

NONPARAMETRIC EMPIRICAL BAYES
SIMULTANEOUS ESTIMATION
FOR MULTIPLE VARIANCES

A Dissertation
Submitted to
the Temple University Graduate Board

In Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
IN STATISTICS

by
Yeil Kwon
May 2018

Examining Committee Members:

Zhigen Zhao, Advisory Chair, Department of Statistical Science
Sanat Sarkar, Department of Statistical Science
Cheng Yong Tang, Department of Statistical Science
Jing Qiu, External Member, University of Delaware

©
Copyright
2018

by

Yeil Kwon
All Rights Reserved

ABSTRACT

The shrinkage estimation has proven to be very useful when dealing with a large number of mean parameters. In this dissertation, we consider the problem of simultaneous estimation of multiple variances and construct a shrinkage type non-parametric estimator. We take the non-parametric empirical Bayes approach by starting with an arbitrary prior on the variances. Under an invariant loss function, the resultant Bayes estimator relies on the marginal cumulative distribution function of the sample variances. Replacing the marginal cdf by the empirical distribution function, we obtain a Non-parametric Empirical Bayes estimator for multiple Variances (NEBV). The proposed estimator converges to the corresponding Bayes version *uniformly* over a large set. Consequently, the NEBV works well in a post-selection setting.

We then apply the NEBV to construct confidence intervals for mean parameters in a post-selection setting. It is shown that the intervals based on the NEBV are shortest among all the intervals which guarantee a desired coverage probability. Through real data analysis, we have further shown that the NEBV based intervals lead to the smallest number of *discordances*, a desirable property when we are faced with the current “replication crisis”.

*For Emily and Catherine,
without whose encouragement
this dissertation would
not have been
written.*

ACKNOWLEDGEMENTS

I would like to thank my precious advisor, Dr. Zhigen Zhao for guiding and supporting me over the years. You have set the perfect example of excellence as a researcher, mentor, instructor, and role model.

I would like to thank my thesis committee members for all of their guidance through this process; your discussion, ideas, and feedback have been absolutely invaluable.

I also would like to thank Dr. Pallavi Chitturi and Dr. Wei-Shih Yang for their constant enthusiasm and encouragement.

I would especially like to thank my amazing family for the love, support, and constant encouragement I have gotten over the years. In particular, I would like to thank my wife Emily and my daughter Catherine. I breathe you in my dreams, and I undoubtedly could not have done this without you.

I'd like to thank my fellow graduate students: Scott, A. Bruce, Zeda Li, Qi Xia, and Shinjini Nandi. Your love and passion have kept me motivated and inspired. You are and always will be my family. I am very grateful to all of you.

Finally, I would like to thank and dedicate this dissertation to my father, Moosa Kwon. It was you who originally generated my love for research with lessons on mathematics and statistics. Although it has been years since you have passed, you live in my heart and I still take your lessons with me, every day.

TABLE OF CONTENTS

ABSTRACT	iii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER	
1. INTRODUCTION	1
2. REVIEW OF EXISTING METHODS	4
2.1 Exponential Lindley-James-Stein Estimator	4
2.2 TW Shrinkage Estimator	6
2.3 Variance Estimator for SAM t -test	6
2.4 Robust Variance Estimator	7
3. LOSS FUNCTIONS OF THE VARIANCE ESTIMATORS	10
3.1 Loss Functions	10
3.2 The Risk of $S_{(1)}^2$ for $\sigma_{(1)}^2$	12
4. NONPARAMETRIC EMPIRICAL BAYES ESTIMATOR FOR THE VARIANCES	14
4.1 Model Assumption	14
4.2 Bayes Decision Rule	15
4.3 Nonparametric Bayes Estimator of the Variances (NEBV)	17
5. ANALYTIC PROPERTIES FOR NEBV	20
5.1 Background	20
5.2 Asymptotic Behavior of NEBV	22

6. APPLICATION WITH NUMERICAL ILLUSTRATIONS	25
6.1 Construction of Confidence Intervals for Selected Parameters	26
6.1.1 Estimation of Hyper-parameters	26
6.1.2 Truncation Method for Estimating τ^2	29
6.1.3 Confidence Interval Based on the Second Order Cor- rection	31
6.2 Numerical Study	32
6.2.1 Independent Case	33
6.2.2 Dependent Case	39
6.2.3 Correlation Matrix generated from Real data	41
6.2.4 Real data Analysis	42
7. A GENERALIZATION OF NEBV	44
7.1 Tweedie's Formula	44
7.2 A Generalized Form of NEBV	45
8. CONCLUSION	49
APPENDIX	50
BIBLIOGRAPHY	62

LIST OF TABLES

Table

6.1	Percentage of discordant genes for various intervals.	42
-----	---	----

LIST OF FIGURES

Figure

3.1	Loss Functions for $\hat{\sigma}^2$ (true value = 0.1)	11
6.1	Plots of $\ln(\text{Risk})$ for estimating $\sigma_{(i)}^2$ under lognormal (the first row) and inverse gamma prior (the second row) for σ_i^2 against $\frac{\tau^2}{\tau^2 + E(\sigma_i^2)}$: From upper-left, each plot corresponds to $CV = 0.25, 0.50, 1.00$ and 1.50 respectively with $E(\sigma_i^2) = 0.5$	34
6.2	Plots of $\ln(\text{Risk})$ for estimating $\sigma_{(i)}^2$ under Inverse Gamma and Gamma mixture prior for σ_i^2 : $0.5 * IG(6, 1) + 0.5 * IG(6, 4)$ for left panels and $0.5 * Gamma(4, 1/16) + 0.5 * Gamma(6, 1/8)$ for right panels, respectively with $E(\sigma^2) = 0.5$	34
6.3	Plots of the coverage probabilities (the first row) and the average length ratio (the second row) for various intervals of $\theta_{(p)}$ for lognormal prior against $\frac{\tau^2}{\tau^2 + E(\sigma_i^2)}$. The CV of the left and right column are $CV = 1.00$ and 1.50 respectively. The error bar represents $0.95 - 2 \times (\text{simulation error})$	35
6.4	Plots of the coverage probabilities (left panels) and the average length ratio (right panels) for various intervals of $\theta_{(p)}$ under inverse gamma prior. The CV of the left and right columns are 1.00 and 1.50 respectively.	36
6.5	Plots of the coverage probabilities and the length ratio of confidence intervals for $\theta_{(p)}$ under Inverse Gamma and Gamma mixture prior for σ_i^2 : $0.5 * IG(6, 1) + 0.5 * IG(6, 4)$ for left panels and $0.5 * Gamma(4, 1/16) + 0.5 * Gamma(6, 1/8)$ for right panels, respectively with $E(\sigma^2) = 0.5$	37
6.6	Plots of the coverage probabilities of four intervals with AR, Banded and Sparse correlation matrices.	38
6.7	Plots of the average length ratio of four intervals with AR, Banded and Sparse correlation matrices.	39
6.8	Plots of the coverage probabilities, length ratio of confidence intervals for $\theta_{(p)}$ based on the Colon Cancer and Leukemia Data.	40
6.9	Distributions for the rate of misclassified rates of 1000 iterations for leukemia and colon cancer data	43

CHAPTER 1

INTRODUCTION

In modern statistical applications, we usually face the situation that there are a large number of parameters to be estimated simultaneously. The shrinkage type methods, initially introduced in *Stein* (1956) and *James and Stein* (1961), have been proved extremely useful in the last six decades. In *Efron and Morris* (1973) and *Efron and Morris* (1975), it is shown that the James-Stein estimator can be interpreted beautifully as an empirical Bayes estimator. The shrinkage estimators have been widely used in various statistical inference (*Morris*, 1983; *He*, 1992; *Hwang et al.*, 2009; *Zhao and Hwang*, 2012; *Hwang and Zhao*, 2013) which can be used in many microarray dataset.

The previous literature focuses mostly on the inference for mean parameters for the normal mean model. The estimation of unknown variances is an important step toward inferring the means. An easy approach is to estimate each variance by the corresponding sample variance, an unbiased estimator. However, in Big Data era, the heteroscedasticity among the variances commonly exists and the sample variances vary dramatically. For instance, in the golden spike-in data set (*Choe et al.*, 2005), the smallest sample variance among all the genes is 6.1×10^{-5} while the largest one is 5.416. Such an easy estimator will produce misleading results because testing the mean effects can be shadowed by extremely small or large values of the sample variances.

To address this issue, various variance estimators are proposed. For instance, *Storey and Tibshirani* (2003) suggested adding a constant to each sample variance in

order to stabilize the variance estimator. *Lin et al.* (2003) suggested a robust estimator using rank order. Empirical Bayesian approaches for estimating the variances have been proposed in *Lönnstedt and Speed* (2002), *Kendzioriski et al.* (2003) and *Smyth* (2004). Approximating the sample variances by lognormal random variables, *Cui et al.* (2005) proposed the exponential Lindley-James-Stein estimator and applied it to the F test. It is demonstrated that the resultant testing method has a high power for capturing the differentially expressed genes. This method has been further investigated by *Tong and Wang* (2007) theoretically. A similar approach has been taken in *Zhao* (2010), *Zhao and Hwang* (2012), and *Hwang and Zhao* (2013) to construct shrinkage confidence intervals for both unselected and the selected mean parameters.

Most of these existing popular empirical Bayes estimators for the variances assume a strong parametric assumption on the prior of the variances. In this dissertation, we assume an arbitrary prior distribution $\pi(\sigma^2)$ for the variances σ_i^2 . When assuming an invariant loss function, the Bayesian decision rule depends on the data through the marginal cumulative distribution function of the sample variances only. As a result, the proposed Non-parametric Empirical Bayes estimator for the Variances (NEBV) relies on the empirical cumulative distribution function, which can reliably estimate the empirical distribution function. Consequently, the corresponding empirical Bayes estimator enjoys good properties such as fast computation and being free from choosing tuning parameters.

In many applications such as the microarray experiments, an investigator might only select those parameter(s) with a large T -statistics in magnitude for further studies. Estimating the corresponding variances post the selection is important toward the inference for the mean. In general, the Bayes estimator is immune to the selection (*Dawid*, 1994) and the corresponding empirical Bayes estimator is a good approximation of the Bayes version when estimating all the variances simultaneously.

When focusing on the selected parameters, the difference between these two becomes substantial. On the other hand, NEBV relies on the estimation of the cumulative distribution function and converges to the Bayes decision rule *uniformly* over a set which covers the selected data with a high probability. As a result, NEBV works well for the selected parameters without any adjustment, which is a striking but extremely useful feature.

The proposed estimator can be used as a step-stool to infer mean parameters. In this dissertation, we use it to construct confidence intervals for the mean parameters in a post selection estimation setting, which is closely related to the *Replication crisis* in scientific investigations. Extensive simulations show that the intervals based on NEBV substantially improve the existing ones. We create a novel simulation using real data sets to demonstrate that the NEBV-based approach results in the smallest number of *discordance*, a term we introduced to denote the number of discoveries based on the current data which cannot be replicated using the newly collected data.

In Chapter 2, we review the representative variance estimators suggested in the existing literature. In Chapter 3, we recall the loss functions and the associated risks in the construction of various commonly used estimators of variance. In Chapter 4, we propose NEBV, a non-parametric empirical Bayes estimator for multiple variances. The analytic properties of the proposed estimator are presented in Chapter 5, which depicts the asymptotic behavior of NEBV. In Chapter 6, we illustrate the results of the numerical study including the extensive simulations and real data analysis to describe the performance of the confidence intervals for the selected parameter based on the proposed variance estimator. Chapter 7 presents the generalization of the NEBV and in Chapter 8, we summarize the features of the proposed estimator NEBV and list the main issues of the further research to complete the dissertation.

CHAPTER 2

REVIEW OF EXISTING METHODS

In this chapter, we review some existing estimators for the variances, which include exponential Lindley-James-Stein Estimator (*Cui et al.*, 2005), optimal shrinkage estimator (*Tong and Wang*, 2007), variance estimator for SAM t -test (*Tusher et al.*, 2001) and robust variance estimator (*Lin et al.*, 2003).

2.1 Exponential Lindley-James-Stein Estimator

Let $S_i^2 = \sigma_i^2 \frac{\chi_k^2}{k}$, $i = 1, \dots, p$, where χ_k^2 is a chi-squared random variable with degree of freedom k . Taking the logarithm transformation on both sides, we have

$$\ln S_i^2 = \ln \sigma_i^2 + \ln \left(\frac{\chi_k^2}{k} \right).$$

Approximate $\ln \left(\frac{\chi_k^2}{k} \right)$ by a normal distribution with mean m , and variance V , with

$$m = E \left[\ln \left(\frac{\chi_k^2}{k} \right) \right] = \psi \left(\frac{k}{2} \right) - \ln \left(\frac{k}{2} \right),$$
$$V = \text{Var} \left[\ln \left(\frac{\chi_k^2}{k} \right) \right] = \psi' \left(\frac{k}{2} \right),$$

where $\psi(\cdot)$ is a digamma function defined as $\psi(x) = \frac{d}{dx} \log \Gamma(x)$. With a normal prior assumption for $\ln \sigma_i^2$, we have the following lognormal model

$$\begin{cases} \ln S_i^2 - m \mid \ln \sigma_i^2 & \sim N(\ln \sigma_i^2, V), \\ \ln \sigma_i^2 & \sim N(\mu_v, \tau_v^2). \end{cases} \quad (2.1-1)$$

The marginal distribution of $\ln S_i^2 - m$ and the posterior distribution of $\ln \sigma_i^2$ given $\ln S_i^2$ are

$$\ln S_i^2 - m \sim N(\mu_v, \tau_v^2 + V),$$

and

$$\ln \sigma_i^2 \mid \ln S_i^2 - m \sim N\left(B(\ln S_i^2 - m) + (1 - B)\mu_v, \frac{\tau_v^2 V}{\tau_v^2 + V}\right),$$

respectively. Here $B = \frac{\tau_v^2}{\tau_v^2 + V}$ is the shrinkage scaling.

Using the method of moments, μ_v and B can be estimated as

$$\begin{aligned} \hat{\mu}_v &= \frac{1}{p} \sum_{i=1}^p (\ln S_i^2 - m) = \ln \left(\prod_{i=1}^p (S_i^2)^{\frac{1}{p}} \right) - m, \\ \hat{B} &= \left(1 - \frac{(p-3)V}{\sum (\ln S_i^2 - \overline{\ln S_i^2})^2} \right)_+, \end{aligned}$$

where $\overline{\ln S_i^2} = \frac{1}{p} \sum \ln S_i^2$. Consequently, *Cui et al.* (2005) proposed the following exponential Lindley-James-Stein estimator for σ_i^2 as

$$\begin{aligned} \hat{\sigma}_{i,ELJS}^2 &= \exp \left\{ \hat{B}(\ln S_i^2 - m) + (1 - \hat{B})\hat{\mu}_v \right\} \\ &= \left(\frac{S_i^2}{e^m} \right)^{\hat{B}} \left(\frac{S_{GM}^2}{e^m} \right)^{1-\hat{B}}, \end{aligned} \quad (2.1-2)$$

where $S_{GM}^2 = \prod_{i=1}^p (S_i^2)^{\frac{1}{p}}$ is a geometric mean of S_i^2 . It is easily seen from (2.1-2) that $\hat{\sigma}_{i,ELJS}^2$ shrinks S_i^2 towards its geometric mean as \hat{B} approaches to zero.

2.2 TW Shrinkage Estimator

The exponential Lindley-James-Stein Estimator (2.1-2) is constructed by combining two estimators, $\frac{S_i^2}{e^m}$ and $\frac{S_{GM}^2}{e^m}$. Note that $\frac{S_i^2}{e^m}$ is biased for σ_i^2 unless $e^{-m} = 1$ or equivalently $k \rightarrow \infty$. Moreover, even when $\sigma_i^2 = \sigma^2$ for all i , $\frac{S_{GM}^2}{e^m}$ is biased as well for estimating σ^2 . *Tong and Wang* (2007) proposed the shrinkage type estimator for σ_i^{2t} using another combination of S_i^2 and S_{GM}^2 as

$$\hat{\sigma}_i^{2t}(\alpha) = (h_p(t)S_{GM}^{2t})^\alpha (h_1(t)S_i^{2t})^{1-\alpha}, \quad 0 \leq \alpha \leq 1. \quad (2.2-3)$$

Here $h_1(t)$ is chosen such that $h_1(t)S_i^{2t}$ is unbiased for σ_i^{2t} , and $h_p(t)$ is chosen such that $h_p(t)S_{GM}^{2t}$ is unbiased for σ^{2t} when $\sigma_i^2 = \sigma^2$ for all i . Indeed, basic calculation shows that

$$h_p(t) = \left(\frac{k}{2}\right)^t \left(\frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{k}{2} + \frac{t}{p}\right)}\right)^p.$$

Let $R(\boldsymbol{\sigma}^{2t}, \hat{\boldsymbol{\sigma}}^{2t}(\boldsymbol{\alpha})) = \frac{1}{p} \sum_i E[L(\sigma_i^{2t}, \hat{\sigma}_i^{2t}(\alpha))]$ be the average risk function where L stands for either the squared loss function or Stein's loss function. In this dissertation, the authors aimed at finding α^* to minimize the risk functions among all the estimator defined in (2.2-3).

2.3 Variance Estimator for SAM t -test

SAM (Significance Analysis Microarrays) t -test is one of the methods for microarrays that identifies genes by detecting significant changes in gene expression through a set of gene-specific t tests (*Tusher et al.*, 2001). SAM assigns a score to each gene based on its change in gene expression relative to the standard deviation of repeated measurements for that gene. Genes assigned scores greater than a threshold are iden-

tified potentially significant. The relative difference $d(i)$ in gene expression is defined as

$$d(i) = \frac{\bar{x}_I(i) - \bar{x}_U(i)}{s(i) + s_0},$$

where $\bar{x}_I(i)$ and $\bar{x}_U(i)$ indicate the average levels of expression for the i -th gene in states I (Irradiated) and U (Unirradiated), respectively. The gene-specific spread measure $s(i)$ is the standard deviation of repeated expression measurements:

$$s^2(i) = a \sum_m (x_m(i) - \bar{x}_I(i))^2 + a \sum_n (x_n(i) - \bar{x}_U(i))^2,$$

where \sum_m and \sum_n stand for the summations of the expression measurements in states I and U , respectively, $a = \left(\frac{1}{n_1} + \frac{1}{n_2}\right) / (n_1 + n_2 - 2)$, and n_1 and n_2 are the numbers of measurements in states I and U . Note that for standardizing $d(i)$, not only $s(i)$ but s_0 is used together by adding them. Here, s_0 is the value selected for minimizing the coefficient of variation. By adding a constant s_0 in the denominator, such a test can avoid false positives resulting from an extremely small $s(i)$'s.

