Summer 2019

Investigations on Multiple Interval Estimators

Taeho Kim

Investigations on Multiple Interval Estimators

by

Taeho Kim

Bachelor of Engineering
Ajou University 2007

Master of Business Administration
Ajou University 2010

Master of Science
Rutgers University 2013

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Statistics

College of Arts and Sciences

University of South Carolina

2019

Accepted by:

Edsel A. Peña, Major Professor

Karl Gregory, Committee Member

Dewei Wang, Committee Member

George Androulakis, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

Acknowledgments

I would like to express the deepest appreciation to my academic advisor Dr. Edsel A. Peña, who has both the ability to construct a broad framework and the eye to study microscopic details. He conveys a spirit of adventure in regard to research and scholarship, and an enthusiasm in regard to teaching. Without his constant help and guidance during my time in graduate school, this dissertation would not have been possible. I feel so lucky to have worked with him at the University of South Carolina.

I would also like to thank my committee members. Dr. Karl Gregory gave me critical comments that have improve my dissertation, and he was also kind enough to share advice based on his personal experiences. Dr. Dewei Wang provided me with consistent support in finishing my dissertation with constructive advice to enhance the procedures in the dissertation. Lastly, Dr. George Androulakis taught me mathematical analysis, which helped me gain mathematical maturity at the beginning of my graduate study; and he also helped me to better understand group theory.

In addition, I want to express my appreciation for Dr. James Lynch, who introduced me to the concept of Haar measures, and who serves as a living example that we can conduct research at any age as long as we have the enthusiasm for it. I thank my research group members: Piaomu Liu, Beidi Qiang, Shiwen Shen, Jeff C. Thompson, Lu Wang, Lili Tong, and Tahmidul Islam. They always gave me great feedback and shared their creative ideas. I also thank the Center for Colon Cancer Research for providing research assistantship opportunities. Lastly, I give my love and appreciation to my family for their support and love during my studies in the United States.

Abstract

Multiple interval estimation for a set of parameters is investigated. To begin, a strategy of optimization for a multiple interval estimator (MIE) is introduced. This approach allocates distinct optimized levels to individual interval estimators so that the global expected content can be minimized while the global coverage probability is still maintained at a global level. This optimal allocation is achieved by a decision theoretic procedure which consists of two global risk functions. The major part of this manuscript is devoted to two multiple interval estimation procedures. Both procedures adopt prior information added to the classical setting, but these procedures do not particularly follow the frequentist or Bayesian approach. The first procedure starts from a practical motivation in the use of prior information. That is, a pair of thresholds is established based on the prior information to discard one side of the interval estimators in a particular subset of an MIE. Through this process, the global expected content of the MIE can be reduced. On the other hand, the second procedure also utilizes prior information, but it focuses more on seeking a coherent structure for an MIE which involves a class of heterogeneous parameters. In particular, the prior information is provided in the form of a non-informative prior distribution. Then the resulting MIE can be viewed from both the frequentist and Bayesian perspectives. An appropriate choice of prior distribution is naturally achieved by assigning group structures over the three fundamental components: the sample space, the parameter space, and the action space. Then the right Haar measure provides the form of non-informative prior distribution. In addition, the left Haar measure can also be exploited to evaluate the expected content of the MIE.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The idea of interval estimation is one of the rare cases among many statistical inferences which is both easy to understand and theoretically profound. It is easy to understand because the idea starts from a motivation to utilize an interval instead of a point for an estimation. It is as natural as using a net instead of a stone for fishing in a river. However, this natural modification provides us additional information from the estimation procedure. Therefore, the accuracy of the estimation can be evaluated by a coverage probability, the probability that the interval covers the true parameter; meanwhile, the precision of the estimation can be evaluated by an expected interval length, the distance between the upper and the lower limits. In addition, the sophistication in hypothesis testing can be equally achieved with interval estimation by exploiting a mathematical duality, providing a structure which can handle profound problems. Based on these advantages, it is not hard to see why interval estimation is considered to be one of the major inferential tools in statistics.

For a single parameter case, the usual optimality condition in constructing an interval estimator (IE) is to minimize an expected length while maintaining a coverage probability at least a given level of $1 - \alpha$. Many problems in this case are considered to be *standard* and appear in major textbooks, e.g., Lehmann and Romano [37]. However, in regard to thinking about an extension to multiple parameters, we can immediately find that the problem becomes complicated. With the multiple parameters, the true difficulty resides in the characteristics of simultaneous decision making. For example, suppose tomorrow is the last day of the finals week in a college. If a stu-

dent has only one exam, then he or she can put all the time and effort into preparing for the exam. However, if two exams are left, then the student should set up a plan to allocate time and effort between the subjects. Resources are always limited and must be allocated in a strategic way to make multiple decisions simultaneously. This adds an additional layer to the existing individual structure of the problem. When we construct a multiple interval estimator (MIE), the limited resource is represented by a global level, $1-q$. Therefore, if we attempt to minimize the global expected content of the MIE, while maintaining the global coverage probability at least a global level of $1-q$, then the problem involves the allocation problem, i.e., how to assign the optimal levels to the corresponding individual IEs.

Likewise, in multiple testing, we attempt to maximize a global power, while controlling the global type-I error rate at most a global size of q. This procedure still entails the allocation problem, i.e., how to assign the optimal sizes to the corresponding individual testings. In comparison with the case of MIEs, a considerable amount of study has been done in constructing multiple testing procedures (MTPs). Accordingly, one of the general approaches in the multiple interval estimation is to invert an existing MTP to construct its matching MIE by utilizing the mathematical duality between hypothesis testing to interval estimation. However, this approach does not work for some cases. This is because there is no information about alternative hypothesis in interval estimation. Therefore, some of the concepts in MTPs, e.g., a power of testing, cannot be directly transferred to MIEs. This implies multiple interval estimation still has its own importance and needs to be investigated independently from multiple testings.

It is important to obtain additional information in simultaneous decision making. In the previous example, suppose the student had additional information about the two exams. For the first subject, the grade so far is 98%, the exam will have non-cumulative coverage and a practice exam will be provided; in contrast, for the second

subject, the grade so far is 89%, the exam will have cumulative coverage and no review material will be provided. In this situation, the student would generally attempt to spend less time preparing for the first exam and invest more time and effort preparing for the second exam. In our study of MIEs, the additional information is available in the form of prior distribution. However, we will not categorize these as either frequentist or Bayesian procedure, considering the prior distribution as just additional information. The resulting MIEs will then be evaluated by the global expected content and the global coverage probability with respect to the *true* parameters in order not to be in favor of either procedure.

The two major topics of this dissertation utilize prior distributions in the MIE constructions to enhance their performances. In chapter 2, we introduce a general optimization strategy which will be used throughout the manuscript with certain variations. The utilization of the additional information is different in the first and second topics. The first project in chapter 3 is more practically oriented, and the information is used to discard a subset of interval estimators to minimize the global expected content. On the other hand, the second project in chapter 4 is more focused on formality and the information is used to build a coherent MIE structure based on the concept of group theory. Lastly, chapter 5 will be devoted to concluding remarks and potential future works.

Chapter 2

Optimization Strategy in MIE

This chapter introduces an optimization procedure which will be utilized throughout the manuscript. This procedure is based on a decision theoretic framework through the two global risk functions which reflect the global coverage probability and global expected content. The optimization procedure searches for the best MIE by allocating the optimal individual levels so that the global expected content can be minimized while maintaining the global coverage probability at least a global level of $1-q$. The general idea is adopted from Peña, Habiger, and Wu [41] which aims at the best MTP by allocating the optimal individual sizes under the family-wise error rate or false discovery rate. The allocation procedure is called *size investing strategy*. Borrowing this term, the goal in this section is to establish a *confidence level investing strategy*.

2.1 Individual Loss

Let $\Theta = (-\infty, \infty)$ be a parameter space and $\mathcal{A} = \{(a_1, a_2) : -\infty < a_1 < a_2 < \infty\}$ be an action space. Now, we define a pair of loss functions L_0 an L_1 as follows:

$$L_0(\theta, w) = \nu(w) \;\&\; L_1(\theta, w) = I\{\theta \notin w\} = I\{\theta \notin (a_1, a_2)\}$$

where $w = (a_1, a_2)$ and ν is the content measure, e.g., $\nu(w) = |a_2 - a_1|$ when θ is a location parameter. Note that the first loss function penalizes intervals with wide lengths and the second loss function penalizes intervals which do not contain true parameters. Given a small positive number $\alpha \in (0, 1)$, we set up an optimization

problem as follows: for every $\theta \in \Theta$,

$$\text{minimize } E_\theta[L_0(\theta, W(X))] \text{ subject to } E_\theta[L_1(\theta, W(X))] \leq \alpha.$$

Note that this setting up represents the usual pursuit of the tightest IE with the coverage probability maintained at least a nominal level of $1 - \alpha$.

One issue with the loss functions is that whereas the range of $L_1(\theta, w)$ is always between zero and one, the range of $L_0(\theta, w)$ is positively unbounded. This lack of balance between two loss functions can cause a statility issue in the optimization procedure. To handle this issue, we adjust the $L_0(\theta, w)$ by adopting a function, $h_\beta(x) = \frac{x}{\beta+x}$, where β is a positive constant as follows:

$$L_0^\beta(\theta, w) = h_\beta(L_0(\theta, w)) = \frac{\nu(w)}{\beta+\nu(w)}.$$

Note that the adjusted loss function $L_0^\beta(\theta, w)$ ranges from zero to one on the positive domain. Moreover, ths function, $h_\beta(x)$, is smooth with the nth derivative, $\frac{d^n}{dx^n}h_\beta(x) = \frac{(-1)^{n+1}n!\beta}{(\beta+x)^{n+1}}$, and easily utilized to the optimization procedure. A similar loss function approach was introduced in Casella and Hwang [9] for a single-dimensional case. We extend our idea to a multi-dimensional case in the next section.

