THE EMPIRICAL BAYES APPROACH FOR SHRINKAGE CONFIDENCE INTERVALS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Zhigen Zhao

August 2009

THE EMPIRICAL BAYES APPROACH FOR SHRINKAGE CONFIDENCE INTERVALS

Zhigen Zhao, Ph.D.

Cornell University 2009

The *parametric empirical Bayes*, introduced by [12], [13], [14], and [34], is gaining more and more attention in both the theoretic and applied statistics. Its ability in borrowing strength makes this idea prevalent in modern technology such as microarray where the number of parameters is very large and the number of observations for each parameter is much smaller. In this dissertation, we will apply this idea into constructing confidence interval for different models and problem settings.

In Chapter 2, we introduce the Log-Normal model and construct the *empirical Bayesian* confidence interval for each parameter by shrinking both means and variances for the very first time. Keeping the coverage probability above the nominal level, the new construction enjoys the shortest average length among all the confidence interval constructed as demonstrated by extensive numerical study as well as in a real data set where the real parameters are known.

In Chapter 3, we deal with the simultaneous interval construction, where the criterion of controlling the simultaneous coverage probability appears to be too conservative. We propose and study the control of the *empirical Bayes* False Coverage Rate (FCR) to address the multiplicity. We further construct intervals which controls the *empirical Bayesian* FCR under the normal-normal model. In Chapter 4, we deal with the model with mixed prior which is more practical in microarray technology and construct intervals which can control the empirical

Bayesian FCR.

All the procedures we have derived in this work based on the *empirical Bayes* approach are explicitly defined and can be computed instantaneously.

BIOGRAPHICAL SKETCH

Zhigen Zhao was born in Shaodong County, Hunan Province, China, on March 29th, 1981. He went to the attached middle school of Hunan Normal University as a senior high school student in 1996 and graduated in 1999. After that, he went to Nankai University in August, 1999 and received his Bachelor of Science in Mathematics from Nankai University, Tianjin, China in 2003. In 2004, he came to Cornell University to pursue his Ph.D in the field of mathematics. In 2007, he married Ms. Xiao Xiao.

To my dear wife Xiao Xiao and our parents

ACKNOWLEDGEMENTS

Many thanks go to my advisor, Professor J. T. Gene Hwang, who has guided me over the past five years. He is my teacher and friend who takes care of me as a father to a son. Not only has he taught me to be a good researcher and scholar, but also he has shown me to be a wise person. He is humble, patient, slow to anger, optimistic, energetic and encouraging. He is there whenever I have any difficulties and problems in either my research or my daily life. He has set me a perfect model as an advisor, a husband, and a family leader.

I would also like to thank Professor Michael Nussbaum for serving as my co-chair. He has showed me how to be a serious and successful researcher and I have been privileged to join his discussion group. I would also like to thank him for his generosity in providing me with financial support several times. I would like to thank Professor James Booth for serving as a committee member and for providing me with many suggestions for my dissertation work.

Many friends have offered me generous help in creating the final version of my dissertation. Especially, I would like to thank my friends, Ms. Xin Ma, Ms. Katherine J. Neafsey, and Dr. Bret Hanlon for spending their precious time editing my dissertation.

I want to thank Ms. Jing-huei Hwang and Brian Hwang for their friendship and hospitality. I would like to further mention my friends in Ithaca: Andrew Eastlick, Edward Hoffman, Pengsheng Ji, Xingqun Jiang, Junran Li, Xin Ma, Roger Miller, Gang Song, and others. There is no end for such a list. My life at Cornell has been full of joy because of them.

My mother and my parents in law have been very encouraging and supportive in the past few years. Their optimism, diligence and initiative have had a great influence on me.

Finally, I would like to thank my dear wife, Ms. Xiao Xiao, for her love, support, understanding and encouragement. She has been with me under all circumstances. We sustain the difficulties, share the joys, and endure the suffering. I am so fortunate to have married her since she is the most wonderful, considerate, and supportive wife.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The *parametric empirical Bayes*, introduced by [12], [13], and [14], has been widely used in theoretical and applied statistics. The basic idea is to assume a Bayesian model and estimate the prior distribution by using the data. Consequently, by using an *empirical Bayes* approach, we can take advantage of Bayes analysis while avoiding the assumption of exact prior distribution of the parameters. A typical setup is as following: assume that $X|\theta \sim f(x|\theta)$ and $\theta \sim \pi(\theta)$, where $\pi \in \Pi$ is a family of prior distribution of θ. If Π consists of only one distribution of θ, then this falls into the category of Bayesian framework. The first important application of such an idea is the estimation of the classical mean problem.

Assume that $X \sim N(\theta, \sigma^2 I_p)$, where σ^2 is assumed to be known. The point estimator $\hat{\theta} = X$ of θ maximizes the joint likelihood function. It is further known that such an estimator is minimax and the Uniformly Minimum Variance Unbiased Estimate (UMVUE), but inadmissible. The celebrated work [25] proved that it can be dominated in terms of mean squared error loss by the James-Stein estimator

$$\delta(X) = (1 - \frac{(p-2)\sigma^2}{\sum X_i^2})X \tag{1.1}$$

when $p \geq 3$ which shrinks the observations X toward 0.

It is well known that this estimator $\delta(X)$ can be derived using the *empirical Bayesian* technique. Assume that $X|\theta \sim N(\theta, \sigma^2 I_p)$ and $\theta \sim N(0, \tau^2 I_p)$ where τ^2 is an unknown hyper-parameter. Then conditioning on τ^2, $E(\theta|X) = MX$ where $M = \frac{\tau^2}{1+\tau^2}$. Since M is unknown, we can replace it by an unbiased estimator $1 - \frac{(p-2)\sigma^2}{||X||^2}$ and thus derive the James-Stein estimator. The estimation of τ^2 or M involves all the observation $X_1, \cdots, X_p$. Consequently, for one specific parameter

θ_i, the corresponding estimator is $(1 - \frac{(p-2)\sigma^2}{||X||^2})_+ X_i$, which depends not only on X_i, but also $X_1, \cdots, X_p$. This is known as the borrowing strength factor. By borrowing strength from seemingly unrelated observations, [25] indicated that one can do better than the maximum likelihood estimator X_i which does not borrow strength from other observations. This phenomenon is called the Stein's paradox. [4] has related this phenomenon to the recurrent diffusion. This paradox indicates that the multiplicity could bring a revolution among the statistical methodology, which is not only restricted to the point estimator, and that the *empirical Bayes* approach could serve as an efficient and important tool.

Inspired by [25], there are several attempts in constructing a new confidence set by using the idea of shrinkage which could dominate the naïve set $CI^z = \{\theta : |\theta - X| \leq c\sigma^2$ where $P(\chi^2_p > c) = \alpha/2$. Here we say CI^{new} is dominating CI^z if

(i) $P(\theta \in CI^{new}) \geq P(\theta \in CI^z)$

(ii) The volume of $CI^{new} \leq$ the volume of CI^z

with strict inequality holding in either (i) or (ii) for a set θ or x with positive Lebesgue measure.

To name a few, [20], [21] and [22] have constructed a new set which keeps the same volume as the naïve set CI^z while improving the coverage probability. [16] applies the decision Bayes theory and constructs a set under the Bayesian framework. In 1983, [5] adapted the decision Bayesian framework proposed by [16] for the Normal-Normal model $X|\theta \sim N(\theta, I_p)$, $\theta \sim N(0, \tau^2 I_p)$. In this paper, they constructed the confidence set by using the *empirical Bayesian* technique. The loss function they have used for the confidence set CI is

$$L(\theta, CI) = kVol(CI) - I_{CI}(\theta),$$

where k is a tuning parameter which is chosen so that the usual $1 - \alpha$ confidence set is minimax against L. The loss function provides a balance between the volume of the confidence set and the true coverage. The decision Bayes confidence set they have derived and the highest posterior density (HPD) region differ in the half radius of the sets. One major issue in deriving their *empirical Bayesian* set concerns the estimation of the hyper-parameter τ^2. After accounting for the estimation error, the resultant set would not have good coverage probabilities. [5] achieved the good coverage probability by taking a loss approach to derive a positive term in the radius and by a truncation so the radius is above a positive lower bound. Their confidence set has smaller volume than the usual confidence set while maintaining the coverage probability to be at least $1-\alpha$. Another advantage of this construction is that it is explicitly defined and thus easy to implement.

The confidence set is very useful; however, in many applications such as microarray experiments, it is more practical to construct confidence intervals for parameters. Direct projection of confidence sets to each individual coordinate will result in a less satisfactory result because the length is too long. Also, based on the frequentist criteria, we are unable to improve the usual z-interval $\{\theta_i : |X_i - \theta_i| \leq z_{1-\alpha/2}\sigma\}$ where $P(|Z| > z_{1-\alpha/2}) = \alpha$. Therefore, the interval construction requires more work and thinking. Not only do we need new methodology other than the projection, we also need new criteria to adjust the frequentist coverage probability. In 1983, [34] introduced the so-called *empirical Bayesian* coverage probability. We now define this concept.

Assume that $X_i|\theta_i \sim f(x_i, \theta_i)$ and θ_i follows the prior distribution $\pi(\theta_i) \in \Pi$, where Π is a family of prior distributions. Then we call CI_i a confidence interval of θ_i that controls the *empirical Bayesian* coverage probability at $(1-\alpha)$ level if

$$E_{\theta_i,X_i}P(\theta_i \in CI_i) \geq 1-\alpha, \text{for all prior distribution of } \pi(\theta_i) \in \Pi.$$

Notice that if the family of prior distributions Π of θ consists of only one prior, then this definition simply corresponds to the Bayesian coverage probability. On the other extreme case when Π consists of all possible distributions including all random variables with point mass, the above criteria is the frequentist coverage probability. If Π is between the above two extreme cases, we are in a paradigm between frequentist and Bayesian. In modern technology especially in microarray, this idea has become quite reasonable since biologists are used to thinking about the distributions of θ_i's. Throughout this dissertation when we mention the coverage probability, we always refer to the *empirical Bayesian* coverage probability.

[33] and [34] have further constructed the *empirical Bayesian* confidence interval for each individual θ_i based on the following canonical model

$$X_i \sim i.i.d. \quad N(\theta_i, \sigma^2) \quad \text{and} \quad \theta_i \sim i.i.d. \quad N(0, \tau^2), i = 1, 2, \cdots, p.$$

He concluded that his interval can control the *empirical Bayesian* coverage probability to be at least $1-\alpha$ and the average length across all the parameters is shorter than that of the usual z-interval.

Inspired by [5], [17] has constructed the confidence interval for each individual θ_i by adapting the decision approach and *empirical Bayesian* idea. It was shown that the length of this interval is always smaller than that of the usual z-interval for each individual.

However, in both of these approaches, they either assume that the variances of each individual θ_i are known and equal or simply replace them by some unbiased estimator S_i^2. The frequentist $1-\alpha$ interval of θ_i is constructed as $\{\theta_i : |X_i - \theta_i| \leq t_{d,1-\alpha/2}\}$ where $X_i \sim N(\theta_i, \sigma_i^2)$, $\frac{S_i^2}{\sigma_i^2} \sim \frac{\chi_d^2}{d}$, and $P(|T_d| > t_{d,1-\alpha/2}) = \alpha$.

When the dimension p is very large, then S_i^2 will have extreme values, either being too large or too small. A large value of S_i^2 hurts the accuracy of the interval while a small S_i^2 hurts the coverage probability. Therefore, it seems that we can do better if we shrink the variances, i.e. if we decrease the large values and increase the small values of S_i^2's. Indeed, this is the case and the variance shrinkage approach is very much needed as demonstrated in Chapter 2. In that chapter, we have proposed a canonical log-normal model where the variances are assumed to be unknown and unequal. We further constructed the confidence interval for each individual when shrinking both means and the variances by using the *empirical Bayesian* idea. The interval construction we have derived is explicitly defined.

In microarray experiments, it is known that most genes have a differentially expression of 0. This ratio could be as large as 90%. Therefore, it is no longer appropriate to simply put a normal prior $N(0, \tau^2)$ for θ_i. As argued in [36], it is necessary to assume with some positive probability π_0, $\theta_0 = 0$ and $\theta \sim N(0, \tau^2)$ with probability $\pi_1 = 1 - \pi_0$. Also, scientists are especially interested in the genes with the largest magnitude of differential expression and wish to construct the confidence interval after such a selection.

The usual z-interval $\{\theta_i : |X_i - \theta_i| < z_{1-\alpha/2}\sigma\}$ fails to control the confidence level due to the selection bias. It is well known that Bayes calculation can "wash away" the selection bias. Based on this, [36] has constructed the confidence interval for selected populations which can control the *empirical Bayesian* coverage probability at $1-\alpha$ level when it is assumed that θ_i's follow a mixture prior. In Chapter 4, we introduce a new loss function to address the issue of a mixture prior and reproduced the procedure from [36]. This idea can be generalized to more practical situations.

In the setting of multiple confidence intervals, the confidence intervals mentioned above control the conservative simultaneous coverage probability. Consequently, the (average) length will be very large especially for a large number of parameters. [2] proposed a new criterion, namely, the False Coverage Rate (FCR), which resembles the FDR for multiple hypothesis testing. They have further constructed the confidence interval which controls the FCR at a given α-level. They have based their argument and reasoning on the frequentist framework. In Chapter 3, we have coined the term *empirical Bayesian* FCR and related the Bayes confidence interval with Bayes FCR. We have further constructed a new *empirical Bayesian* confidence interval which is much shorter than the confidence interval from [2], guaranteeing the control of *empirical Bayesian* FCR at the given α-level.

Through all these chapters, we can see that *empirical Bayesian* methodology plays a central part in defining the criteria and constructing confidence intervals. The ability of washing away selection bias and borrowing strength enables this method to serve as an important and necessary tool in modern statistics.

CHAPTER 2

EMPIRICAL BAYES CONFIDENCE INTERVALS SHRINKING BOTH MEANS AND VARIANCES

2.1 Introduction

Many modern applications of statistics involve simultaneously dealing with a large number p of populations. Microarray data analysis is one such example. Since the empirical Bayes technique is designed to borrow strength from various populations, its improvement over the traditional approach (which treats each population separately) becomes large when p is large. Hence, it is not surprising that the technique is becoming increasingly popular in modern applications.

Although there were many empirical Bayes point estimators proposed in the literature, there are relatively significantly fewer empirical Bayes confidence sets or intervals constructed. This is likely due to the technical difficulty in constructing intervals and not because they are unimportant. In spite of its difficulty, there have been several major attempts since [40]. See [33], [34], [5], [17], [29], [42], [38], [11] and [36], where [36] treated the selected parameters.

As in [33], [34], [29], and [17], we shall focus here on constructing a confidence interval for each mean rather than a simultaneous set for all means. The reason is that in modern applications, we are often interested in assessing each individual population mean and specifying a range for each of the parameters as in the context of multiple testing. If we derive an interval by projecting a set (such as a sphere) to the coordinates of interest, the resultant interval would be very long. All the intervals considered aim at covering the means θ_i's, which are assumed to be random throughout the paper. We construct *empirical Bayes confidence intervals*, or for short, *confidence intervals* whose definition is given at the end of section 2.2.

The existing confidence intervals for the means in literature assume that variances σ_i^2's are either equal or unequal but known. For the situation when variances are unequal and unknown, the suggestion is typically to replace them by the unbiased estimators S_i^2's. See, for example, [33], [34]. The suggestion is not too bad when p is small. However, in modern applications such as microarray data analysis where p is large, there would be advantage to apply the empirical Bayes method to "borrow strength" from other populations, which typically results in shrinking the variances as well as the means. This idea of using shrinkage variance estimators has been carried out in the context of multiple testing to avoid false positives due to small S_i^2. See, for example, the techniques in [43], (known as the SAM technique), [31], [45], [39], [10], [30] and [23]. Even the SAM procedure, which appears to add a data dependent positive constant to the standard deviation estimator, can be viewed as shrinking the standard deviation toward the positive constant when dividing the denominator by 2. [Dividing the denominator by 2 doesn't change the test.]

In this paper, we construct shorter confidence intervals for the means than the t-intervals, which aims at θ_i's also, by "borrowing strength" from the other populations and by shrinking the variances as well as the means. Although we focus on $1-\alpha$ one-dimensional intervals, they can be combined to form simultaneous intervals with simultaneous confidence coefficient $(1-\alpha)^k$ if k intervals are involved. Here the confidence coefficient is defined at the end of Section 2.2. These types of simultaneous confidence intervals can be used to conduct tests for interval hypotheses involving random θ_i's. See [35], page 46, which deals with θ_i's ordered according to the data. Since θ_i's are random, the intervals for θ_i's are not ordinary confidence intervals for fixed parameters. Some readers prefer to call these type of intervals *prediction intervals*, which is especially appropriate in the case of a random effect model where the usual point estimator is called a predictor. See [37].

Specifically, for the model in Section 2.2, we construct in Section 2.3 the double shrinkage confidence intervals when hyper-parameters are known. Section 2.4 gives some analytic and numerical evidence to show that these alternative intervals have better characteristics than the t-interval. Sections 2.5 and 2.6 construct our recommended interval $\hat{C}^{SS}$ for the more realistic situation where the hyper-parameters are unknown and need to be estimated. The obstacle to overcome in Sections 2.5 and 2.6 is to ensure that the length after estimating the hyper-parameters is always greater than a positive quantity, which needs to be determined so as to have coverage probabilities at least $1-\alpha$. The readers who are interested in a direct definition of $\hat{C}^{SS}$ can find a summary in the last paragraph of Section 2.6. The interval $\hat{C}^{SS}$ shrinks both the means and the variances and hence is called a double shrinkage (SS) interval. Numerical studies show that, while maintaining at least $1-\alpha$ coverage probabilities, it is on average , ranging from 40% to 60%, substantially shorter than the $1-\alpha$ t-interval and the shrink-variance-alone intervals. Section 2.7 derives another double shrinkage interval $\hat{C}^{SS_g}$ based on a more traditional model. However $\hat{C}^{SS}$ performs better according to simulation studies. All the alternative procedures derived in this paper are defined explicitly and hence can be computed instantaneously.

We apply in Section 2.8 the intervals to a real data set in which all the "true" parameters are known. We discover that $\hat{C}^{SS}$ and $\hat{C}^{SS_g}$ perform the best in average lengths and their coverage probabilities, approximated by the proportion of covering the means corresponding to the genes, are closest to the nominal level.

2.2 Models and Assumptions

We consider the following canonical model for the observations (X_i, S_i^2), $i = 1, \ldots, p$, where

$$X_i|\theta_i, \sigma_i^2 \overset{ind.}{\sim} N(\theta_i, \sigma_i^2),$$

$$\theta_i \overset{i.i.d.}{\sim} N(\mu, \tau^2),$$

$$\ln(S_i^2) = \ln(\sigma_i^2) + \delta_i,\ \ln(\sigma_i^2) \overset{i.i.d.}{\sim} N(\mu_v, \tau_v^2) \text{ and } \delta_i \overset{i.i.d.}{\sim} N(m, \sigma_{ch}^2). \qquad (2.1)$$

where δ_i is independent of σ_i^2, and μ, τ, μ_v, and τ_v are unknown parameters. However m and σ_{ch}^2 are known and

$$m = E(\ln(\chi_d^2/d)),\ \text{and } \sigma_{ch}^2 = Var(\ln(\chi_d^2/d)), \qquad (2.2)$$

where χ_d^2 is a chi–square random variable with d degrees of freedom. The subscript v refers to the hyper-parameters relating to variances. Furthermore, we assume that X_i's and S_i's conditioning on σ_i's are mutually independent. Traditionally in practice, S_i^2 is the unbiased estimator of σ_i^2 and S_i^2/σ_i^2 is assumed to have the same distribution as χ_d^2/d for some d degrees of freedom. In model (2.1), we assume instead that $e^{\delta_i} = S_i^2/\sigma_i^2$ follows a log-normal distribution. Also we match the mean and variance of δ_i with those of $\ln(\chi_d^2/d)$. These two assumptions whether e^{δ_i} follows log-normal or χ_d^2/d cause little practical difference, which is partly demonstrated in Figure 2.1 and would be described in more details in the paragraph after equation (2.9). Also at the end of Section 2.4, another evidence is presented.