2.4 Robust Variance Estimator

Consider experiments with multiple conditions, multiple readings per gene on a chip and replication of chips with ANOVA model,

$$x_{ijk} = \mu + c_i + g_j + (cg)_{ij} + \epsilon_{ijk},$$

where $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$, and $\epsilon_{ijk} \sim N(0, \sigma_j^2)$, which is the measurement error for k -th replicate. Here I, J , and K represent the number of

conditions, the number of genes, and the number of replicates respectively. Both gene effect, g_j , and interaction $(cg)_{ij}$ are random effects with $(cg)_{ij} \sim N(0, \delta_j^2)$. Then the variance of the differential gene expression score is

$$\begin{cases} \text{Var}(x_{ijk}) = \gamma_j^2 + \delta_j^2 + \sigma_j^2, & \text{if there exists differential expression,} \\ \text{Var}(x_{ijk}) = \gamma_j^2 + \sigma_j^2, & \text{if no differential expression,} \end{cases}$$

where γ_j^2 is the variance for the gene j random effect for $i = 1, \dots, I$, $k = 1, \dots, K$. We can separate $(cg)_{ij}$ differential expression from the g_j gene effect by contrasting conditions. Assume w is the condition contrast with $\sum_i w_i = 0$ and $\sum_i w_i^2 = 1$. Let d_j is the contrast to compare condition 1 with condition 2:

$$d_j = \sqrt{\frac{K}{2}} (\bar{x}_{1j.} - \bar{x}_{2j.}),$$

where $\bar{x}_{1j.} = \sum x_{ijk}/K$ and $\bar{x}_{2j.} = \sum x_{2jk}/K$. The mean and variance of d_j can be written as

$$E(d_j) = \sum_i w_i c_i \sqrt{K},$$

and

$$\begin{cases} \text{Var}(d_j) = \delta_j^2 + \sigma_j^2, & \text{if there exists differential expression,} \\ \text{Var}(d_j) = \sigma_j^2, & \text{if no differential expression,} \end{cases}$$

with more general setting of the contrasts

$$d_{jk} = \sum_i w_i \bar{x}_{ij.} \sqrt{K} = \sum_i w_i \sqrt{K} [c_i + (cg)_{ij} + \bar{\epsilon}_{ij.}],$$

where $\epsilon_{ij.} = \sum_k \epsilon_{ijk}/K$. For the differential expression having the variation across average gene expression $a_j = \sum_{ik} x_{ijk}/IK$, *Lin et al.* (2003) estimate the center and spread using smoothed medians and smoothed median absolute deviations(MAD), respectively. Splines are further used along with the smoothed MAD to get a data-driven, robust estimate of the spread $s(a)$. The standardized test statistic is defined as

$$T_i = \frac{d_j - m(a_j)}{s(a_j)}.$$

CHAPTER 3

LOSS FUNCTIONS OF THE VARIANCE ESTIMATORS

3.1 Loss Functions

Let $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_p^2)$ be the parameters of interest and $\mathbf{S}^2 = (S_1^2, \dots, S_p^2)$ be the corresponding sample variances. Assume that

$$\frac{S_i^2}{\sigma_i^2} \stackrel{iid}{\sim} \frac{\chi_k^2}{k}, \quad i = 1, 2, \dots, p, \quad (3.1-1)$$

where k is the degrees of freedom. Let $\hat{\boldsymbol{\sigma}}^2 = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_p^2)$ be an estimator of the parameter $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$. Consider the following loss function:

$$L_1(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\sigma}}^2) = \sum_{i=1}^p \left(\frac{\sigma_i^2}{\hat{\sigma}_i^2} - 1 \right)^2. \quad (3.1-2)$$

Other commonly used loss functions include

$$L_2(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\sigma}}^2) = \sum_{i=1}^p \left(\frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right)^2, \quad (3.1-3)$$

and

$$L_3(\boldsymbol{\sigma}^2, \hat{\boldsymbol{\sigma}}^2) = \sum_{i=1}^p \left(\frac{\hat{\sigma}_i^2}{\sigma_i^2} - \ln \frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 \right). \quad (3.1-4)$$

The difference between L_1 and L_2 is the position of σ_i^2 and $\hat{\sigma}_i^2$ in the ratio. The loss function L_3 is commonly known as the Stein's loss. All these three loss functions

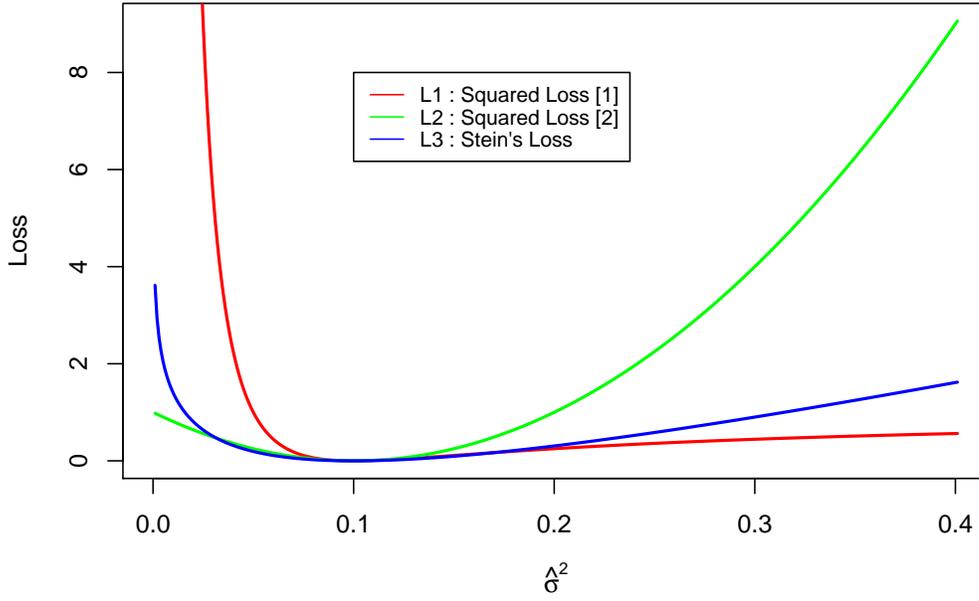


Figure 3.1: Loss Functions for $\hat{\sigma}^2$ (true value = 0.1)

are scale invariant and equals zero when $\hat{\sigma}_i^2 = \sigma_i^2$ for all $i = 1, \dots, p$. The loss increases when the ratio of $\hat{\sigma}_i^2$ to σ_i^2 moves away from one. L_2 imposes more penalty on overestimation but L_1 and L_3 impose more penalty on underestimation. Figure 3.1 displays these loss functions when $p = 1$ and $\sigma^2 = 0.1$. It is shown in the graph that, as the estimator of σ^2 is close to zero, the loss function L_1 and the Stein loss function L_3 assign much larger penalty than the square loss function L_2 . This feature is favored in many application because the primary interest will be those parameters with small variances. Estimating the variance by an even smaller number will result in the inflated probability of committing type I error which should be avoided. Therefore, one should put more penalty on the underestimation (P. 332 of *Casella and Berger* (2001)). The loss function L_2 fails to do so because there is a fixed upper limit for

the penalty at the lower end. On the other hand, by using the loss function L_1 , we can avoid such underestimation. Stein's loss function has also been used in literature (*Stein, 1977; Dey and Srinivasan, 1985; Tong and Wang, 2007*) and the penalty goes to ∞ when the estimation goes to zero. However, as Shown in Chapter 4, the corresponding Bayes rule depends on the marginal density of the sample variances which is difficult to estimate.

In many applications, a practitioner is only interested in a certain set of parameters, denoted as \mathcal{S} , after applying a selection rule. For instance, in microarray analysis, a scientist can select a set of genes based on the magnitude of T -statistics and then zoom in on these parameters for further investigation. Upon selection, the loss function can be naturally modified to

$$L_1^{\mathcal{S}}(\hat{\boldsymbol{\sigma}}^2, \boldsymbol{\sigma}^2) = \sum_{i \in \mathcal{S}} \left(\frac{\sigma_i^2}{\hat{\sigma}_i^2} - 1 \right)^2. \quad (3.1-5)$$

As a special case, one could choose the parameters according to the magnitude of S_i^2 . Let $S_{(1)}^2 \leq S_{(2)}^2 \leq \dots \leq S_{(p)}^2$ be the order statistics of a random sample S_1^2, \dots, S_p^2 . Define $\sigma_{(i)}^2$ as the parameter corresponding to $S_{(i)}^2$, which is called the selected parameter corresponding to $S_{(i)}^2$. For instance, for $p = 3$, if $S_3^2 < S_1^2 < S_2^2$, then $\sigma_{(1)}^2 = \sigma_3^2$, $\sigma_{(2)}^2 = \sigma_1^2$, and $\sigma_{(3)}^2 = \sigma_2^2$. As an extreme case, one might only be interested in $\sigma_{(1)}^2$, the parameter corresponding to the smallest sample variance $S_{(1)}^2$. Note that $\sigma_{(1)}^2$ is not necessarily the smallest among all the σ_i^2 's even though $S_{(1)}^2$ is the smallest among all the S_i^2 's.

3.2 The Risk of $S_{(1)}^2$ for $\sigma_{(1)}^2$

A sample variance S_i^2 is one of the most popular estimators for σ_i^2 since, for each given i , S_i^2 is unbiased and consistent for σ_i^2 under the assumption (3.1-1). When focusing on the selected parameter, say $\sigma_{(1)}^2$, the Bayes risk of the estimator $S_{(1)}^2$

could go to ∞ under loss function L_1^S for some prior distributions.

Assume that σ_i^2 follows an Inverse-Gamma distribution, namely, $\sigma_i^2 \sim IG(\alpha, \beta)$.

Then the posterior distribution of σ_i^2 given S_i^2 is given by

$$\sigma_i^2 | S_i^2 \sim IG\left(\frac{k}{2} + \alpha - 1, \left(\frac{kS_i^2}{2} + \frac{1}{\beta}\right)^{-1}\right).$$

Basic algebra shows that, under the modified squared loss function L_1^S , the risk of $S_{(1)}^2$ for $\sigma_{(1)}^2$ is

$$\begin{aligned} E\left(\frac{\sigma_{(1)}^2}{S_{(1)}^2} - 1\right)^2 &= E\left[E\left(\frac{(\sigma_{(1)}^2 - S_{(1)}^2)^2}{(S_{(1)}^2)^2} \mid \mathbf{S}\right)\right] \\ &= C_1(k, \alpha, \beta)E\left(\frac{1}{S_{(1)}^2}\right)^2 + C_2(k, \alpha, \beta)E\left(\frac{1}{S_{(1)}^2}\right) + C_3(k, \alpha, \beta), \end{aligned} \tag{3.2-6}$$

where $C_j(k, \alpha, \beta)$, $j = 1, 2, 3$, are constants depending on α , β and k only. According to Proposition 3.1, both $E\left(\frac{1}{S_{(1)}^2}\right)^2$ and $E\left(\frac{1}{S_{(1)}^2}\right)$ go to ∞ . As a result, the risk of $S_{(1)}^2$ for estimating $\sigma_{(1)}^2$ based on L_1 diverges as $p \rightarrow \infty$.

Proposition 3.1. *Assume that $S_i^2 | \sigma_i^2 \sim \sigma_i^2 \frac{\chi^2(k)}{k}$ and $\sigma_i^2 \sim IG(\alpha, \beta)$ for $i = 1, \dots, p$, where $\chi^2(k)$ is a chi-squared random variable with degree of freedom k . Let $S_{(1)}^2$ be the first order statistic of S_1^2, \dots, S_p^2 . Then $E\left(\frac{1}{S_{(1)}^2}\right) \rightarrow \infty$ as $p \rightarrow \infty$.*

On the other hand, from the result above, the risk of $S_{(1)}^2$ for $\sigma_{(1)}^2$, under the Stein loss function L_3 , diverge as well. That is,

$$E\left(\frac{S_{(1)}^2}{\sigma_{(1)}^2} - \ln \frac{S_{(1)}^2}{\sigma_{(1)}^2} - 1\right) \rightarrow \infty.$$

CHAPTER 4

NONPARAMETRIC EMPIRICAL BAYES ESTIMATOR FOR THE VARIANCES

4.1 Model Assumption

For the univariate case, the sample variance is a good estimator of σ_i^2 . However, when estimating a large number of σ_i^2 's simultaneously, or when estimating the parameters in a post-selection setting, such an estimator does not provide good estimation. Researches done by *Dawid* (1994), *Senn* (2008) and *Efron* (2011), have demonstrated that such an estimator usually leads to statistical methods which are either invalid or inefficient. The Bayesian/empirical Bayesian approach for this type of inference problem can be one of the attractive solutions, avoiding the selection bias.

In the proposed method, we estimate σ_i^2 based on the non-parametric empirical Bayes approach, assuming an arbitrary prior distribution for σ_i^2 . We consider the following non-parametric Bayesian hierarchical model.

$$\begin{cases} S_i^2 | \sigma_i^2 & \overset{iid}{\sim} \sigma_i^2 \frac{\chi_k^2}{k}, \\ \sigma_i^2 & \overset{iid}{\sim} \pi(\sigma_i^2), \end{cases} \quad (4.1-1)$$

where $\pi(\sigma_i^2)$ is an arbitrary prior of σ_i^2 . Under the model (4.1-1), the marginal dis-

tribution of S_i^2 can be written as

$$m(s_i^2) = \int_0^\infty f(s_i^2|\sigma_i^2)\pi(\sigma_i^2) d\sigma_i^2.$$

Note that the marginal distribution of S_i^2 is quite arbitrary because no particular assumption is imposed for $\pi(\sigma_i^2)$. Furthermore, the posterior distribution of $\sigma_i^2|s_i^2$ can be written as

$$\pi(\sigma_i^2|s_i^2) = \frac{f(s_i^2|\sigma_i^2)\pi(\sigma_i^2)}{\int_0^\infty f(s_i^2|\sigma_i^2)\pi(\sigma_i^2) d\sigma_i^2}.$$

4.2 Bayes Decision Rule

We derive the Bayes decision rule minimizing the conditional risk based on the model (4.1-1) and the loss function L_1 . Given $\mathbf{S}^2 = (S_1^2, \dots, S_p^2)$, the conditional risk based on the loss function L_1 is given by

$$E \left[\sum_{i=1}^p \left(\frac{\sigma_i^2}{\hat{\sigma}_i^2} - 1 \right)^2 \middle| \mathbf{S}^2 \right] = \sum_{i=1}^p \left[\frac{1}{\hat{\sigma}_i^4} E((\sigma_i^2)^2 | \mathbf{S}^2) - \frac{2}{\hat{\sigma}_i^2} E(\sigma_i^2 | \mathbf{S}^2) + 1 \right]. \quad (4.2-2)$$

Therefore, the Bayes decision rule $\hat{\boldsymbol{\sigma}}_B^2 = (\hat{\sigma}_{1,B}^2, \hat{\sigma}_{2,B}^2, \dots, \hat{\sigma}_{p,B}^2)$ which minimizes (4.2-2) is

$$\hat{\sigma}_{i,B}^2 = \frac{E[(\sigma_i^2)^2 | \mathbf{S}^2]}{E[\sigma_i^2 | \mathbf{S}^2]}, \quad i = 1, \dots, p.$$

The following theorem provides us with a closed form of Bayes estimator under L_1 , which depends only on $M(t)$, the marginal cumulative distribution function.

Theorem 4.1. *Assume model (4.1-1). Let $m(s_i^2)$ be a marginal density of S_i^2 and $M(s_i^2)$ be a corresponding cumulative distribution function. Then, given \mathbf{S}_i^2 , the Bayes*

rule of σ_i^2 based on the loss function L_1 is

$$\hat{\sigma}_{i,B}^2 = \frac{k}{2} \left[\frac{\int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-2)} dM(t)}{\int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-1)} dM(t)} - s_i^2 \right], \quad (4.2-3)$$

provided that both

$$\int_0^{\infty} t^{-(\frac{k}{2}-2)} dM(t) \quad \text{and} \quad \int_0^{\infty} t^{-(\frac{k}{2}-1)} dM(t)$$

exist.

It is worth emphasizing again that the Bayes estimator (4.2-3) does not specify any prior distribution for σ_i^2 . In that sense, the Bayes estimator (4.2-3) is very general. However, a comparison of the proposed Bayes estimator with the posterior mean under a specific prior enables us to gain insight about the proposed Bayes estimator.

Example 4.2. For $i = 1, \dots, p$, let $S_i^2 | \sigma_i^2 \stackrel{ind}{\sim} \sigma_i^2 \frac{\chi_k^2}{k}$ and $\sigma_i^2 \stackrel{iid}{\sim} IG(\alpha, \beta)$, with probability density function

$$\pi(\sigma_i^2) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \left(\frac{1}{\sigma_i^2} \right)^{\alpha+1} e^{-\frac{1}{\beta\sigma_i^2}}, \quad \sigma_i^2 > 0.$$

Note that

$$E_\pi [\sigma_i^2] = \frac{1}{\beta(\alpha-1)} \quad \text{and} \quad E_\pi [(\sigma_i^2)^2] = \frac{1}{\beta^2(\alpha-1)(\alpha-2)}$$

Then, from the posterior distribution, we have

$$E(\sigma_i^2 | s_i^2) = \frac{\frac{k}{2}s_i^2 + (\alpha - 1)\frac{1}{\beta(\alpha-1)}}{\frac{k}{2} + \alpha - 1} = B_1 s_i^2 + (1 - B_1)E_\pi(\sigma_i^2)$$

where $B_1 = \frac{\frac{k}{2}}{\frac{k}{2} + \alpha - 1}$. On the other hand, the proposed Bayes Estimator (4.2-3) for σ_i^2 is given by

$$\hat{\sigma}_{i,B}^2 = \frac{\frac{k}{2}s_i^2 + (\alpha - 2)\frac{1}{\beta(\alpha-2)}}{\frac{k}{2} + \alpha - 2} = B_2 s_i^2 + (1 - B_2)\frac{E_\pi[(\sigma_i^2)^2]}{E_\pi[(\sigma_i^2)]}$$

where $B_2 = \frac{\frac{k}{2}}{\frac{k}{2} + \alpha - 2}$.