2.2 Global Loss and Risk

Let M be a positive integer. Then $\boldsymbol{\Theta} = (-\infty, \infty)^M$ is a paramter space with an element $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_M)^T$, and $\boldsymbol{\mathcal{A}} = \{\times_{m=1}^M (a_1^m, a_2^m) : -\infty < a_1^m < a_2^m < \infty,\ m = 1, \ldots, M\}$ is an action space with an element $(\boldsymbol{a_1}, \boldsymbol{a_2}) = [(a_1^1, a_2^1), (a_1^2, a_2^2), \ldots, (a_1^M, a_2^M)]^T$. Now we define two global loss functions as follows:

$$\boldsymbol{L}_0^\beta(\boldsymbol{\theta}, \boldsymbol{w}) = \frac{1}{M}\sum_{m=1}^M L_0^\beta(\theta_m, w_m) \ \&\ \boldsymbol{L}_1(\boldsymbol{\theta}, \boldsymbol{w}) = I\left\{\left(\sum_{m=1}^M L_1(\theta_m, w_m)\right) \geq 1\right\}$$

where $\boldsymbol{w} = (\boldsymbol{a_1}, \boldsymbol{a_2})$. Observe that the interpretations of the individual loss functions are still maintained in the global loss functions. That is, the first global loss function

penalizes multiple intervals with wide expected contents, and the second global loss function penalizes multiple intervals which do not cover at least one true parameter.

Now, let $\mathcal{D}$ be a class of nonrandomized multiple decision functions which consist of $\boldsymbol{\delta : \mathcal{X} \longrightarrow \mathcal{A}}$ where $\boldsymbol{\delta(X;\alpha)} = [\delta_1(X;\alpha) = (LB_1(X;\alpha), UB_1(X;\alpha)), \ldots, \delta_M(X;\alpha) = (LB_M(X;\alpha), UB_M(X;\alpha))]^T$. In order to obtain the risk functions, we need to take expectations on the global loss functions. However, $\boldsymbol{L_0^{\beta}(\theta, w)}$ has a non-linear form with respect to the random variable, so we cannot take that expectation directly. However, this issue can be circumvented by using a linear interpolation, i.e., the risk function can be well approximated. The resulting risk functions are as follows:

$$\boldsymbol{R_0^{\beta}(\theta,\delta) = E_{\theta}\left[L_0^{\beta}(\theta,\delta(X;\alpha))\right]} \;\&\; \boldsymbol{R_1(\theta,\delta) = E_{\theta}\left[L_1(\theta,\delta(X;\alpha))\right].}$$

Notice that the first global risk function is the adjusted global expected content. In addition, the second global risk function is related to the familywise coverage rate (FWCR) which is defined as the probability that an MIE covers all the true parameters:

$$\boldsymbol{R_1(\theta,\delta) = 1 - FWCR(\theta,\delta)}.$$

2.3 Application: Mean of Normal Distribution

2.3.1 Optimization Procedure

Given a small positive number $q \in (0,1)$, we set up an optimization problem as follows:

$$\text{minimize } \boldsymbol{R_0^{\beta}(\theta,\delta)} \text{ subject to } \boldsymbol{R_1(\theta,\delta)} \leq q.$$

Note that the restriction implies the global coverate probability, $\boldsymbol{FWCR(\theta,\delta)}$, is maintained to be at least a global level, $1 - q$. In this subsection, we apply this procedure to an MIE for M normal location parameters with known variances. To simplify the setting, we first adjust the Sufficiency Principle as follows:

$$\bar{X}_m \sim N(\mu_m, \sigma_m^2) \text{ for } m = 1, \ldots, M$$

where the random variables are independent throughout the index m and the variance of $\bar{X}_m$ is set to be σ_m^2 without loss of generality. The form of the mth individual IE, Γ_m, has the usual form as follows:

$$\Gamma_m(X_m;\alpha_m) = [LB_m(X_m;\alpha_m), UB_m(X_m;\alpha_m)] = [\bar{X}_m - z_{\alpha_m/2}\sigma_m, \bar{X}_m + z_{\alpha_m/2}\sigma_m]$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$. The content measure ν is the Lebesgue measure, $v(w) = |a_2 - a_1|$, because the mean is a location parameter. Given this setting, we evaluate the two risk functions as follows:

$$\begin{aligned}
\boldsymbol{R}_{\boldsymbol{0}}^{\beta}(\boldsymbol{\mu},\boldsymbol{\Gamma}) =& \boldsymbol{E}_{\boldsymbol{\mu}}\left[\frac{1}{M}\sum_{m=1}^{M} h_\beta(\nu(LB_m(X_m;\alpha_m), UB_m(X_m;\alpha_m)))\right] \\
\approx& \frac{1}{M}\sum_{m=1}^{M} h_\beta\left(E_{\mu_m}[UB_m(X_m;\alpha_m) - LB_m(X_m;\alpha_m)]\right) \\
=& \frac{1}{M}\sum_{m=1}^{M} h_\beta(2z_{\alpha_m/2}\sigma_m) \\
=& \frac{1}{M}\sum_{m=1}^{M} \frac{2z_{\alpha_m/2}\sigma_m}{\beta + 2z_{\alpha_m/2}\sigma_m} \\
\boldsymbol{R}_{\boldsymbol{1}}(\boldsymbol{\mu},\boldsymbol{\Gamma}) =& 1 - \boldsymbol{P}_{\boldsymbol{\mu}}\left[\left(\sum_{m=1}^{M} L_1(\mu_m, w_m)\right) = 0\right] \\
=& 1 - \boldsymbol{P}_{\boldsymbol{\mu}}\left[\bigcap_{m=1}^{M}\{\mu_m \in (LB_m(X_m;\alpha_m), UB_m(X_m;\alpha_m))\}\right] \\
=& 1 - \prod_{m=1}^{M} P_{\mu_m}\left[\mu_m \in (LB_m(X_m;\alpha_m), UB_m(X_m;\alpha_m))\right] \\
=& 1 - \prod_{m=1}^{M}(1-\alpha_m)
\end{aligned}$$

Note the approximation in the second equality can be achieved by a piece-wise linear interpolation. We reparametrize $\nu_m = z_{\alpha_m/2}$ for a numerical stability. Then the initial optimization problem can be restated as follows:

$$\text{minimize } \frac{1}{M}\sum_{m=1}^{M}\frac{2\nu_m\sigma_m}{\beta + 2\nu_m\sigma_m} \text{ subject to } \sum_{m=1}^{M}\log(2\Phi(\nu_m) - 1) \geq \log(1-q).$$

This problem can be numerically solved by using the Newton-Raphson method after setting up a Lagrange equation. The details are provided in Appendix B.

2.3.2 Optimization Result

The essence of the optimization procedure is the allocation of optimal levels to the individual IEs, called the confidence level investing strategy. This process allows us to reach the smallest global expected content, while maintaining the global level requirement. In this problem, the tuning parameter, β, determines the shape of the function h_β which controls the allocation strategy. <Figure 2.1> illustrates the dynamics due to β in the optimal level allocations of the MIE for $M = 1000$ normal location parameters, μ_m's. Notice that an equi-spaced sequence from 0.01 to 10

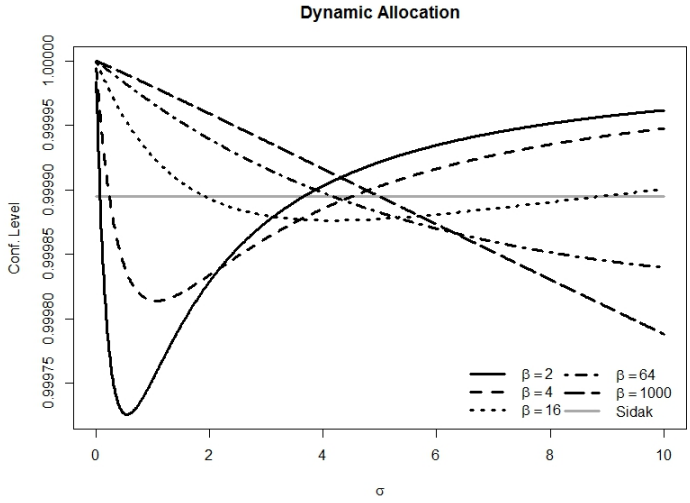


Figure 2.1: Individual Level Allocation: Normal Mean

is assigned for σ_m's. Therefore, the points on the graphs represent the allocated individual levels with respect to the σ_m's on the horizontal line. First, the gray horizontal line represents the allocation based on the Sidak adjustment, i.e., the constant individual levels, $(1-q)^{1/M}$. Compared to this, the black curves represent the allocation results achieved through the optimization procedure. The shape of the curves varies with respect to the value of β. When β is small, the trend shows a highly nonlinear shape, assigning large individual levels to the IEs with very small and large σ's. However, as β becomes larger, the form of the curves reaches an almost linear

line with the negative slope. Eventually, when β is greater than 1000, the shape of the curve remains invariant. In this application, the global expected content defined as a relative expected length (REL), i.e., the average expected length ratio of the MIE with optimal levels to the MIE with Sidak adjustment.