In model (2.1), the distribution of σ_i^2 is usually assumed to be inverse gamma with some parameters a and b. This, combined with the assumption that conditioning on σ_i^2, S_i^2/σ_i^2 is distributed according to χ_d^2/d, is called the *inverse gamma model*, which is spelled out more precisely in the beginning of section 2.7. As opposed to this, model (2.1) is called the *log–normal model*. We focus on model (2.1) instead of the inverse gamma model for two reasons. First, by assuming model (2.1), we can use the traditional normal theory and hence the estimation of hyper-parameters μ, τ^2, μ_v and τ_v^2 are readily available, leading to explicit confidence intervals unlike the inverse gamma model requiring solving an equation to estimate a and b. Second, the numerical evidence shows that the recommended confidence interval $\hat{C}^{SS}$ that we construct using the log-normal model is shorter on average with respect to three models including *log–normal model* and *inverse gamma model* than the confidence interval $\hat{C}^{SS_g}$ constructed using inverse gamma model especially when $d = 2$. This also indicates that the confidence intervals derived are insensitive to the prior distribution assumption of $\ln(\sigma_i^2)$, which is quite comforting.

In multiple testing contexts, model (2.1) is assumed in [23]. Also a similar model is assumed in [27] and [30]. Under model (2.1), we now construct an empirical Bayes confidence interval for θ_i based on the decision theory similar to [5] and [17]. This approach produces shorter intervals than Morris' approach when $1 - \alpha \geq 0.9$ in simple settings studied in [17]. The coverage probabilities for θ_i's studied in this paper refer to the probabilities where θ_i, X_i, S_i, and σ_i are all integrated out and the confidence intervals constructed aim at having the coverage probabilities for θ_i's at least $1-\alpha$ whatever the hyper-parameters μ, τ, μ_v, and τ_v may be. This type of confidence intervals is called by [34] *empirical Bayes intervals having confidence coefficient* $1 - \alpha$. For short, they are called *confidence intervals* in this paper. The empirical Bayes intervals would have coverage probabilities for θ_i at least $1-\alpha$ with respect to a broader class of prior distributions of θ_i and σ_i^2 which consists of mixtures of (2.1) by mixing the hyper-parameters μ, τ, μ_v and τ_v. The coverage probabilities for θ_i's can be interpreted as the frequentist coverage probabilities if the $\theta_i's$ are the random effect in a random effect or more general mixed effect (e.g. ANOVA) model. In such a scenario, some readers prefer to call these intervals *prediction intervals* since they aim at the random θ_i's. In the ANOVA context, X_i is the ANOVA estimator of θ_i and S_i^2 is the mean squared due to error (MSE).

2.3 Construction of the Double Shrinkage Confidence Interval

In this section, we shall consider the loss function for a one–dimensional confidence set C,

$$\frac{k}{\sigma}\text{Len}(C) - I_C(\theta) \tag{2.3}$$

where k is a tuning constant, independent of the parameters, $\text{Len}(\cdot)$ represents the length (or Lebesgue measure) of C and the indicator function $I_C(\theta)$ is one or zero depending on whether $\theta \in C$ or not. This loss function has been used in [26], [5] and [17] when σ is known. For the unknown σ case, a loss function similar to (2.3) except that k/σ is replaced by a constant (independent of σ) has been used in [7], which shows that it leads to many paradoxes. [6] uses a loss function more general than (2.3) (with σ replaced by σ^p and length of C is replaced by its volume) to establish two different kinds of minimaxity properties for the multi–dimensional t–confidence set using a sequence of what we call the inverse gamma models in this paper except that σ_i's are assumed to be identical. This assumption of equal variance is used in all the results described above. In many applications including microarray data analysis, the variances are typically not known to be identical. For such an important case, we construct confidence intervals below. In particular, we construct the individual confidence interval for θ_i under model (2.1).

Under the loss function (2.3), it can be shown that the decision Bayes confidence interval for θ_i is the interval that minimizes

$$E\Big(\frac{k}{\sigma_i}\text{Len}(\mathbf{C}) - I_{\mathbf{C}}(\theta_i)|X_i, S_i^2\Big) = \int_{\mathbf{C}} \Big(kE\Big(\sigma_i^{-1}|X_i, S_i^2\Big) - \pi(\theta_i|X_i, S_i^2)\Big)d\theta_i.$$

Therefore as in [6], the decision-Bayes confidence interval for θ_i under the loss function (2.3) is:

$$C_i^B = \Big\{\theta_i : kE\Big(\sigma_i^{-1}|X_i, S_i^2\Big) < \pi(\theta_i|X_i, S_i^2)\Big\}. \tag{2.4}$$

In the equations above and below, $\pi(\cdot|\cdot)$ and $E(\cdot|\cdot)$ denote the posterior probability density function (pdf) and the posterior expectation. For now, we assume all the hyper-parameters μ, τ, μ_v and τ_v are known and will treat the unknown case in Sections 2.5 and 2.6. In general, $\pi(\sigma_i^2|X_i, S_i^2)$ will depend on X_i. Below and above, when we write the conditional pdf of σ_i^2, it is with respect to the dominating measure $d\sigma_i^2$. Here we approximate it by the conditional p.d.f. of σ_i^2 given S_i^2, $\pi(\sigma_i^2|S_i^2)$, which is easier to compute and is based on the intuition that compared to X_i, S_i^2 contains much more information in estimating σ_i^2. Hence the left hand side of the inequality in (2.4) becomes $kE(\sigma_i^{-1}|S_i^2)$. Note that $E\Big(\sigma_i^{-1}|S_i^2\Big) = E\Big(\exp\Big(-\frac{1}{2}\ln(\sigma_i^2)\Big)|S_i^2\Big)$ and below we approximate it by $\exp\Big(E\Big(-\frac{1}{2}\ln(\sigma_i^2)|S_i^2\Big)\Big)$. The approximation appears to be pretty rough especially because the exact calculation is possible, which leads to an additional multiplication factor. However we deliberately omit the factor so that the resultant interval is shorter and still has good coverage probabilities as testified by Figure 2.3. This is possibly due to the conservative nature of our derivation. Since

$\ln S_i^2 - m | \ln \sigma_i^2 \sim N(\ln(\sigma_i^2), \sigma_{ch}^2)$ and $\ln(\sigma_i^2) \sim N(\mu_v, \tau_v^2)$, the classical Bayesian calculation gives

$$\ln(\sigma_i^2) | \ln(S_i^2) \sim N\Big(M_v(\ln S_i^2 - m) + (1 - M_v)\mu_v, M_v\sigma_{ch}^2\Big),$$

where $M_v = \frac{\tau_v^2}{\tau_v^2 + \sigma_{ch}^2}$. Here all the subscripts "v" stand for "variance". Hence,

$$E\Big(\sigma_i^{-1} | S_i^2\Big) \approx \exp(-\frac{1}{2} E(\ln \sigma_i^2 | \ln S_i^2)) = \frac{1}{\hat{\sigma}_{B,i}}, \text{ where} \tag{2.5}$$

$$\begin{aligned} & \hat{\sigma}_{B,i}^2 = \exp(E(\ln \sigma_i^2 | \ln S_i^2)) \\ = \ & \exp(M_v(\ln S_i^2 - m) + (1 - M_v)\mu_v) = (\frac{S_i^2}{e^m})^{M_v} (e^{\mu_v})^{(1-M_v)}. \end{aligned} \tag{2.6}$$

Note that $\pi(\theta_i | X_i, S_i^2) = \int_0^\infty \pi(\theta_i | X_i, S_i^2, \sigma_i^2) \pi(\sigma_i^2 | X_i, S_i^2) d\sigma_i^2$. Approximating $\pi(\theta_i | X_i, S_i^2, \sigma_i^2)$ by $\pi(\theta_i | X_i, S_i^2, \sigma_i^2 = \hat{\sigma}_{B,i}^2)$, which is denoted for convenience as $\pi(\theta_i | X_i, S_i^2, \hat{\sigma}_{B,i}^2)$, we obtain an approximation of the right hand side of the inequality of (2.4),

$$\pi(\theta_i | X_i, S_i^2) \approx \int_0^\infty \pi(\theta_i | X_i, S_i^2, \hat{\sigma}_{B,i}^2) \pi(\sigma_i^2 | X_i, S_i^2) d\sigma_i^2 = \pi(\theta_i | X_i, S_i^2, \hat{\sigma}_{B,i}^2).$$

Applying this approximation to (2.4), noting the fact that $\theta_i | (X_i, S_i^2, \sigma_i^2) \sim N(M_i X_i + (1 - M_i)\mu, M_i \sigma_i^2)$ where $M_i = \frac{\tau^2}{\tau^2 + \sigma_i^2}$, and substituting M_i by an estimator $\hat{M}_i$ below, we obtain an approximate decision–Bayes interval

$$\begin{aligned} & C_i^{AB} \\ = \ & \Big\{\theta_i : \frac{(\theta_i - \hat{M}_i X_i - (1 - \hat{M}_i)\mu)^2}{\hat{M}_i \hat{\sigma}_{B,i}^2} < -2 \ln k\sqrt{2\pi} - \ln \hat{M}_i\Big\} \end{aligned} \tag{2.7}$$

and $\hat{M}_i = \frac{\tau^2}{\tau^2 + \hat{\sigma}_{B,i}^2}$.

One problem is that k is still unspecified. It seems reasonable to choose k so that the above derivation corresponding to a least favorable prior with $\tau = \infty$ and

$\tau_v = \infty$ (and hence $\hat{M}_i = M_v = 1$) yields the t-interval. This would put C^{AB} on "the equal footing" with the corresponding t interval and one can expect that C^{AB} would have similar coverage probabilities. A similar logic was followed in [5] although they used the less transparent concept of minimaxity. The traditional $1-\alpha$ t-interval has the form $|\theta_i - X_i| \leq tS_i$, where t is chosen to be the $\alpha/2$ upper critical value of a t distribution. The coverage probability of the t-interval under model (2.1) is not exactly $1-\alpha$, but is equal to

$$P(|Z_1|/\exp[(m+\sigma_{ch}Z_2)/2] \leq t), \text{ where} \tag{2.8}$$

$$Z_1 \text{ and } Z_2 \text{ are independent standard normal random variables.} \tag{2.9}$$

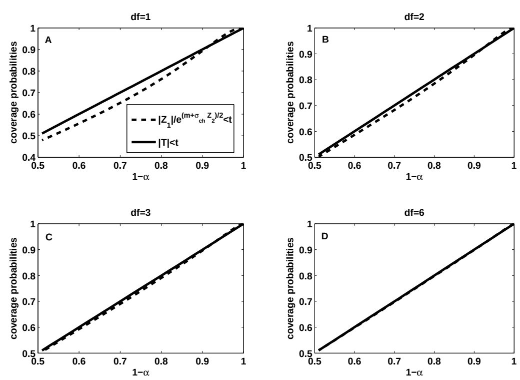

Figure 2.1: (A)-(D): The simulated coverage probabilities of the t-interval $P(|\theta_i - X_i| \leq tS_i)$ are plotted against the nominal level $1-\alpha$ for various degrees of freedom $d = 1, 2, 3$ and 6. The solid 45 degree line in each of the four panels plots these probabilities under the t distribution whereas the dashed curve plots the corresponding probabilities under the model (2.1). Calculations based on 1,000,000 simulations show that the probabilities are close to $1-\alpha$. These probabilities converge to $1-\alpha$ as $d \to \infty$.

Graphs A-D in Figure 2.1 plot (2.8) against $1-\alpha$ in dashed curves for $d = 1, 2, 3$ and 6, which show that when $d \geq 3$, (2.8) is very close to $1-\alpha$, represented by the solid 45 degrees lines. There is no need to show a larger d, since as d increases, (2.8) gets closer to $1-\alpha$ and can be proved to converge to $1-\alpha$ as $d \to \infty$. Even for $d=1$ or 2, these graphs show that (2.8) is greater than or equal to $1-\alpha$, when $1-\alpha \geq 0.9$. Hence the traditional t-interval has good coverage probabilities in many situations including those depicted above even under model (2.1).

Matching the approximate Bayes interval C^{AB} in (2.7) when $M_v = \hat{M}_i = 1$ with the t interval and noting from (2.6) that $\hat{\sigma}^2_{B,i} = S_i^2/e^m$ when $M_v = 1$ require that we set $-2\ln(k\sqrt{2\pi}) = e^m t^2$ or equivalently $k = e^{-t^2 e^m/2}/\sqrt{2\pi}$. Plugging this k into (2.7) leads to the approximate decision–Bayes interval:

$$C_i^{SS} = \{\theta_i : (\theta_i - \hat{M}_i X_i - (1-\hat{M}_i)\mu)^2 < \hat{M}_i \hat{\sigma}^2_{B,i}(t^2 e^m - \ln \hat{M}_i)\}. \tag{2.10}$$

Here SS stands for double shrinkage since in the above interval the center $\hat{M}_i X_i + (1-\hat{M}_i)\mu$ shrinks X_i toward μ and S_i^2 has been replaced by the shrinkage variance estimator $\hat{\sigma}^2_{B,i}$.

We study separately the effects of shrinking the mean (SM) only and of shrinking the variance (SV) only. By setting $\tau^2 = \infty$ and hence $\hat{M}_i = 1$, C_i^{SS} in (2.10) reduces to

$$C_i^{SV} = \{\theta_i : (\theta_i - X_i)^2 < \hat{\sigma}^2_{B,i}(t^2 e^m)\}. \tag{2.11}$$

For SM, we consider

$$C_i^{SM} = \{\theta_i : (\theta_i - \widehat{M_i^0} X_i - (1-\widehat{M_i^0})\mu)^2 < S_i^2 \widehat{M_i^0}(t^2 - \ln(\widehat{M_i^0}))\}, \tag{2.12}$$

where $\widehat{M_i^0} = \frac{\tau^2}{\tau^2 + S_i^2}$. This is identical C_i^{SS} after setting $\tau_v^2 = \infty$ and dropping e^m term and is the interval in [17] with adjustment for the unknown variance case.

2.4 Some Analytic Result

In this section, we derive some analytic properties of the intervals we have constructed. Lemmas 2.4.1 and 2.4.2 show that C_i^{SS} is shorter than C_i^{SV}, which is never longer than the t interval on average. Theorems 2.4.1 and 2.4.2 and numerical evidence show that the coverage probabilities of C_i^{SS} and C_i^{SV} are above the nominal level. The proofs of all the theorems in this paper are given in the Appendix.

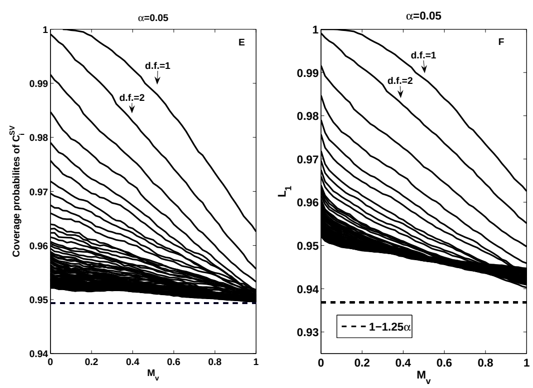

Figure 2.2: (E): The simulated coverage probabilities of C_i^{SV} are plotted against M_v, where M_v runs from 0 to 1 with step size of 0.02. The dotted horizontal line in (E) is $0.95 - 1.65\cdot$(simulation error), the 5% lower bound derived by using the normal approximation of a binomial random variable. This graph shows that the coverage probabilities are above $1-\alpha$. (F): The simulated coverage probabilities L_1 in Theorem 2.4.2 against M_v, where M_v runs from 0 to 1 in step of 0.02. The dotted line is $1 - 1.25\alpha - 1.65\cdot$(simulation error). Hence the graph shows that L_1 is above $1-1.25\alpha$. In both (E) and (F), from top to bottom, the curves correspond to the degrees of freedom from 1 to 100 respectively and the calculations are based on 30,000 simulations.

Theorem 2.4.1 *Under model (2.1), the coverage probabilities of the confidence* *interval* C_i^{SV} *are*

$$P\left((\theta_i - X_i)^2 < \hat{\sigma}_{B,i}^2 t^2 e^m\right) = P(\frac{Z_1^2}{e^{m+\sigma_{ch}\sqrt{M_v}Z_2}} \le t^2) \quad (2.13)$$

where Z_1 *and* Z_2 *are defined in (2.9).*

Remark: Note that when $M_v = 1$, (2.13) is greater than or equal to $1-\alpha$ by the comment after (2.9). For $M_v < 1$, (2.13) can be evaluated by numeral studies since m and σ_{ch} are fixed and the probabilities depend only on the degrees of freedom d, α and M_v, $0 \le M_v \le 1$. Numerical results based on 30,000 simulations show that (2.13) is at least $1-\alpha$ for all M_v varying from 0 to 1 in the step of 0.02, all degrees of freedom within 100, and $\alpha = 0.01, 0.05$ and 0.1. Graph E in Figure 2.2 reports these probabilities for $\alpha = 0.05$. For larger degrees of freedom, t converges to the critical value of Z_1. Also it can be easily seen that as $d \to \infty$, m and σ_{ch} converge to zero. Consequently, (2.13) converges to $1-\alpha$ for all $M_v \in [0,1]$ as also supported by Graph E, where the curves decrease close to $1-\alpha$ as the degrees of freedom increase to 100. In conclusion, the coverage probabilities of the confidence interval C_i^{SV} in (2.11) is greater than or equal to $1-\alpha$ for all practical cases.

Theorem 2.4.2 *Under model (2.1), the coverage probabilities of* C_i^{SS} *satisfy*

$$P(\theta_i \in C_i^{SS}) \ge L_1 \ge L_2, \quad (2.14)$$

where

$$L_1 = P(Z_1^2 \le t^2 e^m \min(e^{\sigma_{ch}\sqrt{M_v}Z_2}, 1)),$$

and

$$L_2 = P(\frac{Z_1^2}{e^{m+\sqrt{M_v}\sigma_{ch}Z_2}} \le t^2) - \frac{1}{2}P(Z_1^2 > t^2 e^m).$$

As $d \to \infty$, L_1 *converges to* $1-\alpha$ *since* m *and* σ_{ch}^2 *converge to 0.*

The lower bounds L_1 and L_2 depend only on two parameters M_v and d and can be extensively examined by simulation. As the remark right under (2.13) stated, (2.13) is higher than $1-\alpha$ for all M_v. Setting $M_v = 1$ implies that $P(Z_1^2 \leq t^2 e^m) \geq 1-\alpha$. Putting these two together show that L_2 is at least $1 - \frac{3}{2}\alpha$. Our numerical studies show that L_1 is at least $1 - 1.25\alpha$ for all degrees of freedom d ranging from 1 to 100, all M_v increasing from 0 to 1 in step of 0.02, and $\alpha = 0.01, 0.05$ and 0.1. Graph F in Figure 2.2 reports only $\alpha = 0.05$ and in such a case $1 - 1.25\alpha = 0.9375$ is very close to the nominal level 0.95. These show that the coverage probabilities of C_i^{SV} and C_i^{SS} are reasonably close to $1-\alpha$ as supported by simulation results to be discussed at the end of this section. Now we compare the half lengths.

