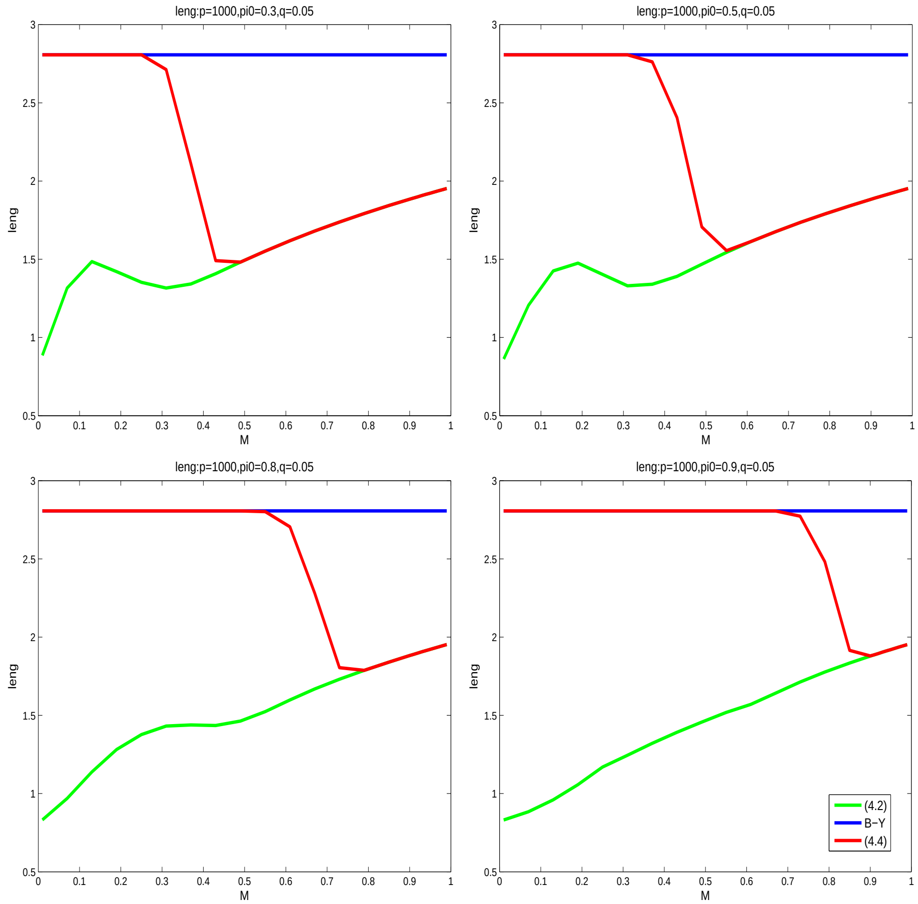

Fig 2. *These figures are the simulated average length of different approaches under the same model setting as figure 1. The average length of our procedure is less than or equal to [2]'s procedure. In some extreme cases, the average length of (12) is only 54% of that of [2]'s procedure.*