<Table 2.1> summarizes the global RELs with respect to β. Note that when β

Table 2.1: Relative Expected Length with respect to β

β	1	2	8	32	1000	Sidak Adj.
REL	1.0317	1.0177	0.9967	0.9888	0.9874	1

is small, the REL is greater than 1, implying the performance is no better than the MIE with Sidak adjustment, the constant allocation. However, the REL becomes smaller and converges to 0.9874 as β reaches 1000 and larger. This overall reduction can be explained through the allocation result. When β = 1000, the optimization procedure assigns smaller individual levels to the IEs with larger σ_m's to counterbalance the sizes of σ_m's with smaller $z_{\alpha_m/2}$'s. To compensate for these *investments*, the procedure matches larger $z_{\alpha_m/2}$'s to the smaller σ_m's by assigning larger levels to the corresponding IEs. These processes are performed simultaneously to minimize the REL, maintaining the FWCR at least the global level of $1 - q = 0.9$. However, the resulting 1.26% reduction in the relative expected length is a quite limited amount. This limitation particularly motivates the thresholding approach in the chapter 3.

2.4 Application: Variance of Normal Distribution

2.4.1 Optimization Procedure

The goal is to determine the optimal allocation of an MIE for M normal scale parameters with known means. The setting is as follows:

$$X_{m1}, \ldots, X_{mn_m} \sim_{iid} N(\mu_m, \sigma_m^2) \text{ for } m = 1, \ldots, M$$

where the random variables are independent throughout the index m. Note that the pivot quantity $\sum_{i=1}^{n_m}(X_{mi}-\mu_m)^2/\sigma_m^2$ follows a $\chi^2(n_m)$ distribution. Then the derivation of the mth individual interval estimator is as usual:

$$\begin{aligned}1-\alpha_m =& P_{\sigma^2}\left[\chi^2_{1-\alpha_m/2,n_m} \le \frac{\sum_{i=1}^{n_m}(X_{mi}-\mu_m)^2}{\sigma_m^2} \le \chi^2_{\alpha_m/2,n_m}\right]\\ =& P_{\sigma^2}\left[\frac{\sum_{i=1}^{n_m}(X_{mi}-\mu_m)^2}{\chi^2_{\alpha_m/2,n_m}} \le \sigma_m^2 \le \frac{\sum_{i=1}^{n_m}(X_{mi}-\mu_m)^2}{\chi^2_{1-\alpha_m/2,n_m}}\right]\end{aligned}$$

where $\chi^2_{\alpha_m/2,n_m} = F_{n_m}(1-\alpha_m/2)$ with F_{n_m}, the cdf of the $\chi^2(n_m)$ distribution. Therefore, the form is as follows:

$$\Gamma_m(X_m;\alpha_m) = [LB_m(X_m;\alpha_m), UB_m(X_m;\alpha_m)] = \left[\frac{\sum_{i=1}^{n_m}(X_{mi}-\mu_m)^2}{\chi^2_{\alpha_m/2,n_m}}, \frac{\sum_{i=1}^{n_m}(X_{mi}-\mu_m)^2}{\chi^2_{1-\alpha_m/2,n_m}}\right].$$

For the content measure, $\nu(w) = \log(a_2) - \log(a_1) = \log(a_2/a_1)$ will be used because it is the legitimate invariant measure for a scale parameter as it will be shown in chapter 4. Given this setting, we evaluate the global risk functions as follows:

$$\begin{aligned}\boldsymbol{R_0^\beta(\sigma^2,\delta)} =& \boldsymbol{E_{\sigma^2}}\left[\frac{1}{M}\sum_{m=1}^{M} h_\beta(\log(UB_m(X_m;\alpha_m)/LB_m(X_m;\alpha_m)))\right]\\ =& \boldsymbol{E_{\sigma^2}}\left[\frac{1}{M}\sum_{m=1}^{M} h_\beta\left(\log\left(\frac{\sum_{i=1}^{n_m}(X_{mi}-\mu_m)^2}{\chi^2_{1-\alpha_m/2,n_m}}\cdot\frac{\chi^2_{\alpha_m/2,n_m}}{\sum_{i=1}^{n_m}(X_{mi}-\mu_m)^2}\right)\right)\right]\\ =& \frac{1}{M}\sum_{m=1}^{M} h_\beta\left(\log\left(\chi^2_{\alpha_m/2,n_m}/\chi^2_{1-\alpha_m/2,n_m}\right)\right)\\ =& \frac{1}{M}\sum_{m=1}^{M}\frac{\log\left(\chi^2_{\alpha_m/2,n_m}\right)-\log\left(\chi^2_{1-\alpha_m/2,n_m}\right)}{\beta+\log\left(\chi^2_{\alpha_m/2,n_m}\right)-\log\left(\chi^2_{1-\alpha_m/2,n_m}\right)}\\ \boldsymbol{R_1(\sigma^2,\delta)} =& 1-\boldsymbol{P_{\sigma^2}}\left[\left(\sum_{m=1}^{M} L_1(\sigma_m^2,w_m)\right)=0\right]\\ =& 1-\boldsymbol{P_{\sigma^2}}\left[\bigcap_{m=1}^{M}\left\{\sigma_m^2\in(LB_m(X_m;\alpha_m),UB_m(X_m;\alpha_m))\right\}\right]\\ =& 1-\prod_{m=1}^{M} P_{\sigma_m^2}\left[\sigma_m^2\in(LB_m(X_m;\alpha_m),UB_m(X_m;\alpha_m))\right]\\ =& 1-\prod_{m=1}^{M}\left[F_{n_m}\left(\chi^2_{\alpha_m/2,n_m}\right)-F_{n_m}\left(\chi^2_{1-\alpha_m/2,n_m}\right)\right]\end{aligned}$$

We reparametrize $\nu_m^U = \chi^2_{\alpha_m/2,n_m}$ and $\nu_m^L = \chi^2_{1-\alpha_m/2,n_m}$ for the numerical stability in the optimization. Then the problem can be restated as follows:

$$\text{Minimize } \frac{1}{M}\sum_{m=1}^{M}\frac{\log\left(\nu_m^U\right)-\log\left(\nu_m^L\right)}{\beta+\log\left(\nu_m^U\right)-\log\left(\nu_m^L\right)}$$

$$\text{subject to } \sum_{m=1}^{M}\log\left(F_{n_m}\left(\nu_m^U\right)-F_{n_m}\left(\nu_m^L\right)\right)\geq\log(1-q)$$

As the interval estimators are non-symmetric, there will be twice as many equations as in the normal mean case. However, these equations can be similarly solved by setting up a Lagrange equation. The details are provided in Appendix B.

2.4.2 Optimization Result

For the illustration, an MIE is constructed for $M = 200$ normal scale parameters with a global level, $1 - q = 0.9$. A sequence of positive integers from 2 to 201 is used for the degrees of freedom of the individual IEs. <Figure 2.2> shows the result of the dynamic allocation. As with the normal mean case, the horizontal line indicates

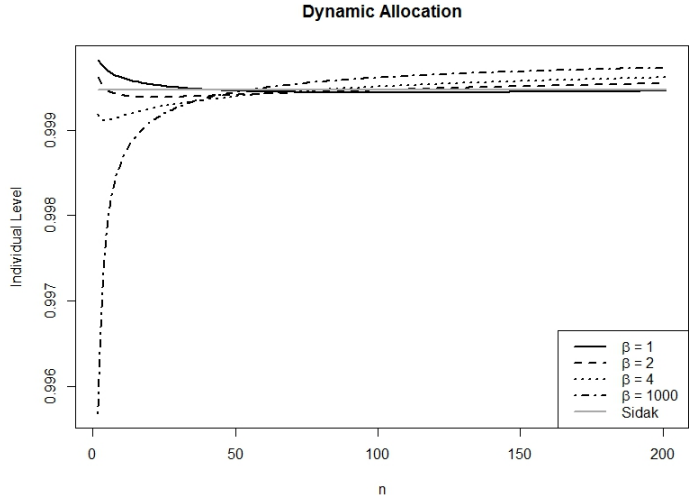


Figure 2.2: Individual Level Allocation: Normal Variance

the constant allocation through the Sidak procedure: $(1-q)^{1/M}$. As β increases, fewer individual levels are assigned to the IEs with smaller degrees of freedom and vice versa. This result can be seen contradictory because the intervals with smaller

degrees of freedom would result in smaller expected interval lengths with respect to the Lebesgue measure. However, the individual expected length is calculated by the invariant measure, $E_{\sigma_m^2}\left[\log(UB_m(X_m;\alpha_m)) - \log(LB_m(X_m;\alpha_m))\right]$. This results in wider expected lengths for the IEs with smaller degrees of freedom as the log value of lower bound becomes negative.

Now, corresponding global REL is provided in <Table 2.2>. The general behavior

Table 2.2: Relative Expected Length with respect to β

β	1	2	4	1000	Sidak Adj.
REL	1.0085	0.9950	0.9848	0.9767	1

is similar to the location case, providing smaller REL as β increases. The REL becomes stable same for $\beta \geq 1000$. Similar to the normal mean case, the reduction is achieved by the process which plugs in smaller individual levels to the IEs with smaller degrees of freedom and vice versa.

2.5 Application: Median of Unknown Distribution

The goal of this application is to construct an optimal MIE for M medians. In particular, we do not assume any distributions; therefore, the MIE is constructed based on a nonparametric procedure. The resulting nonparametric MIE is based on the daily trading price of 29 Dow Jones companies (Visa Inc. was excluded since its initial public offering was in 2008) collected in 2683 consecutive business days from January 3, 2005 to August 29, 2015. The variable we are interested in is the ith daily stock return: $\text{return}_i = \frac{\text{Closing}_i - \text{Closing}_{i-1}}{\text{Closing}_{i-1}}$. The particular nonparametric method we will use is a bootstrap procedure.