Lemma 2.4.1 *Assume*

$$t^2 e^m \geq 1. \tag{2.15}$$

The interval C_i^{SS} is no longer than C_i^{SV} for all observations and is actually shorter with probability one when $\tau^2 > 0$.

Condition (2.15) is weak. For all degrees of freedom, this requires that $1-\alpha$ is at least 68%. The shrink-variances-alone interval C_i^{SV} has shorter expected length than the t interval as shown below.

Lemma 2.4.2 *The expected half-length of the confidence interval C_i^{SV} in each dimension is smaller than that of the t interval, i.e.,*

$$\frac{E\left(\hat{\sigma}_{B,i} t e^{m/2}\right)}{E\left(tS_i\right)} = \exp\left(-\frac{\sigma_{ch}^2}{8}(1+M_v)\right) < 1 \quad \textit{for every } i. \tag{2.16}$$

However, the geometric mean of the lengths of C_i^{SV}'s over the p dimensions is asymptotically equivalent to that of the t intervals.

Lemma 2.4.3 *The geometric mean of the half lengths of the confidence intervals C_i^{SV}'s over the p dimensions is asymptotically equivalent to that of the t intervals as $p \to \infty$.*

The double shrinkage confidence intervals C_i^{SS}'s have smaller length than the t-intervals in both aspects considered in Lemmas 2.4.2 and 2.4.3. We focus on $\hat{M}_i < 1$. Otherwise C_i^{SS} reduces to C_i^{SV} and hence Lemmas 2.4.2 and 2.4.3 apply.

Theorem 2.4.3 *Assume that $\hat{M}_i < 1$ and (2.15) holds. The ratio of the expected half-length of C_i^{SS} to that of the t interval is smaller than the right hand side of (2.16). And the ratio of the geometric mean of the half lengths of C_i^{SS}'s over the p dimensions to that of the t intervals is less than one as $p \to \infty$ if $\tau^2 < \infty$.*

We graphed the coverage probabilities and the expected lengths of the t-interval, C^{SM}, C^{SV} and C^{SS} when the nominal level is .95. Also studied are the "alternative" confidence intervals C^{SV_a} and C^{SS_a}, which are identical to C^{SV} and C^{SS} except that the alternative shrinkage variance estimator, similar to [41]'s proposal, is used to derive the approximate Bayes confidence intervals. The alternative variance estimator is the same as (2.6) except e^m is dropped. Using the alternative variance estimator, we can similarly derive C^{SS_a}, which is the same as (2.10) except e^m is dropped from the length and $\hat{\sigma}^2_{B,i}$ is replaced by the alternative variance estimator in both (2.10) and the definition of $\hat{M}_i$ in (2.7). Similarly, $C_i^{SV_a}$ is C_i^{SV} with the same modifications and hence can also be derived as $C_i^{SS_a}$ with $\hat{M}_i$ taken to be one. Another procedure included is C^{SS_g}, which is derived based on the inverse gamma model. Here g stands for "gamma". This procedure is precisely defined in (2.28) of Section 2.7.

These graphs, similar to Figures 2.3 and 2.4, which deal with unknown hyper-parameter cases, are omitted due to the page limit and are available upon request. For $p = 1000$ and $p = 2000$, and nominal levels 0.95 and $1 - (0.05/p)$, the graphs show that all the six intervals have coverage probabilities no less than the nominal level for various μ, τ, μ_v, τ_v and d. Moreover, the six shrinking procedures have shorter expected lengths than the t-interval, except in a few cases where C^{SS_a}, C^{SV_a}, and C^{SS_g} fail the claim. Except in a few exceptions, C^{SS} has the shortest expected lengths, as anticipated from Theorem 2.4.3 to some extent, and is the best.

Now we discuss the three models studied in the simulations. Model I is the log-normal model (2.1). Model II is the inverse gamma model described in paragraph 2 of Section 2.2, based on which Figures 2.3 and 2.4 are graphed. Model III is a hybrid of Model I and II, and conditioning on σ_i^2, S_i^2/σ_i^2 is assumed to have χ_d^2/d and $log(\sigma_i^2) \sim N(\mu_v, \tau_v^2)$ as in equation (2.1). Simulation results based on the three models all tell the same story. Namely, C^{SS} has coverage probabilities no less than the nominal level, and virtually in all cases, has the shortest expected length. This demonstrate the superiority of C^{SS} over all other intervals with respect to these three models.

2.5 Estimate the Hyperparameters from the Data

In real applications, the hyper-parameters μ and τ, μ_v and τ_v are typically unknown. These parameters need to be estimated as in an empirical Bayes approach. Let $Y_i = \ln(S_i^2) - m$, then $Y_i | \ln\sigma_i^2 \sim N(\ln\sigma_i^2, \sigma_{ch}^2)$. Since $\ln\sigma_i^2 \sim N(\mu_v, \tau_v^2)$, then

$$E(Y_i) = E(\ln\sigma_i^2) = \mu_v, \quad \text{and} \quad E\left[1 - \frac{(p-3)\sigma_{ch}^2}{\sum_i (Y_i - \bar{Y}.)^2}\right] = M_v,$$

where $\bar{Y}. = \sum_i^p Y_i/p$. The above equations suggest two unbiased estimators for μ_v and M_v. After truncation to make sure the estimator for M_v is non-negative, we write the estimators as

$$\hat{\mu}_v = \bar{Y}., \quad \hat{M}_v = \left[1 - \frac{(p-3)\sigma_{ch}^2}{\sum_i (Y_i - \bar{Y}.)^2}\right]_+,$$

where we define $a_+ = max(a, 0)$ throughout the paper. Plugging these two estimators into (2.6) yields the variance estimator proposed by [10]

$$\hat{\sigma}_{EB,i}^2 = \exp(\hat{M}_v Y_i + (1 - \hat{M}_v)\hat{\mu}_v) = (\frac{S_i^2}{e^m})^{\hat{M}_v} e^{\hat{\mu}_v(1-\hat{M}_v)}. \tag{2.17}$$

The variance estimator proposed by [41] is (2.17) with e^m being omitted.

Since

$$X_i|\sigma_i^2 \sim N(\mu, \sigma_i^2 + \tau^2), \text{ for } i = 1, \cdots, p, \tag{2.18}$$

we use the weighted average estimator for μ

$$\hat{\mu} = \sum_i \frac{X_i/\hat{\sigma}_{EB,i}^2}{\sum_i 1/\hat{\sigma}_{EB,i}^2}. \tag{2.19}$$

Since $E(\frac{\sum_i [(X_i-\mu)^2 - \sigma_i^2]}{p}|\sigma_1^2, \cdots, \sigma_p^2) = \tau^2$, and by replacing σ_i^2 by $\hat{\sigma}_{EB,i}^2$ and μ by $\hat{\mu}$, we derive the estimators of τ^2 and M_i as

$$\hat{\tau}^2 = (\sum_i \left[(X_i - \hat{\mu})^2 - \hat{\sigma}_{EB,i}^2\right]/p)_+, \text{ and } \hat{M}_i^{EB} = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}_{EB,i}^2}. \tag{2.20}$$

Plugging these estimators for the hyper-parameters into (2.10) leads to the resultant double shrinkage (SS) empirical Bayes confidence interval for θ_i

$$\{\theta_i : (\theta_i - \hat{M}_i^{EB} X_i - (1 - \hat{M}_i^{EB})\hat{\mu})^2 < \hat{M}_i^{EB} \hat{\sigma}_{EB,i}^2 (t^2 e^m - \ln \hat{M}_i^{EB})\}. \tag{2.21}$$

The other four forms of empirical Bayes confidence intervals (SM, SV, SV_a and SS_a) can be obtained similarly by plugging these estimators for the hyper-parameters into the corresponding approximate Bayes intervals. In particular, SV_a and SS_a are the same as SV and SS except Tong and Wang's variance estimator is used.

2.6 Truncate Both the Length and the Center

In (2.20), it is possible that $\hat{\tau} = 0$ implying $\hat{M}_i^{EB} = 0$ and hence the length in (2.20) is zero, resulting in poor coverage probabilities. To improve on them, it is necessary to truncate $\hat{\tau}$, or equivalently to replace $\hat{\tau}$ by

$$\hat{\tau}_T = \max(\hat{\tau}, \tau_0), \tag{2.22}$$

where τ_0 is a positive number to be specified. As in [36], we would choose τ_0 to make the probability of zero length to be as small as α, i.e.,

$$P_{\tau_0}(\hat{\tau}^2 = 0) = P_{\tau_0}(\sum \left[(X_i - \hat{\mu})^2 - \hat{\sigma}_{EB,i}^2\right] \geq 0) = \alpha, \tag{2.23}$$

where P_{τ_0} denotes the probability corresponding to $\tau = \tau_0$. We shall find τ_0 such that the approximate formula of (2.23), i.e., (2.23) with $\hat{\mu}$ and $\hat{\sigma}_{EB,i}$ replaced by μ and σ_i is solved. Unlike in [36], there is no exact solution to the approximated (2.23). Hence we apply the central limit theorem as $p \to \infty$. It is straightforward to show that the mean and the variance of $(X_i - \mu)^2 - \sigma_i^2$ equal τ_0^2 and $2(\sigma_i^2 + \tau_0^2)^2$. These and the central limit theorem yield a solution of τ_0 to the approximated (2.23) in terms of σ_i, which after being replaced by $\hat{\sigma}_{EB,i}$ gives

$$\tau_0^2 = \frac{2z^2 \sum_i \hat{\sigma}_{EB,i}^2 + z\sqrt{4z^2(\sum_i \hat{\sigma}_{EB,i}^2)^2 + 2(p^2 - 2pz^2)\sum_i \hat{\sigma}_{EB,i}^4}}{p^2 - 2pz^2}. \tag{2.24}$$

We applied this truncation first to the length of (2.21) only and found that the coverage probabilities could drop to about 0.75, significantly smaller than the nominal level 0.95 (when $d = 2$, $p = 1000$ and $\tau_v = 1$). The end of the Appendix has some analytical calculation to explain this. The explanation also gives an insight why truncating the center is necessary. That is, we need to replace not only the length but all the four $\hat{M}_i^{EB}$ terms in (2.21) by its truncation version

$$\hat{M}_i^{EB,T} = \frac{\hat{\tau}_T^2}{\hat{\tau}_T^2 + \hat{\sigma}_{EB,i}^2}. \tag{2.25}$$

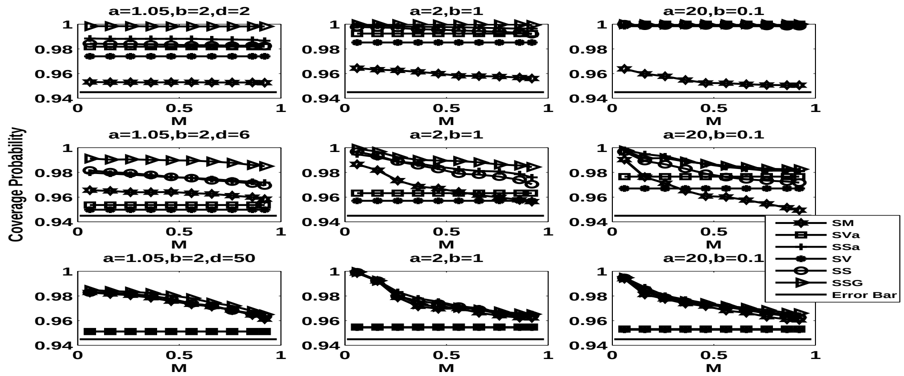

Figure 2.3: The coverage probabilities of the six confidence intervals, $\hat{C}^{SM}, \hat{C}^{SV}, \hat{C}^{SV_a}, \hat{C}^{SS_a}, \hat{C}^{SS_g}$ and $\hat{C}^{SS}$ for θ_1, are plotted against M for $p = 1000$ and $1-\alpha = .95$ under the inverse gamma model (Model II) for various combinations of the parameter a, b and d. Each row corresponds to different degrees of freedom d and each column corresponds to different values of (a, b). Their coverage probabilities are shown to be at least .95 or higher, since they are all above the solid line, which represents the simulation 5% confidence lower bound constructed using the binomial approximation with $p = .95$, the targeted confidence level. Similar graphs based on model I and model III for both $p = 1000$ and $p = 2000$ also demonstrate that these intervals' coverage probabilities are about or above the nominal level.

The resultant double shrinkage confidence interval is denoted as $\hat{C}_i^{SS}$, i.e.,

$$\hat{C}_i^{SS} = (2.21) \text{ with all four } \hat{M}_i^{EB} \text{ replaced by } \hat{M}_i^{EB,T}. \tag{2.26}$$

This truncation is applied to all the shrinkage confidence intervals reported below. In each procedure, the truncation value τ_0^2 is (2.24) with $\hat{\sigma}_{EB,i}^2$ replaced by the corresponding variance estimator for that procedure. The resultant empirical Bayes intervals with truncation for θ_i are similarly denoted by $\hat{C}_i^{SM}$, $\hat{C}_i^{SV}$, $\hat{C}_i^{SV_a}$ and $\hat{C}_i^{SS_a}$.

The numerical simulations reported in Figure 2.3 plot the coverage probabilities of all the six empirical Bayes intervals for θ_1 at the nominal level 0.95, which demonstrates that these intervals all have coverage probabilities above the nominal level

$1 - \alpha = 0.95$ similar to their counter-parts in Section 2.4. Also Figure 2.4 shows that the expected half–lengths of the six empirical Bayes intervals are shorter than those of the t-interval except in a few cases involving $\hat{C}_i^{SS_a}$, $\hat{C}_i^{SV_a}$, and $\hat{C}_i^{SS_g}$. The double shrinkage confidence interval $\hat{C}_i^{SS}$ is always shorter than the t-interval in all situations and is the shortest in virtually all situations. This is the recommended interval of this paper.

The simulation model comments in the last paragraph of Section 2.4 apply here. The figures report calculations based on Model II, the inverse gamma model. Numerical studies not reported here show that the conclusion in the last paragraph about the superiority of $\hat{C}_i^{SS}$ remains true even with two other models, models I and III, discussed at the end of Section 2.4. The conclusion is also numerically established, although not reported, for $p = 1000$ and 2000 and the nominal level $1 - \alpha/p$.

Following the referees' suggestions, we now summarize the definition of $\hat{C}_i^{SS}$ for θ_i in (2.26) with coverage probabilities numerically shown to be above $1 - \alpha$. We first recount the definition of (2.21). Note that X_i's are point estimators of θ_i's (e.g. X_i's are the ANOVA estimators of θ_i's) and t is the $\alpha/2$ upper critical point of a t-distribution. Further $m, \hat{\mu}$ and $\hat{\sigma}^2_{EB}$ in (2.21) are defined in (2.2), (2.19) and (2.20). Finally, $\hat{C}_i^{SS}$ is a modification of (2.21) in that each of the four $\hat{M}_i^{EB}$'s is replaced by the truncation version $\hat{M}_i^{EB,T}$, which is defined in (2.22) through (2.25) with z being the α upper critical point of a standard normal random variable.

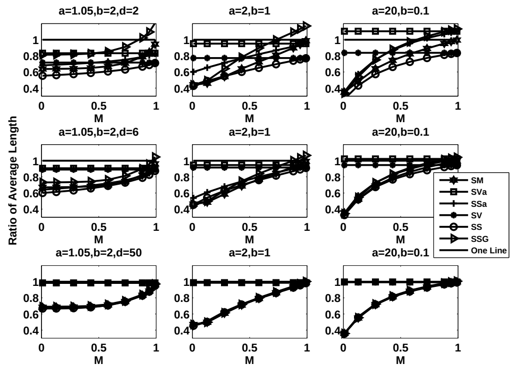

Figure 2.4: Under the same setting as in Figure 3, the ratios of average lengths of six intervals to the t-interval are plotted for various combinations of parameters a, b and d. Each row corresponds to different degrees of freedom d and each column corresponds to different values of (a, b). Similar graphs were plotted based on models I and III, with $p = 1000$ and 2000, which all lead to the same conclusion that $\hat{C}_i^{SS}$ has coverage probabilities above $1 - \alpha$ along with expected lengths less than the t-interval. With only a few exceptions, $\hat{C}_i^{SS}$ performs the best among all the intervals considered in the figure.

2.7 Results for the Inverse Gamma Model

In Section 2.2, we depicted a model called the inverse gamma model, which is the same as model (2.1) except that S_i^2 and σ_i^2 are assumed to have the following distributions:

$$\text{Conditioning on } \sigma_i^2, (S_i^2/\sigma_i^2) \sim \chi_d^2/d \text{ and } \sigma_i^2 \sim \textit{inverse gamma}(a, b). \quad (2.27)$$

Hence $(\sigma_i^2)^{-1}$ has a gamma distribution with parameters a and b. See [3](p. 561). Consequently, $\sigma_i^2|S_i^2 \sim \textit{inverse gamma}(a', b')$, where $a' = d/2 + a$ and $b' = (1/b + dS_i^2/2)^{-1}$. We may then approximate σ_i^2 by

$$\hat{\sigma}_i^2 = E(\sigma_i^2|S_i^2) = ((a-1)b)^{-1} = (d/2 + a - 1)^{-1}(1/b + dS_i^2/2).$$

Following the similar argument leading to (2.7), we derive an approximate decision-Bayes interval that is identical to (2.7) except that $\hat{\sigma}^2_{B,i}$ is replaced by $\hat{\sigma}_i^2$ and $\hat{M}_i$ by $\tau^2/(\tau^2 + \hat{\sigma}_i^2)$. To choose k, consider the least informative prior as $\tau \to \infty$, $a \to 1$ and $b \to \infty$ (which implies that $\hat{\sigma}_i^2 \to S_i^2$ and $\hat{M}_i \to 1$), and choose k so that the resultant confidence interval becomes the t-interval. Hence we need to choose k so that $-2\ln(k\sqrt{2\pi}) = t^2$, leading to the interval

$$C_i^{SS_g} = \{\theta_i : (\theta_i - \hat{M}_i X_i - (1 - \hat{M}_i)\mu)^2 < \hat{M}_i \hat{\sigma}_i^2 (t^2 - \ln \hat{M}_i)\} \tag{2.28}$$

Here the subscript g stands for "gamma". Parallel to Theorem 2.4.2, we have the following Theorem 2.7.1, whose proof is omitted.

Theorem 2.7.1 *Under the Inverse Gamma model (2.27), the coverage probabilities of the approximate decision-Bayes confidence interval (2.28) satisfies*

$$P(\theta_i \in C_i^{SS_g}) \geq P(T^2_{d+2a} < t^2) - \alpha,$$

where t is chosen such that $P(T_d^2 < t^2) = 1 - \alpha$ and T_d denotes a t random variable with d degrees of freedom.

The numerical evidence shows that $P(T^2_{d+2a} < t^2) \geq P(T_d^2 < t^2) \geq 1 - \alpha$ when $a > 0$. Therefore, the lower bound for the coverage probabilities of (2.28) is $1 - 2\alpha$ when the nominal level is $1 - \alpha$.

To construct a useful interval in practice, namely, the empirical Bayes confidence interval out of C^{SSg}, we need to estimate the hyper-parameters τ^2, μ, a and b. The estimation of τ^2 and μ can follow what was done between (2.18) and (2.20). However, we would also need to do the truncation depicted in Sections 2.5 and 2.6. One small problem is the estimation of a and b, which can be done using only S_i^2. The maximum likelihood estimators of a and b have been proposed in [45]. Method of moments based on $\ln(S_i^2)$ solving just one equation numerically has been proposed in [39]. Applying Smyth's technique lead to a confidence interval denoted by $\hat{C}^{SS_g}$. One may expect that $\hat{C}^{SS_g}$ should perform best under the inverse gamma model. Numerical study indicates surprisingly that $\hat{C}^{SS}$ still perform better than $\hat{C}^{SS_g}$ in that the former has shorter average length and coverage probabilities above $1 - \alpha$. See Figures 2.3 and 2.4. The improvement is substantially especially when d is small, say smaller than 7. Searching for the explanation for the superiority of $\hat{C}^{SS}$ over C^{SSg} would be an interesting future project. However, two of the authors (Qiu and Zhao) have independently programmed and have arrive the same conclusion that $\hat{C}^{SSg}$ may become longer than t-interval in expected length.