The example above shows us that the proposed Bayes estimator is the weighted average of S_i^2 and the ratio of the second moment and the first moment of the prior distribution, while the posterior mean $E(\sigma_i^2 | S_i^2)$ is the weighted average of S_i^2 and the prior mean $E_\pi(\sigma_i^2)$. In case that the degree of freedom k is relatively smaller than the hyperparameter α , the proposed Bayes estimator shrinks S_i^2 towards $\frac{E_\pi[(\sigma_i^2)^2]}{E_\pi[(\sigma_i^2)]}$.

4.3 Nonparametric Bayes Estimator of the Variances (NEBV)

The Bayes decision rule (4.2-3) relies on the marginal cumulative distribution function $M(t)$ which can be easily and reliably estimated using the empirical distribution function $M_p(t) = \frac{1}{p} \sum_i I(S_i^2 \leq t)$. When replacing $M(t)$ with $M_p(t)$, we have

the following *Non-parametric Empirical Bayes estimator of the Variances (NEBV)*.

$$\begin{aligned}\hat{\sigma}_{i,NEBV}^2 &= \frac{k}{2} \left[\frac{\int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-2)} dM_p(t)}{\int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-1)} dM_p(t)} - s_i^2 \right] \\ &= \frac{k}{2} \left[\frac{\sum_{s_j^2 \geq s_i^2} (s_j^2)^{-(\frac{k}{2}-2)}}{\sum_{s_j^2 \geq s_i^2} (s_j^2)^{-(\frac{k}{2}-1)}} - s_i^2 \right].\end{aligned}\tag{4.3-4}$$

Let $A_i = \sum_{s_j^2 \geq s_i^2} (s_j^2)^{-(\frac{k}{2}-1)}$, and

$$a_{ij} = \begin{cases} (s_j^2)^{-(\frac{k}{2}-1)}, & \text{if } s_j^2 \geq s_i^2, \\ 0, & \text{otherwise.} \end{cases}$$

By setting $w_{ij} = \frac{a_{ij}}{A_i}$, we have $\sum_j w_{ij} = 1$ and the NEBV can be written as

$$\hat{\sigma}_{i,NEBV}^2 = \frac{k}{2} \left(\sum_j w_{ij} s_j^2 - s_i^2 \right).\tag{4.3-5}$$

Note that $w_{ij} > 0$ if and only if $s_j^2 \geq s_i^2$, implying that when estimating σ_i^2 , the estimator (4.3-5) only relies on the data s_j^2 's which are greater than or equal to s_i^2 . Second, for $k > 2$, w_{ij} decreases with respect to s_j^2 , implying that $\hat{\sigma}_{i,NEBV}^2$ puts less weight on those observations which are farther away from s_i^2 .

For a given selection rule, let $S_{(i)}^2$ be the sample variance of the i -th selected data. Then the corresponding variance, $\sigma_{(i)}^2$, is estimated by

$$\hat{\sigma}_{(i),NEBV}^2 = \frac{k}{2} \left(\sum_j w_{(i)j} s_j^2 - s_{(i)}^2 \right),\tag{4.3-6}$$

where $w_{(i)j}$ is defined similar to w_{ij} with s_i^2 being replaced by $s_{(i)}^2$.

We would like to point out that, due to the nature of the NEBV, it suffers when

estimating parameters with the largest sample variances. When implementing this method, we estimate those parameters corresponding to the top b (say $b = 5$) largest sample variances by these sample variances. Note that such parameters are usually not interesting to scientific investigators in most applications.

Recently, Efron revived, in *Efron* (2011), the importance of Tweedie's formula for estimating the mean parameter for a normal model where

$$X_i|\theta_i \sim N(\theta_i, \sigma^2), \quad \theta_i \sim \pi(\theta_i).$$

The non-parametric empirical Bayes estimator for the θ_i 's is

$$\hat{\theta}_{i,B} = E(\theta_i|X_i = x_i) = x_i + \sigma^2 \frac{m'(x_i)}{m(x_i)} \quad (4.3-7)$$

where $m(x_i)$ is the marginal density function of X_i . This method has been extended to the exponential family estimation in *Muralidharan* (2012).

When assuming Stein's loss function, basic calculation shows that

$$\hat{\sigma}'_{i,B} = \frac{1}{E\left(\frac{1}{\sigma_i^2} \middle| S_i^2\right)} = \left(\frac{k-2}{k s_i^2} - \frac{2}{k} \cdot \frac{m'(s_i^2)}{m(s_i^2)}\right)^{-1}.$$

The empirical Bayes version of this requires the estimation of the marginal density and its derivative. However, the estimation of the density function is a challenging problem itself and the performance of various techniques for density estimation further deteriorates at the tail. Additionally, choosing a tuning parameter optimally in kernel density estimation is a daunting task, not even to mention that the optimal tuning parameters for estimating the density and its derivative do not even agree.

On the contrary, the NEBV relies on the sample variances via its marginal cumulative distribution function, which can be easily estimated.

CHAPTER 5

ANALYTIC PROPERTIES FOR NEBV

5.1 Background

Since the NEBV proposed in (4.3-4) includes the ratio of two empirical processes, its analytic properties are connected with the theory of empirical process. For decades, empirical process theory has played an important role to enrich the classical limiting theorems such as LLN and CLT as a counterpart in the stochastic process. In particular, there is no doubt Glivenko-Cantelli theorem (*Glivenko*, 1933; *Cantelli*, 1933) and Donsker theorem (*Donsker*, 1952), which have been extended and applied to plenty of areas, belong to the seminal achievements igniting a series of researches of the empirical process. Glivenko-Cantelli theorem is associated with the statistic suggested by *Kolmogorov* (1933) and *Smirnov* (1939):

$$D_p = \sup_{t \in \mathbb{R}} \left| \mathbb{F}_p(t) - F(t) \right|,$$

where F is a distribution function and $\mathbb{F} = \frac{1}{p} \sum_{j=1}^p 1\{X_j \leq t\}$ is an empirical distribution function. D_p measures the supremum of the distance between \mathbb{F} and F and Glivenko-Cantelli theorem states not only D_p converges to zero almost surely but also it takes place uniformly over $t \in \mathbb{R}$.

Theorem 5.1. (*Glivenko-Cantelli, 1933*) *Let X_1, \dots, X_p be a random sample from a distribution function F and \mathbb{F} be an empirical distribution function. Then, for every*

$t \in \mathbb{R}$,

$$\sup_{t \in \mathbb{R}} \left| \mathbb{F}_p(t) - F(t) \right| \xrightarrow{a.s.} 0.$$

Glivenko-Cantelli theorem can be generalized beyond the distribution function. Let P be a distribution function and \mathbb{P}_p be an empirical measure on sample space \mathcal{X} such that $\mathbb{P}_p(A) = \frac{1}{p} \sum_{j=1}^p 1\{1 \leq j \leq p : X_j \in A\}$ with $A \in \mathcal{X}$. Suppose that \mathcal{F} is a collection of real-valued measurable functions f on \mathcal{X} . Then $\mathbb{P}_p f = \int f d\mathbb{P}_p$. This enables us to interpret, more generally, \mathbb{P}_p as a real-valued stochastic process indexed by \mathcal{F} . Based on this perspective, we introduce a well-known special class as follows.

Definition 5.2. (Glivenko-Cantelli class) A class \mathcal{F} of measurable functions $f : \mathcal{X} \rightarrow \mathbb{R}$ is said to be a P -Glivenko-Cantelli class if $\|\mathbb{P}_p - P\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{P}_p f - P f| \xrightarrow{a.s.} 0$.

Blum (1955), *DeHardt* (1971), *Vapnik and Chervonenkis* (1968), *Gine and Zinn* (1984) provided the methods finding *Glivenko-Cantelli class* described in Definition 5.2 based on the covering number, $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$, and the bracketing number, $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$. Let $\mathcal{F} = \{f : f(t) = 1(x \leq t), t \in \mathbb{R}\}$. Then, from Definition 5.2 and Glivenko-Cantelli Theorem, \mathcal{F} is a *P-Glivenko-Cantelli class*. In the next section, we will make use of the Definition 5.2 and relevant theory of *Glivenko-Cantelli class* to verify that, under some condition, the proposed estimator converges to the Bayes rule (4.2-3).

On the other hand, we have another important class, *Donsker class*, which provides the limiting distribution of empirical process. Let $\mathbb{G}_p = \sqrt{p}(\mathbb{P}_p - P)$ and \mathbb{G} be a mean zero empirical process with covariance $Pfg - PfPg$, for any $f, g \in \mathcal{F}$, both of which are the empirical processes indexed by some class of functions, \mathcal{F} . Then *Donsker class* is defined as follows.

Definition 5.3. (Donsker Class) Supposed that \mathcal{F} is a class of functions, for all t , $\sup_{f \in \mathcal{F}} |f - Pf| < \infty$. Then \mathcal{F} is a P -Donsker class if $\mathbb{G}_p \xrightarrow{d} \mathbb{G}$ in $l^\infty(\mathcal{F})$, where

\mathbb{G} is a tight Borel measurable random element.

Ossiander (1987) suggested an equivalent condition to being *P-Donsker class* based on the bracketing entropy integral. We also can find other important results on *P-Donsker class* from *Van der Vaart and Wellner* (1996), *Dudley* (1956) and *Talagrand* (1987). Recall $\mathcal{F} = \{f : f(t) = 1(x \leq t), t \in \mathbb{R}\}$ belongs to *Glivenko-Cantelli class*. In fact, we can see that it is also *P-Donsker class* from *Donsker* (1952). Furthermore, for $f(t) = 1(x \leq t)$ and $g(s) = 1(x \leq s)$, its limiting empirical process is the Brownian bridge with mean zero and covariance $F(t \wedge s) - F(t)F(s)$.

5.2 Asymptotic Behavior of NEBV

Let $f(t, u) = t^{-(k/2-2)}\mathbb{I}(t > u)$ and $g(t, u) = t^{-(k/2-1)}\mathbb{I}(t > u)$. Then the Bayes estimator and the NEBV can be written as

$$\hat{\sigma}_{i,B}^2 = \frac{k}{2} \left[\frac{\int_0^\infty f(t, s_i^2) dM(t)}{\int_0^\infty g(t, s_i^2) dM(t)} - s_i^2 \right],$$

and

$$\hat{\sigma}_{i,NEBV}^2 = \frac{k}{2} \left[\frac{\int_0^\infty f(t, s_i^2) dM_p(t)}{\int_0^\infty g(t, s_i^2) dM_p(t)} - s_i^2 \right],$$

respectively. First, we study the asymptotic behavior for the numerator and denominator separately.

Theorem 5.4. *Suppose that $\int_0^\infty t^{-(k/2-2)} dM(t) < \infty$ and $\int_0^\infty t^{-(k/2-1)} dM(t) < \infty$ and $M(t)$ is continuous, then*

$$\sup_u \left| \int_0^\infty f(t, u) dM_p(t) - \int_0^\infty f(t, u) dM(t) \right| \xrightarrow{a.s.} 0,$$

and

$$\sup_u \left| \int_0^\infty g(t, u) dM_p(t) - \int_0^\infty g(t, u) dM(t) \right| \xrightarrow{a.s.} 0,$$

This theorem implies that both the numerator and the denominator of NEBV converge to those of the Bayes estimator uniformly in S_i^2 .

Though both the numerator and the denominator of NEBV estimator converge to the corresponding part of Bayes rule uniformly, unfortunately, it does not guarantee that the ratio converges uniformly. The reason is that the denominator $\int_0^\infty g(t, u) dM(t)$ can be arbitrarily small when u goes to ∞ . For any arbitrary small number $\delta > 0$, let D^δ be a set defined as

$$D^\delta \equiv \left\{ u \mid \int_u^\infty t^{-(\frac{k}{2}-1)} dM(t) > \delta \right\}. \quad (5.2-1)$$

Then for any $u \in D^\delta$, there exists a constant λ such that the denominator in the Bayes rule is greater than or equal to λ . We then have the following theorem:

Theorem 5.5. *Assume the same condition as Theorem 5.4, then*

$$\sup_{s_i^2 \in \mathcal{D}^\delta} \left| \sigma_{i,NEBV}^2 - \sigma_{i,B}^2 \right| \xrightarrow{a.s.} 0.$$

Remark 5.6. Note that the set $D^\delta = (0, U(\delta))$ where $U(\delta)$ is decreasing with respect to δ . When $\delta \rightarrow 0$, $U(\delta) \rightarrow \infty$. We only show that $\hat{\sigma}_{i,NEBV}^2$ converges to $\hat{\sigma}_{i,B}^2$ uniformly on the set D^δ . This result is sufficient in many applications because the sample variances for those parameters which are selected, for instance using the T -statistic, are small with high probability.

An immediate corollary follows.

Corollary 5.7. *Assume the same condition as Theorem ??, then*

$$\sup_{s_{(i)}^2 \in D^\delta} |\hat{\sigma}_{(i),NEBV}^2 - \hat{\sigma}_{(i),B}^2| \xrightarrow{a.s.} 0.$$

This striking results shows that NEBV, when applied to the post-selection setting without modification, converges to the Bayes estimator as long as $S_{(i)}^2 \in D^\delta$. One of the key reason is that NEBV depends on the empirical distribution function and the empirical distribution converges to the distribution function uniformly according to the famous DKW's inequality (*Dvoretzky et al.*, 1956).

CHAPTER 6

APPLICATION WITH NUMERICAL ILLUSTRATIONS

In the analysis of the data with a large number of parameters, such as microarray experiments, a researcher is interested in investigating the most significant ones. Thus, in modern statistics, it is urgent to do the inference, for instance, construct confidence intervals for these selected parameters. Some approaches, such as that of *Giovannucci et al.* (1995) and *Rossouw et al.* (2002), treat this problem as if there is no selection. These approaches, however, result in severe selection bias and are thus misleading. Bonferroni procedure is valid, yet extremely conservative.

Another alternative is the Bayesian approach, which is known to be immune to the selection bias. (See *Dawid* (1994), *Senn* (2008), *Efron* (2011) and *Zhao and Sarkar* (2015)). In *Zhao and Hwang* (2012) and *Hwang and Zhao* (2013), they have studied the empirical Bayes confidence intervals for the selected parameters. It is shown that the proposed approach is much more informative in providing information for the selected parameters. However, all these approaches assume either known variances or a parametric prior for the variances.

In this section, we will briefly review the existing selective confidence intervals proposed in *Zhao and Hwang* (2012) and *Hwang and Zhao* (2013) and apply the NEBV to these approaches in order to show the advantages of the proposed method.

6.1 Construction of Confidence Intervals for Selected Parameters

Assume a normal-normal hierarchical model as

$$\begin{aligned} X_i | \theta_i &\stackrel{ind}{\sim} N(\theta_i, \sigma_i^2), \\ \theta_i &\stackrel{iid}{\sim} N(\mu, \tau^2), \end{aligned} \tag{6.1-1}$$

for $i = 1, \dots, p$. The posterior distribution of θ_i is given by

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

where $B = \frac{\tau^2}{\sigma_i^2 + \tau^2}$. The $100(1 - \gamma)\%$ credible interval is

$$CI_i^B = BX_i + (1 - B)\mu \pm Z_{\gamma/2} \sqrt{B\sigma_i^2}. \tag{6.1-2}$$

The hyper-parameters (μ, τ^2) can be estimated using the method of moments. For unknown σ_i^2 , the empirical Bayes estimator for the posterior mean, $E(\theta_i | X_i)$, is $\hat{B}X_i + (1 - \hat{B})\mu$ with $\hat{B} = \frac{\tau^2}{\hat{\sigma}_i^2 + \tau^2}$.

6.1.1 Estimation of Hyper-parameters

The natural estimator for μ is \bar{X} or the weighted average:

$$\hat{\mu} = \bar{X} \quad \text{or} \quad \hat{\mu} = \sum_{i=1}^p \frac{v_i X_i}{\sum_{i=1}^p v_i},$$

where $v_i = \frac{1}{\sigma_i^2}$. For estimating τ^2 , since we have the marginal distribution of X_i as $N(\mu, \sigma_i^2 + \tau^2)$ with $E(X_i - \mu)^2 = \sigma_i^2 + \tau^2$, we can estimate τ^2 as

$$\hat{\tau}^2 = \left(\frac{1}{p} \sum_{i=1}^p [(X_i - \mu)^2 - \sigma_i^2] \right)_+ = \left(\frac{1}{p} \sum_{i=1}^p (X_i - \mu)^2 - \frac{1}{p} \sum_{i=1}^p \sigma_i^2 \right)_+, \quad (6.1-3)$$

provided μ and σ_i^2 is known. When μ is unknown, the estimator of τ^2 approaches to (6.1-3) as p is large as follows. Under the same model as (6.1-1), we have the distribution of $\hat{\mu}$ and $X_i - \hat{\mu}$ as

$$\begin{aligned} \hat{\mu} &= \frac{1}{p} \sum_{i=1}^p X_i \sim N \left(\theta_i, \frac{1}{p^2} \sum_{i=1}^p (\tau^2 + \sigma_i^2) \right), \\ X_i - \hat{\mu} &\sim N \left(0, \left(\frac{p-1}{p} \right) \tau^2 + \left(\frac{p-1}{p} \right)^2 \sigma_i^2 + \frac{1}{p^2} \sum_{j=1, j \neq i}^p \sigma_j^2 \right), \end{aligned}$$

with

$$E(X_i - \hat{\mu})^2 = \left(\frac{p-1}{p} \right) \tau^2 + \left(\frac{p-1}{p} \right)^2 \sigma_i^2 + \frac{1}{p^2} \sum_{j=1, j \neq i}^p \sigma_j^2,$$

and

$$\tau^2 = \left(\frac{p}{p-1} \right) E \left[(X_i - \hat{\mu})^2 - \left(\frac{p-1}{p} \right)^2 \sigma_i^2 - \frac{1}{p^2} \sum_{j=1, j \neq i}^p \sigma_j^2 \right],$$

since

$$\begin{aligned} \text{Var}(X_i - \hat{\mu}) &= \text{Var}(X_i) + \text{Var}(\hat{\mu}) - 2\text{Cov}(X_i, \hat{\mu}) \\ &= (\tau^2 + \sigma_i^2) + \frac{1}{p^2} \sum_{i=1}^p (\tau^2 + \sigma_i^2) - 2\text{Cov} \left(X_i, \frac{X_i}{p} \right) \\ &= (\tau^2 + \sigma_i^2) + \frac{1}{p^2} \sum_{i=1}^p (\tau^2 + \sigma_i^2) - \frac{2}{p} (\tau^2 + \sigma_i^2) \\ &= \left(\frac{p-1}{p} \right) \tau^2 + \left(\frac{p-1}{p} \right)^2 \sigma_i^2 + \frac{1}{p^2} \sum_{j=1, j \neq i}^p \sigma_j^2. \end{aligned}$$

Here, note that $\text{Var}(X_i - \hat{\mu}) \rightarrow \tau^2 + \sigma_i^2$ as $p \rightarrow \infty$ under the assumption $E(\sigma_i^2) < \infty$. Now, for estimating τ^2 ,

$$\begin{aligned}\hat{\tau}^2 &= \left(\frac{p}{p-1}\right) \frac{1}{p} \sum_{i=1}^p \left[(X_i - \hat{\mu})^2 - \left(\frac{p-1}{p}\right)^2 \sigma_i^2 - \frac{1}{p^2} \sum_{j=1, j \neq i}^p \sigma_j^2 \right] \\ &= \left(\frac{p}{p-1}\right) \frac{1}{p} \sum_{i=1}^p (X_i - \hat{\mu})^2 - \left(\frac{p-1}{p}\right) \frac{1}{p} \sum_{i=1}^p \sigma_i^2 - \frac{1}{p^2(p-1)} \sum_{i=1}^p \sum_{j=1, j \neq i}^p \sigma_j^2.\end{aligned}$$

If p is large, since the last term above approaches to zero, we have

$$\begin{aligned}\hat{\tau}^2 &\doteq \left(\frac{p}{p-1}\right) \frac{1}{p} \sum_{i=1}^p (X_i - \hat{\mu})^2 - \left(\frac{p-1}{p}\right) \frac{1}{p} \sum_{i=1}^p \sigma_i^2 \\ &\doteq \frac{1}{p} \sum_{i=1}^p (X_i - \hat{\mu})^2 - \frac{1}{p} \sum_{i=1}^p \sigma_i^2.\end{aligned}\tag{6.1-4}$$

Hwang et al. (2009) estimated τ^2 by substituting $\hat{\sigma}_{i,EB}^2$ for $\hat{\sigma}_i^2$ in (6.1-4) but *Zhao* (2010) suggested using the unbiased estimator S_i^2 for $\hat{\sigma}_i^2$ rather than the biased estimator $\hat{\sigma}_{i,EB}^2$ in order to make $\hat{\tau}^2$ consistent.