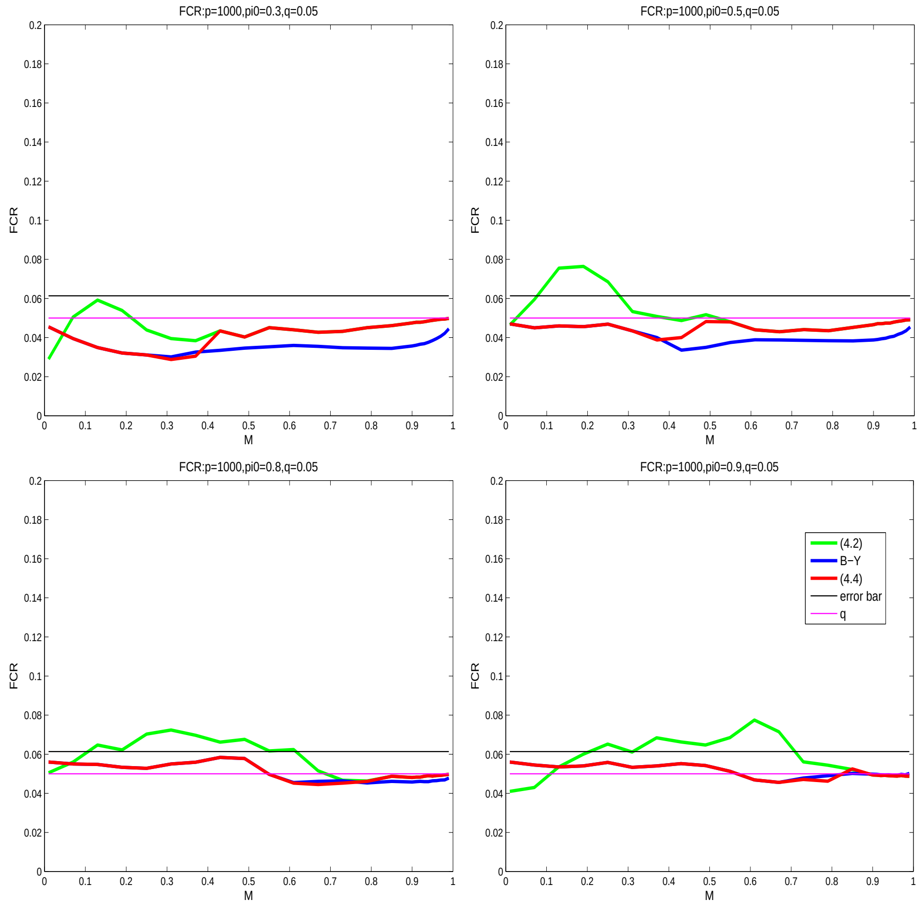

FIG 3. *These figures are the simulated Bayes FCR under different model settings against* $M = \frac{\tau^2}{1+\tau^2}$. *The dimension is set to be 1000. The selection rule is based on [1] which aims at controlling the False Discovery Rate to be less or equal than* 5%. *The hyper parameter* π_0 *varies among* $0.3, 0.5, 0.8$ *and* 0.9. *The Bayes FCR level that we aim for is* 5%, *which is represented by the magenta line. When* τ^2 *is small, (10) doesn't control the Bayes FCR. However, the Bayesian FCR of the mixed procedure (12) and [2]'s procedure are always less than or equal to the error bar, which equals to q plus the simulation error.*

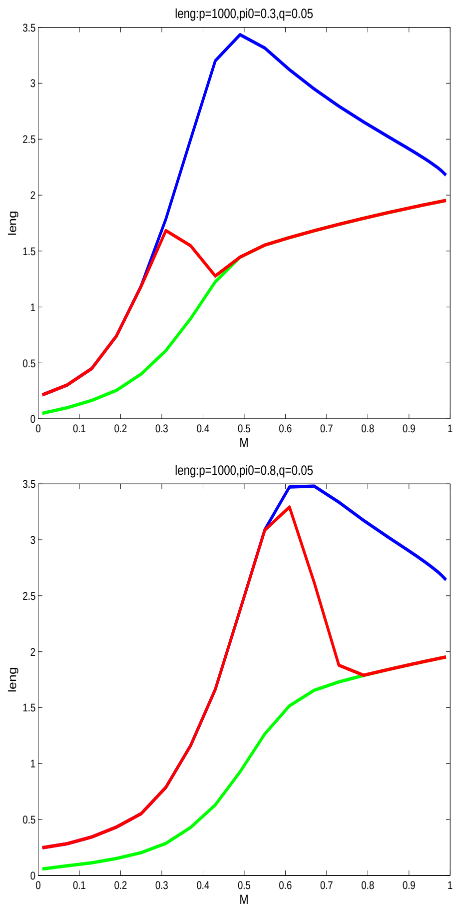

Fig 4. *These figures are the simulated average length of different approaches under the same model as figure 1. The average length of our procedure is less than [2]'s procedure. In some extreme cases, the average length of (12) is only 44% of that of [2]'s procedure.*

5. Real Data Analysis

In this section, we apply different intervals to a microarray data set, the Synteni data of [10], which was revisited by [7] and [11]. The description of the data set can be found in [10]. Figure 6 of [11] is a Q-Q plot of the ANOVA estimator X_g, which shows that the normal-mixture model (1) fits the data well.

In [7], they use simultaneous confidence intervals to detect genes with an expression level of $\Delta = 3$ or more. We will first apply the procedure of [1] to select parameters with expression levels significantly larger than or equal to $\log_2 3$, and then construct the simultaneous interval for such selected observations. B-H's procedure declares that the first 89 genes are significant.

In figure 5, we construct the confidence intervals for these 89 genes by using [11], [2], and (12). Our confidence interval (12) for $\theta_{(g)}$ is $0.93X_{(g)} \pm 0.96$. Compared with the interval $X_{(g)} \pm 1.47$ of BY's procedure, $0.93X_{(g)} \pm 1.67$ of [11], our intervals enjoy great length reduction.

6. Discussion

In this chapter, we have defined a new loss function for confidence interval construction when assuming the mixed prior model (1). We use two different ways to choose the tuning parameter in the loss function to obtain [11]'s procedure and the empirical Bayesian FCR controlling intervals. Since [11] controls the simultaneous coverage coefficient by using Bonferroni's correction, their lengths are much larger than (12) where we aim at controlling the *empirical Bayes* FCR.

However, there is still much need for further research. In model (1), we assume equal and known variance σ^2. In many applications, σ^2 are unknown and unequal. [9] proposed a double shrinkage empirical Bayesian interval for a single parameter without selection under the normal-lognormal model. Therefore, one natural extension of this work is to consider the mixture-prior model when variances are unequal and unknown. The loss function (2) provides us with a potential tool to construct corresponding intervals.