2.8 Data Analysis

We apply our confidence interval procedures to an Affymetrix control data set, the golden spike-in data set of [8]. The most striking feature of this data set is that all the true parameters are pre-chosen and known and hence the data presents an opportunity that statisticians rarely have to check against any proposed procedures. For the details of the data, see the paper above. It seems interesting to apply the shrinkage intervals to this data set to demonstrate the advantage of shrinkage. For a more convincing data analysis, one would need to apply $\hat{C}_i^{SS}$ to an individual subgroup (containing the ith gene) that has expression levels homogeneous in means and variances. We take a naïve approach to apply $\hat{C}_i^{SS}$ to the whole group. We download data from the website http://www.elwood9.net/spike, which are already processed by using MAS background correction, the constant subset normalization at the probe level, PM adjustment for nonspecific signal by MAS (v5), MAS (v5) for expression summary, and finally LOWESS normalization at the

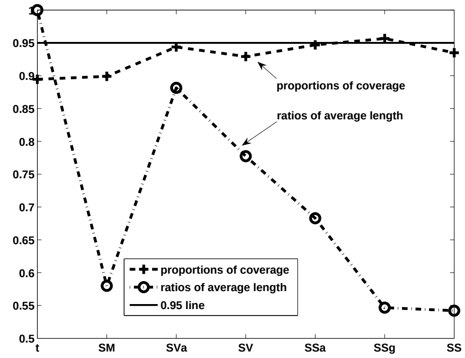

Figure 2.5: This figure plots in the upper curve the proportions of coverage across the genes of seven confidence intervals with $1 - \alpha = .95$, indicated by the horizontal axis, after being applied to the spike-in data set of [8]. The lower curve plots the ratios of the average lengths over the genes of these intervals to that of the t–interval. The recommended confidence interval $\hat{C}^{SS}$ has the proportion of coverage approximately .95. And as demonstrated in the upper curve, it has average length only 55% of that of the t–interval. The interval $\hat{C}^{SS_g}$ works similarly to, but slightly better than $\hat{C}^{SS}$.

probe set level. After taking log_2 transformation, the data of size 14010-by-6 are fit to 14010 gene-specific one-way ANOVA models where $p = 14010$ corresponds to the number of genes and six replicates include three from each of the control and treatment groups. The residual plots show that the variances of the control and treatment groups are quite different from each other for most of genes. Hence we construct the t interval for each of the genes (see, e.g., [32]) using Satterwaite approximation, $X_i \pm tS_i$, $i = 1, \cdots, p$, where

$$X_i = \overline{Y}_{1i} - \overline{Y}_{2i}, \quad S_i = \sqrt{s_{1i}^2/n_1 + s_{2i}^2/n_2} \text{ and } \quad d = \frac{(s_{1i}^2/n_1 + s_{2i}^2/n_2)^2}{\frac{(s_{1i}^2/n_1)^2}{n_1-1} + \frac{(s_{2i}^2/n_2)^2}{n_2-1}}.$$

Also $\overline{Y}_{1i}$ and $\overline{Y}_{2i}$ are the sample means of the control and treatment groups for the ith gene, each based on $n_1 = n_2 = 3$ samples, and s_{1i}^2 and s_{2i}^2 are the corresponding unbiased sample variances. The degrees of freedom d for all genes after omitting the decimal places are either 2 or 3. We use $d = 2$ for all genes to be conservative.

It would be interesting to generalize our procedure to the case where the degrees of freedom vary according to different genes. We also apply the six empirical Bayes confidence intervals to the data and plot the proportions of coverage and average lengths relative to that of the t interval in Figure 2.5. The proportion of coverage can be viewed as an approximation of the coverage probability since $p = 14010$ is large.

The proportions of coverage are plotted in "+" symbols and the average lengths relative to that of the t interval are plotted in "O" symbols. The t interval has low proportion of coverage of about 0.89 at the nominal level 0.95. The shrink-means-alone interval $\hat{C}^{SM}$ reduces the average length but does not improve on the proportion of coverage. The shrink-variances-alone interval with Tong and Wang's modified variance estimator $\hat{C}^{SV_a}$ improves the proportion of coverage but only reduces the average length to about 88%. The double shrinkage confidence intervals $\hat{C}^{SS_a}$, $\hat{C}^{SSg}$ and $\hat{C}^{SS}$ have good proportion of coverage and reduce the average lengths to about 68%, 55% and 54%, respectively. The intervals $\hat{C}^{SS}$ and $\hat{C}^{SSg}$ seem the most attractive. Also $\hat{C}^{SSg}$ has a slightly better proportion of coverage.

2.9 Conclusion

In this paper, we construct empirical Bayes confidence intervals that shrink both the means and the variances. These intervals are on average much shorter than the t-intervals and have higher coverage probabilities both in simulations and in real data. They are better than the shrink-means-alone or shrink-variances-alone confidence intervals. We made parametric assumption in this paper. However, by bootstrapping the statistics (as in [15]), we would likely have intervals that can be applied to a more nonparametric setting. The fact that $\hat{C}^{SS}$ is explicitly defined and can be computed instantaneously will facilitate the Bootstrap approach. In this paper, we set the nominal level $1-\alpha$ for a single interval to be 95%. In the situation of the multiple inference, there is a need to consider nominal level to be higher than 95%. For the nominal level $1-0.05/p$ and for $p=1000$ and $p=2000$, the numerical studies show that $\hat{C}^{SS}$ always perform better than the t-interval, with largest saving in expected length ranging from 40% to 60%. Except in a few exceptions, $\hat{C}^{SS}$ always has the shortest expected lengths among all the shrinkage intervals considered. In all cases, $\hat{C}^{SS}$ has coverage probabilities above the nominal level.

However, this interval construction only aims at one individual. As soon as simultaneous confidence interval is concerned, we can only apply Bonferroni's correction at this stage. We will address this multiplicity in the next chapter.

CHAPTER 3

EMPIRICAL BAYES FCR CONTROLLING CONFIDENCE INTERVALS

3.1 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.

[2], 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 the confidence intervals for the parameters corresponding to these selected genes, what should be done? This is the question raised in [2]. Their proposed criterion is to examine the FCR, which is the average rate of false coverage (i.e. not covering θ_i) among the selected intervals. They demonstrate 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 is not controlled. They then constructed confidence intervals that have a controlled FCR.

In the approach of [2], the FCR is defined in the frequentist's sense and is required to be less than q for every 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 the FCR over such a distribution of θ_i's. 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 distributions is considered instead and we require, for every distribution in the class, that the Bayes FCR of confidence intervals is controlled to be less than or equal to q. In such a case, we say that the confidence intervals have *empirical Bayes FCR* controlled at the level q.

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 3.2, 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. In section 3.3, we apply the theorem in section 3.2 to a class of prior distributions. We establish that under certain settings, 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 [33] and [34] are used. The asymptotic property holds regardless of the selection rule. In section 3.4, we construct the empirical Bayes confidence intervals which are numerically shown to have the empirical Bayes FCR controlled when p is finite. Moreover, the empirical Bayes intervals are always shorter in average length than and could be one third as long as the [2]'s intervals. Some general conclusions are written in section 3.5.

3.2 General Theorem on Bayes Intervals

We begin by giving the definition of *False Coverage Rate(FCR)* of confidence intervals, a term coined in [2]. Consider one-dimensional parameters θ_i, $i = 1, \cdots, p$. Assume that X_i is an estimator of θ_i. 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 [2] is

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

where the expectation integrates out X, under the assumption that θ_i's are fixed. [2] 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 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 θ_i's with respect to their distribution π and define the Bayes FCR as

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

In this paper, we aim at constructing confidence intervals such that the Bayes FCR is controlled by some pre-specified level q for any selection rule and any $\pi \in \Pi$ for some set Π. We call such a procedure an *empirical Bayes FCR-controlling intervals.*

Note that, in agreement with the finding in [2] for the frequentist FCR, the classical 95% t-intervals have Bayes FCR much larger than 5% and fail drastically to control it at 5% level as demonstrated in Figures 3.1, 3.3, and 3.5 by black dotted lines. This is due to the fact that these parameters have been preselected - they are declared to be significantly different than zero when applying Benjamini and Hochberg'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 $\pi(\theta)$.

Lemma 3.2.1 *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 lemma and all the other theorems below are given in the Appendix unless it is obvious from the context. Given the lemma, the following theorem is obvious.

Theorem 3.2.1 *If* $P(\theta_i \notin CI_i|X) \leq q$, $\forall i$, *then* $FCR_\pi \leq qP(R > 0) \leq q$, *for any selection rule based on* X *leading to* R.

In both Lemma 3.2.1 and Theorem 3.2.1 above, there is no independent assumption needed among $\{X_i\}$ or among $\{\theta_i\}$. 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.

Assume that the observations X_i's follow a distribution as $X_i \sim N(\theta_i, \sigma_i^2)$, $i = 1, 2, \cdots, p$ where σ_i^2 are known constants. Assume that the (prior) distribution of θ_i's is $N(\mu, \tau^2)$. We assume that (X_i, θ_i), $1 \leq i \leq p$, are independent.

Theorem 3.2.2 *Define the confidence interval* CI_i^B *as*

$$CI_i^B = [M_i X_i + (1 - M_i)\mu] \pm z\sqrt{M_i}\sigma_i, \tag{3.1}$$

where $P(|Z| > z) = q$.

Having posterior coverage probability $1 - q$, the above procedure, by Theorem 3.2.1, controls the Bayes FCR at the q-level for any selection rule.

Theorem 3.2.1 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. If we assume a class of priors, indexed by some hyper-parameters, it seems reasonable to use data to estimate them as in the empirical Bayes approach. However, the resultant intervals no longer satisfy the assumption in Theorem 3.2.1 because of the estimation error. The following theorems provide us tools to show the asymptotic properties.

Theorem 3.2.3 *Assume that* $\max_{1 \leq i \leq p} P(\theta_i \notin CI_i | X) = \alpha(p, X)$ *and*

$$\lim_{p \to \infty} P(\alpha(p, X) \leq q + \epsilon) \to 1, \forall \epsilon > 0. \tag{3.2}$$

Then

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

When condition (3.2) 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 3.2.4 *Assume that $\frac{R}{p} \to \eta > 0$, and*

$$\frac{1}{p}\sum_i |P(\theta \notin CI_i|X) - q| \to 0, \text{ almost surely,} \tag{3.3}$$

where q is any number independent of i. Then

$$\lim_{p\to\infty} FCR_\pi = f(q).$$

We also have the version of the theorem for the one-sided case.

Theorem 3.2.5 *Assume that all other assumptions except (3.3) is replaced by*

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

where for a number a, $(a)_+$ stands for the positive part of a and equals $\max(a, 0)$. *Then $\limsup_{p\to\infty} FCR_\pi \le f(q)$.*

3.3 Empirical Bayes Approach

In section 3.2, we have showed that the confidence intervals can typically control the Bayes FCR. However, these results are for the case when a precise prior is used. In real application, we do not know the prior and it seems more reasonable to consider a class of priors with unknown hyper-parameters which should be estimated by using data as in the empirical Bayes approach.

Recall that $X_i|\theta_i \sim N(\theta_i, \sigma_i^2)$, and $\theta_i \sim N(\mu, \tau^2)$; therefore

$$EX_i = E(EX_i|\theta_i) = E(\theta_i) = \mu, \text{ and}$$

$$EX_i^2 = E(E(X_i^2|\theta_i)) = E(\theta_i^2 + \sigma_i^2) = \mu^2 + \tau^2 + \sigma_i^2.$$

Now we estimate μ by $\hat{\mu} = \bar{X}$, and τ^2 by

$$\hat{\tau}^2 = \left(\frac{\sum_{i=1}^p (X_i^2 - \sigma_i^2)}{p} - \hat{\mu}^2\right)_+, \tag{3.4}$$

where we take the positive part in (3.4) to ensure that the estimator is not negative. Naturally, 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 (3.1) by their estimators above results in the so-called empirical Bayesian interval

$$CI_i^{EB} = [\hat{M}_i X_i + (1 - \hat{M}_i)\hat{\mu}] \pm z\sqrt{\hat{M}_i}\sigma_i. \tag{3.5}$$

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 (3.5) would behave like the Bayes procedure (3.1), having the asymptotic Bayes FCR controlled. This indeed can be proved as in the two theorems below under the assumption that $\tau > 0$.

Theorem 3.3.1 *For any $\epsilon > 0$, if $\sum_{i=1}^p \sigma_i^4 = o(\frac{p^2}{(\log p)^{1+\epsilon}})$, then*

$$\limsup_{p \to \infty} FCR \leq q, \forall \pi.$$

Alternatively, an application of theorem 3.2.4 provides us the asymptotic property under a less restrictive condition when the number of selection R increases as p increases.

Theorem 3.3.2 *If* $\frac{R}{p} \to \eta > 0$*, and* $\sum_{i=1}^{p} \sigma_i^4 = o(p^2)$*, then* $\lim_{p\to\infty} FCR_\pi = q$.

Both conditions on the order of $\sum_i \sigma_i^4$ are mild, much weaker than the result from the law of large number. More specifically, when σ_i^2's are generated as samples from a population with the mean and variance both finite, these two conditions are satisfied. Both Theorems 3.3.1 and 3.3.2 under their assumptions imply that the empirical Bayesian interval (3.5) controls the FCR_π at the q-level for any normal prior π when $p \to \infty$. However, when p is finite, the FCR_π can be higher than q. In figures 3.1, 3.3, and 3.5, we simulate the FCR_π when the dimension is $p = 2000$ with $q = 5\%$. Cyan solid lines represent the FCR_π of the interval (3.5), and demonstrate the failure of controlling FCR_π especially when τ^2 is close to zero. From the point of view of Theorem 3.2.1, this is likely due to the fact that (3.5) does not have the empirical Bayes non-coverage probabilities controlled to be less than 5%. [34], [5], and [17] for the non-selecting setting, and [36] for the selecting setting, have discovered the similar phenomenon and they all have applied the truncation to boost the coverage probability. We will address such modifications in the next section which leads to a controlled FCR even for moderate p.

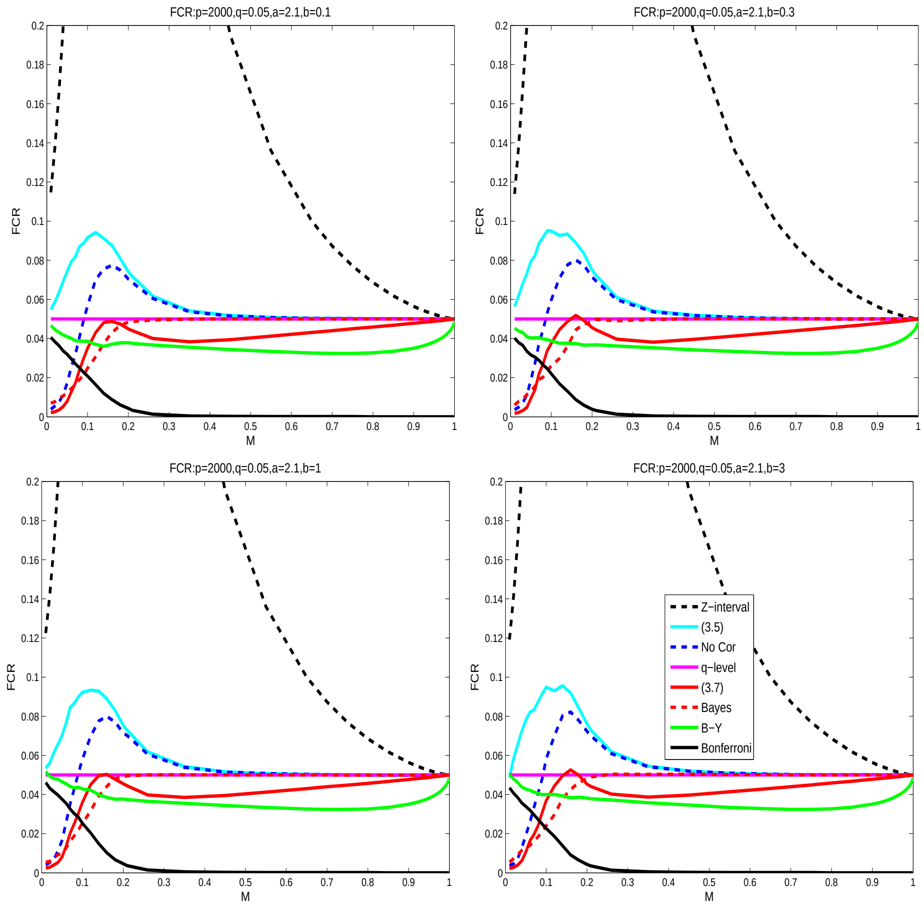

Figure 3.1: The Bayesian FCR of different interval constructions are plotted against $M = \frac{\tau^2}{1+\tau^2}$ for $p = 2000$ and $q = 0.05$ under the model Normal-Normal Model when assuming the unequal but known variances. The variances are sampled independently from the inverse gamma random variable for various combinations of parameters a and b. The a is chosen to be 2.1 in this figure, and $2.5, 3$ respectively in figure 3.3 and 3.5. The b varies among $0.1, 0.3, 1$, and 3, corresponding to the four pictures above. The parameters are selected according to [1]'s FDR procedure at 5% level. The naïve t-interval fails to control the FCR at the q-level; [2]'s procedure, Bonferroni correction, and our empirical decision Bayes confidence intervals (3.7) all control Bayes FCR at q-level. However, the empirical Bayes intervals having no correction or no truncation fail to control the FCR when τ^2 is moderately small.