Now, we construct the $100(1 - \gamma)\%$ confidence interval for θ_i as

$$\left(\hat{\theta}_i - z_{\gamma/2} \sqrt{\frac{\hat{\sigma}_i^2 \hat{\tau}^2}{\hat{\sigma}_i^2 + \hat{\tau}^2}}, \hat{\theta}_i + z_{\gamma/2} \sqrt{\frac{\hat{\sigma}_i^2 \hat{\tau}^2}{\hat{\sigma}_i^2 + \hat{\tau}^2}} \right),\tag{6.1-5}$$

where

$$\hat{\theta}_i = \frac{\hat{\tau}^2}{\hat{\sigma}_i^2 + \hat{\tau}^2} X_i + \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^2 + \hat{\tau}^2} \hat{\mu}.$$

Hence, $100(1 - \gamma)\%$ confidence interval for selected parameter, $\theta_{(i)}$ is given by

$$\left(\hat{\theta}_{(i)} - z_{\gamma/2} \sqrt{\frac{\hat{\sigma}_{(i)}^2 \hat{\tau}^2}{\hat{\sigma}_{(i)}^2 + \hat{\tau}^2}}, \hat{\theta}_{(i)} + z_{\gamma/2} \sqrt{\frac{\hat{\sigma}_{(i)}^2 \hat{\tau}^2}{\hat{\sigma}_{(i)}^2 + \hat{\tau}^2}} \right),\tag{6.1-6}$$

where

$$\hat{\theta}_{(i)} = \frac{\hat{\tau}^2}{\hat{\sigma}_{(i)}^2 + \hat{\tau}^2} X_{(i)} + \frac{\hat{\sigma}_{(i)}^2}{\hat{\sigma}_{(i)}^2 + \hat{\tau}^2} \hat{\mu}. \quad (6.1-7)$$

6.1.2 Truncation Method for Estimating τ^2

For constructing the confidence interval in (6.1-6), we encounter a problem in the case that τ^2 is very close to zero because it causes the length of the confidence interval unreasonably short. With very small τ^2 , the coverage probability of (6.1-6) is likely to be lower than the nominal level, and worse, τ^2 can be zero in some cases. Thus we need to adjust the estimator τ^2 so that we have, at least, a positive length of the confidence interval. A few truncation methods have been suggested in researches done by *He* (1992), *Qiu and Hwang* (2007) and *Hwang et al.* (2009). Here we use the method of truncation in *Hwang et al.* (2009).

First, since $X_i - \mu \sim N(0, \tau^2 + \sigma_i^2)$, we have $(X_i - \mu)^2 / (\tau^2 + \sigma_i^2) \sim \chi_1^2$ with $E(X_i - \mu)^2 = \tau^2 + \sigma_i^2$ and $\text{Var}(X_i - \mu)^2 = 2(\tau^2 + \sigma_i^2)$. Then, for large p , by CLT,

$$\sum_{i=1}^p (X_i - \mu)^2 \sim N\left(\sum_{i=1}^p (\tau^2 + \sigma_i^2), 2 \sum_{i=1}^p (\tau^2 + \sigma_i^2)\right),$$

approximately. Thus $\sum_{i=1}^p (X_i - \mu)^2 = \sum_{i=1}^p (\tau^2 + \sigma_i^2) + Z \sqrt{2 \sum_{i=1}^p (\tau^2 + \sigma_i^2)}$, where Z is a standard normal random variable. Now, let γ^* is a given significant level. From (6.1-3), we like to find τ^2 such that the probability of $\hat{\tau}^2 > 0$ is less than equal to γ^* .

$$\begin{aligned}
\gamma^* &\geq \Pr(\hat{\tau}^2 = 0) = \Pr\left(\frac{1}{p} \sum_{i=1}^p (X_i - \mu)^2 < \frac{1}{p} \sum_{i=1}^p \sigma_i^2\right) \\
&= \Pr\left(\frac{1}{p} \left\{ \sum_{i=1}^p (\tau^2 + \sigma_i^2) + Z \sqrt{2 \sum_{i=1}^p \tau^2 + \sigma_i^2} \right\} < \frac{1}{p} \sum_{i=1}^p \sigma_i^2\right) \\
&= \Pr\left(\tau^2 + \frac{1}{p} \sum_{i=1}^p \sigma_i^2 + \frac{Z}{p} \sqrt{2 \sum_{i=1}^p (\tau^2 + \sigma_i^2)} < \frac{1}{p} \sum_{i=1}^p \sigma_i^2\right) \\
&= \Pr\left(\tau^2 + \frac{Z}{p} \sqrt{2 \sum_{i=1}^p (\tau^2 + \sigma_i^2)} < 0\right) \\
&= \Pr\left(Z < -\frac{p\tau^2}{\sqrt{2 \sum_{i=1}^p (\tau^2 + \sigma_i^2)}}\right).
\end{aligned}$$

Therefore, we have

$$-\frac{p\tau^2}{\sqrt{2 \sum_{i=1}^p (\tau^2 + \sigma_i^2)}} \leq -z_{\gamma^*},$$

and

$$p^2(\tau^2)^2 - 2z_{\gamma^*}^2 p\tau^2 - 2z_{\gamma^*}^2 \sum_{i=1}^p \sigma_i^2 \geq 0.$$

By solving this inequality, since τ^2 should be positive, we obtain,

$$\tau^2 \geq \frac{1}{p} \left(z_{\gamma^*}^2 + z_{\gamma^*} \sqrt{z_{\gamma^*}^2 + 2 \sum_{i=1}^p \sigma_i^2} \right).$$

Hence, the truncated estimator for τ^2 is given by

$$\hat{\tau}^2 = \max\left(\hat{\tau}_0^2, \frac{1}{p} \sum_{i=1}^p (X_i - \mu)^2 - \frac{1}{p} \sum_{i=1}^p \sigma_i^2\right), \quad (6.1-8)$$

where

$$\hat{\tau}_0^2 = \frac{1}{p} \left(z_{\gamma^*}^2 + z_{\gamma^*} \sqrt{z_{\gamma^*}^2 + 2 \sum_{i=1}^p \hat{\sigma}_i^2} \right).$$

As mentioned above, we use an unbiased estimator S_i^2 for replacing $\hat{\sigma}_i^2$ with.

6.1.3 Confidence Interval Based on the Second Order Correction

The confidence interval constructing problem under normality assumption with a large sample has been discussed for decades and we can find considerable results from the articles including *Cohen and Sackrowitz* (1982), *Venter* (1988), *Venter and Steel* (1991), and *Qiu and Hwang* (2007). In particular, *Qiu and Hwang* (2007) suggested sharp individual and simultaneous intervals for the selected parameters of the populations with equal variances. In case that variances of the populations are not identical to each other, *Hwang and Zhao* (2013) developed the outstanding procedure for constructing the confidence interval for selected parameter by introducing a concept of the second order correction. This interval is constructed under the assumption (6.1-1) and (2.1-1). The confidence interval for $\theta_{(i)}$ is given by

$$\left(\hat{\theta}_{(i)} + L, \hat{\theta}_{(i)} + U \right), \quad (6.1-9)$$

where

$$L = -az_1 + (1 - e^{cz_2})b \quad \text{and} \quad U = az_1 + (1 - e^{-cz_2})b$$

with

$$a = \sqrt{\hat{A}_{1,i}^2 + \hat{A}_{2,i}^2}, \quad b = \frac{\tau^2 X_i}{(1 + \tau^2/\hat{\sigma}_i^2)(\hat{\sigma}_i^2 + \hat{\tau}^2)}, \quad c = \sqrt{\left(1 - \frac{(p-3)V}{\sum(\ln S_i^2 - \overline{\ln S_i^2})^2}\right)_+} V$$

$$A_{1,i} = \sqrt{\frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2}}, \quad A_{2,i} = \frac{\sigma_i^2 X_i}{p(\tau^2 + \sigma_i^2)} \sqrt{2 \sum (\tau^2 + \sigma_i^2)^2},$$

and z_1 and z_2 are the numbers such that $\Pr(|Z_1| < z_1, |Z_2| < z_2) = 1 - \gamma$ with two independent standard normal random variables Z_1 and Z_2 .

When σ_i^2 is finite for all i , the distribution of $\theta_i - \frac{\hat{\tau}^2}{(\hat{\tau}^2 + \sigma_i)} X_i$ has the same distribution as $A_{1,i}Z_1 + A_{2,i}Z_2 + O_p\left(\frac{\ln p}{p}\right)$. Since $A_{1,i}Z_1$ is of order $O_p\left(\frac{1}{\sqrt{p}}\right)$, the coverage probability for the confidence interval based on $A_{1,i}Z_1$ only is of order $O_p\left(\frac{1}{\sqrt{p}}\right)$, while the confidence interval constructed by (6.1-9) makes the order of the coverage probability equal to $1 - \gamma + O_p\left(\frac{\ln p}{p}\right)$ as $p \rightarrow \infty$.

6.2 Numerical Study

In this section, we conduct the simulation to investigate the performance of the proposed NEBV based on the confidence interval for the selected parameter. In order to find out the most significant case for non-zero θ_i , it is natural to focus on the largest t statistic, which is given by $t_i = \frac{|X_i|}{S_i}$. Let $\theta_{(p)}$ and $\sigma_{(p)}^2$ represent the θ_i and σ_i^2 corresponding to the largest t -statistic. For instance, $\theta_{(p)} = \theta_4$ and $\sigma_{(p)}^2 = \sigma_4^2$, if $\frac{|X_4|}{S_4} = \max_{1 \leq i \leq p} \left(\frac{|X_i|}{S_i}\right)$. After obtaining the statistic (X_i, S_i^2) , $i = 1, \dots, p$, we made the inference for the parameters $(\theta_{(p)}, \sigma_{(p)}^2)$ corresponding to the observation with the largest t -statistic in magnitude.

We considered the following three variance estimators: $\hat{\sigma}_{(p),NEBV}^2$, $\hat{\sigma}_{(p),ELJS}^2$ (Cui *et al.*, 2005), and $\hat{\sigma}_{(p),TW}^2$ (Tong and Wang, 2007). We wanted to point out that these methods are not designed for estimating the variance post the selection. Indeed, to the best of our knowledge, there is no methods designed for estimating the variances post the selection in the existing literature.

We first compared the risk of different estimators for estimating the variance $\sigma_{(p)}^2$. For the comparison of the average risk, we make use of the squared loss function

$L_1(\sigma^2, \hat{\sigma}^2) = (\frac{\sigma^2}{\hat{\sigma}^2} - 1)^2$ for $\sigma_{(p)}^2$ and calculate average risk for each estimator as

$$\text{Average Risk} = \frac{1}{q} \sum_{j=1}^q \left(\frac{\sigma_{(p),j}^2}{\hat{\sigma}_{(p),j}^2} - 1 \right)^2,$$

where q is the number of repeats of simulations.

Next, we considered the confidence interval for $\theta_{(p)}$. In addition to the three intervals by replacing $\hat{\sigma}_{(i)}^2$ as one of the three variance estimators above, we include the second-order corrected (SOC) interval (*Hwang and Zhao, 2013*) and Bonferroni's corrected T -interval $X_{(p)} \pm t_{k, \frac{\alpha}{2p}} S_{(p)}$. The SOC intervals are designed for the log-normal model where $\log S_i^2 | \log \sigma_i^2 \sim N(\log \sigma_i^2 + m_k, \sigma_k^2)$ and $\log \sigma_i^2 \sim N(\mu_v, \tau_v^2)$ where m_k and σ_k^2 are two constants. We set the desired coverage probability as 95%. For the estimators of the hyper parameters, $\bar{X} = \frac{1}{p} \sum_{i=1}^p X_i$ is used for $\hat{\mu}$ and the truncated estimator in (6.1-8) is used for $\hat{\tau}^2$ by replacing $\hat{\sigma}_i^2$ with the unbiased estimator S_i^2 .

We compare three quantities: (1) the average risks of the estimators for $\sigma_{(p)}^2$ in log-scale. (2) the coverage probabilities of the confidence intervals for $\theta_{(p)}$ (3) the average lengths of confidence intervals for $\theta_{(p)}$. The average lengths of confidence intervals are compared by using the ratio of the length of each interval considered to the length of the Bonferroni interval.

6.2.1 Independent Case

In this section, we generate (X_i, S_i^2) 's according to following model

$$\begin{aligned} X_i | \theta_i, \sigma_i^2 &\stackrel{ind}{\sim} N(\theta_i, \sigma_i^2), & \theta_i &\stackrel{iid}{\sim} N(\mu, \tau^2), \\ kS_i^2 | \sigma_i^2 &\stackrel{ind}{\sim} \sigma_i^2 \chi_k^2, & \sigma_i^2 &\stackrel{iid}{\sim} \pi(\cdot), \end{aligned} \tag{6.2-10}$$

with different choices of the prior for σ_i^2 's including Gamma ($Gamma(\alpha, \beta)$), inverse Gamma ($IG(\alpha, \beta)$), mixture of Gamma/inverse Gamma prior, and lognormal

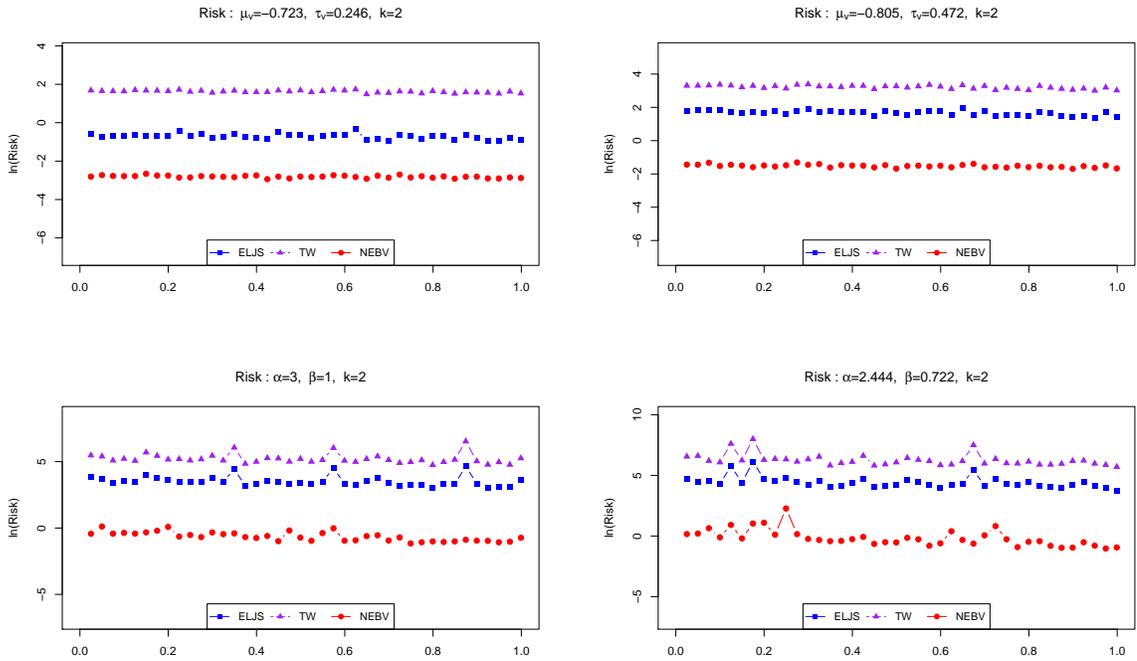


Figure 6.1: Plots of $\ln(\text{Risk})$ for estimating σ_i^2 under lognormal (the first row) and inverse gamma prior (the second row) for σ_i^2 against $\frac{\tau^2}{\tau^2 + E(\sigma_i^2)}$: From upper-left, each plot corresponds to $CV = 0.25, 0.50, 1.00$ and 1.50 respectively with $E(\sigma_i^2) = 0.5$.