2.5.1 Bootstrap-t MIE for Medians

Suppose the mth company's stock return data are i.i.d. and follows from unknown distribution F_m with the corresponding median θ_m for $m = 1, 2, \ldots, 29$. We start the

procedure by drawing the sample median, $\hat{\theta}_m$, from this original mth company's data:

$$X_{m1}, X_{m2}, \ldots, X_{mn_m} \;\Rightarrow\; \hat{\theta}_m.$$

Among the bootstrap methods, the bootstrap-t method allows us to apply the optimization procedure by providing bootstrap-t sampling distributions as well as estimated standard errors. In this method, along with the usual B number of the bootstrap resampling, we also draw additional b number of sub-resampling to obtain the estimated standard error. Using these estimated standard errors, we derive the standardized medians and collect them to create the bootstrap sampling distributions. The procedure for the mth bootstrap-t IE is as follows:

$$X_{m1}, X_{m2}, \ldots, X_{mn_m} \;\Rightarrow\; \hat{\theta}_m$$

(Resample)

$$\begin{cases}
X^{1^*}_{m1}, X^{1^*}_{m2}, \ldots, X^{1^*}_{mn_m} \;\Rightarrow\; \hat{\theta}^*_{m1} \\
\quad \text{(Sub-resample)} \\
\quad \begin{cases}
X^{1^*1^*}_{m1}, X^{1^*1^*}_{m2}, \ldots, X^{1^*1^*}_{mn_m} \;\Rightarrow\; \hat{\theta}^{**}_{m11} \quad \Rightarrow\; t^*_{m1} = \frac{\hat{\theta}^*_{m1} - \hat{\theta_m}}{\hat{s.e.}^*_{m1}} \\
\qquad \vdots \qquad\qquad\qquad \vdots \quad \Rightarrow\; \hat{s.e.}^*_{m1} \\
X^{1^*b^*}_{m1}, X^{1^*b^*}_{m2}, \ldots, X^{1^*b^*}_{mn_m} \;\Rightarrow\; \hat{\theta}^{**}_{m1b}
\end{cases} \\
\quad \vdots \qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots \\
X^{B^*}_{m1}, X^{B^*}_{m2}, \ldots, X^{B^*}_{mn_m} \;\Rightarrow\; \hat{\theta}^*_{mB} \\
\quad \text{(Sub-resample)} \\
\quad \begin{cases}
X^{B^*1^*}_{m1}, X^{B^*1^*}_{m2}, \ldots, X^{B^*1^*}_{mn_m} \;\Rightarrow\; \hat{\theta}^{**}_{mB1} \quad \Rightarrow\; t^*_{mB} = \frac{\hat{\theta}^*_{mB} - \hat{\theta_m}}{\hat{s.e.}^*_{mB}} \\
\qquad \vdots \qquad\qquad\qquad \vdots \quad \Rightarrow\; \hat{s.e.}^*_{mB} \\
X^{B^*b^*}_{m1}, X^{B^*b^*}_{m}, \ldots, X^{B^*b^*}_{mn_m} \;\Rightarrow\; \hat{\theta}^{**}_{mBb}
\end{cases}
\end{cases}$$

<Figure 2.3> represents an example of the bootstrap-t sampling distributions for Apple Inc. In order to utilize the sampling distributions for the optimization, we need to achieve a certain degree of smoothnesses. To do so, we use the kernel density

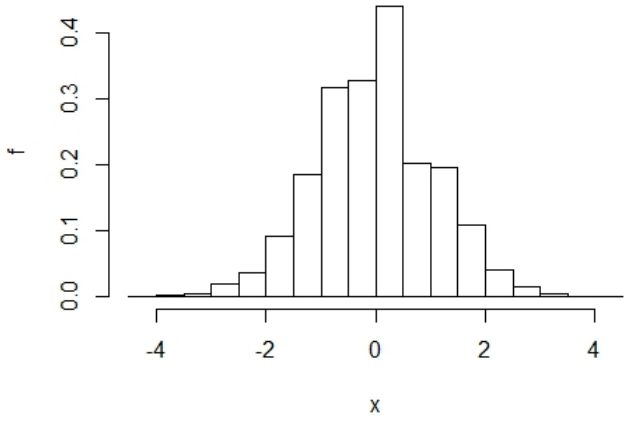


Figure 2.3: Bootstrap-t Sampling Distribution with B=10000 and b=30: Apple

estimator and increase the tuning parameter, h, up to the point that the empirical densities achieve the sufficient smoothness.

2.5.2 Kernel Density Estimation

The usual kernel density estimator is as follows:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

where $K(\cdot)$ is a non-negative function integrated to one with mean zero. In this application, we use the standard normal density, $\phi(\cdot)$, for K. Then the smoothing parameter h will be increased to obtain the necessary smoothness for the optimization. Once h is chosen, then we utilize the same value for $\hat{F}_h(x)$ and $\hat{f}'_h(x)$ which are the smoothed kernel estimators for the distribution function and the derivative of the density function, respectively:

$$\hat{F}_h(x) = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{x - x_i}{h}\right);$$
$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} \phi\left(\frac{x - x_i}{h}\right);$$

$$\hat{f'}_h(x) = \frac{1}{nh^2}\sum_{i=1}^{n} -(x - x_i)\phi\left(\frac{x - x_i}{h}\right).$$

The example of Apple Inc. is in <Figure 2.4>.

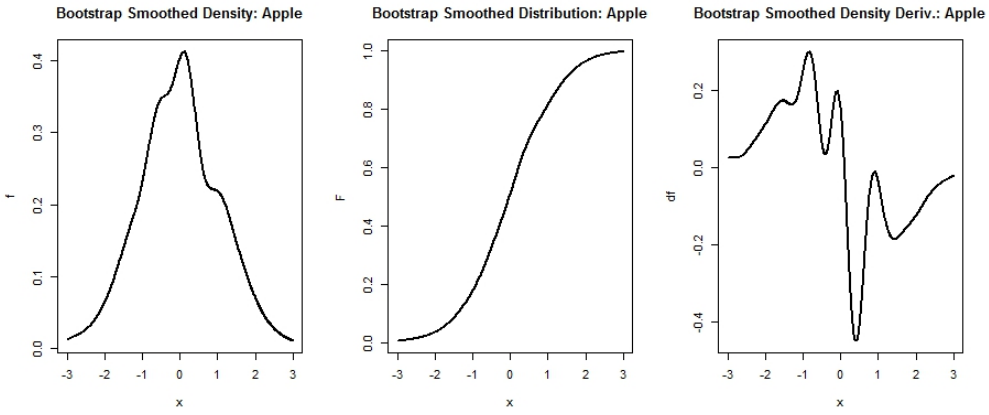


Figure 2.4: Example: Smoothed Function Estimators for Apple Inc.

2.5.3 Optimization Procedure

The bootstrap sampling distributions are, in general, asymmetric. Therefore, we use a similar approach as in the scale parameters case:

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{M}\sum_{m=1}^{M} \frac{\left\{\hat{F}_h^{-1}(1-\alpha_m/2) - \hat{F}_h^{-1}(\alpha_m/2)\right\}\hat{\sigma}_m/\sqrt{n_m}}{\beta + \left\{\hat{F}_h^{-1}(1-\alpha_m/2) - \hat{F}_h^{-1}(\alpha_m/2)\right\}\hat{\sigma}_m/\sqrt{n_m}} \\ \text{subject to} \quad & \sum_{m=1}^{M} \log(1-\alpha_m) \geq \log(1-q) \end{aligned}$$

As with the normal case, we do the following reparametrizations: $v_m^l = \hat{F}_h^{-1}(\alpha_m/2)$, $v_m^u = \hat{F}_h^{-1}(1-\alpha_m/2)$, and $\sigma_m = \hat{\sigma}_m/\sqrt{n_m}$. Then the problem becomes

$$\begin{aligned} \text{minimize} \quad & \frac{1}{M}\sum_{m=1}^{M} \frac{(\nu_m^u - \nu_m^l)\sigma_m}{\beta + (\nu_m^u - \nu_m^l)\sigma_m} \\ \text{subject to} \quad & \sum_{m=1}^{M} \log\left\{\hat{F}_h(\nu_m^u) - \hat{F}_h(\nu_m^l)\right\} \geq \log(1-q) \end{aligned}$$

This problem can be numerically solved, and the details are provided in Appendix B.