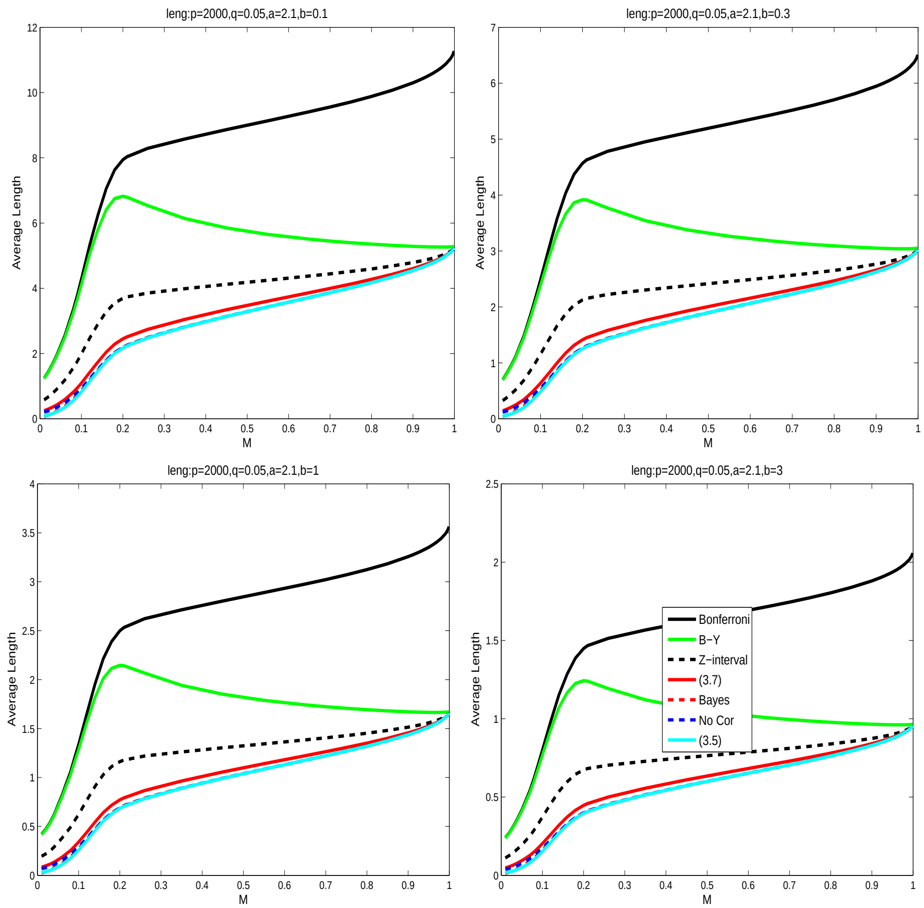

Figure 3.2: Under the same setting as in figure 3.3, the average length of intervals that are constructed are plotted for various combinations of parameters a and b. It is shown that the empirical decision Bayes intervals (3.7) enjoys huge reduction of the average length. The price paid for the truncation and correction which ensure that the FCR of (3.7) is controlled, is small. The average length that corresponds to Bonferroni's correction is way too large due to its extreme conservativeness. The average half-length of [2]'s procedure is uniformly larger than ours and could go up to three times larger.

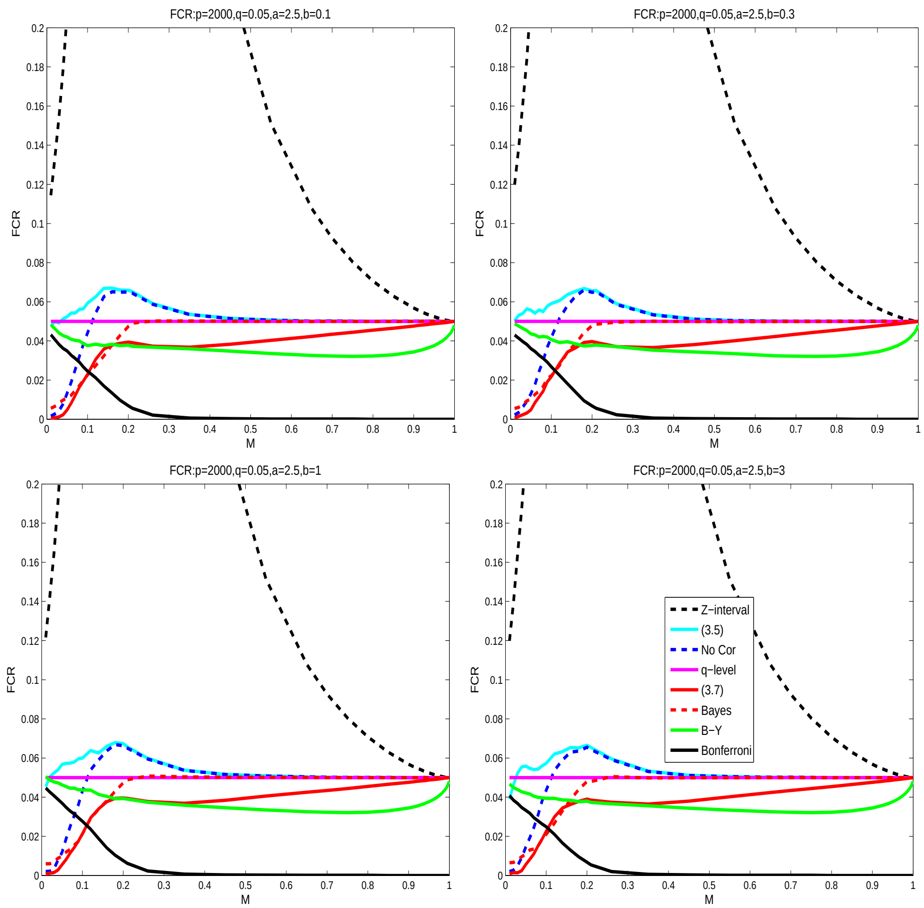

Figure 3.3: This simulation setting is the same as that of the figure 3.1 except that a is chosen to be 2.5.

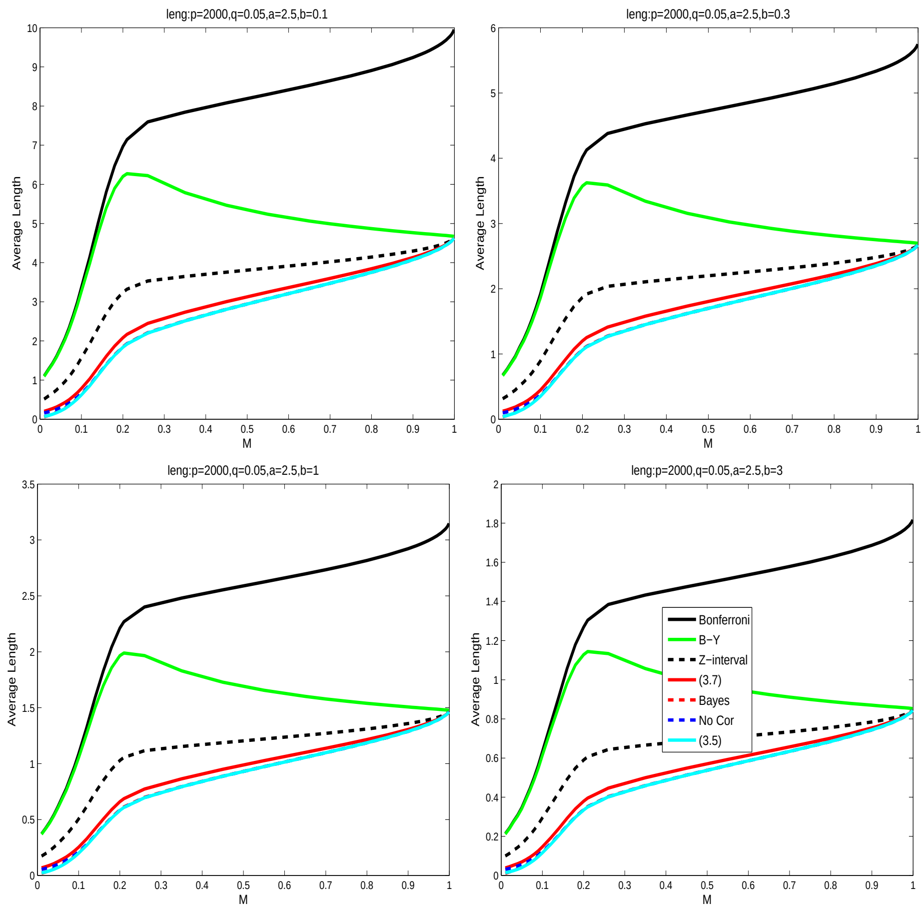

Figure 3.4: This simulation setting is the same as that of the figure 3.2 except that a is chosen to be 2.5.

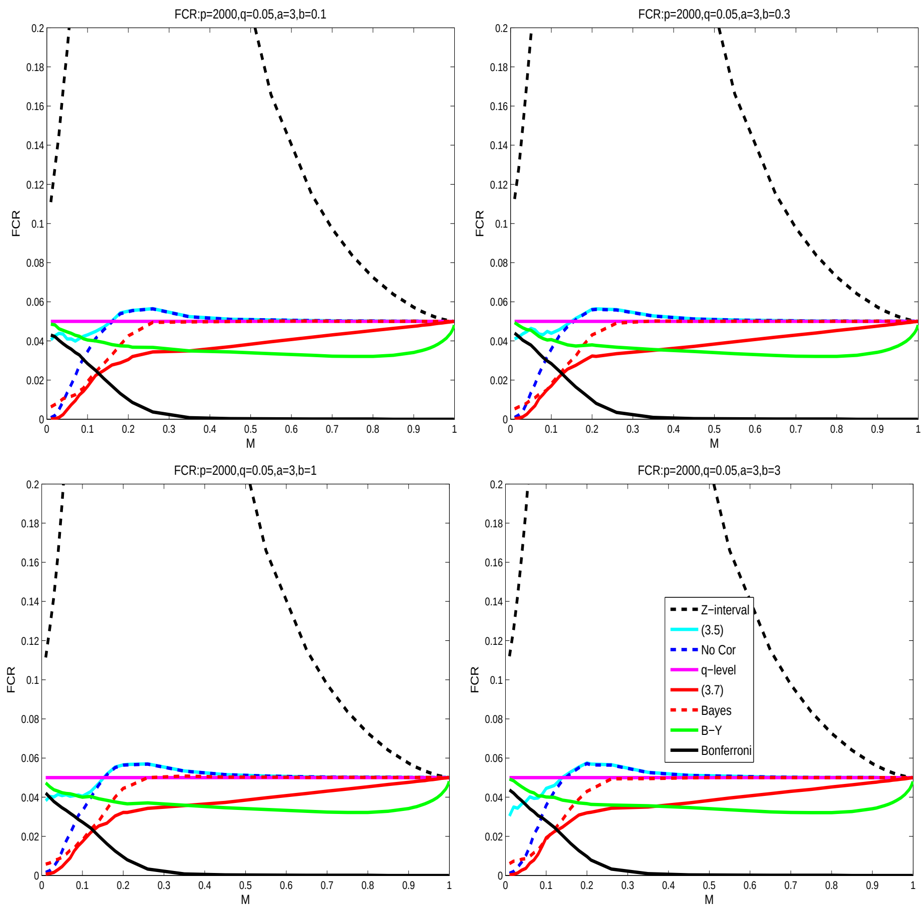

Figure 3.5: This simulation setting is the same as that of the figure 3.1 except that a is chosen to be 3.

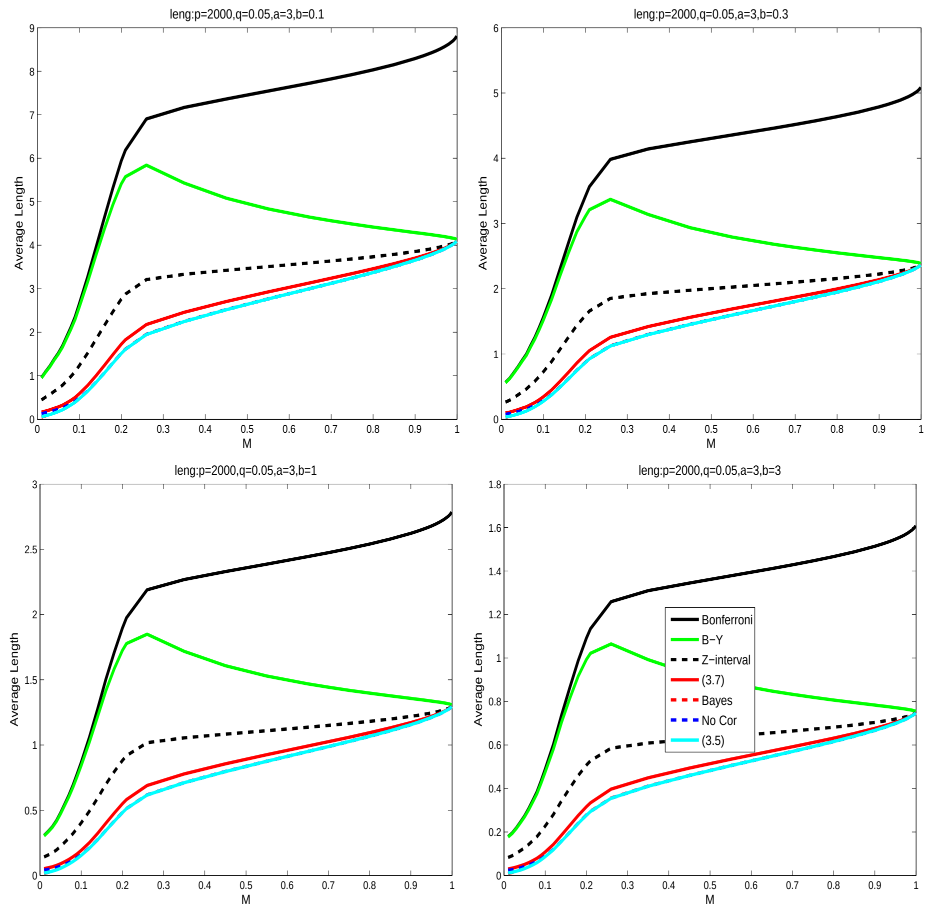

Figure 3.6: This simulation setting is the same as that of the figure 3.2 except that a is chosen to be 3.

3.4 Modification

Adapting the arguments in [36] and [24] to this case, we truncate the estimator $\hat{\tau}^2$ at τ_0^2 where τ_0^2 satisfies

$$P_{\tau^2=\tau_0^2}(\hat{\tau}^2 = 0) \leq q.$$

Let $\bar{\sigma}^2 = \frac{\sum_i \sigma_i^2}{p}$, $\bar{\sigma}^4 = \frac{\sum_i \sigma_i^4}{p}$. The central limit theorem yields an approximation of τ_0^2 as

$$\tau_0^2 = \frac{2z^2\bar{\sigma}^2 + \sqrt{4z^4\bar{\sigma}^4 + 2z^2\bar{\sigma}^4(p - 2z^2)}}{p - 2z^2}.$$

Now, replacing $\hat{\tau}^2$ by $\hat{\tau}_*^2 = max(\hat{\tau}^2, \tau_0^2)$ and M_i in the half length of (3.5) by $\hat{M}_i^*$ where $\hat{M}_i^* = \frac{\hat{\tau}_*^2}{\hat{\tau}_*^2 + \sigma_i^2}$, we obtain the following interval

$$CI_i^* = [\hat{M}_i X_i + (1 - \hat{M}_i)\hat{\mu}] \pm z\sqrt{\hat{M}_i^*}\sigma_i. \tag{3.6}$$

In figures 3.1, 3.3, and 3.5, blue dotted lines represent the FCR_π corresponding to the empirical interval (3.6) for different parameter settings. It can be easily seen that interval (3.6) performs significantly better than (3.5) especially when τ^2 is small. However, the intervals (3.6) still fail to control the FCR_π especially when τ^2 is moderately small.

As in [5], [17], and [36], we apply another correction to the length as could be derived by decision Bayes approach. Let $V(\hat{M}_i^*) = \sqrt{z^2 - log(\hat{M}_i^*)}$. The empirical decision Bayes interval is defined as

$$CI_i^F = [\hat{M}_i X_i + (1 - \hat{M}_i)\hat{\mu}] \pm \sqrt{\hat{M}_i^*} V(\hat{M}_i^*)\sigma_i. \tag{3.7}$$

The intervals (3.6) and (3.7) are wider than (3.5) which has FCR_π controlled asymptotically. Consequently, the FCR_π of (3.6) and (3.7) are both smaller than

q asymptotically as long as the conditions on σ_i's in Theorems 3.2.4 and 3.2.5 in Section 3.2 are satisfied.

Red solid lines in figures 3.1, 3.3, and 3.5 graph the Bayes FCR of (3.7), demonstrating that their FCR_π is less or equal than $q = 0.05$ for any τ^2. The variances σ_i^2's are simulated from inverse gamma distribution with parameters a and b. For various choices of a, b, and p, our simulations always show that the recommended interval (3.7) has FCR_π controlled at $q = 0.05$.

In these three figures, the FCR_π of [2]'s procedure are plotted by green solid lines. Note that [2]'s procedure is analytically proved to control the frequentist FCR. Hence it is not surprising that their Bayes FCR is always controlled, as demonstrated by our figures as well. However, they are very conservative especially when there are many θ_i's. The price for the [2]'s confidence intervals to pay is that they are very long. Figure 3.2, 3.4, and 3.6 show that their average lengths, although shorter than the Bonferroni's intervals, are longer than other intervals, including (3.7). In some situations, [2]'s average length is three times as long as (3.7). Note the (3.7) has average length almost identical to the Bayes intervals, which have the minimum lengths. However, the Bayes intervals assume the knowledge of τ^2, unrealistic in real applications.

Interestingly, the empirical Bayes interval (3.7) does not adjust the level q which affects directly its length. Its length is even shorter than the t-interval. This may call into question as to whether the empirical Bayes FCR criterion requires the multiplicity correction. There are two reasons that indicate it does require the multiplicity correction in some sense. First of all, the traditional t-interval still fail terribly, having a large Bayes FCR as demonstrated in figures 3.1, 3.3, and 3.5. Hence the empirical Bayes criterion do not wash away the selection bias. Secondly, the empirical Bayes intervals do adjust the answer when p is different, since then $X_1, \cdots, X_p$ would all be different. The empirical Bayes borrows the strength from all $X_1, \cdots, X_p$ to do the "multiplicity" adjustment.

Fundamentally, the empirical Bayesian approach does such an amazing job that it bases its interval on a center (a point estimator) whose selection bias has been corrected. See [19]. As a result, there is no need to make the length longer. In contrast, the [2]'s interval does not correct the center and expand its length to have a good FCR. Thus the length becomes very large in order to cover the bias of its center.

3.5 Conclusion

In this paper, we propose controlling the empirical Bayes FCR as a criterion alternative to the (frequentist) FCR proposed by [2]. 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 prior, or the empirical Bayes FCR, we derive sharper confidence intervals.

We constructed empirical Bayes intervals under the normal-normal model with confidence coefficient $1-q$ and demonstrated that the intervals have the empirical Bayes FCR controlled at level q. The classical frequentist t-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.

However, in microarray experiments, it is well known that for most genes, say more than 90%, the corresponding differential expression is identically zero (see [36]). Therefore, it is important to construct confidence intervals when accounting for such information. We will combine the decision approach and the *empirical Bayes* approach to deal with such a more practical situation in the next chapter.

CHAPTER 4

DECISION APPROACH AND EMPIRICAL BAYES FCR-CONTROLLING INTERVAL FOR MIXED PRIOR MODEL

4.1 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 the selection bias (see [36] and [19] 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 of selected populations (see, for example, [9] and [19]). However, only a few confidence intervals have been constructed for selected populations. One exciting work is [36], which offers a way to construct intervals that can control the 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 with an unknown mean and variance and a single point *zero*. Because they control the simultaneous coverage probability, their criterion is more stringent than the FCR discussed below. Their intervals are much shorter than the intervals constructed using Bonferroni's method.

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 non-coverage coefficient. They constructed confidence intervals for multiple selected parameters which can control the FCR at a specified q-level, typically 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 can be substantially improved compared to the intervals we shall propose.

Later, [46] introduced the Bayes FCR and connected Bayes confidence interval which aims to control Bayesian non-coverage coefficients with the Bayesian FCR controlling intervals. They applied this general theorem to the normal-normal setting where the observations follow a normal distribution with unequal but known variances and the parameters 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 because they addressed the multiplicity by reducing the bias of the point estimator, the center of their intervals. The result is reported in Chapter 3 of this dissertation.