References

[1] Y. Benjamini and Y. Hochberg. 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, 1995.

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

[3] G. Casella and J. T. Hwang. Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *Journal of the American Statistical Association*, 78(383):688–698, 1983.

[4] A. Cohen and H. B. Sackrowitz. Estimating the mean of the selected population,. *In Statistical Decision Theory and Related Topics III (S. S. Gupta, J. O. Berger, eds.)*, 1:243–270, 1982.

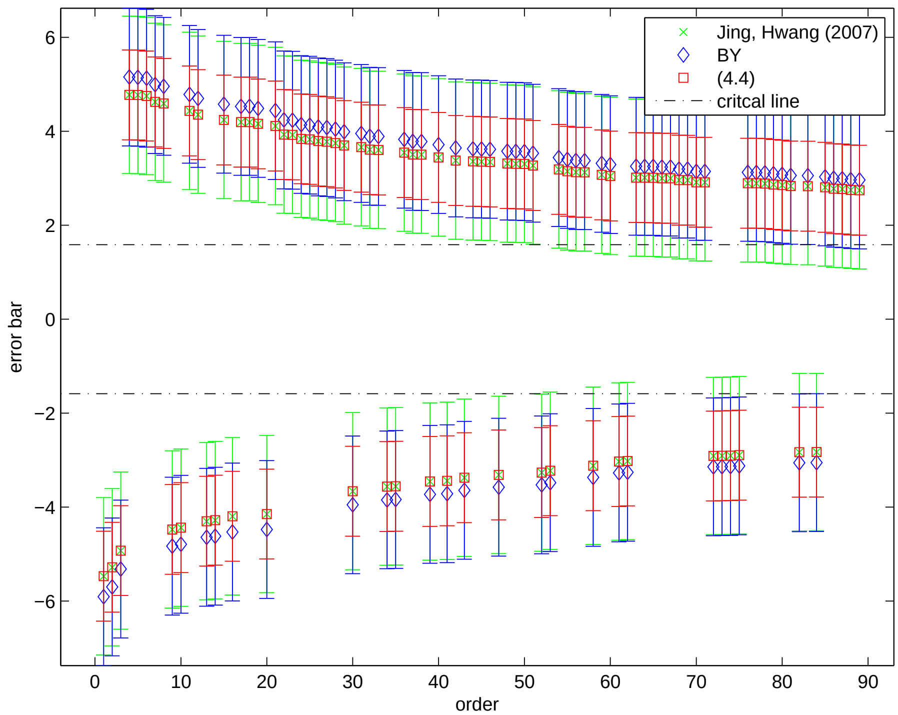

FIG 5. *Three different interval approaches, [11], [2], and (12) are applied to the Synteni data of [10]. The FDR procedure of [1], aiming at finding the genes with differentially expressed levels that are significantly larger than or equal to* $\log_2 3$ *while controlling the False Discovery Rate to be at most 5%, is applied to select genes for interval estimation. Among 1285 genes, 89 of them are declared significant and the corresponding intervals are constructed and plotted in this figure. From the figure, one can see that the center of the procedure in [11] is the same as in (12). However, since they aim to control the simultaneous coverage coefficient by using Bonferroni's correction, lengths of their intervals are much larger than that of (12). [2] centers their intervals at the biased estimator* $X_{(i)}$*'s. Thus they end up correcting the selection bias by increasing the length. As a result, their lengths are much larger than that of (12). However, the length of the procedure from [2] is slightly smaller than that of the procedure in [11].*

[5] R. E. Faith. Minimax Bayes point and set estimators of a multivariate normal mean. *Unpublished Ph.D. dissertation, Department of Statistics, University of Michigan*, 1976.
[6] K. He. Parametric empirical Bayes confidence intervals based on James-Stein estimator. *Statist. Decisions*, 10(1-2):121–132, 1992.
[7] J. C. Hsu, J. Y. Chang, and T. Wang. Simultaneous confidence intervals for differential gene expressions. *J. Statist. Plann. Inference*, 136(7):2182–2196, 2006.
[8] J. T. Hwang. Empirical Bayes estimation for the means of the selected populations. *Sankhyā Ser. A*, 55(2):285–304, 1993.
[9] J. T. Hwang, J. Qiu, and Z. Zhao. Empirical Bayes confidence intervals shrinking both means and variances. *Journal of the Royal Statistical Society, Series B*, 71(1):265–285, 2009.
[10] K. Kerr, M. Martin, and G. Churchill. Analysis of variance for gene expression microarray data. *Journal of Computational Biology*, 7(6):819–837, 2000.
[11] J. Qiu and J. T. Hwang. Sharp simultaneous intervals for the means of selected populations with application to microarray data analysis. *Biometrics*, 63(3):767–776, 2007.
[12] Z. Zhao and J. T. Hwang. Empirical Bayes FCR controlling confidence interval. Submitted, 2008.

Appendix: Sketch of Technical Arguments

Proof of Theorem 2.1.