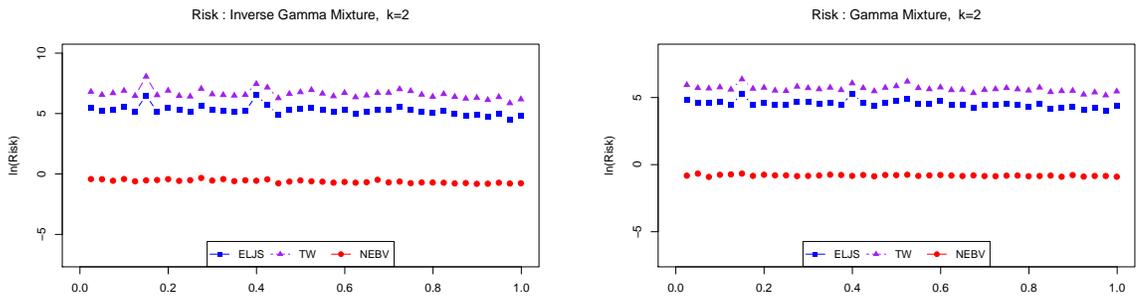


Figure 6.2: Plots of $\ln(\text{Risk})$ for estimating σ_i^2 under Inverse Gamma and Gamma mixture prior for σ_i^2 : $0.5 * IG(6, 1) + 0.5 * IG(6, 4)$ for left panels and $0.5 * Gamma(4, 1/16) + 0.5 * Gamma(6, 1/8)$ for right panels, respectively with $E(\sigma^2) = 0.5$.

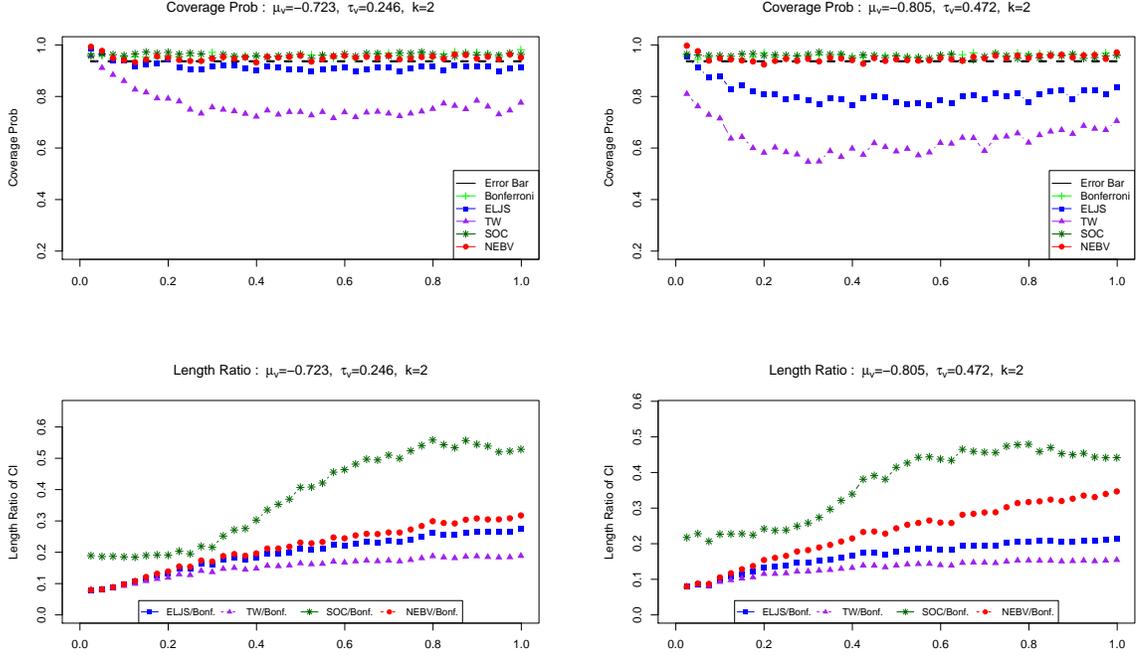


Figure 6.3: Plots of the coverage probabilities (the first row) and the average length ratio (the second row) for various intervals of $\theta_{(p)}$ for lognormal prior against $\frac{\tau^2}{\tau^2 + E(\sigma_i^2)}$. The CV of the left and right column are $CV = 1.00$ and 1.50 respectively. The error bar represents $0.95 - 2 \times (\text{simulation error})$.

($LN(\mu_v, \tau_v^2)$) prior. With gamma prior, we set α , the shape parameter as 16, 4, 1 and $\frac{4}{9}$, respectively, and β , the scale parameter, as $\frac{1}{2\alpha}$ so that we have $E(\sigma_i^2) = 0.5$ and the corresponding coefficient of variations (CV) 0.25, 0.5, 1.0 and 1.5, respectively. For inverse gamma prior and lognormal prior, we set the hyper-parameters such that $E(\sigma_i^2) = 0.5$ and $CV = 0.25, 0.5, 1.0$ and 1.5. With the mixture priors, σ_i^2 's are generated from $0.5 * IG(6, 1) + 0.5 * IG(6, 4)$ and $0.5 * Gamma(4, \frac{1}{16}) + 0.5 * Gamma(6, \frac{1}{8})$ so that we have $E(\sigma_i^2) = 0.5$. The dimension for the numerical study is set as $p = 3,000$ and the number of replication $q = 1,000$ with $k = 2$, the degree of freedom for chi-square distribution in (6.2-10).

The logarithm of the risks based on the loss function (3.1-5) are reported in Figure 6.1. It is clear that the proposed NEBV has the smallest risk for all parameter settings.

Figure 6.3 shows the simulated results of the lognormal prior for σ_i^2 's. The cov-

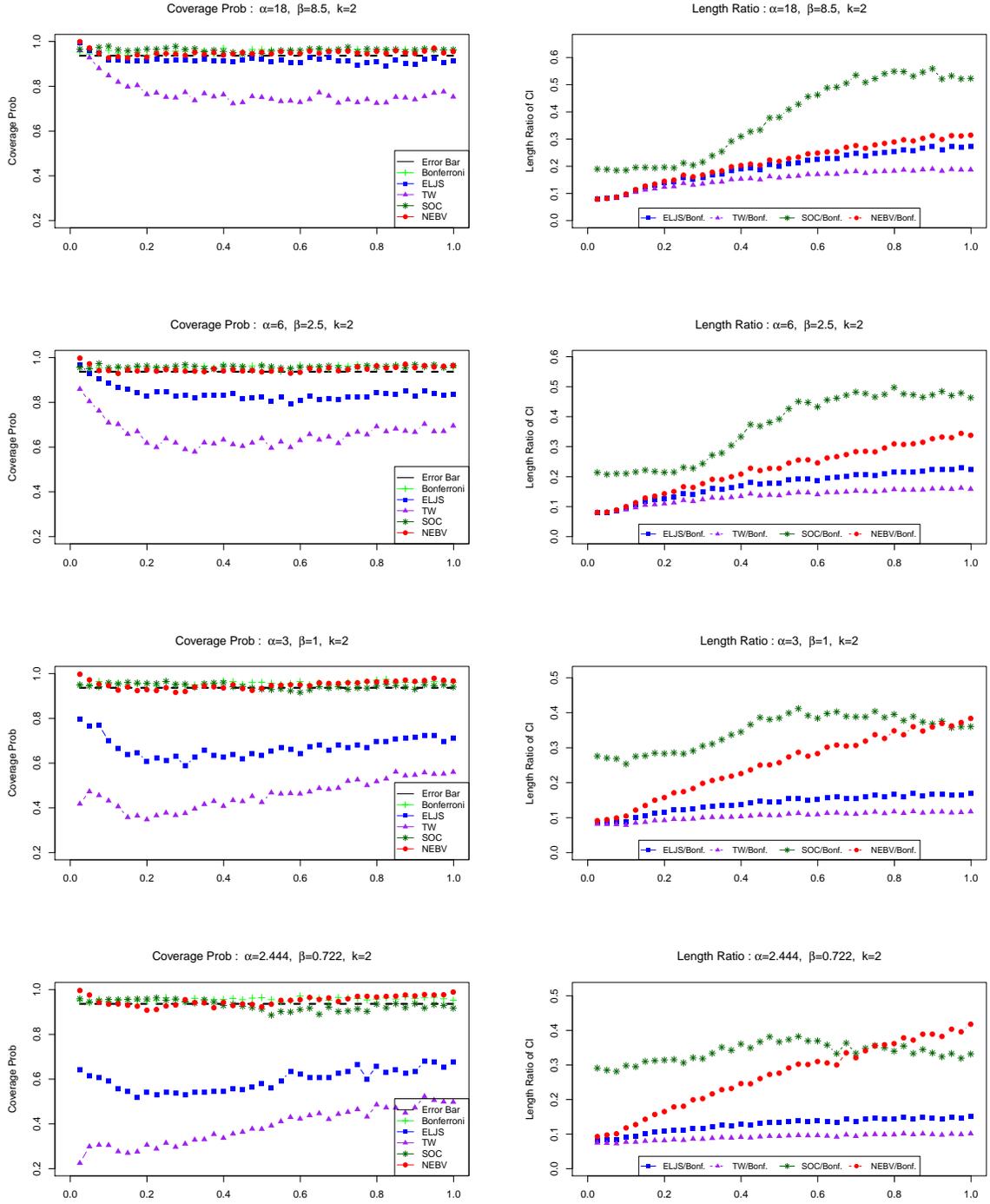


Figure 6.4: Plots of the coverage probabilities (left panels) and the average length ratio (right panels) for various intervals of $\theta_{(p)}$ under inverse gamma prior. The CV of the left and right columns are 1.00 and 1.50 respectively.

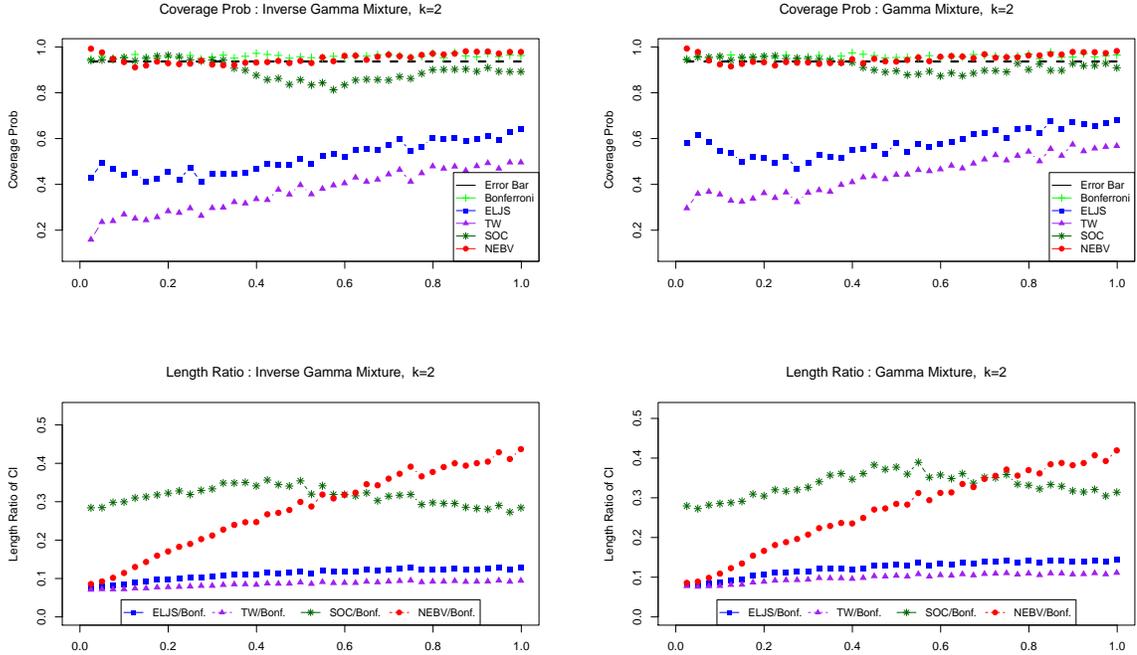


Figure 6.5: Plots of the coverage probabilities and the length ratio of confidence intervals for $\theta_{(p)}$ under Inverse Gamma and Gamma mixture prior for σ_i^2 : $0.5 * IG(6, 1) + 0.5 * IG(6, 4)$ for left panels and $0.5 * Gamma(4, 1/16) + 0.5 * Gamma(6, 1/8)$ for right panels, respectively with $E(\sigma^2) = 0.5$.

coverage probabilities are reported in the first row of the figure. The intervals based on NEBV, the Bonferroni's corrected t -interval, and SOC intervals guarantee a good coverage probability for all hyperparameters, and the intervals based on the other two estimators, ELJS and TW, have substantially low coverage probabilities. For instance, when using the TW's estimator for the lognormal model with $\mu_v = -0.805$ and $\tau_v = 0.472$, the coverage probability can be as low as 60%. Figure 6.4 shows the similar behavior of ELJS and TW intervals with low coverage probability under inverse gamma prior. We also reported the average ratios of the lengths of all the confidence intervals to the length of t -interval based on the Bonferroni's correction in the second row of the figure 6.3. It is seen that among these three intervals which guarantee a desired coverage probability, the one based on NEBV is the shortest one.

In Figure 6.4, we reported the coverage probability and average length ratio for the

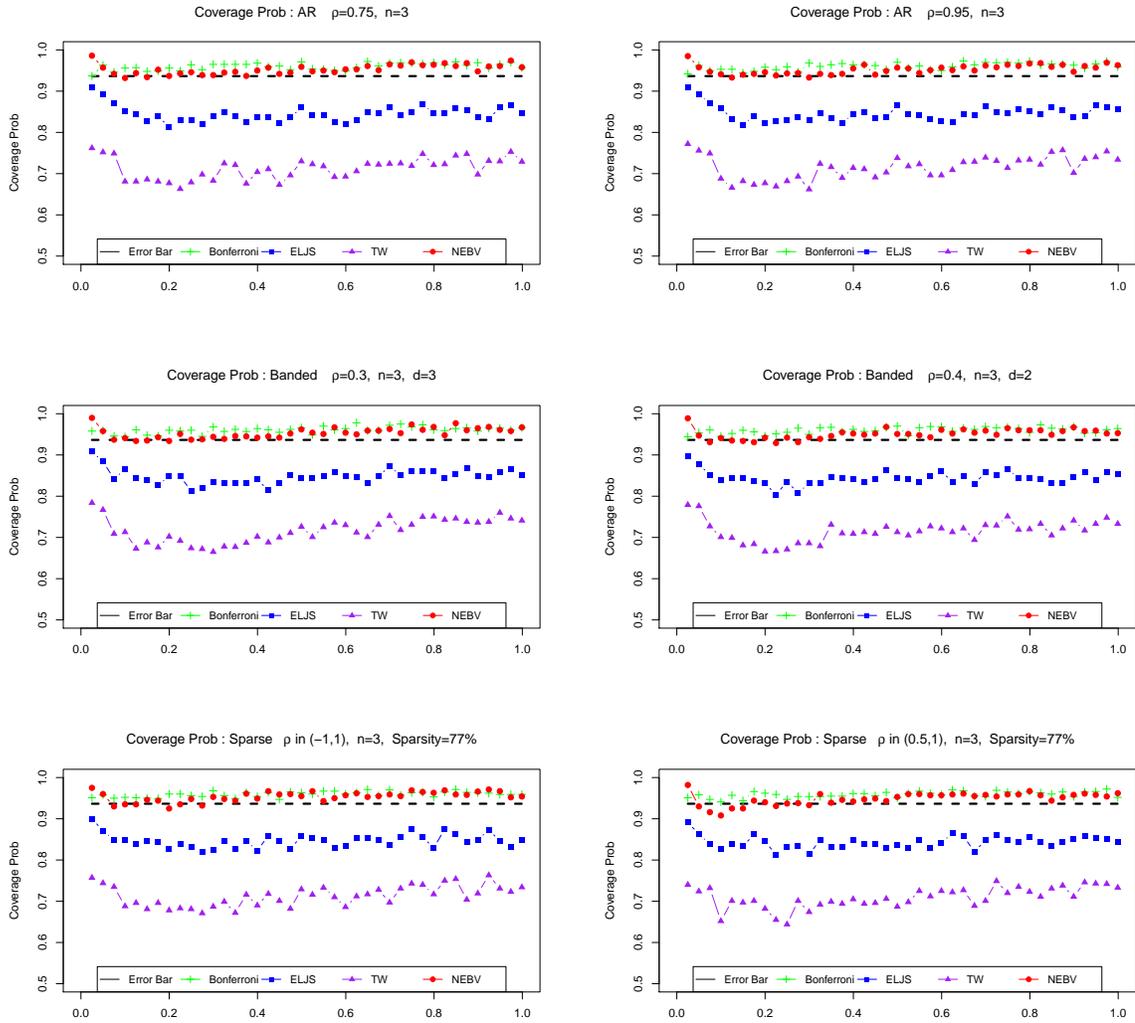


Figure 6.6: Plots of the coverage probabilities of four intervals with AR, Banded and Sparse correlation matrices.

various approach for the inverse gamma prior. The intervals based on NEBV and the Bonferroni's corrected t -interval still have good coverage probability. However, the coverage probability of SOC interval dropped below the nominal level. SOC intervals are designed to work well for the lognormal-model and does not work well when this assumption fails.