Here, we use the decision approach and empirical Bayes to construct intervals for selected populations under the same model setting of [36]. Application of decision approach to interval/set estimation has a long history which dates back to [16], [5], and [17]. Recently, [24] have constructed the double shrinkage empirical confidence interval for a one dimensional parameter when assuming the variances to be unequal and unknown. The result is reported in Chapter 2 of this dissertation. However, the loss functions they have used need adjustment for the mixed model we consider here (detailed argument is in section 4.2.2). Thus a new loss function with two tuning parameters k_1 and k_2 is proposed. One specific choice of k_2 results in [36]'s procedure. 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 4.2, we introduce the model setting and the decision Bayes rule based on our new loss function. In section 4.3, we will connect the decision Bayesian rule with [36]'s procedure first and then derive a procedure which can control the Bayes FCR. In section 4.4, empirical Bayesian approach is constructed and evaluated both numerically and analytically. In section 4.5, we apply the confidence intervals constructed in section 4.4 to a real microarray data set and compare it with those of [2] and [36]. It turns out that our procedure out-performs theirs. The average length of our interval is only 57% of that of [36]'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 Bayes FCR. However, this seems a more realistic criterion.

4.2 Normal-Mixture Model for the means

4.2.1 Model Assumption

In microarray, it is generally assumed that observed differentially expressed levels X_i's are normally distributed with true means θ_i's, $i = 1, 2, \cdots, p$, where the dimension p is very large. Due to the extremely large number of dimensions, it seems natural for statisticians to model the true means θ_i's. A conventional choice is the normal prior where $\theta_i \overset{i.i.d.}{\sim} N(0, \tau^2)$.

However, in [36], they applied the Q-Q plot to a microarray data and showed that a normal-normal model cannot fit the data well. To remedy this, they introduced the *normal-mixture model* as following. Assume that $X_i|\theta_i \sim N(\theta_i, \sigma^2)$, and

$$\pi(\theta_i) \begin{cases} = 0 & \text{with probability } \pi_0, \\ \sim N(0, \tau^2) & \text{with probability } \pi_1 = 1 - \pi_0. \end{cases} \tag{4.1}$$

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 4.4, we estimate them through data by using consistent estimators and derive an empirical Bayesian procedure.

4.2.2 Bayes Interval

Historically, there have been many attempts to apply the decision Bayes approach to construct confidence sets/intervals. [16] considered a linear loss function for confidence set CI of the parameter θ as $L(\theta, CI) = kVolume(CI) - I_{CI}(\theta)$. Also [5] uses the same loss where the tuning parameter k was determined so the usual $1 - \alpha$ confidence set is minimax. [17] 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. [24] 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 (4.1). 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 added extra terms which influence the loss function only when the point *zero* is included and thus define the loss function as,

$$L(\theta_i, CI_i) = k_1 Len(CI_i) I_i - I_{CI_i}(\theta_i) I_i + I_{CI_i}(0)(k_2 - (1 - I_i)), 0 \leq k_2 \leq 1. \quad (4.2)$$

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*. In such a 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. Thus, appropriate choice of the tuning parameter k_2 guides us to decide when *zero* should be included.

Furthermore, the flexibility of choosing k_2 offers us constructions under different settings. For example, when assuming the normal-normal model, the loss function (4.2) reduces to [17]'s if we set $k_2 = 0$. In section 4.3, we apply two different choices of k_2, one of which will reproduce [36]'s procedure, 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 (4.1) and the loss function

(4.2).

Theorem 4.2.1 *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*

$$EL(\theta_i, CI_i|X) = \pi_i^1(X)\int_{CI_i}(k_1 - \pi(\theta_i|X, I_i = 1))d\theta_i + I_{CI_i}(0|X)(k_2 - \pi_i^0(X)). \quad (4.3)$$

The Bayes interval is

$$CI_i = \begin{cases} \{\theta_i : k_1 < \pi(\theta_i|X_i, I_i = 1)\} \setminus \{0\} & \text{if } k_2 > \pi_i^0(X), \\ \{\theta_i : k_1 < \pi(\theta_i|X_i, I_i = 1)\} \cup \{0\} & \text{if } k_2 \leq \pi_i^0(X). \end{cases} \quad (4.4)$$

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 4.2.1 shows that the decision Bayes interval uses k_2 as the threshold value.

Under model (4.1), $\pi(\theta_i|X, I_i = 1) \sim N(MX_i, M\sigma^2)$ where $M = \frac{\tau^2}{\tau^2+\sigma^2}$, therefore

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

As in the Section 3 of [24], 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 (4.4). Then the decision Bayes interval becomes

$$CI_i^B = \begin{cases} \{\theta_i : (\theta_i - MX_i)^2 < M\sigma^2(z_{q/2}^2 - \log M)\} \setminus \{0\} & \text{if } k_2 > \pi_i^0(X), \\ \{\theta_i : (\theta_i - MX_i)^2 < M\sigma^2(z_{q/2}^2 - \log M)\} \cup \{0\} & \text{if } k_2 \leq \pi_i^0(X). \end{cases} \quad (4.5)$$

Unlike the interval $MX_i \pm \sqrt{M}\sigma z_{q/2}$, which is directly derived from the posterior distribution, (4.5) has an extra positive term $M\sigma^2(-\log M)$ which is necessary to boost the coverage probability when the hyper parameters are estimated through the data in section 4.4. 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.

4.3 Choose k_2

4.3.1 Qiu and Hwang (2007)

[36] constructed the interval for K parameters $\theta_{(p-K+1)}, \cdots, \theta_{(p)}$ under the model (4.1) 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

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

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 $\max\limits_{1 \leq j \leq p} |\theta_{(j)}|$. We construct the interval for $\theta_{(j)}$ where $p - K + 1 \leq j \leq p$ as

$$CI^B_{(j)} = \begin{cases} \{\theta_{(j)} : (\theta_{(j)} - MX_{(j)})^2 < M\sigma^2(z^2_{q/2K} - \log M)\} \setminus \{0\} & \text{if } k_2 > \pi^0_{(j)}(X), \\ \{\theta_{(j)} : (\theta_{(j)} - MX_{(j)})^2 < M\sigma^2(z^2_{q/2K} - \log M)\} \cup \{0\} & \text{if } k_2 \leq \pi^0_{(j)}(X). \end{cases} \tag{4.6}$$

When compared with (4.5), the major difference is that we use the critical value $z_{q/2K}$ to address the multiplicity.

Direct calculation shows that for each j,

$$P(\theta_{(j)} \notin CI^B_{(j)}|X) \leq q/K + \pi^0_{(j)}(X)((\pi^0_{(j)}(X) < k_2) - q/K).$$

Consequently, the simultaneous non-coverage coefficient satisfies

$$P(\theta_{(j)} \notin CI_{(j)}, j = p-K+1, \cdots, p) \leq q + E \sum_{j=p-K+1}^{p} \pi^0_{(j)}(X)(I(\pi^0_{(j)}(X) < k_2) - q/K). \tag{4.7}$$

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

$$k_2 = \arg\max_k \{E \sum_{j=p-K+1}^{p} \pi^0_{(j)}(X)(I(\pi^0_{(j)}(X) < k) - q/K) \leq 0\}, \tag{4.8}$$

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, (4.6) is identical to [36]'s Bayes procedure, hence providing a surprising satisfaction of [36]. This always indicates that the loss function (4.2) is reasonable and useful.

4.3.2 Bayes FCR Controlling Interval

[2] initiated the concept of FCR, which is much less conservative than the simultaneous non-coverage coefficients. [46], as reported in Chapter 3, has extended this idea to the Bayesian framework through a new concept, Bayes FCR. They have shown that there is a natural connection between the Bayes FCR and the Bayes non-coverage probability. In this subsection, we will show that (4.5) can control the Bayes FCR at the q-level if k_2 is chosen appropriately.

Theorem 4.3.1 *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 (4.5) satisfies*

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

In other words, the Bayes FCR of the intervals (4.5) is controlled at the q level.

Now assume that the selection rule in [36] is applied in Theorem 4.3.1, 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 (4.8) 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 4.3.1 is larger than or equal to k_2'. Therefore, the frequency that (4.5) includes *zero* is less than that of [36]. 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 (4.5). The discrepancy becomes large when K is large. These two facts imply that the Bayes FCR controlling interval is less conservative than [36]. However, the construction of [36] could control the 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 (4.5) 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, as shown below.

4.4 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 [46].

Recall the model 4.1. 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,

$$\hat{\pi}_1 = \frac{(m_2 - \sigma^2)^2}{m_4/3 + \sigma^4 - 2\sigma^2 m_2}, \hat{\tau}^2 = \frac{(m_2 - \sigma^2)}{\hat{\pi}_1}. \tag{4.9}$$

Plug these two estimators back to the function of f as in Theorem 4.3.1 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 (4.9). Then we can construct the empirical Bayes interval as,

$$\begin{aligned} & CI_i^{EB} \qquad\qquad (4.10) \\ = & \begin{cases} \{\theta_i : (\theta_i - \hat{M}X_i)^2 < \hat{M}\sigma^2(z_{q/2}^2 - \log \hat{M})\} \setminus \{0\} & \text{if } \hat{k}_2 > \hat{\pi}_i^0(X), \\ \{\theta_i : (\theta_i - \hat{M}X_i)^2 < \hat{M}\sigma^2(z_{q/2}^2 - \log \hat{M})\} \cup \{0\} & \text{if } \hat{k}_2 \le \hat{\pi}_i^0(X). \end{cases} \end{aligned}$$

The following theorem describes the asymptotic property of the construction.

Theorem 4.4.1 *Assume that* $0 < \pi_0 < 1$, $\tau^2 > 0$. *For any* $\epsilon > 0$, *if there always exists* $\delta, N > 0$ *such that*

$$|f(p, \tau'^2, \pi_0', k') - f(p, \tau^2, \pi_0, k)| < \epsilon. \tag{4.11}$$

when given $(\tau'^2 - \tau^2)^2 + (\pi_0' - \pi_0)^2 + (k' - k)^2 < \delta$ *for all* $\forall p > N, k, k^{'} > 0$*, then under the model (4.1), the empirical Bayes interval (4.10) satisfies*

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

Proposition 4.4.1 *If we select the first* R *parameters with* $R \to \infty$ *when* $p \to \infty$*, then* f *satisfies the condition in Theorem 4.4.1.*

This proposition implies that when all observations are selected for interval estimation, (4.10) can control the empirical Bayes FCR asymptotically.

However, like all other existing constructions such as [5], [36], and [24], the interval (4.10) cannot provide a satisfactory answer automatically for the finite sample case.

In figure 4.1, 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 too conservative in the sense that the Bayes FCR is very low when M is close to 1 and that they have a large average length. The green line, corresponding to the construction (4.10), performs well when τ^2 is relatively large; however some modifications are required when τ^2 is small.

[36] has argued that π_0 is nearly unidentifiable when τ is small. This will cause the estimator (4.9) to be very inaccurate. Therefore, they use the Bonferroni's correction $(X_i - z_{q/2p}\sigma, X_i + z_{q/2p}\sigma)$ if $\hat{p}\hat{\tau}^2 < \min(\sqrt{720/p}, 0.6)$, a threshold obtained from extensive numerical calculations. It also seems necessary to mix the procedure

(4.10) with the interval $(X_i - z_{Rq/2p}\sigma, X_i + z_{Rq/2p}\sigma)$, which is inspired by [2]. Below, we have an analytic argument that helps us to find the proposed threshold value.

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, which is controlled at the level of q. Therefore, set the threshold value τ_0^2 to satisfy $P_{\tau^2=\tau_0^2}(\frac{m_4/3-\sigma^4}{m_2-\sigma^2} \leq 2\sigma^2) \leq q$.

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

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

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

$$CI_i^{Final} = \begin{cases} X_i \pm z_{Rq/(2p)}\sigma & \text{if } m_2 - \sigma^2 < \tau_0^2, \\ CI_i^{EB}, & \text{if } m_2 - \sigma^2 > \tau_0^2. \end{cases} \tag{4.12}$$

In figure 4.1, the red solid line corresponds to the above empirical Bayes intervals. They perform the same as [2] when τ^2 is very small because of the mixed procedure. The portion of the mixture increases when π_0 increases. However, (4.12) 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 4.2 that corresponds to the same model settings in figure 4.1. The average length of (4.12) 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 4.3 and 4.4, 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 (4.12) 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%.

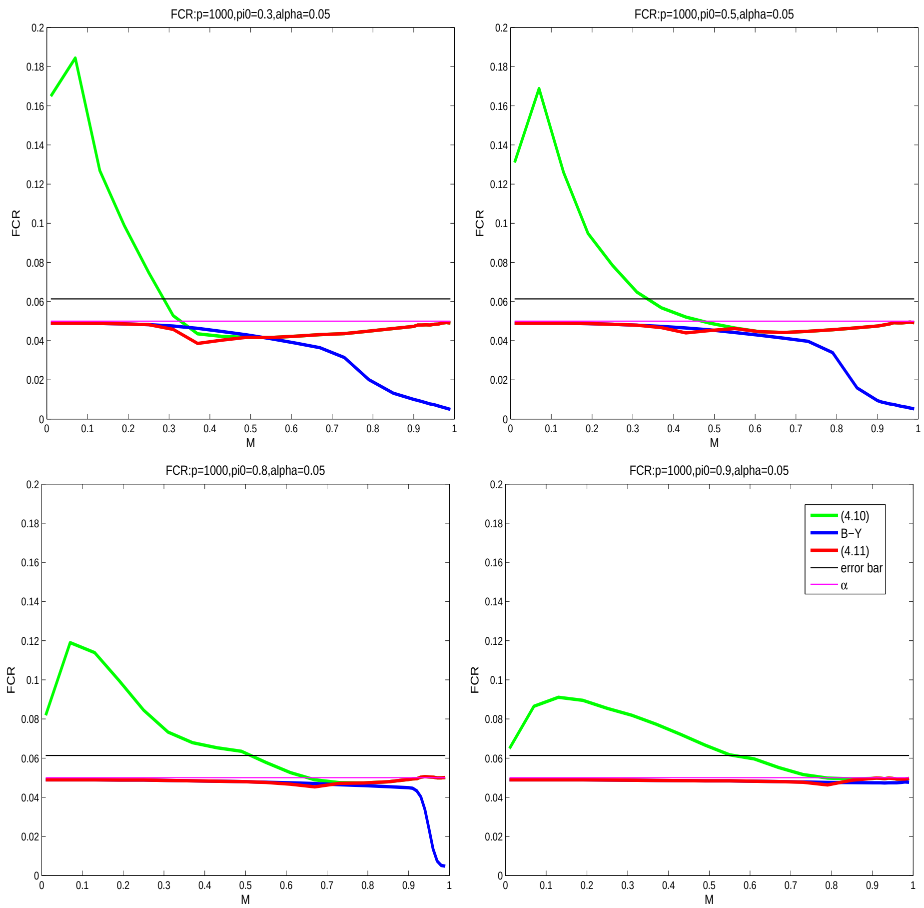

Figure 4.1: These figures are the simulated Bayes FCR under different model settings against $M = \frac{\tau^2}{1+\tau^2}$. The dimension is set to be 1000, and top 100 observations after ordering all X_i's according to their magnitude are selected for confidence interval construction. The hyper parameter 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, (4.10) doesn't control the Bayes FCR at 5%. However, the mixed procedure (4.12) does control the Bayes FCR for any hyper parameters. The portion of the mixture increases as π_0 increases.

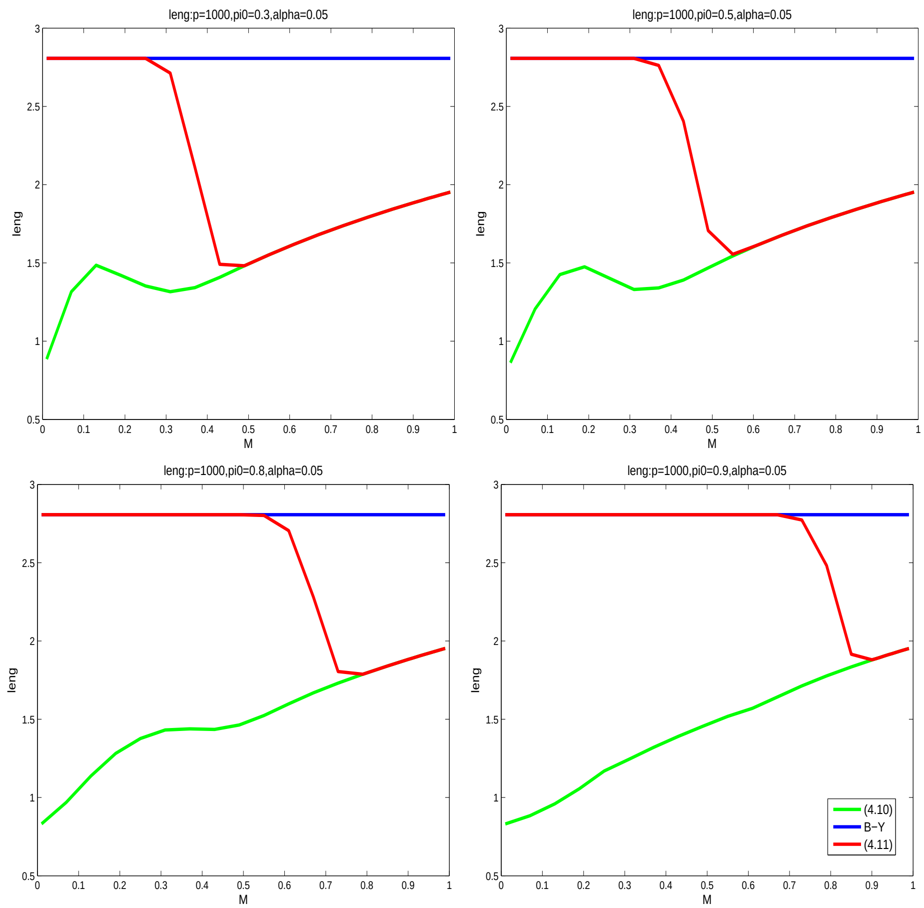

Figure 4.2: These figures are the simulated average length of different approaches under the same model setting as figure 4.1. The average length of our procedure is less than or equal to [2]'s procedure. In some extreme cases, the average length of (4.12) is only 54% of that of [2]'s procedure.

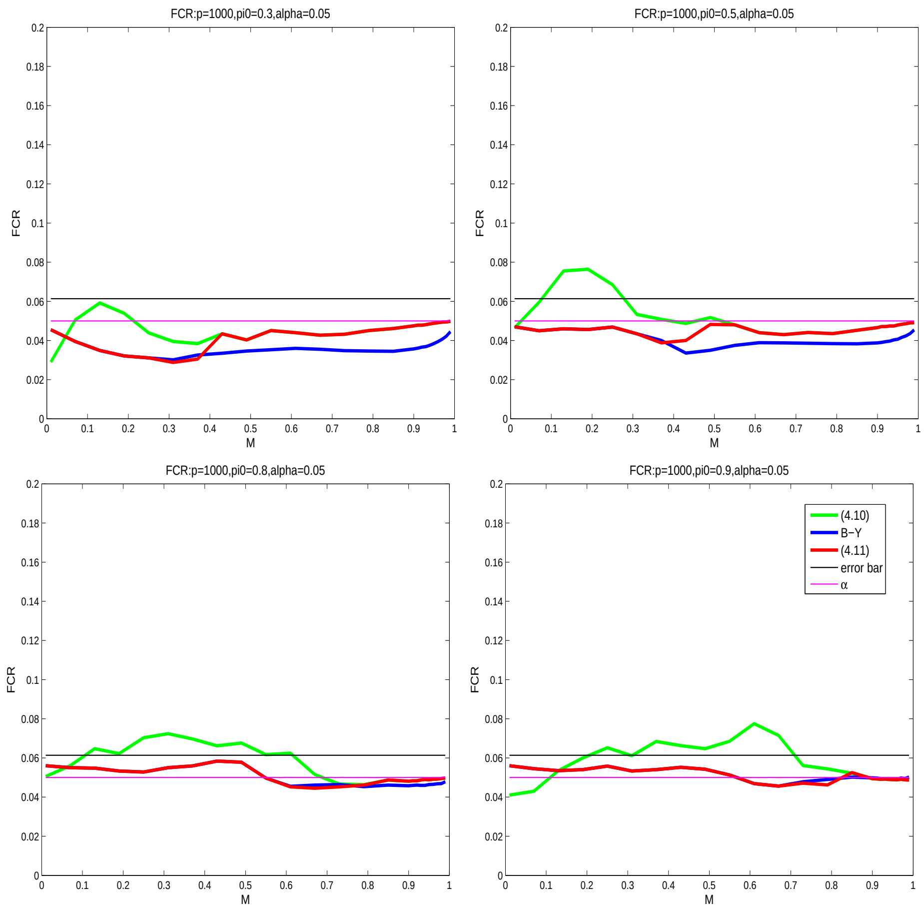

Figure 4.3: These figures are the simulated Bayes FCR under different model settings against $M = \frac{\tau^2}{1+\tau^2}$. The dimension is set to be 1000. The selection rule is based on [1] which aims at controlling the False Discovery Rate to be less or equal than 5%. The hyper parameter π_0 varies 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, (4.10) doesn't control the Bayes FCR. However, the Bayesian FCR of the mixed procedure (4.12) and [2]'s procedure are always less than or equal to the error bar, which equals to q plus the simulation error.