Firstly,

$$\begin{aligned} &EL(\theta_i, CI_i|X) = k_1 Len(CI_i) P(I_i = 1|X) \\ - &\int I(\theta_i \in CI_i, I_i = 1) m(\theta_i|X) d\theta_i + I_{CI_i}(0|X)(k_2 - \pi_i^0(X)). \end{aligned} \tag{A.1}$$

The integration $\int I_{CI_i}(\theta_i, I_i = 1) m(\theta_i|X) d\theta_i$ can be written as $\int_{CI_i} m(\theta_i, I_i = 1|X) d\theta_i$ where $\pi(\theta_i, I_i = 1|X) = \pi_i^1(X)\pi(\theta_i|I_i = 1, X)$. Write $Len(CI_i)$ as $\int_{CI_i} 1 d\theta_i$. Then (A.1) equals to

$$\pi_i^1(X) \int_{CI_i} (k_1 - \pi(\theta_i|X, I_i = 1)) d\theta_i + I_{CI_i}(0|X)(k_2 - \pi_i^0(X)). \tag{A.2}$$

Now consider two intervals CI_i^1 and CI_i^2 where $CI_i^1 = \{\theta_i : k_1 < \pi(\theta_i|X, I_i = 1)\} \setminus \{0\}$ and $CI_i^2 = \{\theta_i : k_1 < \pi(\theta_i|X, I_i = 1)\} \cup \{0\}$. Then both CI_i^1 and CI_i^2 minimize the first term of the formula (A.2). Since $0 \in CI_i^2$ and $0 \notin CI_i^1$, then

$$EL(CI_i^2|X) = EL(CI_i^1|X) + (k_2 - \pi_i^0(X)).$$

Consequently, the Bayes interval includes 0 if and only if $k_2 < \pi_i^0(X)$, i.e. it is the one that is defined in (4).

Proof of Theorem 3.1.
According to [12],

$$FCR_\pi = E\frac{\sum_{i\in\mathcal{R}} P(\theta_i \notin CI_i|X)}{R} I(R>0).$$

Since

$$\begin{aligned} & P(\theta_i \notin CI_i^B|X) \\ = \;& P(\theta_i \notin CI_i^B|X, I_i=0)P(I_i=0|X) + P(\theta_i \notin CI_i^B|X, I_i=1)P(I_i=1|X) \\ = \;& \pi_i^0(X)I(\pi_i^0(X)<k_2) + (1-\pi_i^0(X))P(\theta_i \notin CI_i^B|X, I_i=1), \end{aligned} \quad (A.3)$$

and $P(\theta_i \notin CI_i^B|X, I_i=1) \le q$,

$$\begin{aligned} & FCR_\pi \\ \le \;& qE(I(R>0)) + E\frac{\sum_{i\in\mathcal{R}} \pi_i^0(X)(I(\pi_i^0(X)<k_2)-q)}{R} I(R>0) \\ = \;& qP(R>0) + f(k_2). \end{aligned}$$

The choice of k_2 ensures that $f(k_2) \le 0$. Consequently,

$$FCR_\pi \le qP(R>0).$$

Proof of theorem 4.1.
Before the proof, we will state and prove the following lemma.

Lemma A.1 *Assume that $\hat\tau^2$ and $\hat\pi_0$ are consistent estimators of τ^2 and π_0, then for any $\delta > 0$, there $\exists P_0 > 0$ such that $\forall p > P_0$,*

$$|\hat\pi_i^0 - \pi_i^0| \le \delta, \text{ for all } i = 1, 2, \cdots, p.$$

Direct calculation shows that $\pi_i^0 = \frac{\pi_0}{\pi_0 + \pi_1 \frac{\sigma}{\sqrt{\sigma^2+\tau^2}}\exp(\frac{MX_i^2}{2\sigma^2})}$ and $\hat\pi_i^0$ has the same form as π_i^0 except that π_0 and τ^2 are replaced by their estimators $\hat\pi_0$ and $\hat\tau^2$. Now, we introduce an intermediate estimator $\tilde\pi_i^0$ where π_0 is assumed known. We shall prove that the lemma holds for $\tilde\pi_i^0$ first.

Since $\hat\tau^2$ is consistent, $\hat M = \frac{\hat\tau^2}{\hat\tau^2+\sigma^2}$ is also a consistent estimator of M. Then, for $\epsilon = \frac{1}{k} < \min(\frac{1-M}{M}\delta, \frac{\pi_1\sigma}{\pi_0\sqrt{\sigma^2+\tau^2}}\delta)$, there exists N, such that $\forall p > N$, $|\hat M - M| < \epsilon M$.