2.5.4 Optimization Result

We first present the optimal allocation of individual sizes, i.e., 1 - confidence level, with respect to estimated standard errors for different β in <Figure 2.5>. Observe

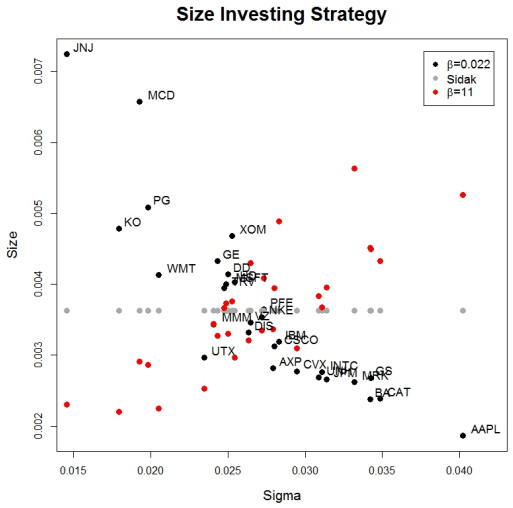


Figure 2.5: Size Allocation of MIE for Stock Return Data

that the negative relation between size and estimated standard error for β = 0.022. On the other hand, we can see the positive relation between size and estimated standard error for large β = 11. Although these relations do not follow perfect curves, the tendencies are similar to <Figure 2.1>. The first and second panel of <Figure 2.6>

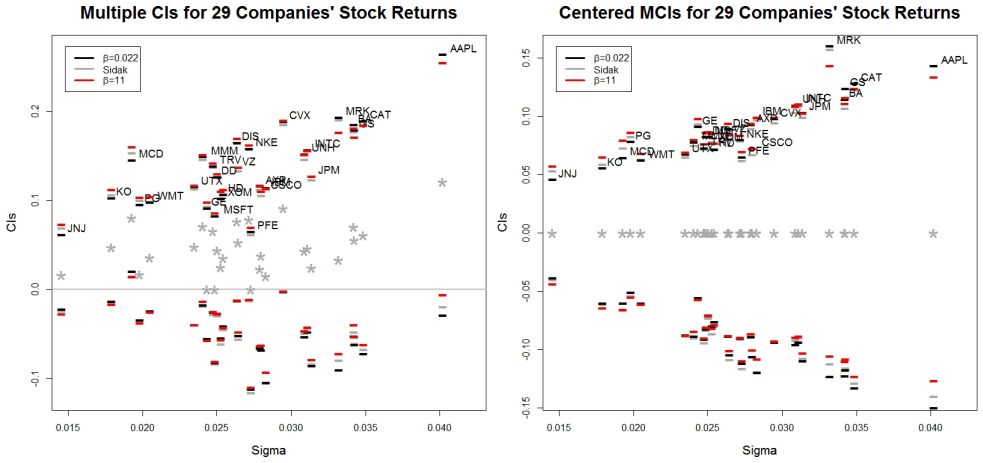


Figure 2.6: Bootstrap-t MIE for Stock Return Data

show the actual and centered individual IEs along with the IEs based on the Sidak procedure, respectively. With β = 0.022, the IEs with smaller standard errors have smaller relative expected lengths, and the relative expected length increase as the standard error increase. With β = 11, on the other hand, the IEs with smaller standard errors have larger relative expected lengths and these decrease when standard errors increase. The total lengths of the MIE for the stock return data are 5.216 and 5.186 for the β = 0.022 and β = 11, respectively. In addition, the baseline criterion Sidak procedure provides the total length 5.203. This also coincides with the normal mean application in <Table 2.2>.

Chapter 3

Bayes MIE with Thresholding

3.1 Overview

Suppose an interval estimator (IE) is to be developed for a single parameter. Then one could evaluate the estimator by investigating two quantities: the expected interval length and coverage probability with respect to the true parameter. Except in some special cases, the forms of these quantities, especially the coverage probability, would depend on the true parameter value. A classical approach to summarize this information is through the confidence coefficient, the infimum of coverage probabilities throughout the parameter values. However, this is not the only way of summarization. In particular, when prior information is available, we can summarize the information by integrating the expected interval length and coverage probability with respect to the prior distribution. These quantities are denoted as Bayes expected length (BEL) and Bayes coverage probability (BCP), since this particular integration is reminiscent of the derivation of the Bayes risk. The procedure, constructed based on these quantities, is called Bayes MIE (BMIE), and is not within the class of classical confidence intervals. This is because the coverage probability at a specific parameter value may be lower than a given level, although the procedure maintains the BCP at least the nominal level. However, the modification through the integration allows us to reflect the prior information to the expected interval length and coverage probability without losing their own characteristics. Therefore, an optimality of a single IE can be sought by minimizing the BEL, while maintaining the BCP to be at least a nominal

level.

When the dimension of an estimation problem becomes higher, i.e., multiple parameters (or multiple functions of parameters) are targeted, an optimality of the corresponding multiple interval estimator (MIE) can be achieved by replacing the BEL and BCP with their relevant extensions. Under an independence assumption throughout the individual IEs, it is reasonable to choose the Bayes average expected length (BAEL) for the extension of the BEL. On the other hand, the extension of the BCP can be accomplished in various ways. In this study, we start from an idea widely used in the field of multiple testing and transfer it to multiple interval estimation by using the duality between a α-size hypothesis testing and a 1-α confidence interval estimation.

Family-wise error rate (FWER) is one of the most well-established extensions of type-I error rate in the field of multiple testing. It is defined as the probability of committing at least one type-I error. (Hochberg and Tamhane [27]) Therefore, in a multiple testing problem, one attempts to build up a procedure which controls FWER at most a global level of q, while trying to maximize a global power, pertaining to the average number of true rejections. Many multiple testing procedures (MTPs) are constructed based on FWER because it is conceptually intuitive and mathematically manageable. However, it tends to lead relatively to a lower global power, i.e., an inefficiency in practical terms. Therefore, the improvement upon the global power has been one of the main goals for MTPs using FWER.

Now, we define an extension of the coverage probability by transferring the idea of the FWER to multiple interval estimation.

Definition 3.1. Let M be the number of target parameters. Suppose we construct an MIE and an MTP which are related through the duality between individual IEs and testings for these parameters. Then we define family-wise coverage rate (FWCR)

as follows:

$$\begin{aligned}\text{FWCR} &= Pr[\text{MIE contains all true parameters}] \\ &= 1 - Pr[\text{MTP causes at least one type-I error}] = 1 - \text{FWER}\end{aligned}$$

This definition implies that an MIE which maintains the FWCR at least a global level of $1 - q$ is the mathematical dual of an MTP which controls the FWER at most a global size of q. Due to the way that the FWCR is defined, it naturally inherits the characteristics of the FWER; as a result, while the FWER tends to lead to a lower global power, the FWCR tends to result in a wider global expected content. Our goal is to overcome this limitation of the multiple interval estimation using FWCR by adopting a thresholding approach. The thresholds require additional information to be established, and we utilize prior information for the purpose. Under the existence of the prior information, we introduce a Bayes FWCR (BFWCR) to summarize global coverage information by integrating FWCR with respect to the prior distribution. Note that the FWCR can be expressed as the product of M BCPs under the independent assumption. In conclusion, our MIE attempts to minimize the BAEL while maintaining the BFWCR at least the global level of $1 - q$.

To motivate the thresholding approach, let us consider that an MIE consists of two-sided individual IEs. If we can devise a reasonable way to ensure that a subset of the target parameters resides on one side of the corresponding IEs, then the other side of the IEs could be removed in order to minimize the global expected content, i.e., the BAEL, of the MIE. This removal process can be executed by using a pair of thresholds which can be set up with additional information. To provide the additional information, we adopt a prior distribution in addition to the classical setting: $\boldsymbol{X}_m|\theta_m \sim F_m(\cdot)$ and $\theta_m \sim \Pi_m(\cdot)$ where $\boldsymbol{X}_m$ is the mth random sample and the variables are independent throughout $m = 1, 2, \ldots, M$.

The optimal threshold can be obtained through a modification of the optimization procedure in chapter 2. In this study, the global loss functions are defined to reflect

the BAEL and BFWCR, so that the resulting risk functions are plugged into the optimization procedure. The resulting procedure then provides not only the best threshold but also the the optimal individual levels concurrently.

We review related studies in section 2. The Bayes MIE with thresholding (BMIE Thres) under the normal-normal model is introduced in section 3, with its analytic properties and behaviors with respect to the thresholds. In section 4, the modified optimization procedure is presented, and the optimal thresholding parameter is investigated. In section 5 and 6, the performance of the BMIE Thres is demonstrated by data applications under the known and unknown standard deviations, respectively. In-season baseball batting average data is applied to the procedure for the former case, and leukemia gene expression data is applied for the latter case. In section 7, we perform a simulation study to take care of the prior misspecification, as well as to compare BMIE Thres with Bayesian credible MIEs. Lastly, discussions and future works are presented in the last section.

3.2 Related Studies

The multiplicity issue is a fundamental problem whenever an inferential procedure attempts to handle a set of parameters simultaneously. Due to the issue, we cannot simply assign the usual 0.05 individual size to MTPs or 0.95 individual level to MIEs because the global type-I error increases or the global coverage probability decreases as the number of parameters increases. (Lehmann and Romano [37]) In earlier studies, such as Miller Jr [38], one of the first published books on multiple inferences, the meaning of *multiple* was usually about less than 10 parameters. However, in the 1990's, the issue was magnified in earnest as the dimensions of problems became much higher due to the influence of high-throughput data. (Efron [13]) As a result, the amount of research in MTPs was largely boosted.

The FWER is one of the classical global type-I error rate to involve the mul-

tiplicity issue in MTPs. Suppose we consider a problem of multiple testing under the FWER for M number of target parameters with a global size of q. An intuitive multiple adjustment for the individual sizes would be the Bonferroni approach, q/M. If an independence assumption is satisfied, the Sidak approach, $1-(1-q)^{1/M}$, would also be a valid adjustment. This means, by applying these individual sizes to the MTP, the FWER can be controlled to be less than or equals to the global size of q. However, these one-step approaches generally result in low global powers, e.g., a limited number of rejections. (Shaffer [50]) In order to overcome this limitation, the step-wise approaches were introduced by Holm [29] and Hochberg [28]. These procedures utilize the information of ordered p-values to assign particular sizes to the corresponding individual testings. Also, Westfall and Young [54] suggested a resampling procedure in order to use the dependence structure of p-values to increase the global power. Another general approach is a p-value weighting. This approach seeks optimal weights for p-values to achieve a higher global power. Naturally, the issue is how to choose the optimal weights. To handle this, Westfall, Krishen, and Young [53] and Dobriban et al. [10] assumed prior information. Although the procedures utilize prior information to choose the weights, they never claimed their approaches to be Bayesian, as the procedures are aimed to control the FWER, a frequentist global type-I error rate. Instead, Dobriban called their approach quasi-Bayesian. Peña, Habiger, and Wu [41] consider the problem of multiple testing as a general problem under a decision theoretic framework. Their MTP allocates optimal sizes to individual tests to maximize a global power under the FWER and false discovery rate (FDR) by Benjamini and Hochberg [1].