In Figure 6.5, the variances σ_i^2 's are generated from a mixture Gamma/inverse Gamma distribution. Although SOC intervals can be shorter than the NEBV-based

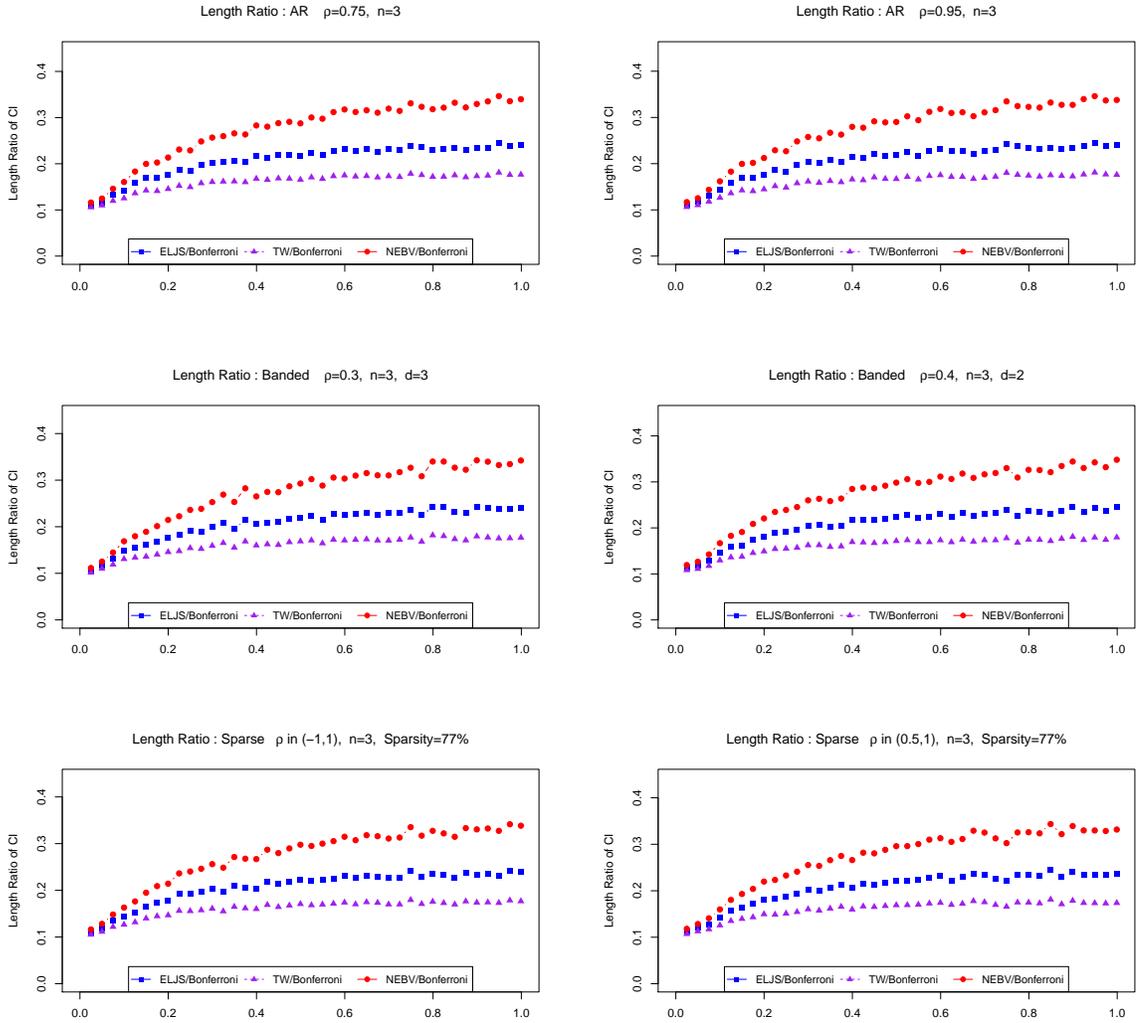


Figure 6.7: Plots of the average length ratio of four intervals with AR, Banded and Sparse correlation matrices.

intervals, but the coverage probability of SOC intervals is much lower than 95% for some parameter settings. The NEBV-based intervals are robust to the prior of σ_i^2 's and still works well.

6.2.2 Dependent Case

The NEBV estimator and the confidence interval (6.1-5) are based on the independence assumption. In this section, we assess their performance when the independence

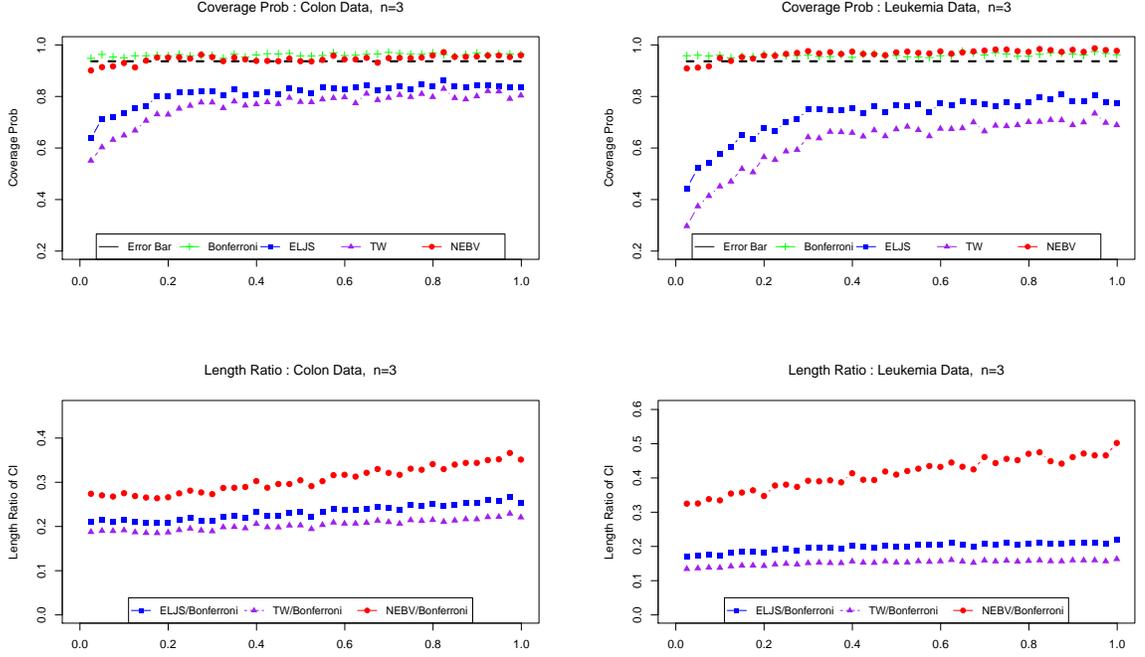


Figure 6.8: Plots of the coverage probabilities, length ratio of confidence intervals for $\theta_{(p)}$ based on the Colon Cancer and Leukemia Data.

assumption is violated. We generate the data independently from the multivariate normal distribution, namely, for $i = 1, 2, \dots, n$,

$$\mathbf{X}_{i,\cdot} = (x_{i1}, \dots, x_{ip}) \mid \boldsymbol{\theta}, \Sigma \sim N(\boldsymbol{\theta}, \Sigma), \quad \theta_i \stackrel{iid}{\sim} N(\mu, \tau^2). \quad (6.2-11)$$

Here, the variance-covariance matrix Σ is given as $D^{1/2}RD^{1/2}$ where $D = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ with $\sigma_i^2 \stackrel{iid}{\sim} IG(\alpha, \beta)$. Four types of correlation matrix R are considered:

(i) Banded Correlation matrix:

$$\begin{cases} \rho_{ij} = 0, & \text{if } |j - i| > d, \\ \rho_{ij} = \rho, & \text{if } 0 < |j - i| \leq d \end{cases}$$

(ii) AR Correlation matrix : $\rho_{ij} = \rho^{|i-j|}$

(iii) Sparse Correlation matrix

In the banded case, (d, ρ) are set as $(5, 0.1)$, $(4, 0.2)$, $(3, 0.3)$, $(2, 0.4)$, and $(1, 0.5)$ respectively. For the AR case, ρ is chosen as 0.5, 0.75, 0.9, 0.95, and 0.99 respectively with $n = 3, 6$ and 10. To get a sparse correlation matrix, we start with a $p \times p$ matrix A by setting 1.5% of randomly chosen entries as numbers from $uniform(a, b)$ and setting the rest as zero. We then convert all the diagonal elements to one. Let $R = B^{-\frac{1}{2}} A' A B^{-\frac{1}{2}}$ where B is a diagonal matrix with the diagonal-line elements being the diagonal line of $A' A$. Approximately, 77% of the entries of sparse correlation matrix R are zero. We consider the following choices of (a, b) : $(-1, 1)$, $(0, 1)$, $(0.5, 1)$, $(0.9, 1)$, and $(0.9, 0.9)$. We set $p = 1000$ and $q = 1000$.

We have done extensive simulation studies and only report representative results in Figure 6.6. Under all these dependence, the coverage probability of the NEBV-based intervals for $\theta_{(p)}$ guarantee the desired coverage probability. On the other hand, the intervals based on ELJS and TW fail to achieve the satisfactory coverage probability. In Figure 6.7, we reported the average ratio of the length.

6.2.3 Correlation Matrix generated from Real data

In the last case, we consider a correlation matrix estimated from real data from two microarray experiments: Leukemia (*Golub et al.*, 1999) and Colon Cancer (*Alon et al.*, 1999). The leukemia data includes the expressions of genes extracted from patients with two types of leukemia ($p = 7129$): Acute Lymphoblastic Leukemia (ALL, 47 patients) and Acute Myeloid Leukemia (AML, 25 patients). The colon cancer data contains gene expressions of 2000 genes for 22 patients normal people and 40 patients. Since leukemia data and colon cancer data have 72 and 62 observations with 7128 and 2000 variables, respectively. We randomly select 70 and 60 variables respectively from each dataset and calculate the sample covariance matrices which are treated as Σ in the simulation.

Method	Mean	Median	STD
<i>Leukemia</i>			
S^2	0.211	0.209	0.0164
LJS	0.166	0.164	0.0154
TW	0.169	0.167	0.0159
$NEBV$	0.149	0.146	0.0144
<i>Colon</i>			
S^2	0.238	0.232	0.0391
LJS	0.176	0.170	0.0357
TW	0.178	0.172	0.0366
$NEBV$	0.156	0.152	0.0333

Table 6.1: Percentage of discordant genes for various intervals.

The results of coverage probabilities, and length of the intervals for $\theta_{(p)}$ are shown in Figure 6.8. What we have observed is similar to that for the other settings. The intervals based on NEBV guarantees a desired coverage probability; while the coverage probabilities of the intervals based on ELJS and TW could be much lower than the nominal level.

6.2.4 Real data Analysis

According to a 2016 poll of 1,500 scientists reported in the journal *Nature*, 70% of them had failed to reproduce at least one other scientist’s experiment (*Baker (2016)*). *Replication crisis*, which has long-standing roots, catches growing awareness in the last decades. A natural question is in what degree, the statistical conclusion based on certain methods agrees when provided with different/new data set. In this section, we use the above two microarray data to examine the performance of various statistical methods.

For the Leukemia data set, we first randomly split ALL and AML group into two subgroups with approximately the same size. We then constructed confidence intervals based on one sub-group. We declared a gene to be significant if the corresponding interval does not enclose zero. We did the same for the other sub-group. We call a

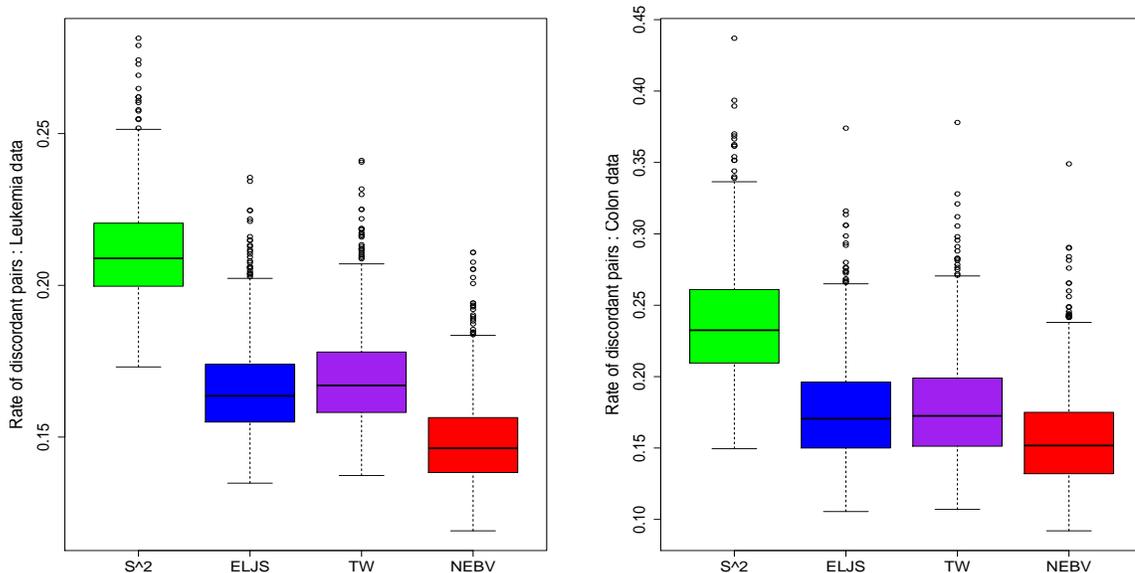


Figure 6.9: Distributions for the rate of misclassified rates of 1000 iterations for leukemia and colon cancer data

gene *discordant* if the intervals based on these two sub-groups lead to contradicting conclusion. Namely, one interval enclose zero while the other does not. A gene is *discordant* implies that the conclusion that it is significant in one sample can not be replicated in the second sample. We replicated these steps 1,000 times to calculate the average proportions of discordant genes. We did the same thing for the colon cancer data and reported the result in Table 6.1 and Figure 6.9.

In Figure 6.9, we plot the box-plot based on four interval constructions: intervals based on ELJS, the sample variance, TW, and NEBV. In Table 6.1, we summarize the summary statistic based on 1,000 replications. It is seen that the intervals based on NEBV commits the smallest number of *discordance* and the difference is statistical significant.

In summary, as shown in both simulations and real data analysis, the performance of the intervals based on NEBV is better than its competitors and is thus recommended.

CHAPTER 7

A GENERALIZATION OF NEBV

7.1 Tweedie's Formula

Efron (2011) considered a non-parametric empirical Bayes estimator for a natural parameter in exponential family. Assume, for $i = 1, \dots, p$,

$$\eta_i \sim \pi(\cdot) \quad \text{and} \quad f(x_i|\eta_i) = e^{\eta_i x_i - A(\eta_i)} f_0(x_i).$$

Here η_i is the natural parameter of the family with and $f_0(z_i)$ the density when $\eta_i = 0$. Then the posterior is given by

$$g(\eta_i|x_i) = \frac{f(x_i|\eta_i)}{f(x_i)} = e^{\eta_i x_i - \lambda(x_i)} [\pi(\eta_i) e^{-A(\eta_i)}],$$

where $f(x_i) = \int_{\mathcal{X}} f(x_i|\eta_i)\pi(\eta_i)d\eta_i$ and $\lambda(z_i) = \ln\left(\frac{f(x_i)}{f_0(x_i)}\right)$. Hence the posterior mean and variance of η_i given z_i is

$$E(\eta_i|x_i) = \lambda'(x_i) = \frac{f'(x_i)}{f(x_i)} - \frac{f'_0(x_i)}{f_0(x_i)} \quad \text{and} \quad V(\eta_i|x_i) = \lambda''(x_i).$$

In particular, for the normal family, $X_i|\eta_i \sim N(\theta_i, \sigma^2)$ with $\eta_i = \theta_i/\sigma^2$, we have the posterior mean as

$$E(\theta_i|x_i) = x_i + \sigma^2 \frac{f'(x_i)}{f(x_i)}.$$

Following example gives us another application of Tweedie's formula for estimating σ_i^2 . We can find the posterior mean of $1/\sigma_i^2$ under chi-squared model.

Example 7.1. Let $S_i^2 | \sigma_i^2 \sim \sigma_i^2 \frac{\chi_k^2}{k}$. Then the pdf of S_i^2 given σ_i^2 is

$$f(s_i^2 | \sigma_i^2) = \frac{\left(\frac{ks_i^2}{\sigma_i^2}\right)^{\frac{k}{2}-1} e^{-\frac{ks_i^2}{2\sigma_i^2}}}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}} \cdot \frac{k}{\sigma_i^2} = e^{\eta_i s_i^2 + \frac{k}{2} \ln(1-2\eta_i)} \frac{(s_i^2)^{\frac{k}{2}-1} e^{-\frac{s_i^2}{2}}}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}},$$

where $\eta_i = -\frac{1}{2} \left(\frac{k}{\sigma_i^2} - 1\right)$, $A(\eta_i) = -\frac{k}{2} \ln(1-2\eta_i)$, and $f_0(s_i^2) = \frac{(s_i^2)^{\frac{k}{2}-1} e^{-\frac{s_i^2}{2}}}{\Gamma(\frac{k}{2}) 2^{\frac{k}{2}}}$. Thus the posterior mean of η_i given s_i^2 is

$$E(\eta_i | s_i^2) = \frac{f'(s_i^2)}{f(s_i^2)} - \frac{f'_0(s_i^2)}{f_0(s_i^2)} = \frac{1}{2} - \frac{k-2}{2s_i^2} + \frac{f'(s_i^2)}{f(s_i^2)},$$

and it turns out

$$\hat{\eta}_i = \left[E\left(\frac{1}{\sigma_i^2} \mid s_i^2\right) \right]^{-1} = \left[\frac{k-2}{k} \left(\frac{1}{s_i^2}\right) + \frac{2}{k} \left(-\frac{f'(s_i^2)}{f(s_i^2)}\right) \right]^{-1}.$$

Note that the posterior mean of $\frac{1}{\sigma_i^2}$ given s_i^2 is a form of the weighted average of $\frac{1}{s_i^2}$ and $-\frac{f'(s_i^2)}{f(s_i^2)}$ and $\hat{\eta}_i$ is the Bayes decision rule under Stein's loss function L_3 .

7.2 A Generalized Form of NEBV

As one of the non-parametric empirical Bayes estimators, NEBV has the generalized form as Tweedie's formula does. For $i = 1, \dots, p$, consider an exponential family with a form of

$$f(x_i | \eta_i) = e^{-\frac{x_i}{\eta_i} - A(\eta_i)} h(x_i), \quad x_i > 0, \quad (7.2-1)$$

with $\eta_i > 0$ and an arbitrary prior $\pi(\cdot)$ for η_i .

Theorem 7.2. *Assume model (7.2-1). Let $m(x_i)$ be a marginal density of X_i and $M(x_i)$ be a corresponding cumulative distribution function. Then, given x_i , the Bayes rule of η_i based on the loss function L_1 is*

$$\hat{\eta}_{i,B} = \frac{\int_{x_i}^{\infty} \frac{t}{h(t)} dM(t)}{\int_{x_i}^{\infty} \frac{1}{h(t)} dM(t)} - x_i, \quad (7.2-2)$$

provided that both

$$\int_0^{\infty} \frac{t}{h(t)} dM(t) \quad \text{and} \quad \int_0^{\infty} \frac{1}{h(t)} dM(t)$$

exist.