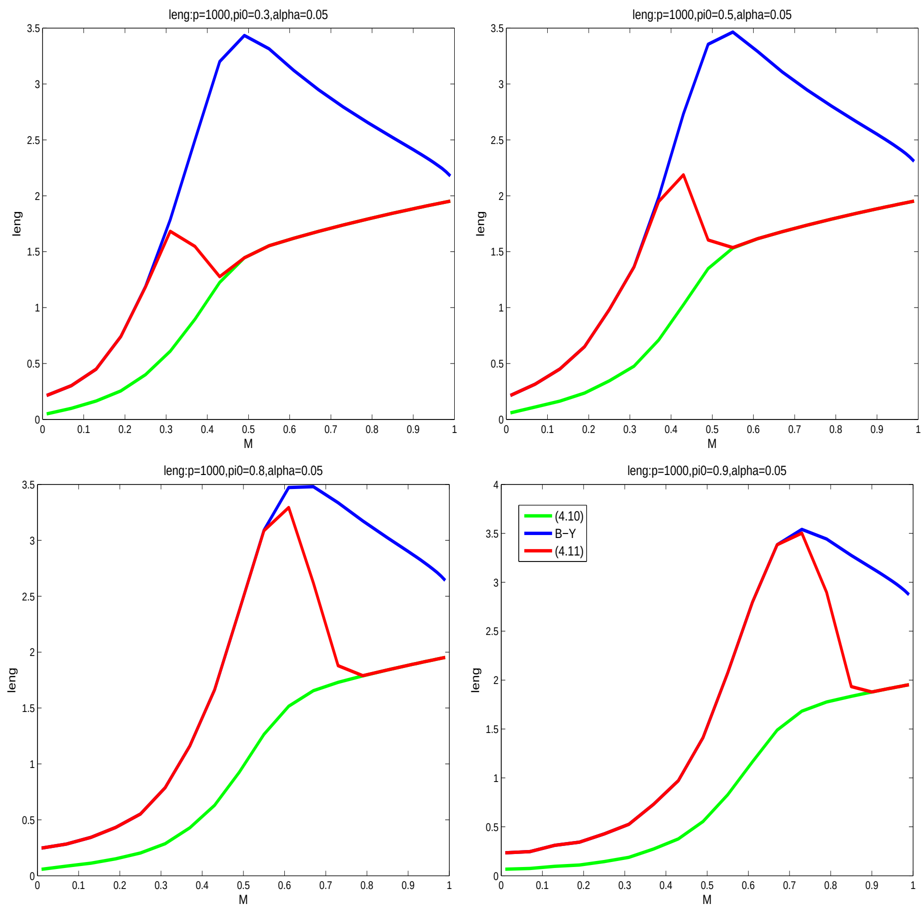

Figure 4.4: These figures are the simulated average length of different approaches under the same model as figure 4.1. The average length of our procedure is less than [2]'s procedure. In some extreme cases, the average length of (4.12) is only 44% of that of [2]'s procedure.

4.5 Real Data Analysis

In this section, we apply different intervals to a microarray data set, the Synteni data of [28], which was revisited by [18] and [36]. The description of the data set can be found in [28]. Figure 6 of [36] is a Q-Q plot of the ANOVA estimator X_g, which shows that the normal-mixture model (4.1) fits the data well.

In [18], 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 4.5, we construct the confidence intervals for these 89 genes by using [36], [2], and (4.12). Our confidence interval (4.12) 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 [36], our intervals enjoy great length reduction.

4.6 Discussion

In this chapter, we have defined a new loss function for confidence interval construction when assuming the mixed prior model (4.1). We use two different ways to choose the tuning parameter in the loss function to obtain [36]'s procedure and the empirical Bayesian FCR controlling intervals. Since [36] controls the simultaneous coverage coefficient by using Bonferroni's correction, their lengths are much larger than (4.12) where we aim at controlling the *empirical Bayes* FCR.

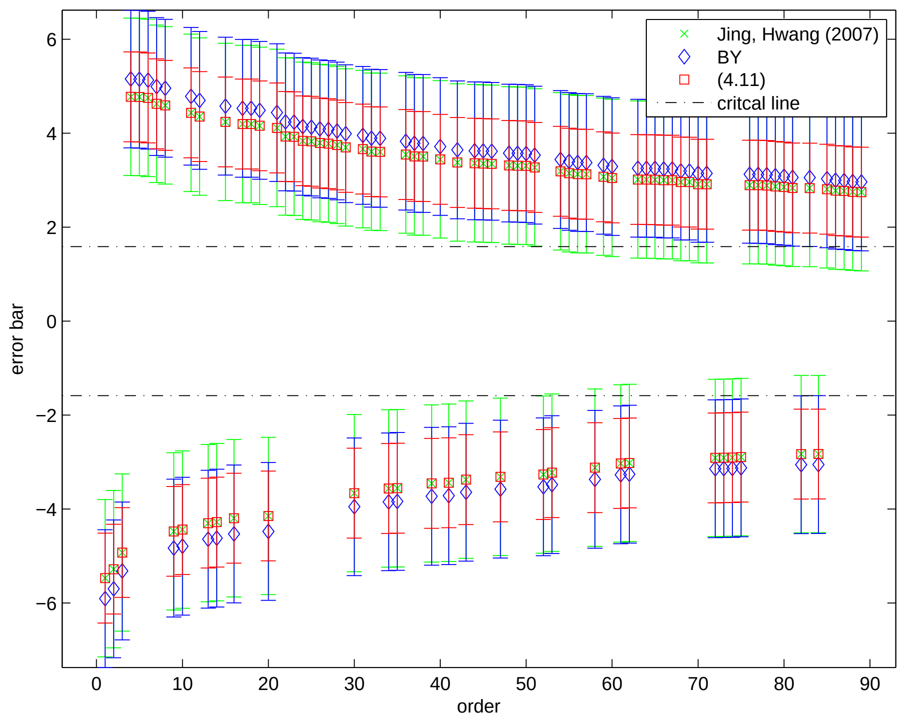

Figure 4.5: Three different interval approaches, [36], [2], and (4.12) are applied to the Synteni data of [28]. The FDR procedure of [1], aiming at finding the genes with differentially expressed levels that are significantly larger than or equal to $\log_2 3$ while controlling the False Discovery Rate 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 [36] is the same as in (4.12). 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 (4.12). [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 (4.12). However, the length of the procedure from [2] is slightly smaller than that of the procedure in [36].

However, there is still much need for further research. In model (4.1), we assume equal and known variance σ^2. In many applications, σ^2 are unknown and unequal. [24] 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 (4.2) provides us with a potential tool to construct corresponding intervals.

APPENDIX A

TECHNICAL PROOF

Proof of Theorem 2.4.1:

Under model (2.1), it is easy to see that (2.13) equals

$$P(\sigma_i^2 Z_1^2 \leq \hat{\sigma}_{B,i}^2 t^2 e^m) = P(Z_1^2 \leq \frac{\hat{\sigma}_{B,i}^2}{\sigma_i^2} t^2 e^m),$$

where the conditional distribution of $Z_1 = (X_i - \theta_i)/\sigma_i$ is $N(0,1)$, not depending on σ_i and hence Z_1 is statistically independent of σ_i. Replacing $\hat{\sigma}_{B,i}^2$ by (2.6) gives

$$\begin{aligned} \ln(\frac{\hat{\sigma}_{B,i}^2}{\sigma_i^2}) &= M_v(\ln S_i^2 - m) + u_v(1 - M_v) - \ln \sigma_i^2 \\ &= M_v(\ln \sigma_i^2 + \eta_i) + \mu_v(1 - M_v) - \ln \sigma_i^2 = M_v \eta_i - (1 - M_v)(\ln \sigma_i^2 - \mu_v), \end{aligned}$$

where we introduce the new notation $\eta_i = \delta_i - m$, which has a $N(0, \sigma_{ch}^2)$ and is independent of σ_i. Noting $\ln(\sigma_i^2) - \mu_v$ has a $N(0, \sigma_{ch}^2)$ and combing the two independent normal random variables, we may write

$$\ln(\frac{\hat{\sigma}_{B,i}^2}{\sigma_i^2}) = \sigma_{ch} \sqrt{M_v} Z_2, \tag{A.1}$$

because of the identity $M_v^2 \sigma_{ch}^2 + \tau_v^2 (1 - M_v)^2 = \sigma_{ch}^2 M_v$. Here Z_2 is a standard normal random variable. Also since both η_i and σ_i are independent of Z_1 under model (2.1), $\ln(\frac{\hat{\sigma}_{B,i}^2}{\sigma_i^2})$ and Z_2 are independent of Z_1. Putting all these together establishes (2.13).

We need Lemma A.0.1 below for proving Theorem 2.4.2.

Lemma A.0.1 *The function*

$$Q(\tau^2) = \frac{\hat{M}_i \sigma_i^2}{(M_i - \hat{M}_i)^2 (\sigma_i^2 + \tau^2) + M_i \sigma_i^2},$$

is bounded between $\min(1, \frac{\sigma_i^2}{\hat{\sigma}_{B,i}^2})$ *and* $\max(1, \frac{\sigma_i^2}{\hat{\sigma}_{B,i}^2})$.

Proof of Lemma A.0.1: Direct substitution and a little algebra show that

$$Q(x) = \frac{(x+y)(x+y')y}{x(y'-y)^2 + y(x+y')^2},$$

where $x = \tau^2, y = \sigma_i^2$ and $y' = \hat{\sigma}_{B,i}^2$. It is easy to see that

$$\lim_{x \to 0} Q(x) = y/y', \quad \text{and} \quad \lim_{x \to \infty} Q(x) = 1.$$

With a little algebra, the derivative of $Q(x)$ can be shown to equal $\frac{(y'-y)yy'(x+y)^2}{(x(y'-y)^2+y(x+y')^2)^2}$. Hence Q is increasing (or decreasing) in x if $y' > y$ (or $y' < y$). Hence whether $y' > y$ or not, $Q(x)$ is between 1 and y/y', concluding the lemma.

Proof of Theorem 2.4.2: To establish the first inequality, it suffices to prove that

$$P((\theta_i - \hat{M}_i X_i - (1 - \hat{M}_i)\mu)^2 \leq \hat{M}_i \hat{\sigma}_{B,i}^2 (t^2 e^m)) \geq L_1. \tag{A.2}$$

The point is that $-\ln(\hat{M}_i) \geq 0$ can be dropped from (2.14) or (2.10). Classical Bayesian theory calculations show that

$$\theta_i | X_i, \sigma_i^2 \sim N(M_i X_i + (1 - M_i)\mu, M_i \sigma_i^2),$$

$$Z_i^0 \equiv \frac{\theta_i - M_i X_i - (1 - M_i)\mu}{\sigma_i \sqrt{M_i}} | X_i, \sigma_i^2 \sim N(0, 1).$$

Let $Z_i = (X_i - \mu)/\sqrt{\sigma_i^2 + \tau^2}$. Since the conditional distribution of Z_i^0 given X_i and σ_i is $N(0, 1)$, not depending on σ_i and X_i and hence Z_i^0 is statistically independent of X_i and σ_i and hence independent of Z_i. Similar argument shows that conditioning on σ_i, $Z_i \sim N(0, 1)$, independent of σ_i^2. One can solve θ_i and X_i in terms of Z_i and Z_i^0 in the last two displayed equations. The solutions can replace θ_i and X_i in (A.2) establishing

$$\begin{aligned} & \theta_i - \hat{M}_i X_i - (1 - \hat{M}_i)\mu \\ = \; & \sigma\sqrt{M}_i Z_i^0 + (M_i - \hat{M}_i)(\mu + \sqrt{\sigma_i^2 + \tau^2} Z_i) - (M_i - \hat{M}_i)\mu. \end{aligned} \tag{A.3}$$

Conditioning on σ_i^2 and S_i^2, M_i and $\hat{M}_i$ are constants and the only random variables above are the independent standard normal variables Z_i^0 and Z_i. Simple evaluation shows that (A.3) has $N(0, \sigma_i^2 M_i + (\sigma_i^2 + \tau^2)(M_i - \hat{M}_i)^2)$. Now the left hand side of (A.2) equals

$$P\left(\left[\sigma_i^2 M_i + (\sigma_i^2 + \tau^2)(M_i - \hat{M}_i)^2\right] Z_1^2 \leq \hat{M}_i \hat{\sigma}_{B,i}^2 (t^2 e^m)\right) = P\left(\frac{\sigma_i^2 Z_1^2}{\hat{\sigma}_{B,i}^2 t^2 e^m} \leq Q(\tau^2)\right)$$

By Lemma A.0.1, $Q(\tau^2)$ is bounded below by $\min(1, \frac{\sigma_i^2}{\hat{\sigma}_{B,i}^2})$. Substituting $Q(\tau^2)$ by this lower bound, the last displayed equation is bounded below by

$$P\left(\frac{Z_1^2 \sigma_i^2}{\hat{\sigma}_{B,i}^2 t^2 e^m} \leq \min(1, \frac{\sigma_i^2}{\hat{\sigma}_{B,i}^2})\right) = P\left(Z_1^2 \leq t^2 e^m \min(1, \frac{\hat{\sigma}_{B,i}^2}{\sigma_i^2})\right).$$

This and (A.1) imply (A.2), establishing the first inequality. To prove the second inequality, $L_1 \geq L_2$, note that

$$\begin{aligned} L_1 &= P(\frac{Z_1^2}{e^{m+\sqrt{M_v}\sigma_{ch} Z_2}} \leq t^2, Z_2 < 0) + P(Z_1^2 \leq t^2 e^m, Z_2 > 0) \\ &\geq P(\frac{Z_1^2}{e^{m+\sqrt{M_v}\sigma_{ch} Z_2}} \leq t^2) - P(Z_2 > 0) + \frac{1}{2} P(Z_1^2 \leq t^2 e^m), \end{aligned}$$

which equals obviously L_2, establishing the second inequality.

Proof of Lemma 2.4.1:

It suffices to show that $R_i(\hat{M}_i) \leq 1$ where $R_i(\hat{M}_i) = \hat{M}_i(t^2 e^m - \ln(\hat{M}_i))/t^2 e^m$ is the ratio of the length of C^{SS} to C^{SV}. This can be proved easily by showing that R_i has a positive derivative under (2.15), implying that $R_i(\hat{M}_i)$ is maximized at $\hat{M}_i = 1$. Since $R_i(1) = 1$, the lemma is established. When $\tau^2 > 0$, the strict inequality follows from the fact that $P(\hat{M}_i < 1) = 1$.

Proof of Lemma 2.4.2:

Proof: Since $\ln S_i^2 \sim N(\mu_v + m, \tau_v^2 + \sigma_{ch}^2)$, then $\ln(S_i) \sim N(\frac{1}{2}(\mu_v + m), \frac{1}{4}(\tau_v^2 + \sigma_{ch}^2))$. Using the generating functions of a normal random variable,

$$E(S_i) = E(e^{\ln(S_i)}) = \exp[\frac{1}{2}(\mu_v + m) + \frac{\tau_v^2 + \sigma_{ch}^2}{8}].$$

Applying similar calculation, one can evaluate $E(S_i)^{M_v}$ and a little algebra shows that

$$E(\hat{\sigma}_{B,i}te^{m/2}) = t\exp(\frac{\mu_v + m}{2} + \frac{M_v^2(\tau_v^2 + \sigma_{ch}^2)}{8}).$$

These two displayed equations show that

$$\frac{E(\hat{\sigma}_{B,i}te^{m/2})}{E(S_it)} = \exp(\frac{(M_v^2 - 1)(\tau_v^2 + \sigma_{ch}^2)}{8}) = \exp(-\frac{\sigma_{ch}^2}{8}(M_v + 1)) < 1,$$

where the last equation follows from the fact that $M_v = \frac{\tau_v^2}{\tau_v^2 + \sigma_{ch}^2}$.

Proof of Lemma 2.4.3:

Write the ratio of $\left(\prod_i^p \sqrt{\hat{\sigma}_{B,i}^2 t^2 e^m}\right)^{1/p}$ to $(\prod_i^p (tS_i))^{1/p}$ as $(\prod_{i=1}^p r_i)^{1/p}$, where $r_i = \frac{\hat{\sigma}_{B,i}e^{m/2}}{S_i} = (\frac{e^{m+\mu_v}}{S_i^2})^{\frac{1-M_v}{2}}$, where the last equation follows from replacing $\hat{\sigma}_{B,i}$ by (2.6). Taking the logarithm of the geometric mean of the ratio gives

$$\ln(\prod_{i=1}^p r_i)^{1/p} = \frac{(1 - M_v)}{2}(\mu_v + m) + \frac{(M_v - 1)}{2}\frac{\sum_{i=1}^p \ln(S_i^2)}{p}. \tag{A.4}$$

Note that the marginal distribution of $\ln S_i^2$ is $N(\mu_v + m, \sigma_{ch}^2 + \tau_v^2)$, and $\ln(S_i^2)$ are i. i. d. Strong law of large number implies that the last expression converges to zero as $p \to \infty$. Hence the ratio converges to 1, completing the proof.

Proof of Theorem 2.4.3:

The first statement follows directly from Lemmas 2.4.1 and 2.4.2. Also, Lemmas 2.4.1 and 2.4.3 imply that R, which denotes the ratio of the geometric means of the lengths of C^{SS} to that of C^{SV}, converges to a quantity less than or equal to 1. To prove that $\lim R < 1$, it is equivalent to show that $\lim \ln(R) < 0$. Follow closely the proof of Lemma 2.4.1 and note that $R = (\prod R_i(\hat{M}_i))^{1/p}$ and apply law of large numbers to establish that $\ln(R)$ converges to $E(\ln(R_i(\hat{M}_i))$. Also when $\tau^2 > 0$, $P(\hat{M}_i < 1) = 1$. Consequently $R_i(\hat{M}_i) < 0$ with probability one and hence $E(\ln(R_i(\hat{M}_i)) < 0$.