Without loss of generality, assume that $M > \hat M$, i.e. $0 < M - \hat M < \epsilon M = \frac{M}{k}$. Since M is a increasing function with respect to τ^2 when σ^2 is fixed, therefore $\tau^2 > \hat\tau^2$. Direct calculation shows that

$$\tilde\pi_i^0 - \pi_i^0 = \frac{\pi_0\pi_1\sigma(\sqrt{\frac{\sigma^2+\hat\tau^2}{\sigma^2+\tau^2}}\exp(\frac{(M-\hat M)X_i^2}{2\sigma^2}) - 1)}{(\pi_0\sqrt{\sigma^2+\hat\tau^2}\exp(-\frac{\hat M X_i^2}{2\sigma^2}) + \pi_1\sigma)(\pi_0 + \pi_1\frac{\sigma}{\sqrt{\sigma^2+\tau^2}}\exp(\frac{MX_i^2}{2\sigma^2}))} \quad (A.4)$$

Since $0 < \hat{M} < M$,

$$0 < \frac{\sigma^2 + \hat{\tau}^2}{\sigma^2 + \tau^2} = \frac{1 - M}{1 - \hat{M}} < 1.$$

Consequently,

$$\sqrt{\frac{\sigma^2 + \hat{\tau}^2}{\sigma^2 + \tau^2}} > \frac{\sigma^2 + \hat{\tau}^2}{\sigma^2 + \tau^2} = \frac{1 - M}{1 - \hat{M}}.$$

Therefore, (A.4) implies that

$$\tilde{\pi}_i^0 - \pi_i^0 > \frac{\pi_0 \pi_1 \sigma(\frac{1-M}{1-\hat{M}} - 1)}{(\pi_0 \sqrt{\sigma^2 + \hat{\tau}^2} \exp(-\frac{\hat{M} X_i^2}{2\sigma^2}) + \pi_1 \sigma)(\pi_0 + \pi_1 \frac{\sigma}{\sigma^2 + \tau^2} \exp(\frac{M X_i^2}{2\sigma^2}))}$$

Since the numerator is negative and the denominator is larger than $\pi_0 \pi_1 \sigma$,

$$\tilde{\pi}_i^0 - \pi_i^0 > \frac{\pi_0 \pi_1 \sigma \frac{\hat{M} - M}{1 - \hat{M}}}{\pi_0 \pi_1 \sigma} > \frac{\hat{M} - M}{1 - M}.$$

Furthermore, $\hat{M} - M > -\epsilon M$ implies that

$$\tilde{\pi}_i^0 - \pi_i^0 > \frac{M}{1 - M}(-\epsilon) > -\delta. \tag{A.5}$$

On the other hand,

$$\tilde{\pi}_i^0 - \pi_i^0 \leq \frac{\pi_0 \pi_1 \sigma (\exp(\frac{\epsilon M X_i^2}{2\sigma^2}) - 1)}{\frac{\pi_1^2 \sigma^2}{\sqrt{\sigma^2 + \tau^2}} \exp(\frac{M X_i^2}{2\sigma^2})} = \frac{\pi_0 \sqrt{\sigma^2 + \tau^2}}{\pi_1 \sigma} \cdot \frac{exp(\frac{\epsilon M X_i^2}{2\sigma^2}) - 1}{\exp(\frac{M X_i^2}{2\sigma^2})}.$$

We use C to denote the constant $\frac{\pi_0 \sqrt{\sigma^2 + \tau^2}}{\pi_1 \sigma}$, and let $y = \exp(\frac{\epsilon M X_i^2}{2\sigma^2})$, then $\exp(\frac{M X_i^2}{2\sigma^2}) = y^k$. If $X_i = 0$, then $y = 1$,

$$\tilde{\pi}_i^0 - \pi_i^0 \leq 0.$$

Otherwise, if $X_i \neq 0$, then $y > 1$, and

$$\tilde{\pi}_i^0 - \pi_i^0 \leq C \frac{y - 1}{y^k} = C \frac{y - 1}{(y - 1 + 1)^k} \leq C \frac{y - 1}{k(y - 1)} < C\epsilon. \tag{A.6}$$

Combine (A.5) and (A.6), then

$$|\tilde{\pi}_i^0 - \pi_i^0| \leq max(\delta, C\epsilon) < \delta. \tag{A.7}$$

Now, assume that π_0 is also estimated by $\hat{\pi}_0$. Let $A = \frac{\sigma}{\sqrt{\sigma^2 + \hat{\tau}^2}} \exp(\frac{\hat{M} X_i^2}{2\sigma^2})$, then

$$|\hat{\pi}_i^0 - \tilde{\pi}_i^0| = |\frac{\hat{\pi}_0}{\hat{\pi}_0 + \hat{\pi}_1 A} - \frac{\pi_0}{\pi_1 + \pi_1 A}| = |\frac{(\hat{\pi}_0 - \pi_0) A}{(\hat{\pi}_0 + \hat{\pi}_1 A)(\pi_0 + \pi_1 A)}|$$

The denominator greater than $\hat{\pi}_0\pi_1 A$ implies that $|\hat{\pi}_i^0 - \tilde{\pi}_i^0| < |\frac{\hat{\pi}_0-\pi_0}{\hat{\pi}_0\pi_1}|$. Since $\hat{\pi}_0$ is consistent for π_0, for any $\delta > 0$, there $\exists P_0$ such that $\forall p > P_0$, $|\hat{\pi}_0 - \pi_0| < \delta$, then

$$|\hat{\pi}_i^0 - \tilde{\pi}_i^0| \le D\delta,$$

where D is a constant that only depends on π_0. Combining this with (A.7), one can get that

$$|\hat{\pi}_i^0 - \pi_i^0| \le (1+D)\delta, \text{ for all } i = 1, 2, \cdots, p$$

and completes the proof.