In early works on multiple interval estimation, e.g., Scheffé [48] and Roy and Bose [47], the researchers mainly considered it as a part of multiple comparisons among a few number of parameters within the setting of ANOVA or regression. In addition, Benjamini and Yekutieli [2] pointed out that the cases of ignoring the multiplicity

adjustment in multiple interval estimations were more frequent than the cases in multiple testings even after the influence of the high-throughput data in the 1990's. Furthermore, some existing studies introduced MIEs as the concomitant procedures of established MTPs. However, with the lack of information from alternative hypotheses in MIEs, there exists no explicit relation between the global power in the MTPs and the global expected length in the MIEs. This implies that the multiple interval estimation is of its own area and worth investigating as equivalent of the multiple testing. A good example is the relation between the FDR and the false coverage rate (FCR) by Benjamini and Yekutieli [2]. Although two concepts represent the global type-I error rate and global coverage probability developed by the same group of researchers, the lack of alternative information forced the authors to bring in the concept of parameter selection in relation to another topic, a selective inference. (Fithian, Sun, and Taylor [19])

There are several studies that investigated MIEs using the empirical Bayes framework, and these are closely related to our study in terms of the modeling perspective. Morris [39] investigated the empirical Bayes interval estimation under the same setting as Efron and Morris [15], which studied the point estimation. Casella and Hwang [8] studied parametric empirical Bayes confidence sets for multivariate normal means. Due to the shrinkage effect, the empirical Bayes confidence interval provides shifted estimates, which result in better coverage probability. In the next section, we exploit the model setup of the parametric empirical Bayes for our MIE to establish a pair of thresholds. A similar thresholding idea was considered in Habiger and Peña [24] for MTPs to maximize a global power.

3.3 Bayes Multiple Interval Estimator with Thresholding

3.3.1 Mathematical Framework

Consider a statistical model $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{P}})$ where $\boldsymbol{\mathcal{X}} = \times_{m=1}^{M} \mathcal{X}_m$, $\boldsymbol{\mathcal{F}} = \sigma\left(\times_{m=1}^{M} \mathcal{F}_m\right)$, and $\boldsymbol{\mathcal{P}}$ is the collection of probability measures on the product space. In addition, $(\boldsymbol{\Theta}, \boldsymbol{\mathcal{T}})$ is a measurable space with $\boldsymbol{\Theta} = \times_{m=1}^{M} \Theta_m$ and $\boldsymbol{\mathcal{T}} = \sigma\left(\times_{m=1}^{M} \mathcal{T}_m\right)$ where θ_m and $\mathcal{T}_m$ are the mth parameter space and its corresponding sigma field, respectively. Suppose $\boldsymbol{P_\theta}$ is the family of probability distribution functions with respect to $\boldsymbol{\mathcal{P}}$ on $\boldsymbol{\mathcal{X}}$. A random quantity $\boldsymbol{X}$ is generated from $\boldsymbol{P_\theta}$; moreover, another random quantity $\boldsymbol{U}$, which is independent of $\boldsymbol{X}$, is generated from M standard uniform distributions to consider the randomized procedure. Note that $\boldsymbol{X}$ consists of X_ms which are independent throughout the index $m = 1, 2, \ldots, M$, so that $\boldsymbol{P_\theta}(\boldsymbol{x}) = \prod_{m=1}^{M} P_{\theta_m}(x_m)$. Moreover, let $\mathscr{P}_m$ be a probability measure on Θ_m with the corresponding distribution function Π_m for $m = 1, 2, \ldots, M$. For each fixed $\theta_m \in \Theta_m$, let f_{θ_m} be the density with respect to the Lebesgue measure on $\mathcal{X}_m$ and assume the mapping $(\theta_m, x_m) \mapsto f_{\theta_m}(x_m)$ is product-measurable.

Given $\alpha_m \in (0, 1)$, we summarize the information of coverage probability for the mth IE, Γ_m, by integrating of the coverage probability with respect to the prior distribution, Π_m: $\int_{\Theta_m} P_{\theta_m}[\theta_m \in \Gamma_m(X_m, U_m; \alpha_m)] d\Pi_m(\theta_m)$. This is the Bayes coverage probability (BCP) of the mth IE. Now, we define a Bayes MIE as follows:

Definition 3.2. Given $q \in (0, 1)$, $100 \times (1 - q)\%$ Bayes Multiple Interval Estimator (BMIE) for $\boldsymbol{\theta}$ is a map, $\boldsymbol{\Gamma(\cdot)} : \boldsymbol{\mathcal{X}} \times (0, 1)^M \longrightarrow \boldsymbol{\mathcal{T}}$, such that

$$\prod_{m=1}^{M} \int_{\Theta_m} P_{\theta_m}[\theta_m \in \Gamma_m(X_m, U_m; \alpha_m)] d\Pi_m(\theta_m) \geq 1 - q.$$

The left-hand side quantity is called the Bayes family-wise coverage rate (BFWCR) of $\boldsymbol{\Gamma(X, U)}$. In addition, an individual interval estimator in a BMIE is called a Bayes interval estimator (BIE).

In a classical interval estimation for a location parameter, the expected content can be quantified by the Lebesgue measure, λ. In general, however, there exist compatible measures for distinct parameters, e.g., $\frac{1}{\theta}d\lambda(\theta)$ for a scale parameter. Therefore, we let ν be a measure for the general content. Then, $\int_{\theta_m} E_{\theta_m}[\nu(\Gamma_m(X_m, U_m; \alpha_m))]d\Pi_m(\theta_m)$ is the expected content of mth individual IE and called the Bayes expected length (BEL). Based on this, we define the global expected content of a BMIE.

Definition 3.3. The Bayes average expected length (BAEL) of a BMIE is defined as follows:

$$\frac{1}{M}\sum_{m=1}^{M}\int_{\Theta_m} E_{\theta_m}[\nu(\Gamma_m(X_m, U_m; \alpha_m))]d\Pi_m(\theta_m).$$

Initially, the mth individual IE, $\Gamma_m(X_m, U_m; \alpha_m)$, is simple as it depends only on X_m and U_m with respect to the given level, α_m. However, if a BMIE procedure involves hyper-parameter estimations and optimizations as in the later sections, then the individual estimator becomes composite for $\boldsymbol{X}$ and $\boldsymbol{U}$. Given this setting, the goal of study is to seek a BMIE, $\boldsymbol{\Gamma(X, U)}$, which minimizes the corresponding BAEL.

3.3.2 Thresholding Idea and Form of BMIE Thres

In this subsection, we introduce a particular BMIE equipped with a pair of thresholds to reduce its BAEL. This procedure is called a BMIE with thresholding (BMIE Thres). Suppose we are interested in M normal means with known standard deviations. Random samples are observed from the normal distributions and these are independent throughout the index m from 1 to M. In addition, the prior information of the location parameters is available; therefore, a normal prior distribution is assumed with the hyper-parameters, η and τ:

$$\bar{X}_m|\mu_m \sim \mathcal{N}\left(\mu_m, \sigma_m^2\right) \;\; \&; \; \mu_m \sim \mathcal{N}\left(\eta, \tau^2\right) \text{ for } m = 1, 2, \ldots, M.$$

Here, the problem is simplified by the Sufficiency Principle and σ_m's, which depend on the sample sizes, denote the standard errors without loss of generality. Since we

assume a common prior distribution for every location parameter, the setting becomes identical to the parametric empirical Bayes framework in Efron and Morris [15] and Casella and Hwang [8].

Recall the well known fact that the posterior distribution of the normal-normal model also follows a normal distribution (Berger [3]) as follows:

$$\mu_m|\bar{x}_m \overset{\text{ind.}}{\sim} \mathcal{N}\left(\hat{\mu}_m, \beta_m\sigma_m^2\right) \text{ for } m = 1, 2, \ldots, M$$

where $\hat{\mu}_m = \beta_m\bar{x}_m + (1-\beta_m)\eta$ and $\beta_m = \frac{\tau^2}{\tau^2+\sigma_m^2}$. In particular, the posterior mean, $\hat{\mu}_m$, is a convex combination of the maximum likelihood estimate, $\bar{x}_m$, and the prior mean, η. Now, to illustrate the idea of thresholding, suppose σ_m's are identical and τ is fixed so that all the posterior distributions have a constant dispersion. Next,

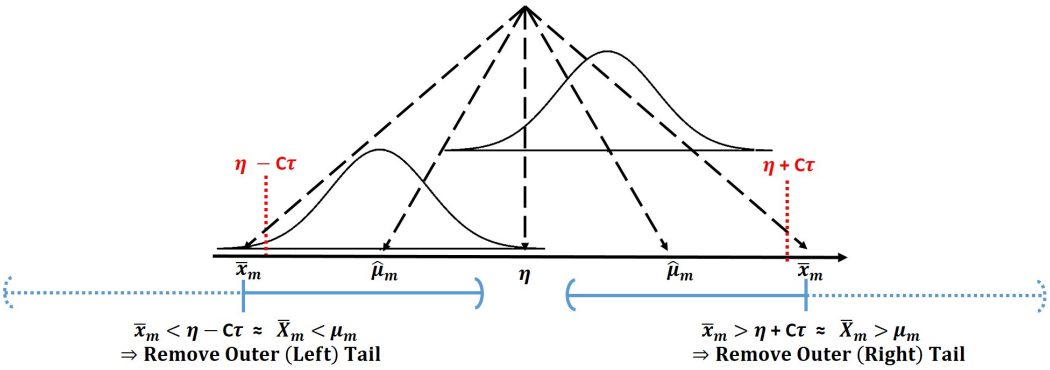


Figure 3.1: Idea of Thresholding Approach

envision a situation in which an observed $\bar{x}_m$ deviates from the prior mean, η. Then the distance between $\bar{x}_m$ and $\hat{\mu}_m$ becomes larger in proportion to the increase of the distance between $\bar{x}_m$ and η. Still, we construct an IE in a classical way, letting $\bar{x}_m$ to be the center of the mth observed IE. Now let us call the inner tail for the side of IE close to the η and the outer tail for the opposite side. Note that as the estimate $\bar{x}_m$ reaches far outside from the η, the posterior probability that μ_m resides in the inner tail becomes greater. In this situation, we would remove the outer tail to reduce the global expected length.