Explaining why truncation of the center is necessary even after the

length is truncated

We consider a confidence interval $\hat{C}^{SS}_{i,T}$, which only truncate the length, i.e.,

$$\begin{aligned}\hat{C}^{SS}_{i,T} &= \{\theta_i : |\theta_i - \hat{M}^{EB}_i X_i - (1-\hat{M}^{EB}_i)\hat{\mu}|^2 < RHS\}, \text{ where} \\ RHS &= \hat{M}^{EB,T}_i \hat{\sigma}^2_{EB,i}(t^2 e^m - ln(\hat{M}^{EB,T}_i)) \leq \hat{\sigma}^2_{EB,i} t^2 e^m.\end{aligned}$$

We shall consider the case where p, τ_0, τ and τ_v are all pretty large. Hence we shall approximate $\hat{\mu}$ by μ, $\hat{\sigma}_{EB,i}$ by S_i and hence the upperbound (A.5) by $S_i^2 t^2$ since $\hat{M}_v \approx 1$. Hence we only need to show that the interval

$$\hat{M}^{EB}_i X_i + (1-\hat{M}^{EB}_i)\mu \pm S_i t \tag{A.5}$$

has low coverage probability. Note that τ_0 can obviously be large as long as $E(\sigma_i^2)$ is large (in extreme simulation case, although in practice τ_0 is usually small). Also since $P_\tau(\hat{\tau}=0)$ is decreasing in τ and $P_{\tau_0}(\hat{\tau}=0)=\alpha$ by (2.23), $P_\tau(\hat{\tau}=0)$ can be somewhat larger than α as long as $\tau < \tau_0$ (numerical evidence shows that it can be higher than 0.3 where $\alpha = 0.05$). Therefore there could be a wide range of τ, $\tau < \tau_0$, such that $P_\tau(\hat{\tau}=0)$ is larger than α. Now

$$P(\theta_i \in (A.5)) \leq P(\theta_i \in (A.5), \hat{\tau}=0) + P(\hat{\tau} \neq 0). \tag{A.6}$$

The second term of the upper bound could be as low as 0.7. The first term can be written as

$$P(|(M_i - \hat{M}^{EB}_i)\frac{X_i}{\sigma_i} + \sqrt{M_i}Z_i| \leq \frac{S_i}{\sigma_i}t, \hat{\tau}=0), \tag{A.7}$$

where Z_i is a standard normal random variable independent of X_i, σ_i and S_i. Note that since $\hat{\tau}=0$, $\hat{M}^{EB}_i = 0$. However, when τ is large, M_i is close to 1. Now the left hand side in (A.7) becomes $|\frac{X_i}{\sigma_i} + Z_i|$. Since the variance of X_i/σ_i is $(\sigma_i^2+\tau^2)/\sigma_i^2$ which converges to infinity as τ goes to infinity, the probability (A.7) converges to zero. This concludes by (A.6) that the interval (A.5) could have low probability.

On the other hand if we use $\hat{M}_i^{EB,T}$ in place of $\hat{M}_i^{EB}$ in the center, $M_i \approx 1$ and $\hat{M}_i^{EB,T} \approx 1$, hence $M_i - \hat{M}_i^{EB,T} \approx 0$. Hence, the probability in (A.7) won't converge to zero and can be big. Hence truncating the center would help as also confirmed by numerical studies.

Proof of Lemma 3.2.1.

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

Proof of Theorem 3.2.1.

If $P(\theta_i \notin CI_i)|X \leq q$, then $EV|X \leq q \sum_{i \in A} 1 = Rq$. Consequently

$$FCR_\pi = \int_{\{R>0\}} \frac{EV|X}{R} m(X) dX \leq \int_{\{R>0\}} q m(X) dX = qP(R > 0) \leq q.$$

Proof of Theorem 3.2.2.

Since $X_i|\theta_i \sim N(\theta_i, \sigma_i^2)$ and $\theta_i \sim N(\mu, \tau^2)$,

$$\theta_i|X_i \sim N(M_i X_i + (1 - M_i)\mu, M_i \sigma_i^2) \quad \text{where} \quad M_i = \frac{\tau^2}{\tau^2 + \sigma_i^2}. \tag{A.8}$$

Note that $P(|Z| > z) = q$. Consequently

$$P(|\theta_i - (M_i X_i + (1 - M_i)\mu)| > z\sqrt{M_i \sigma_i^2}|X_i) = q.$$

Applying Theorem 3.2.1 concludes that the interval (3.1) has the Bayes FCR at the q-level.

Proof of Theorem 3.2.3.

Lemma 3.2.1 and the first assumption of this theorem imply that FCR_π is bounded above by $E\alpha(p, X)$. Now

$$E(\alpha(p,X)) = \int_{\alpha(p,X)>q+\epsilon} \alpha(p,X)m(X)dX + \int_{\alpha(p,X)\leq q+\epsilon} \alpha(p,X)m(X)dX$$

The fact that $\alpha(p, X) \leq 1$ implies that

$$\int_{\alpha(p,X)>q+\epsilon} \alpha(p,X)m(X)dX \leq \int_{\alpha(p,X)>q+\epsilon} m(X)dX = P(\alpha(p,X) > q+\epsilon).$$

Combining this with the fact that $\int_{\alpha(p,X)\leq q+\epsilon} \alpha(p,X)m(X)dX \leq q+\epsilon$, we obtain that $E(\alpha(p,X)) \leq P(\alpha(p,X) > q+\epsilon) + q + \epsilon$. As $p \to \infty$,

$$\limsup_{p\to\infty} E(\alpha(p,X)) \leq \limsup_{p\to\infty} P(\alpha(p,X) > q+\epsilon) + q + \epsilon = q + \epsilon,$$

for every $\epsilon > 0$ and hence the same inequality is true for $\epsilon = 0$. We now conclude that

$$\limsup_{p\to\infty} FCR_\pi \leq \limsup_{p\to\infty} E(\alpha(p,X)) \leq q.$$

Proof of Theorem 3.2.4.

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

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

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

$$\lim_{p\to\infty} |FCR - q| \leq \frac{1}{\eta} E \lim_{p\to\infty} \frac{1}{p} \sum |P(\theta_i \notin CI_i|X) - q|, \qquad \text{(A.9)}$$

which by (3.3) equals zero, completing the proof.

Proof of Theorem 3.3.1.

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

Lemma A.0.2 *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 $p \to \infty$. A result needed for the next proof is as following:

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

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 $\hat{\mu}+\mu$, one of which converges to zero in probability and the other to a constant in probability.

Now to prove the second part, by (3.4), we have

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

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

$$Var(X_i^2) = 2(\sigma_i^2+\tau^2)(\tau^2+\sigma_i^2+2\mu^2).$$

Therefore

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

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 3.3.1.

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 [44].) 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 A.0.2, 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 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 $P(A_p) \to 0$ as $p \to \infty$.

According to the construction of CI_i^{EB} in (3.5), $P(\theta_i \notin CI_i|X) = P(|\theta_i - (\hat{M}_iX_i + (1-\hat{M}_i)\hat{\mu})| > z\sqrt{\hat{M}_i}\sigma_i|X)$. 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. As a result, the conditional non-coverage probability $P(\theta \notin CI_i^{EB}|X)$ can be written as

$$P(|Z - \frac{(\hat{M}_i - M_i)(X_i - \mu) + (1-\hat{M}_i)(\hat{\mu} - \mu)}{\sqrt{M_i}\sigma_i}| > z\sqrt{\frac{\hat{M}_i}{M_i}})$$

which is bounded above by $P(|Z-(g_1+g_2)|>z\sqrt{\frac{\hat{M}_i}{M_i}})$ where $g_1=|\frac{(\hat{M}_i-M_i)(X_i-\mu)}{\sqrt{M_i}\sigma_i}|$ and $g_2=|\frac{(1-\hat{M}_i)(\hat{\mu}-\mu)}{\sqrt{M_i}\sigma_i}|$.

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

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

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

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

$g_2\leq C_2\delta_1$. We can choose the constants C_1 and C_2 such that they 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 $i=1,2,\cdots,p$,

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

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 apply Theorem 3.2.3 and we finish the proof.

Proof of Theorem 3.3.2:

It suffices to show that condition (3.3) holds. According to Lemma A.0.2, 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.

By (A.8), we may write $\theta_i = M_i X_i + (1-M_i)\mu + Z(M_i\sigma_i^2)^{1/2}$, 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 (A.10). 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 3.3.1,

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

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:

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

Thus

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

Summing over i on both sides, we have

$$\frac{1}{p}\sum|P(\theta_i \notin CI_i|X) - q| \le max(A, B), \tag{A.10}$$

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

$$B = \frac{1}{p}\sum_i |q - P(|Z - C_1\delta_2 Z_i - C_2\delta_1| > z(1 - \frac{\delta_1}{\tau^2}))|. \tag{A.11}$$

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 (A.10) converges to zero as $p \to \infty$. Condition (3.3) is established and so is the theorem.

Proof of Theorem 4.2.1.

Firstly,

$$\begin{aligned} EL(\theta_i, CI_i|X) &= k_1 Len(CI_i)P(I_i = 1|X) \qquad \text{(A.12)}\\ &- \int I(\theta_i \in CI_i, I_i = 1)m(\theta_i|X)d\theta_i + I_{CI_i}(0|X)(k_2 - \pi_i^0(X)). \end{aligned}$$

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 $\int_{CI_i} 1d\theta_i$. Then (A.12) equals to

$$\pi_i^1(X)\int_{CI_i}(k_1 - \pi(\theta_i|X, I_i = 1))d\theta_i + I_{CI_i}(0|X)(k_2 - \pi_i^0(X)). \tag{A.13}$$

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.13). 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 (4.4).

Proof of Theorem 4.3.1.

According to [46],

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

Since

$$\begin{aligned} & P(\theta_i \notin CI_i^B|X) \\ = \; & P(\theta_i \notin CI_i^B|X, I_i = 0)P(I_i = 0|X) + P(\theta_i \notin CI_i^B|X, I_i = 1)P(I_i = 1|X) \\ = \; & \pi_i^0(X)I(\pi_i^0(X) < k_2) + (1 - \pi_i^0(X))P(\theta_i \notin CI_i^B|X, I_i = 1), \end{aligned} \tag{A.14}$$

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

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

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

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

Proof of theorem 4.4.1.

Before the proof, we will state and prove the following lemma.

Lemma A.0.3 *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 $\forall p > P_0$,*

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

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

$$\tilde{\pi}_i^0 - \pi_i^0 = \frac{\pi_0 \pi_1 \sigma (\sqrt{\frac{\sigma^2+\hat{\tau}^2}{\sigma^2+\tau^2}} \exp(\frac{(M-\hat{M})X_i^2}{2\sigma^2}) - 1)}{(\pi_0 \sqrt{\sigma^2+\hat{\tau}^2} \exp(-\frac{\hat{M}X_i^2}{2\sigma^2}) + \pi_1 \sigma)(\pi_0 + \pi_1 \frac{\sigma}{\sqrt{\sigma^2+\tau^2}} \exp(\frac{MX_i^2}{2\sigma^2}))} \tag{A.15}$$

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.15) implies that

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

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

$$\tilde{\pi}_i^0 - \pi_i^0 > \frac{M}{1-M}(-\epsilon) > -\delta. \tag{A.16}$$

On the other hand,

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

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

$$\tilde{\pi}_i^0 - \pi_i^0 \leq C\frac{y-1}{y^k} = C\frac{y-1}{(y-1+1)^k} \leq C\frac{y-1}{k(y-1)} < C\epsilon. \tag{A.17}$$

Combine (A.16) and (A.17), then

$$|\tilde{\pi}_i^0 - \pi_i^0| \leq max(\delta, C\epsilon) < \delta. \tag{A.18}$$

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

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

where D is a constant that only depends on π_0. Combining this with (A.18), one can get that

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

and completes the proof.

Proof of the theorem

According to [46], $FCR_\pi = E\frac{\sum_{i\in\mathcal{R}} P(\theta_i \notin CI_i^{EB}|X)}{R}(R>0)$ 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.14) in the proof of theorem 4.3.1,

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

In the empirical Bayes interval (4.10), 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.0.3), 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,

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

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

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

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

$$E\Delta_1 \le \epsilon. \tag{A.19}$$

For the second term Δ_2,

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

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,

$$P((\theta_i - \hat{M}X_i)^2 > \hat{M}\sigma^2 z_{q/2}^2 | X, I_i = 1) = P(|Z - \frac{(\hat{M} - M)X_i}{\sqrt{M}\sigma}| > \sqrt{\frac{\hat{M}}{M}} z_{q/2} | X). \tag{A.20}$$

Assume that $X_{(p)}$ is the observation that has the largest absolute value, then $0 \le |\frac{(\hat{M}-M)X_i}{\sqrt{M}\sigma}| \le |\frac{(\hat{M}-M)X_{(p)}}{\sqrt{M}\sigma}|$. Consequently, for any $i = 1, 2, \cdots, p$, (A.20) falls into the range

$$[P(|Z - \frac{|(\hat{M} - M)X_{(n)}|}{\sqrt{M}\sigma}| \ge \sqrt{\frac{\hat{M}}{M}} z_{q/2}), P(|Z| \ge \sqrt{\frac{\hat{M}}{M}} z_{q/2})]. \tag{A.21}$$

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.21) can be rewritten as

$$[P(|Z - |\frac{\sigma(\hat{\tau}^2 - \tau^2)}{\tau(\hat{\tau}^2 + \sigma^2)} Z_{(p)}|| \ge |\sqrt{\frac{\hat{M}}{M}} z_{q/2}), P(|Z| \ge \sqrt{\frac{\hat{M}}{M}} z_{q/2})]. \tag{A.22}$$

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

$$|\Delta_2| \le max(|q - P(|Z - |\frac{\sigma(\hat{\tau}^2 - \tau^2)}{\tau(\hat{\tau}^2 + \sigma^2)} Z_{(p)}|| \ge |\sqrt{\frac{\hat{M}}{M}} z_{q/2})|, |q - P(|Z| \ge \sqrt{\frac{\hat{M}}{M}} z_{q/2})|). \tag{A.23}$$

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

$$|\frac{\sigma(\hat{\tau}^2 - \tau^2)}{\tau(\hat{\tau}^2 + \sigma^2)} Z_{(p)}| = o_p(1). \tag{A.24}$$

The dominated convergence theorem implies that

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

and

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

Applying the dominated convergence theorem again, one can deduce from (A.23) that

$$\limsup_{p\to\infty} E|\Delta_2| \leq 0. \tag{A.25}$$

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.24) implies that

$$\limsup_{p\to\infty} E|\Delta_3| \leq \lim_{p\to\infty} EP(|Z| \geq z_{q/2}) = q. \tag{A.26}$$

(A.19), (A.25), and (A.26) imply that

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

Proof of the proposition 4.4.1.

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

$$qE(\pi_1^0(X) - \pi_1'^0(Y)) = q(\pi_0 - \pi_0'). \tag{A.27}$$

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.27), one obtains that $|\Delta| < \epsilon$ when δ is sufficiently small, which completes the proof.

BIBLIOGRAPHY

[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] J. O. Berger. *Statistical decision theory and Bayesian analysis.* Springer-Verlag, New York., second edition, 1985.

[4] L. D. Brown. Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *The Annals of Mathematical Statistics*, 42(3):855–903, 1971.

[5] 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.

[6] G. Casella and J. T. Hwang. Evaluating confidence sets using loss functions. *Statist. Sinica*, 1(1):159–173, 1991.

[7] G. Casella, J. T. Hwang, and C. Robert. A paradox in decision-theoretic interval estimation. *Statist. Sinica*, 3(1):141–155, 1993.

[8] S. E. Choe, M. Bouttros, A. M. Michelson, G. M. Chruch, and M.S. Halfon. Preferred analysis methods for affymetrix genechips revealed by a wholly defined control dataset. *Genome Biology* , 6(2):R16.1–16, 2005.

[9] 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.

[10] X. Cui, J. T. Hwang, J. Qiu, N. J. Blades, and G. A. Churchill. Improved statistical tests for differential gene expression by shrinking variance components estimates. *BIOSTATISTICS*, 6:59–75, 2005.

[11] B. Efron. Minimum volume confidence regions for a multivariate normal mean vector. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(4):655–670, 2006.

[12] B. Efron and C. N. Morris. Limiting the risk of Bayes and empirical Bayes estimators. I. The Bayes case. *J. Amer. Statist. Assoc.*, 66:807–815, 1971.

[13] B. Efron and C. N. Morris. Limiting the risk of Bayes and empirical Bayes estimators. II. The empirical Bayes case. *J. Amer. Statist. Assoc.*, 67:130–139, 1972.

[14] B. Efron and C. N. Morris. Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.*, 68:117–130, 1973.

[15] B. Efron and R. Tibshirani. *An introduction to the bootstrap.* London:Chapmans and Hall, 1993.

[16] 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.

[17] K. He. Parametric empirical Bayes confidence intervals based on James-Stein estimator. *Statist. Decisions*, 10(1-2):121–132, 1992.

[18] 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.

[19] J. T. Hwang. Empirical Bayes estimation for the means of the selected populations. *Sankhyā Ser. A*, 55(2):285–304, 1993.

[20] J. T. Hwang and G. Casella. Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.*, 10(3):868–881, 1982.

[21] J. T. Hwang and G. Casella. Improved set estimators for a multivariate normal mean. *Statist. Decisions*, (suppl. 1):3–16, 1984. Recent results in estimation theory and related topics.

[22] J. T. Hwang and J. Chen. Improved confidence sets for the coefficients of a linear model with spherically symmetric errors. *Ann. Statist.*, 14(2):444–460, 1986.

[23] J. T. Hwang and P. Liu. Optimal tests shrinkage both means and variances applicable to microarray data analysis. *Submitted*, 2007.

[24] 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.

[25] W. James and C. Stein. Estimation with quadratic loss. *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 361–379, 1961.

[26] V. M. Joshi. Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.*, 40:1042–1067, 1969.

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

[28] K. Kerr, M. Martin, and G. Churchill. Analysis of variance for gene expression microarray data. *Journal of Computational Biology*, 7(6):819–837, 2000.

[29] Nan M. Laird and Thomas A. Louis. Empirical Bayes confidence intervals based on bootstrap samples. *J. Amer. Statist. Assoc.*, 82(399):739–757, 1987. With discussion and with a reply by the authors.

[30] K. Lo and R. Gottardo. Flexible empirical Bayes models for differential gene expression. *Bioinformatics*, 23(3):328–335, 2007.

[31] I. Lönnstedt and T. Speed. Replicated microarray data. *Statist. Sinica*, 12(1):31–46, 2002. Special issue on bioinformatics.

[32] J. S. Milton and Arnold J. C. *Introduction to probability and statistics: Principles and applications for engineering and the computing sciences.* McGraw-Hill Companies, fourth edition, 2002.

[33] C. N. Morris. Parametric empirical Bayes confidence intervals. *Scientific inference, data analysis, and robustness (Madison, Wis., 1981)*, 48:25–50, 1983.

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

[35] J. Qiu. Sharp simultaneous intervals for the means of selected populations with application to microarray data analysis. *Ph.D. dissertation, Cornell University*, 2004.

[36] 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.

[37] J. Qiu and J. T. Hwang. Reply to Dr. Ekkehard Glimm's letter. *Biometrics*, page to appear, 2008.

[38] R. Samworth. Small confidence sets for the mean of a spherically symmetric distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(3):343–361, 2005.

[39] G. K. Smyth. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 3., 2004.

[40] C. M. Stein. Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. Ser. B*, 24:265–296, 1962.

[41] T. Tong and Y. Wang. Optimal shrinkage estimation of variances with applications to microarray data analysis. *J. Amer. Statist. Assoc.*, 102(477):113–122, 2007.

[42] Y. Tseng and L. D. Brown. Good exact confidence sets for a multivariate normal mean. *Ann. Statist.*, 25(5):2228–2258, 1997.

[43] V. G. Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Science.*, 98:5116–5121, 2001.

[44] M. Woodroofe. *Probability with applications.* New York, McGraw-Hill., 1975.

[45] G. W. Wright and R. M. Simon. A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics*, 19:2448–2455, 2003.

[46] Z. Zhao and J. T. Hwang. Empirical Bayes FCR controlling confidence interval. Submitted, 2008.