Furthermore, based on the Theorem 7.2, we have a generalized version of NEBV for θ_i as

$$\hat{\eta}_{i,NEBV} = \frac{\int_{x_i}^{\infty} \frac{t}{h(t)} dM_p(t)}{\int_{x_i}^{\infty} \frac{1}{h(t)} dM_p(t)} - x_i. \quad (7.2-3)$$

Example 7.3. Let z_i be an exponential random variable with parameter λ_i , for $1, \dots, p$. Then $f(z_i|\lambda_i) = \frac{1}{\lambda_i} \exp\left(-\frac{z_i}{\lambda_i}\right)$ with $\eta_i = \lambda_i$, and $h(z_i) = 1$. Hence

$$\hat{\lambda}_{i,NEBV} = \frac{\int_{z_i}^{\infty} t dM_p(t)}{\int_{z_i}^{\infty} dM_p(t)} - z_i = \frac{\sum_{z_j > z_i} z_j}{\sum_{z_j > z_i} \mathbb{I}(z_j > z_i)} - z_i.$$

Moreover, the following example shows that we can regard the Bayes estimator (4.2-3) as a special case of (7.2-3).

Example 7.4. Let $S_i^2 | \sigma_i^2 \sim \sigma_i^2 \frac{\chi_k^2}{k}$. Then the pdf of S_i^2 given σ_i^2 is

$$f(s_i^2 | \sigma_i^2) = \frac{\left(\frac{ks_i^2}{\sigma_i^2}\right)^{\frac{k}{2}-1} e^{-\frac{ks_i^2}{2\sigma_i^2}}}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}} \cdot \frac{k}{\sigma_i^2},$$

where $\eta_i = \frac{2\sigma_i^2}{k}$, and $h(s_i^2) = \frac{(s_i^2)^{\frac{k}{2}-1} k^{\frac{k}{2}}}{\Gamma(\frac{k}{2}) 2^{\frac{k}{2}}}$. Thus

$$\hat{\eta}_{i,B} = \frac{\int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-2)} dM_p(t)}{\int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-1)} dM_p(t)} - s_i^2,$$

and

$$\hat{\sigma}_{i,B}^2 = \frac{k}{2} \left[\frac{\int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-2)} dM(t)}{\int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-1)} dM(t)} - s_i^2 \right].$$

On the other hand, when considering another form of exponential family with a pdf as

$$f(x_i | \eta_i) = e^{\eta_i \ln x_i - A(\eta_i)} h(x_i), \quad x > 0, \quad (7.2-4)$$

with $\eta_i > 0$ and an arbitrary prior $\pi(\cdot)$ for η_i . Note that, in (7.2-6), the sufficient statistic for η_i is $\ln x_i$. The next theorem shows that we have a Bayes estimator for η_i with a slightly different form.

Theorem 7.5. *Under the same condition as Theorem 7.2, the Bayes rule of η_i based on the loss function L_1 is*

$$\hat{\eta}_{i,B}^{-1} = x_i - \frac{\int_0^{x_i} \frac{\ln t}{h(t)} dM(t)}{\int_0^{x_i} \frac{1}{th(t)} dM(t)}. \quad (7.2-5)$$

Therefore, we have a corresponding empirical Bayes estimator for η_i as

$$\hat{\eta}_{i,NEBV}^{-1} = x_i - \frac{\int_0^{x_i} \frac{\ln t}{h(t)} dM_p(t)}{\int_0^{x_i} \frac{1}{th(t)} dM_p(t)}. \quad (7.2-6)$$

In (7.2-6), the NEBV for η_i depends on x'_j 's, $i = 1, \dots, p$, which are less than or equal to x'_i 's.

CHAPTER 8

CONCLUSION

We develop a novel method of non-parametric empirical Bayes estimation of the variances assuming an arbitrary prior distribution. The Bayes estimator derived under invariant squared loss function L_1 relies on the data via the marginal distribution function of sample variance only. The NEBV, the empirical Bayes version, thus, relies on the empirical distribution function. As a result, it avoids any estimation of the density function, a known challenging problem, and converges to Bayes estimator over a large set of the selected data with a high probability.

We applied such an estimator to the construction of confidence intervals in the post selection setting. Extensive numerical study which covers simulations and real data analysis have shown noticeable evidences that NEBV provides much better inference procedures.

NEBV is an empirical Bayes estimator designed to estimate multiple variances simultaneously. In the numerical study, the interval based on the proposed estimator NEBV outperforms other intervals based on the existing variance estimators under a few of well-known dependent data structures. However, beyond the estimation of variances, it would be an interesting problem to find a non-parametric empirical Bayes estimator for the variance-covariance matrix. In addition, in this work, the inference for the mean and variance parameters are splitted as two steps. Another natural yet challenging question is how to obtain a non-parametric estimator of the means assuming unknown and unequal variances. We leave all these in future research.

APPENDICES

Technical Details

Proof of Proposition 3.1.

The marginal distribution of S_i^2 is

$$f(s_i^2) = \frac{\beta k}{2} \frac{\Gamma\left(\frac{k}{2} + \alpha\right)}{\Gamma\left(\frac{k}{2}\right) \Gamma(\alpha)} \left(\frac{\beta k}{2} s_i^2\right)^{\frac{k}{2}-1} \left(\frac{\beta k}{2} s_i^2 + 1\right)^{-\frac{k}{2}-\alpha}, \quad 0 < s_i^2 < \infty.$$

Let $Y = \frac{\frac{\beta k}{2} S_i^2}{\frac{\beta k}{2} S_i^2 + 1}$, then

$$f(y_i) = \frac{\Gamma\left(\frac{k}{2} + \alpha\right)}{\Gamma\left(\frac{k}{2}\right) \Gamma(\alpha)} y_i^{\frac{k}{2}-1} (1 - y_i)^{\alpha-1}, \quad 0 < y_i < 1.$$

In other words, $Y \sim \text{Beta}\left(\frac{k}{2}, \alpha\right)$. Let $Y_{(1)}^p$ the smallest order statistics among Y_1, \dots, Y_p .

It suffices to show that $Y_{(1)}^p \rightarrow 0$ a.s.

For any $\epsilon > 0$,

$$P(Y_{(1)}^p > \epsilon) = \prod_{i=1}^p P(Y_i > \epsilon) = a(\epsilon)^p,$$

where $0 < a(\epsilon) < 1$. Consequently,

$$\sum_{p=1}^{\infty} P(Y_{(1)}^p > \epsilon) = \sum_{p=1}^{\infty} a(\epsilon)^p < \infty.$$

By Borel-Cantelli Lemma, $Y_{(1)}$ converges to zero almost surely. \square

Proof of Theorem 4.1.

To ease the notation, we drop the subscript “ i ” in the proof. Recall that $f(s^2|\sigma^2)$ and $\pi(\sigma^2)$ are the density function of $s^2|\sigma_i^2$ and the prior of σ^2 respectively. Define $m(s^2)$, $n(s^2)$ and $h(s^2)$ as

$$m(s^2) = \int_0^\infty f(s^2|\sigma^2)\pi(\sigma^2) d\sigma^2, \quad (.1)$$

$$n(s^2) = \int_0^\infty \sigma^2 f(s^2|\sigma^2)\pi(\sigma^2) d\sigma^2, \quad (.2)$$

and

$$h(s^2) = \int_0^\infty (\sigma^2)^2 f(s^2|\sigma^2)\pi(\sigma^2) d\sigma^2. \quad (.3)$$

Note that $m(s^2)$ is the marginal distribution of S^2 . Then

$$\hat{\sigma}_B^2 = \frac{E[(\sigma^2)^2|S^2]}{E[\sigma^2|S^2]} = \frac{\int_0^\infty (\sigma^2)^2 \pi(\sigma^2|s^2) d\sigma^2}{\int_0^\infty \sigma^2 \pi(\sigma^2|s^2) d\sigma^2} = \frac{h(s^2)}{n(s^2)}.$$

Take the derivatives of $n(s^2)(s^2)^{-(\frac{k}{2}-1)}$ with respect to s^2 ,

$$\left[n(s^2)(s^2)^{-(\frac{k}{2}-1)} \right]' = -\frac{k}{2} m(s^2)(s^2)^{-(\frac{k}{2}-1)}. \quad (.4)$$

Consequently,

$$n(s^2)(s^2)^{-(\frac{k}{2}-1)} = \int_0^{s^2} -\frac{k}{2} m(t) t^{-(\frac{k}{2}-1)} dt + C, \quad (.5)$$

for some constant C .

Note that the left hand side can be expressed as

$$\begin{aligned} n(s^2)(s^2)^{-(\frac{k}{2}-1)} &= \int_0^\infty (s^2)^{-(\frac{k}{2}-1)} \sigma^2 f(s^2|\sigma^2) \pi(\sigma^2) d\sigma^2 \\ &= \int_0^\infty \frac{k \left(\frac{k}{\sigma^2}\right)^{\frac{k}{2}-1} e^{-\frac{k s^2}{2\sigma^2}}}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}} \cdot \pi(\sigma^2) d\sigma^2. \end{aligned} \quad (.6)$$

From (.5) and (.6), as s^2 approaches to zero,

$$C = \lim_{s^2 \rightarrow 0} n(s^2)(s^2)^{-(\frac{k}{2}-1)} = \frac{(k/2)^{k/2}}{\Gamma(k/2)} E \left(\frac{1}{\sigma^2} \right)^{k/2-1} = \frac{k}{2} E \left(\frac{1}{S^2} \right)^{\frac{k}{2}-1}.$$

Therefore,

$$n(s^2)(s^2)^{-(\frac{k}{2}-1)} = \frac{k}{2} \int_{s^2}^\infty t^{-(\frac{k}{2}-1)} dM(t).$$

Now, we consider the first and second order derivatives of $h(s^2)(s^2)^{-(\frac{k}{2}-1)}$ with respect to s^2 ,

$$\left[h(s^2)(s^2)^{-(\frac{k}{2}-1)} \right]' = -\frac{k}{2} n(s^2)(s^2)^{-(\frac{k}{2}-1)}, \quad (.7)$$

$$\left[h(s^2)(s^2)^{-(\frac{k}{2}-1)} \right]'' = \frac{k^2}{4} m(s^2)(s^2)^{-(\frac{k}{2}-1)}. \quad (.8)$$

Consequently,

$$\left(h(s^2)(s^2)^{-(\frac{k}{2}-1)} \right)' = \int_0^{s^2} \frac{k^2}{4} m(t) t^{-(\frac{k}{2}-1)} dt + C_1, \quad (.9)$$

and

$$\begin{aligned} h(s^2)(s^2)^{-\left(\frac{k}{2}-1\right)} &= \int_0^{s^2} \int_0^y \frac{k^2}{4} m(t) t^{-\left(\frac{k}{2}-1\right)} dt dy + C_1 s^2 + C_2 \\ &= \frac{k^2}{4} \int_0^{s^2} m(t) t^{-\left(\frac{k}{2}-1\right)} (s^2 - t) dt + C_1 s^2 + C_2, \end{aligned} \quad (.10)$$

for some constants C_1 and C_2 .

From (.7) and (.9), as s^2 approaches to zero, similar argument shows that

$$C_1 = \lim_{s^2 \rightarrow 0} \left(h(s^2)(s^2)^{-\left(\frac{k}{2}-1\right)} \right)' = -\frac{k^2}{4} E \left(\frac{1}{S^2} \right)^{\frac{k}{2}-1}.$$

Similarly, combine equations (.8) and (.10) and let s^2 approach to zero,

$$C_2 = \lim_{s^2 \rightarrow 0} h(s^2)(s^2)^{-\left(\frac{k}{2}-1\right)} = \frac{k^2}{4} E \left(\frac{1}{S^2} \right)^{\frac{k}{2}-2}.$$

Thus,

$$\begin{aligned} &h(s^2)(s^2)^{-\left(\frac{k}{2}-1\right)} \\ &= \frac{k^2}{4} \int_0^{s^2} m(t) t^{-\left(\frac{k}{2}-1\right)} (s^2 - t) dt - \frac{k^2}{4} E \left(\frac{1}{S^2} \right)^{\frac{k}{2}-1} s^2 + \frac{k^2}{4} E \left(\frac{1}{S^2} \right)^{\frac{k}{2}-2} \\ &= \frac{k^2}{4} \left[- \left(s^2 \int_{s^2}^{\infty} m(t) t^{-\left(\frac{k}{2}-1\right)} dt \right) + \left(\int_{s^2}^{\infty} m(t) t^{-\left(\frac{k}{2}-2\right)} dt \right) \right] \\ &= \frac{k^2}{4} \int_{s^2}^{\infty} t^{-\left(\frac{k}{2}-1\right)} (t - s^2) dM(t). \end{aligned}$$

Therefore,

$$\hat{\sigma}_B^2 = \frac{h(s^2)}{n(s^2)} = \frac{k}{2} \left[\frac{\int_{s^2}^{\infty} t^{-\left(\frac{k}{2}-2\right)} dM(t)}{\int_{s^2}^{\infty} t^{-\left(\frac{k}{2}-1\right)} dM(t)} - s^2 \right].$$

□

Proof of Theorem 5.4.

We introduce one of the monumental theorems in the empirical process, on which our proof is based, contributed by *Blum* (1955) and *DeHardt* (1971).

Theorem .1. (*Blum-DeHardt*) *Let \mathcal{F} be a class of measurable functions such that $N_{[]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$, for every $\epsilon > 0$. Then \mathcal{F} is P -Glivenko-Cantelli.*

The bracket $[l, u]$ is the set of all the functions $f \in \mathcal{F}$ with $l \leq f \leq u$. An ϵ -bracket is a bracket with $\|u - l\| \leq \epsilon$. The bracketing number $N_{[]}(\epsilon, \mathcal{F}, L_1(P))$ is the minimum number of ϵ -brackets with which \mathcal{F} can be covered.

We only prove the part for the numerator and the denominator can be similarly done. Let $\mathcal{F} = \{f : f(t, b) = t^{-(k/2-2)}\mathbb{I}(t > b), b > 0\}$ and $Pf(t, b) = \int_b^\infty t^{-(k/2-2)} dM(t)$. It suffices to show that \mathcal{F} is a P -Glivenko-Cantelli class of functions. Since $\int_0^\infty t^{-(k/2-2)} dM(t) < \infty$, for any $\epsilon > 0$, an infinite collection of real numbers $0 = v_0 < v_1 < v_2 < \dots < v_m = \infty$ can be found such that

$$\begin{aligned} Pf(t, v_{j-1}) - Pf(t, v_j) &= \int_{v_{j-1}}^\infty t^{-(k/2-2)} dM(t) - \int_{v_j}^\infty t^{-(k/2-2)} dM(t) \\ &= \int_{v_{j-1}}^{v_j} t^{-(k/2-2)} dM(t) \\ &\leq \epsilon \end{aligned}$$

for all $1 \leq j \leq m$, with

$$Pf(t, v_m^-) = \lim_{v_m \uparrow \infty} Pf(t, v_m) = \lim_{v_m \uparrow \infty} \int_{v_m}^\infty t^{-(k/2-2)} dM(t) = 0.$$

Consider the collection of brackets $\{[l_j, u_j], 1 \leq j \leq m\}$, with $l_j(t) = t^{-(k/2-i)} 1\{t \geq v_j\}$ and $u_j(t) = t^{-(k/2-i)} 1\{t > v_{j-1}\}$. Now each $f \in \mathcal{F}$ is in at least one bracket and $|u_j - l_j|_P = Pf(t, v_{j-1}) - Pf(t, v_j^-) \leq \epsilon$ for all $1 \leq j \leq m$. Thus, by Theorem .1, \mathcal{F} is a *P-Glivenco-Cantelli Class* of functions. \square

Proof of Theorem 5.5.

Let

$$A_p(s_i^2) = \int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-2)} dM_p(t), A(s_i^2) = \int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-2)} dM(t),$$

and

$$B_p(s_i^2) = \int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-1)} dM_p(t), B(s_i^2) = \int_{s_i^2}^{\infty} t^{-(\frac{k}{2}-1)} dM(t).$$

According to the proof of Theorem ??, $\sup_{s_i^2 \in R} |A_p(s_i^2) - A(s_i^2)| \rightarrow 0$ and $\sup_{s_i^2 \in R} |B_p(s_i^2) - B(s_i^2)| \rightarrow 0$ a.s.. Let $L = \inf_{s_i^2 \in D^\delta} \{B(s_i^2)\}$. Then for any $\epsilon > 0$, when p is sufficiently large

$$\inf_{s_i^2 \in D^\delta} B_p(s_i^2) \geq L - \epsilon \quad a.s.,$$

and $\sup_{s_i^2 \in R} A_p(s_i^2) \leq C$, a.s. for some constant C . Then

$$\begin{aligned} & \sup_{s_i^2 \in D^\delta} |\hat{\sigma}_{i,NEBV}^2 - \hat{\sigma}_{i,B}| \\ &= \sup_{s_i^2 \in D^\delta} \left| \frac{A_p(s_i^2)}{B_p(s_i^2)} - \frac{A(s_i^2)}{B(s_i^2)} \right| \\ &= \sup_{s_i^2 \in D^\delta} \left| \frac{A_p(s_i^2)(B(s_i^2) - B_p(s_i^2))}{B_p(s_i^2)B(s_i^2)} + \frac{A_p(s_i^2) - A(s_i^2)}{B(s_i^2)} \right| \\ &\leq \frac{C}{L^2} \sup_{s_i^2 \in D^\delta} |B(s_i^2) - B_p(s_i^2)| + \frac{1}{L} \sup_{s_i^2 \in D^\delta} |A(s_i^2) - A_p(s_i^2)| \rightarrow 0, \quad a.s.. \end{aligned}$$

\square

Proof of Theorem 7.2.