Proof of the theorem

According to [12], $FCR_\pi = E\frac{\sum_{i\in\mathcal{R}} P(\theta_i \notin CI_i^{EB}|X)}{R}(R > 0)$ where $\mathcal{R}$ is the set of index of parameters that are selected and R is the number of selected parameters, i.e. $R = \#\mathcal{R}$. Similarly as formula (A.3) in the proof of theorem 3.1,

$$\begin{aligned} & P(\theta_i \notin CI_i^{EB}|X) \\ = \ & \pi_i^0(X) I(\hat{\pi}_i^0(X) < \hat{k}_2) + (1 - \pi_i^0(X)) P(\theta_i \notin CI_i^{EB}|X, I_i = 1) \end{aligned}$$

In the empirical Bayes interval (10), there exists a positive correction term $-\hat{M}\log\hat{M}\sigma^2$. Dropping this term results in a short interval which enlarges the non-coverage probability, i.e.

$$P(\theta_i \notin CI_i^{EB}|X) \le P(|\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2).$$

Consequently,

$$\begin{aligned} & P(\theta_i \notin CI_i^{EB}|X) \\ \le \ & \pi_i^0(X) I(\hat{\pi}_i^0(X) < \hat{k}_2) + (1 - \pi_i^0(X)) P((\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2|X, I_i = 1). \end{aligned}$$

Rearrange the terms in the above formula, one can simply the conditional non-coverage probability $P(\theta_i \notin CI_i^{EB}|X)$ as

$$\begin{aligned} & \pi_i^0(X)(I(\hat{\pi}_i^0(X) < \hat{k}_2) - q) + \pi_i^0(X)(q - P((\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2|X, I_i = 1)) \\ & + P((\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2|X, I_i = 1). \end{aligned}$$

Let

$$\Delta_1 = \frac{\sum_{i\in A} \pi_i^0(X)(I(\hat{\pi}_i^0(X) < \hat{k}_2) - q)}{R},$$

$$\Delta_2 = \frac{\sum_{i\in A} \pi_i^0(X)(q - P((\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2|X, I_i = 1))}{R},$$

and

$$\Delta_3 = \frac{\sum_{i\in A} P((\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2|X, I_i = 1)}{R},$$

then FCR_π can be controlled from above by $E(\Delta_1 + \Delta_2 + \Delta_3)$.

Since $\hat{\pi}_0$ and $\hat{\tau}^2$ are obtained by using the method of moments, Delta method implies that $\hat{\pi}_0 - \hat{\pi} = O_p(\frac{1}{\sqrt{p}})$ and $\hat{\tau}^2 - \tau^2 = O_p(\frac{1}{\sqrt{p}})$.

According to Lemma (A.1), for any $\epsilon > 0$, we can always find sufficiently large P_0, such that for any $p > P_0$, $(\hat{\tau}^2 - \tau^2)^2 < \delta/3$ and $(\hat{\pi}_i^0(X) - \hat{\pi}_i^0(X))^2 < \delta/3$. Consequently,

$$E\Delta_1 \leq E\frac{\sum_{i\in A}\pi_i^0(X)(I(\pi_i^0(X) < \hat{k}_2 + \sqrt{\delta/3}) - q)}{R} = f(p, \tau^2, \pi_0, \hat{k}_2 + \sqrt{\delta/3}).$$

Since $(\hat{\tau}^2 - \tau^2)^2 + (\hat{\pi}_i^0(X) - \pi_i^0(X))^2 + (\delta/3)^2 \leq \delta$, therefore according to the property of the function f,

$$f(p, \tau^2, \pi_0, \hat{k}_2 + \sqrt{\delta/3}) \leq f(p, \hat{\tau}^2, \hat{\pi}_0, \hat{k}_2) + \epsilon \leq \epsilon,$$

Since $\hat{k}_2$ is simulated as the maximum k_2 such that $f(p, \hat{\tau}^2, \hat{\pi}_0, k_2) \leq 0$,

$$E\Delta_1 \leq \epsilon. \tag{A.8}$$

For the second term Δ_2,

$$\begin{aligned} |\Delta_2| &\leq \frac{\sum_{i\in A}\pi_i^0(X)|q - P((\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2 | X, I_i = 1)|}{R} \\ &\leq \frac{\sum_{i\in A}|q - P((\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2 | X, I_i = 1)|}{R}. \end{aligned}$$