The decision on the removal can be made by using a pair of thresholds on both sides: $\eta - C\tau$ and $\eta + C\tau$. That is, once an estimate $\bar{x}_m$ falls outside of the thresholds, then we discard the outer tail and keep only the inner tail. In actual situations, σ_m would vary; however, the idea remains the same although the effect of thresholding is affected by the size of σ_m in relation to the size of τ. With this motivation, the mth BIE with thresholding (BIE Thres) has the following form equipped with indicator functions on both sides:

$$\Gamma_m(\boldsymbol{X}_m; \alpha_m) = \left(\bar{X}_m - z_{\alpha_m/2}\sigma_m I\left\{\bar{X}_m > \eta - C\tau\right\}, \bar{X}_m + z_{\alpha_m/2}\sigma_m I\left\{\bar{X}_m < \eta + C\tau\right\}\right)$$

where $z_{\alpha_m/2} = \Phi^{-1}(1 - \alpha_m/2)$. Note that η and τ will be estimated in the actual implementation. In addition, the BIE Thres does not require a randomizer, U_m, as the distribution of $\bar{X}_m$ is continuous.

3.3.3 Individual Performance Quantities

In this subsection, we ascertain the forms of the performance quantities of a single BIE Thres under the normal-normal model. In addition, their properties with respect to C is investigated in relation to the classical z-based confidence interval estimator for a normal location parameter (z-based IE). First, the following proposition shows the form and property of the BEL of the mth BIE Thres:

Proposition 3.4. *The Bayes expected length (BEL) of the mth BIE Thres has the form of*

$$BEL[C, \alpha_m; \sigma_m, \tau] = 2z_{\alpha_m/2}\sigma_m\Phi(C_m)$$

where $C_m = C\tau/\sqrt{\sigma_m^2 + \tau^2}$, and the BEL approaches the expected length of the corresponding z-based IE as C goes to infinity.

Notice that the BEL consists of two parts: the first part, $2z_{\alpha_m/2}\sigma_m$, is the same as the BEL of the z-based IE; and the second part, $\Phi(C_m)$, reflects the thresholding

effect. From the form of C_m, it is evident that the BEL of the BIE Thres approaches the BEL of the z-based IE as C goes to infinity. Next, we introduce the form and property of the BCP of the mth BIE Thres in the following proposition:

Proposition 3.5. *The Bayes coverage probability (BCP) of the mth BIE Thres has the form of*

$$BCP[C, \alpha_m; \sigma_m, \tau] = 2 \int_{-\infty}^{C_m} \left\{ \Phi\left(\frac{\sigma_m}{\tau} y + \sqrt{1 + \frac{\sigma_m^2}{\tau^2}} z_{\alpha_m/2} \right) - \Phi\left(\frac{\sigma_m}{\tau} y \right) \right\} d\Phi(y)$$

where $C_m = C\tau/\sqrt{\sigma_m^2 + \tau^2}$, and the BCP approaches the coverage probability of the corresponding z-based IE as C goes to infinity.

Although the BCP of the BIE Thres has no closed form, it is at least second differentiable with respect to α_m so that it can be applied to the optimization in the next section. It is also not difficult to show that the BCP of BIE Thres approaches $1 - \alpha_m$, the BCP of the z-based IE.

3.3.4 Global Performance Quantities

In the previous subsection, we studied the performance quantities of a single BIE Thres. Now, we investigate global performance quantities for a BMIE Thres. All the quantities are derived by assuming M target parameters. First, we introduce a Bayes threshold ratio (BTR). It is a ratio of the number of the thresholded BIEs to the total number of the BIEs in a BMIE Thres, i.e., the proportion of the one-sided BIEs. The form of the BTR is presented in the following proposition:

Proposition 3.6. *The Bayes threshold ratio (BTR) has the form of*

$$BTR[C; \boldsymbol{\sigma}, \tau] = \frac{2}{M} \sum_{i=1}^{M} \Phi(-C_m)$$

where $C_m = C\tau/\sqrt{\sigma_m^2 + \tau^2}$. As C increases from zero to infinity, the BTR decreases from 1 to 0.

In addition to the BTR, we have two more global quantities: a Bayes relative expected length (BREL) and a Bayes family-wise coverage rate (BFWCR). The BREL is a ratio of the BAEL of a BMIE Thres to the BAEL of the corresponding z-based MIE, and it is the global expected content of a BMIE in this study. The BFWCR is essentially the product of the M individual BCPs, and it reflects the global coverage probability of a BMIE.

Corollary 3.7. *Given M parameters and a global level $1-q$, the Bayes relative expected length (BREL), Bayes family-wise coverage rate (BFWCR) and Bayes thresholding ratio (BTR) have the following forms:*

$$\begin{aligned}
BREL[C, \boldsymbol{\alpha}; \boldsymbol{\sigma}, \tau] &= \frac{1}{M}\sum_{m=1}^{M}\left[2z_{\alpha_S}\sigma_m\Phi(C_m)\right] \Big/ \frac{1}{M}\sum_{m=1}^{M}\left[2z_{\alpha_S}\sigma_m\right]; \\
BFWCR[C, \boldsymbol{\alpha}; \boldsymbol{\sigma}, \tau] &= \prod_{m=1}^{M}\left[2\int_{-\infty}^{C_m}\left\{\Phi\left(\frac{\sigma_m}{\tau}y + \sqrt{1+\frac{\sigma_m^2}{\tau^2}}z_{\alpha_S}\right) - \Phi\left(\frac{\sigma_m}{\tau}y\right)\right\}d\Phi(y)\right]; \\
BTR[C; \boldsymbol{\sigma}, \tau] &= \frac{2}{M}\sum_{m=1}^{M}\Phi(-C_m)
\end{aligned}$$

where $C_m = C\tau/\sqrt{\sigma_m^2+\tau^2}$, and α_S is the Sidak adjustment, $1-(1-q)^{1/M}$.

From the previous propositions, it is straightforward to show the BREL converges to 1 and the BFWCR converges to $1-q$ as C goes to infinity. In addition, the BTR converges to 0 as C goes to infinity, implying the BMIE Thres gets closer and closer to the corresponding z-based MIE.

The global quantities depend on the thresholding parameter C. Thus, we graphically ascertain their behaviors with respect to C in <Figure 3.2>. It is based on a BMIE Thres for $M = 1000$ normal location parameters under the global level $1-q = 0.9$. For the setting, we assign an equi-spaced sequence from 0.01 to 10 for the standard errors, σ_m's and three different value 2, 3, and 5 for the prior standard deviation, τ. Then we plot the global quantities in (Corollary 3.7) for increasing C from 0 to 6. The solid, dashed, and dotted lines indicate the different τ's: 2, 3, and 5, respectively; and dark red, black, and light red color lines indicate the BREL,

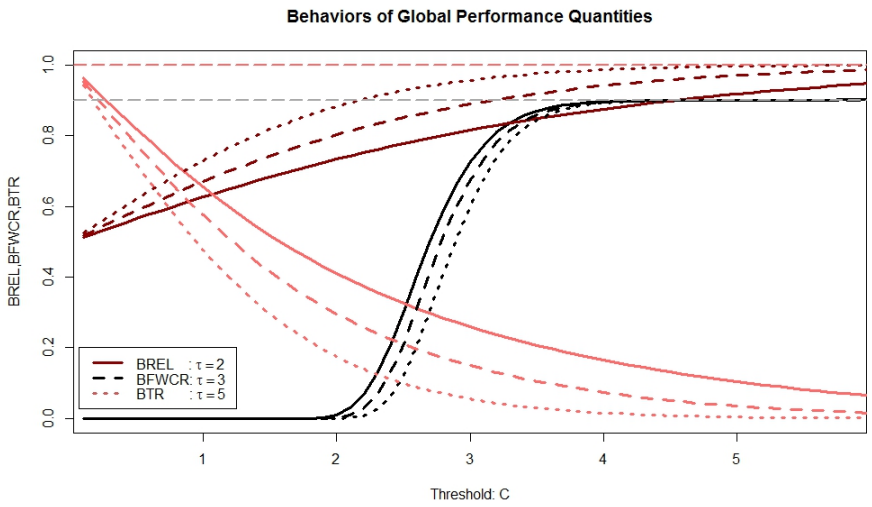


Figure 3.2: Three Global Quantities of BMIE Thres for different τ's

BFWCR, and BTR of the BMIE Thres. First, note that the opposite behaviors of the BTR and BREL. When C equals zero, every BIE lacks an outer tail, so the BTR is one and the BREL is one half. As C increases, the number of one sided BIEs reduces, and this results in the decrease of the BTR and the increase of the BREL. The rates of decrease and increase depend on the value of τ because the smaller τ of the true distributions implies stronger information to remove one side of the BIEs in general. Now, the BFWCR shows an interesting behavior. It remains almost zero up to a C of around 2. This is because each BCP is quite less than the individual level; as a result, the BFWCR, i.e., the product of $M = 1000$ BCPs, becomes almost zero. However, the BFWCR rapidly recovers the value as C increases from 2 to 3, and it almost reach the global level $1 - q = 0.9$ around 3.5 for any τ's. It is important to observe that there exists certain reduction in the BREL when this occurs. This implies the BMIE Thres performs better than the z-based MIE.