As the proof for Theorem 4.1, for making the notations simple, we drop the subscript “ i ” in the proof. Recall the pdf of x_i given η_i is

$$f(x|\eta) = e^{-\frac{x}{\eta} - A(\eta)} h(x), \quad x > 0.$$

Define $p(x)$, $q(x)$, and $m(x)$ as

$$q(x) = \int_0^\infty \eta^2 e^{-\frac{x}{\eta} - A(\eta)} h(x) \pi(\eta) d\eta,$$

$$p(x) = \int_0^\infty \eta e^{-\frac{x}{\eta} - A(\eta)} h(x) \pi(\eta) d\eta,$$

and

$$m(x) = \int_0^\infty e^{-\frac{x}{\eta} - A(\eta)} h(x) \pi(\eta) d\eta.$$

Taking derivatives of the $p(x)$ and $\frac{p(x)}{h(x)}$, we have

$$p'(x) = \frac{h'(x)}{h(x)} p(x) - m(x) \quad \text{and} \quad m(x) = \frac{h'(x)}{h(x)} p(x) - p'(x).$$

and

$$\left[\frac{p(x)}{h(x)} \right]' = \frac{p'(x)}{h(x)} - \frac{p(x)h'(x)}{h^2(x)} = \frac{1}{h(x)} \left[p'(x) - \frac{h'(x)}{h(x)} p(x) \right] = -\frac{m(x)}{h(x)}.$$

Integrating both sides yields that

$$\frac{p(x)}{h(x)} = \int_0^x -\frac{m(t)}{h(t)} dt + C,$$

and by taking a limit, we have

$$C = \lim_{x \rightarrow 0} \frac{p(x)}{h(x)} = \lim_{x \rightarrow 0} \int_0^\infty \frac{1}{\eta} e^{-\eta x - A(\eta)} \pi(\eta) d\eta = \int_0^\infty \frac{1}{\eta} e^{-A(\eta)} \pi(\eta) d\eta = E_\pi \left(\frac{1}{\eta} e^{-A(\eta)} \right).$$

Therefore,

$$\begin{aligned} \frac{p(x)}{h(x)} &= \int_0^x -\frac{m(t)}{h(t)} dt + C \\ &= \int_0^x -\frac{m(t)}{h(t)} dt + E \left(\frac{1}{h(X)} \right) \\ &= \int_0^x -\frac{m(t)}{h(t)} dt + \int_0^\infty \frac{m(t)}{h(t)} dt \\ &= \int_x^\infty \frac{m(t)}{h(t)} dt, \end{aligned}$$

since

$$E \left(\frac{1}{h(X)} \right) = \int_0^\infty \frac{m(x)}{h(x)} dx = \int_0^\infty \int_0^\infty e^{-\eta x - A(\eta)} \pi(\eta) d\eta dx = \int_0^\infty \frac{1}{\eta} e^{-A(\eta)} \pi(\eta) d\eta = C.$$

On the other hand, the derivatives of the $q(x)$ and $\frac{q(x)}{h(x)}$ give us

$$q'(x) = \frac{h'(x)}{h(x)} q(x) - p(x),$$

and

$$\left[\frac{q(x)}{h(x)} \right]' = \frac{q'(x)}{h(x)} - \frac{q(x)h'(x)}{h^2(x)} = \frac{1}{h(x)} \left[q'(x) - \frac{h'(x)}{h(x)} q(x) \right] = -\frac{p(x)}{h(x)}.$$

Thus, we have

$$\left[\frac{q(x)}{h(x)} \right]'' = \left[-\frac{p(x)}{h(x)} \right]' = \frac{m(x)}{h(x)}.$$

Integrating both sides yields that

$$\left[\frac{q(x)}{h(x)} \right]' = \int_0^x \frac{m(t)}{h(t)} dt + C_1,$$

and

$$\frac{q(x)}{h(x)} = \int_0^x \int_0^y \frac{m(t)}{h(t)} dt dy + C_1 x + C_2.$$

Note that C_1 and C_2 are the limits of $\left[\frac{q(x)}{h(x)} \right]'$ and $\frac{q(x)}{h(x)}$, respectively, as

$$C_1 = \lim_{x \rightarrow 0} \left[\frac{q(x)}{h(x)} \right]' = \lim_{x \rightarrow 0} -\frac{p(x)}{h(x)} = \lim_{x \rightarrow 0} \int_0^\infty -\frac{1}{\eta} e^{-\eta x - A(\eta)} \pi(\eta) d\eta = -E_\pi \left(\frac{1}{\eta} e^{-A(\eta)} \right),$$

and

$$C_2 = \lim_{x \rightarrow 0} \frac{q(x)}{h(x)} = \lim_{x \rightarrow 0} \int_0^\infty \frac{1}{\eta^2} e^{-\eta x - A(\eta)} \pi(\eta) d\eta = \int_0^\infty \frac{1}{\eta^2} e^{-A(\eta)} \pi(\eta) d\eta = E_\pi \left(\frac{1}{\eta^2} e^{-A(\eta)} \right).$$

Furthermore,

$$E \left(\frac{1}{h(X)} \right) = -E_\pi \left(\frac{1}{\eta^2} e^{-A(\eta)} \right) = -C_1,$$

and

$$\begin{aligned} E \left(\frac{x}{h(X)} \right) &= \int_0^\infty \frac{x}{h(x)} m(x) dx \\ &= \int_0^\infty \int_0^\infty x e^{-\eta x - A(\eta)} \pi(\eta) d\eta dx \\ &= \int_0^\infty \frac{1}{\eta^2} e^{-A(\eta)} \pi(\eta) d\eta = E_\pi \left(\frac{1}{\eta^2} e^{-A(\eta)} \right) = C_2. \end{aligned}$$

Hence,

$$\begin{aligned}
\frac{q(x)}{h(x)} &= \int_0^x \int_0^y \frac{m(t)}{h(t)} dt dy + C_1 x + C_2 \\
&= \int_0^x \int_t^x \frac{m(t)}{h(t)} dy dt - E\left(\frac{1}{h(X)}\right)x + E\left(\frac{x}{h(X)}\right) \\
&= \int_0^x \frac{x-t}{h(t)} m(t) dt - x \int_0^\infty \frac{1}{h(t)} m(t) dt + \int_0^\infty \frac{t}{h(t)} m(t) dt \\
&= \int_x^\infty \frac{t}{h(t)} m(t) dt - x \int_x^\infty \frac{1}{h(t)} m(t) dt.
\end{aligned}$$

Therefore, the generalized NEBV is given by

$$\hat{\eta} = \frac{E[\eta^2|x]}{E[\eta|x]} = \frac{q(x)}{p(x)} = \frac{\int_x^\infty \frac{t}{h(t)} dM(t)}{\int_x^\infty \frac{1}{h(t)} dM(t)} - x.$$

□

Proof of Theorem 7.5.

Once again, as the proof for Theorem 4.1, for using ease the notation, we drop the subscript “ i ” in the proof. Define $p(x)$, $q(x)$, and $m(x)$ as

$$q(x) = \int_0^\infty \frac{1}{\eta^2} e^{\eta \ln x - A(\eta)} h(x) \pi(\eta) d\eta,$$

$$p(x) = \int_0^\infty \frac{1}{\eta} e^{\eta \ln x - A(\eta)} h(x) \pi(\eta) d\eta,$$

and

$$m(x) = \int_0^\infty e^{\eta \ln x - A(\eta)} h(x) \pi(\eta) d\eta.$$

Taking derivatives of the $p(x)$ and $\frac{p(x)}{h(x)}$, we have

$$p'(x) = \frac{m(x)}{x} + \frac{h'(x)}{h(x)} p(x),$$

and

$$\left[\frac{p(x)}{h(x)} \right]' = \frac{p'(x)}{h(x)} - \frac{p(x)h'(x)}{h^2(x)} = \frac{1}{h(x)} \left[p'(x) - \frac{h'(x)}{h(x)} p(x) \right] = \frac{m(x)}{xh(x)}.$$

Integrating both sides yields that

$$\frac{p(x)}{h(x)} = \int_0^x \frac{m(t)}{th(t)} dt + C.$$

Since $\eta > 0$,

$$C = \lim_{x \rightarrow 0} \frac{p(x)}{h(x)} = \lim_{x \rightarrow 0} \int_0^\infty \frac{1}{\eta} e^{\eta \ln x - A(\eta)} \pi(\eta) d\eta = 0.$$

On the other hand, the derivatives of the $q(x)$ and $\frac{q(x)}{h(x)}$ give us

$$q'(x) = \frac{p(x)}{x} + \frac{h'(x)}{h(x)} q(x),$$

and

$$\left[\frac{q(x)}{h(x)} \right]' = \frac{q'(x)}{h(x)} - \frac{q(x)h'(x)}{h^2(x)} = \frac{1}{h(x)} \left[q'(x) - \frac{h'(x)}{h(x)} q(x) \right] = \frac{p(x)}{xh(x)}.$$

Thus,

$$\left[\frac{q(x)}{h(x)} \right]'' = \left[\frac{p(x)}{xh(x)} \right]' = -\frac{p(x)}{x^2h(x)} + \frac{m(x)}{x^2h(x)}.$$

Integrating both sides yields that

$$\begin{aligned}
\left[\frac{q(x)}{h(x)}\right]' &= \int_0^x -\frac{p(t)}{t^2h(t)} + \frac{m(t)}{t^2h(t)} dt + C_1 \\
&= \int_0^x -\frac{p(t)}{t^2h(t)} dt + \int_0^x \frac{m(t)}{t^2h(t)} dt + C_1 \\
&= \int_0^x -\frac{1}{y^2} \int_0^y \frac{m(t)}{th(t)} dt dy + \int_0^x \frac{m(t)}{t^2h(t)} dt + C_1 \\
&= \int_0^x \int_t^x -\frac{1}{y^2} dy \frac{m(t)}{th(t)} dt + \int_0^x \frac{m(t)}{t^2h(t)} dt + C_1 \\
&= \int_0^x \left(\frac{1}{x} - \frac{1}{t}\right) \frac{m(t)}{th(t)} dt + \int_0^x \frac{m(t)}{t^2h(t)} dt + C_1 \\
&= \frac{1}{x} \int_0^x \frac{m(t)}{th(t)} dt + C_1,
\end{aligned}$$

where

$$C_1 = \lim_{x \rightarrow 0} \left[\frac{q(x)}{h(x)}\right]' = \lim_{x \rightarrow 0} \left[\frac{p(x)}{xh(x)}\right]' = \lim_{x \rightarrow 0} \int_0^\infty \frac{1}{x\eta} e^{\eta \ln x - A(\eta)} \pi(\eta) d\eta = 0.$$

Now,

$$\frac{q(x)}{h(x)} = \int_0^x \frac{1}{y} \int_0^y \frac{m(t)}{th(t)} dt dy + C_2 = \int_0^x \int_t^x \frac{1}{y} dy \frac{m(t)}{th(t)} dt + C_2 = \int_0^x (\ln x - \ln t) \frac{m(t)}{th(t)} dt,$$

since $C_2 = \lim_{x \rightarrow 0} \frac{q(x)}{h(x)} = \lim_{x \rightarrow 0} \int_0^\infty \frac{1}{\eta^2} e^{\eta \ln x - A(\eta)} \pi(\eta) d\eta = 0$. Therefore,

$$\hat{\eta}_{NEBV}^{-1} = \frac{E\left[\frac{1}{\eta^2} \middle| x\right]}{E\left[\frac{1}{\eta} \middle| x\right]} = \frac{p(x)}{q(x)} = \ln x - \frac{\int_0^x \frac{\ln t}{th(t)} dM(t)}{\int_0^x \frac{1}{th(t)} dM(t)}.$$

□

BIBLIOGRAPHY

- Alon, U., N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine (1999), Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, *Proc. Natl. Acad. Sci.*, *96*(2), 6745–6750.
- Baker, M. (2016), 1,500 scientists lift the lid on reproducibility, *Nature News*, *533*(7604), 452.
- Blum, J. (1955), On the convergence of empiric distributions, *The Annals of Mathematical Statistics*, *26*, 527–529.
- Cantelli, F. P. (1933), Sulla determinazione empirica delle leggi di probabilita, *Giorn. Ist. Ital. Attuari*, *4*, 415–420.
- Casella, G., and R. Berger (2001), *Statistical Inference*, Second ed., Duxbury Press.
- Choe, S. E., M. Bouttros, A. M. Michelson, G. M. Chruch, and M. Halfon (2005), Preferred analysis methods for affymetrix genechips revealed by a wholly defined control dataset, *Genome Biology*, *6*(2), R16.1–16.
- Cohen, A., and H. B. Sackrowitz (1982), 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.
- Cui, X., J. T. Hwang, J. Qiu, N. J. Blades, and G. A. Churchill (2005), Improved statistical tests for differential gene expression by shrinking variance components estimates, *Biostatistics*, *6*, 59–75.
- Dawid, A. P. (1994), Selection paradoxes of Bayesian inference, *Lecture Notes-Monograph Series*, pp. 211–220.
- DeHardt, J. (1971), Generalizations of the Glivenko-Cantelli theorem, *The Annals of Mathematical Statistics*, *42*, 2050–2055.
- Dey, D. K., and C. Srinivasan (1985), Estimation of a covariance matrix under Stein’s loss, *The Annals of Statistics*, pp. 1581–1591.
- Donsker, M. D. (1952), Justification and extension of Doob’s heuristic approach to the Kolmogorov-Smirnov theorems, *The Annals of Mathematical Statistics*, *23*, 277–281.

- Dudley, R. M. (1956), Universal Donsker classes and metric entropy, *The Annals of Probability*, 15, 1306–1326.
- Dvoretzky, A., J. Kiefer, and J. Wolfowitz (1956), Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, *The Annals of Mathematical Statistics*, 27(3), 642–669.
- Efron, B. (2011), Tweedie’s formula and selection bias, *Journal of the American Statistical Association*, 106(496), 1602–1614.
- Efron, B., and C. N. Morris (1973), Stein’s estimation rule and its competitors—an empirical Bayes approach, *Journal of the American Statistical Association*, 68, 117–130.
- Efron, B., and C. N. Morris (1975), Data analysis using Stein’s estimator and its generalizations, *Journal of the American Statistical Association*, pp. 311–319.
- Gine, E., and J. Zinn (1984), Some limit theorems for empirical processes, *The Annals of Probability*, 12, 929–989.
- Giovannucci, E., A. Ascherio, E. B. Rimm, M. J. Stampfer, G. A. Colditz, and W. C. Willett (1995), Intake of Carotenoids and Retino in Relation to Risk of Prostate Cancer, *Journal of the National Cancer Institute*, 87(23), 1767–1776.
- Glivenko, V. (1933), Sulla determinazione empirica della legge di probabilita, *Giorn. Ist. Ital. Attuari*, 4, 92–99.
- Golub, T. R., et al. (1999), Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, 286, 531–537.
- He, K. (1992), Parametric empirical Bayes confidence intervals based on James-Stein estimator, *Statistics & Decisions*, 10(1-2), 121–132.
- Hwang, J. T., and Z. Zhao (2013), Empirical Bayes confidence intervals for selected parameters in high dimension with application to microarray data analysis, *Journal of the American Statistical Association*, 108(502), 607–618.
- Hwang, J. T., J. Qiu, and Z. Zhao (2009), Empirical Bayes confidence intervals shrinking both means and variances, *Journal of the Royal Statistical Society. Series B*, 71(1), 265–285.
- James, W., and C. Stein (1961), Estimation with quadratic loss, *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, Vol. I, pp. 361–379.
- Kendzioriski, C., M. Newton, H. Lan, and M. Gould (2003), On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles., *Statistics in Medicine*, 22, 3899–3914.
- Kolmogorov, A. N. (1933), Sulla determinazione empirica di una legge di distribuzione, *Giorn. Ist. Ital. Attuari*, 4, 83–91.

- Lin, Y., S. T. Nadler, A. D. Attie, and B. S. Yandell (2003), Adaptive gene picking with microarray data: detecting important low abundance signals, *The Analysis of Gene Expression Data: Methods and Software*.
- Lönnstedt, I., and T. Speed (2002), Replicated microarray data, *Statistica Sinica*, 12(1), 31–46, special issue on bioinformatics.
- Morris, C. N. (1983), Parametric empirical Bayes confidence intervals, *Scientific inference, data analysis, and robustness (Madison, Wis., 1981)*, 48, 25–50.
- Muralidharan, O. (2012), High dimensional exponential family estimation via empirical Bayes, *Statistica Sinica*, pp. 1217–1232.
- Ossiander, M. (1987), A central limit theorem under metric entropy with l2 bracketing, *The Annals of Probability*, 15, 897 – 919.
- Qiu, J., and J. T. Hwang (2007), Sharp simultaneous intervals for the means of selected populations with application to microarray data analysis, *Biometrics*, 63(3), 767–776.
- Rossouw, J. E., G. L. Anderson, R. L. Prentice, and et al. (2002), Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results From the Women’s Health Initiative Randomized Controlled Trial, *Journal of the American Medical Association*, 288(3), 321–333.
- Senn, S. (2008), A note concerning a selection paradox of Dawid’s, *The American Statistician*, 62(3), 206–210.
- Smirnov, N. V. (1939), Sur les ecarts de la courbe de distribution empirique, *Mat. Sbornik (N.S.)*, 6, 3–26.
- Smyth, G. K. (2004), Linear models and empirical Bayes methods for assessing differential expression in microarray experiments, *Statistical Applications in Genetics and Molecular Biology*, 3(1), Article 3.
- Stein, C. (1956), Inadmissibility of the usual estimator for the mean of a multivariate distribution, *Proceedings of 3rd Berkeley Symposium on Probability and Statistics, Vol. I*, pp. 197–206.
- Stein, C. (1977), Lectures on the theory of estimation of many parameters, *Studies in the Statistical Theory of Estimation I*, 74, 4–65.
- Storey, J., and R. Tibshirani (2003), SAM thresholding and false discovery rates for detecting differential gene expression in dna microarrays, *The Analysis of Gene Expression Data: Methods and Software*, pp. 272–290.
- Talagrand, M. (1987), Donsker classes and random geometry, *The Annals of Probability*, 15, 1327–1338.

- Tong, T., and Y. Wang (2007), Optimal shrinkage estimation of variances with applications to microarray data analysis, *Journal of the American Statistical Association*, 102(477), 113–122.
- Tusher, V. G., R. Tibshirani, and G. Chu (2001), Significance analysis of microarrays applied to the ionizing radiation response., *Proceedings of the National Academy of Science.*, 98, 5116–5121.
- Van der Vaart, A. W., and J. A. Wellner (1996), *Weak Convergence and Empirical Processes.*, Springer-Verlag, New York.
- Vapnik, V. N., and A. Y. Chervonenkis (1968), On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability and Its Applications*, 16, 264–280.
- Venter, J. H. (1988), Confidence bounds based on the largest treatment mean, *South African Journal of Science*, 84, 340–342.
- Venter, J. H., and S. J. Steel (1991), Estimation of the mean of the population selected from k populations, *Journal of Statistical Computation and Simulation*, 38, 1–14.
- Zhao, Z. (2010), Double shrinkage empirical Bayesian estimation for unknown and unequal variances, *Statistics and Its Interface*, 3, 533–541.
- Zhao, Z., and J. T. Hwang (2012), Empirical Bayes false coverate rate controlling confidence interval, *Journal of the Royal Statistical Society. Series B*, 74(5), 871–891.
- Zhao, Z., and S. K. Sarkar (2015), On the credible interval under the Zero-Inflated Mixture Prior in high dimensional inference, *Statistica Sinica*, 25(2), 725–742.