Taking a close look at the term $P((\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2 | X, I_i = 1)$, one knows that $(\theta_i | X_i, I_i = 1) \sim N(MX_i, M\sigma^2)$. Therefore one can replace θ_i by $MX_i + \sqrt{M}\sigma Z$ where Z is a standard normal random variable which is independent of X_i. Consequently,

$$P((\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2 | X, I_i = 1) = P(|Z - \frac{(\hat{M} - M)X_i}{\sqrt{M}\sigma}| > \sqrt{\frac{\hat{M}}{M}} z_{q/2} | X). \tag{A.9}$$

Assume that $X_{(p)}$ is the observation that has the largest absolute value, then $0 \leq |\frac{(\hat{M}-M)X_i}{\sqrt{M}\sigma}| \leq |\frac{(\hat{M}-M)X_{(p)}}{\sqrt{M}\sigma}|$. Consequently, for any $i = 1, 2, \cdots, p$, (A.9) falls into the range

$$[P(|Z - \frac{|(\hat{M} - M)X_{(n)}|}{\sqrt{M}\sigma}| \geq \sqrt{\frac{\hat{M}}{M}} z_{q/2}), P(|Z| \geq \sqrt{\frac{\hat{M}}{M}} z_{q/2})]. \tag{A.10}$$

Let $X_i = \sqrt{\sigma^2 + \tau^2} Z_i$, then $Z_i = \pi_0 N(0, \frac{\sigma^2}{\sigma^2 + \tau^2}) + \pi_1 N(0, 1)$. Furthermore

$$|\frac{(\hat{M} - M)X_{(p)}}{\sqrt{M}\sigma}| = |\frac{\sigma(\hat{\tau}^2 - \tau^2)}{\tau(\hat{\tau}^2 + \sigma^2)} Z_{(p)}|,$$

As a result, the range (A.10) can be rewritten as

$$[P(|Z-|\frac{\sigma(\hat{\tau}^2-\tau^2)}{\tau(\hat{\tau}^2+\sigma^2)}Z_{(p)}||\geq|\sqrt{\frac{\hat{M}}{M}}z_{q/2}), P(|Z|\geq\sqrt{\frac{\hat{M}}{M}}z_{q/2})]. \tag{A.11}$$

Since the above range applies for all i's, one knows that

$$|\Delta_2|\leq max(|q-P(|Z-|\frac{\sigma(\hat{\tau}^2-\tau^2)}{\tau(\hat{\tau}^2+\sigma^2)}Z_{(p)}||\geq|\sqrt{\frac{\hat{M}}{M}}z_{q/2})|, |q-P(|Z|\geq\sqrt{\frac{\hat{M}}{M}}z_{q/2})|). \tag{A.12}$$

Since $\hat{\tau}^2-\tau^2=O_p(\frac{1}{\sqrt{p}}), Z_{(p)}=O(\sqrt{2\log p})$,

$$|\frac{\sigma(\hat{\tau}^2-\tau^2)}{\tau(\hat{\tau}^2+\sigma^2)}Z_{(p)}|=o_p(1). \tag{A.13}$$

The dominated convergence theorem implies that

$$P(|Z-|\frac{\sigma(\hat{\tau}^2-\tau^2)}{\tau(\hat{\tau}^2+\sigma^2)}Z_{(p)}||\geq|\sqrt{\frac{\hat{M}}{M}}z_{q/2}))\to P(|Z|>z_{q/2})=q,$$

and

$$P(|Z|\geq\sqrt{\frac{\hat{M}}{M}}z_{q/2})\to q.$$

Applying the dominated convergence theorem again, one can deduce from (A.12) that

$$\limsup_{p\to\infty}E|\Delta_2|\leq 0. \tag{A.14}$$

Similar arguments apply to Δ_3 and one can show that

$$\begin{aligned}\Delta_3 &\leq P(|Z-\frac{|(\hat{M}-M)X_{(p)}|}{\sqrt{M}\sigma}|\geq\sqrt{\frac{\hat{M}}{M}}z_{q/2}|X)\\ &= P(|Z-|\frac{\sigma(\hat{\tau}^2-\tau^2)}{\tau(\hat{\tau}^2+\sigma^2)}Z_{(p)}||\geq\sqrt{\frac{\hat{M}}{M}}z_{q/2}).\end{aligned}$$

Dominated convergence theorem and (A.13) implies that

$$\limsup_{p\to\infty}E|\Delta_3|\leq\lim_{p\to\infty}EP(|Z|\geq z_{q/2})=q. \tag{A.15}$$

(A.8), (A.14), and (A.15) imply that

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

Proof of the proposition 4.1.

Assume that $X_i \sim \pi_0 N(0,\sigma^2) + (1-\pi_0)N(0,\tau^2+\sigma^2)$ and $Y_i \sim \pi_0' N(0,\sigma^2) + (1-\pi_0')N(0,\tau'^2+\sigma^2)$ where $i = 1, 2, \cdots, p$. Then

$$\begin{aligned} & |f(p,\pi_0,\tau^2,k) - f(p,\pi_0',\tau'^2,k')| \\ = \;& E\frac{\sum \pi_i^0(X)((\pi_i^0(X)<k)-q) - \pi_i'^0(Y)(I(\pi_i'^0(Y)<k')-q)}{R} \\ = \;& E\frac{q\sum(\pi_i^0(X)-\pi_i'^0(Y)) + \sum(\pi_i^0(X)I(\pi_i^0(X)<k) - \pi_i'^0(Y)I(\pi_i'^0(Y)<k'))}{R} \end{aligned}$$

where the summation is taken from 1 to R. Since R goes to ∞ as $p \to \infty$, therefore by using the law of large number, the inside function of the above expectation converges to $\Delta = qE(\pi_1^0(X) - \pi_1'^0(Y)) + E(\pi_1^0(X)I(\pi_1^0(X) < k) - \pi_1'^0(Y)I(\pi_1'^0(Y) < k'))$ in probability. Since the integral is a bounded function, it is sufficient to show that $\forall \epsilon > 0$, there exists δ, such that $(k'-k)^2 + (\tau'^2 - \tau^2)^2 + (\pi_0' - \pi_0)^2 < \delta$ implies that $|\Delta| < \epsilon$.

In fact, $E\pi_1^0(X)E(P(\theta_0 = 0|X)) = P(\theta_0 = 0) = \pi_0$. This implies that

$$qE(\pi_1^0(X) - \pi_1'^0(Y)) = q(\pi_0 - \pi_0'). \tag{A.16}$$

Furthermore, direct calculation shows that

$$\begin{aligned} & E(\pi_1^0(X)I(\pi_1^0(X)<k)) = \int_{\pi_1^0(X)<k} \pi_1^0(X)m(X)dX \\ = \;& \pi_0 \int_{\pi_1^0(X)<k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2})dx. \end{aligned}$$

Since $\{\pi_1^0(X) < k\}$ implies that $|X|^2 > \frac{2\sigma^2}{M}(\log\frac{1-k}{k} + \log\frac{\pi_0}{\pi_1\sqrt{1-M}})$,

$$\begin{aligned} & E(\pi_1^0(X)I(\pi_1^0(X)<k) - \pi_1'^0(Y)I(\pi_1'^0(Y)<k')) \\ = \;& P(|N|^2 > \frac{2\sigma^2}{M}(\log\frac{1-k}{k} + \log\frac{\pi_0}{\pi_1\sqrt{1-M}})) \\ - \;& P(|N|^2 > \frac{2\sigma^2}{M'}(\log\frac{1-k'}{k'} + \log\frac{\pi_0'}{\pi_1'\sqrt{1-M'}})), \end{aligned}$$

where N is a standard normal random variable. When k, k' are close to 1, then $\log\frac{1-k}{k} \to -\infty$, therefore, $|E(\pi_1^0(X)I(\pi_1^0(X) < k) - \pi_1'^0(Y)I(\pi_1'^0(Y) < k'))| = 0$ if $k, k' > \epsilon_1$ where $\epsilon_1 < 1$ is close to 1 sufficiently. Similarly, if k, k' are close to 0, then $\log\frac{1-k}{k} \to \infty$. We can choose sufficiently small ϵ_0, such that when $k, k' < \epsilon_0$,

$$P(|N|^2 > \frac{2\sigma^2}{M}(\log\frac{1-k}{k} + \log\frac{\pi_0}{\pi_1\sqrt{1-M}})) < \frac{\epsilon}{2}$$

and

$$P(|N|^2 > \frac{2\sigma^2}{M'}(\log\frac{1-k'}{k'} + \log\frac{\pi_0'}{\pi_1'\sqrt{1-M'}})) < \frac{\epsilon}{2}.$$

Consequently, $|E(\pi_1^0(X)I(\pi_1^0(X) < k) - \pi_1'^0(Y)I(\pi_1'^0(Y) < k'))| < \epsilon$ when k, k' are either close to 0 or 1.

Furthermore, assume that $0 < \epsilon_0 < k, k' < \epsilon_1 < 1$, then by the continuity of $E(\pi_1^0(X)I(\pi_1^0(X) < k) - \pi_1'^0(Y)I(\pi_1'^0(Y) < k'))$, there exists a small $\delta < \epsilon$, such that $(k' - k)^2 + (\tau'^2 - \tau^2)^2 + (\pi_0' - \pi_0)^2 < \delta$ implies that $|E(\pi_1^0(X)I(\pi_1^0(X) < k) - \pi_1'^0(Y)I(\pi_1'^0(Y) < k'))| < \epsilon$. Combining this with (A.16), one obtains that $|\Delta| < \epsilon$ when δ is sufficiently small, which completes the proof.