At this point, it is natural to ask about the optimal value for C. The optimal C^* is the value which provides the shortest BREL of the BMIE Thres, while at the same time maintaining the BFWCR at a higher-than-global level. However, <Figure 3.2> shows that not only the BREL but also the BFWCR change as C varies. Therefore,

in order to choose the optimal thresholding parameter, C^*, we must first arrange a way to adjust the BFWCR to match the global level so that the BRELs can be fairly compared throughout the values of C. This process can be achieved using the optimization method in the next section.

3.4 Optimization

In this subsection, we perform an optimization procedure which determines the best C^* for a BMIE Thres. In the previous subsection, the BMIE Thres shows the limited ability in maintaining the BFWCR at least the global level when C is small. The optimization approach will provide a way to push the BFWCR up to the global level so that we can determine the optimal C^* which provides the minimum BREL, or equivalently the minimum BAEL. Furthermore, the approach also allows us to assign optimal levels, α_m^*'s, to individual BIEs. Note that this optimization procedure is a variation of the optimization strategy in chapter 2.

3.4.1 Optimization Procedure

The optimization procedure consists of two global risk functions, $\boldsymbol{R_0^\beta}$ and $\boldsymbol{R_1}$, which represent the adjusted BREL and 1-BFWCR, respectively. We first set up the following optimization problem:

$$\text{minimize } \boldsymbol{R_0^\beta(\theta, \delta)} \text{ subject to } \boldsymbol{R_1(\theta, \delta)} \leq q$$

where the components are as follows:

$$\begin{aligned}
\boldsymbol{R_0^\beta(\theta, \delta)} &= \frac{1}{M} \sum_{m=1}^{M} \frac{BEL[C, \alpha_m; \sigma_m, \tau]}{\beta + BEL[C, \alpha_m; \sigma_m, \tau]}; \\
\boldsymbol{R_1(\theta, \delta)} &= 1 - \prod_{i=1}^{M} BCP[C, \alpha_m; \sigma_m, \tau]
\end{aligned}$$

where $C_m = C\tau/\sqrt{\sigma_m^2+\tau^2}$. Due to a numerical stability of the procedure, we reparametrize α_m's to ν_m's. Then the restated optimization problem is as follows:

$$\text{minimize } \frac{1}{M}\sum_{m=1}^{M}\frac{BEL[C,\nu_m;\sigma_m,\tau]}{\beta+BEL[C,\nu_m;\sigma_m,\tau]}$$
$$\text{subject to } \sum_{m=1}^{M}\log\left(BCP[C,\nu_m;\sigma_m,\tau]\right)\geq\log(1-q)$$

$$\text{where } \begin{cases} BEL[C,\nu_m;\sigma_m,\tau] = & 2\nu_m\sigma_m\Phi(C_m); \\ BCP[C,\nu_m;\sigma_m,\tau] = & 2\int_{-\infty}^{C_m}\left\{\Phi\left(\frac{\sigma_m}{\tau}z_m+\frac{\sqrt{\tau^2+\sigma_m^2}}{\tau}\nu_m\right)-\Phi\left(\frac{\sigma_m}{\tau}z_m\right)\right\}d\Phi(z_m); \\ C_m=\frac{C\tau}{\sqrt{\sigma_m^2+\tau^2}}; \quad \nu_m = & z_{\alpha_m/2}=\Phi^{-1}\left(1-\alpha_m/2\right). \end{cases}$$

This optimization is feasible because $BEL[C,\nu_m;\sigma_m,\tau]$ and $BCP[C,\nu_m;\sigma_m,\tau]$ are at least second differentiable with respect to ν_m for any fixed C. Thus, we can set up a Lagrange equation, and solve it numerically by using the Newton-Raphson method.

3.4.2 Optimization Result

The result of the optimization is presented in this subsection. We exploit the setting for <Figure 3.2> in which we ascertained the behaviors of the global quantities. After the optimization, the BFWCRs are always equal to the global level, $1-q=0.9$, for any given τ and C. Therefore, we only present the resulting BRELs for different τ's, 2, 3, and 5, with respect to C in <Figure 3.3>. In the plot, the trajectories show similar inverted arch shapes on the left hand side. The lowest points represent the smallest BRELs which determine the optimal C^*'s. For the different τ's, 2, 3, and 5, the minimum BRELs are 84.7%, 92.0%, and 97.4%, with the corresponding optimal C^*'s are equal to 3.4, 3.5, and 3.8, respectively. From each optimal C^*, the corresponding BREL increases as C increases or decreases. When C increases, the increase of the BREL is evident due to the decrease of the BTR, i.e., the ratio of one-sided BIEs. When C decreases, the increase of the BREL is due to the optimization procedure which lifts up the BFWCR to the global level $1-q=0.9$. This is also apparent from

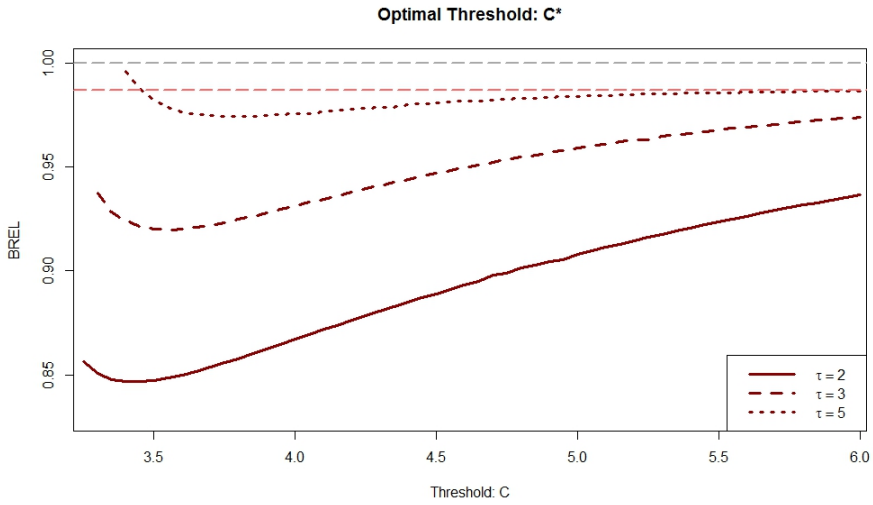


Figure 3.3: C^* obtained from Optimization

a single dimensional situation wherein we get a wider expected interval length by increasing its confidence level.

It is worth mentioning that the BRELs converge to the light red dashed line instead of 1, the gray dashed line. This is because the reduction between the gray and light red dashed lines is solely achieved by the optimal allocation of the individual levels, α_m^*'s, as shown in chapter 2. That is, the amount of reduction can also be

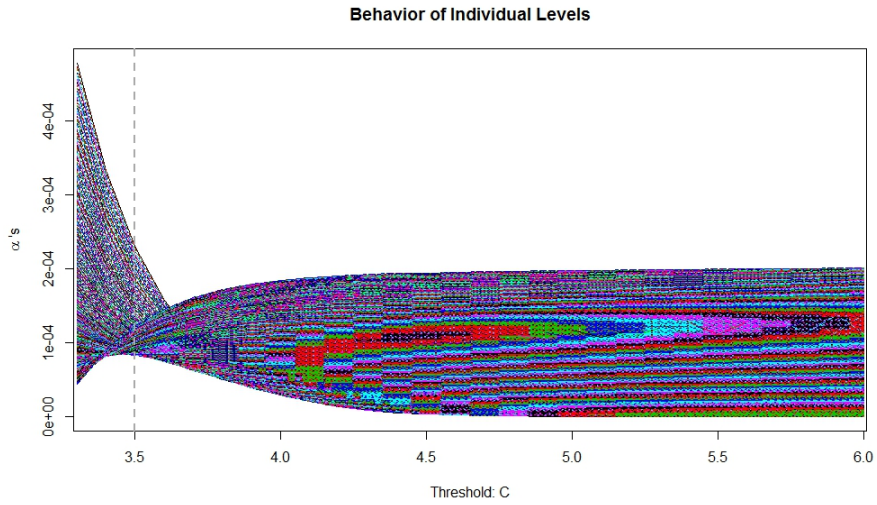


Figure 3.4: Behavior of Individual Levels, α_m's, for $\tau = 3$

obtained from the optimization procedure with the classical z-based MIE. Therefore, the rest of reductions between the minimum BRELs and the light red dashed line measures the actual thresholding effect of the BMIE Thres. <Figure 3.4> shows the optimal individual level allocation result with respect to C for $\tau = 3$. Note that the optimal $C^* = 3.5$ is the position where the behaviors of individual levels α_m's become inverted. In conclusion, the result suggests that meaningful amounts of the BREL reduction can be achieved when τ is relatively small, and the reduction is contributed by two parts: the reduction from the thresholding which can be optimized by C^* and the reduction from the optimal individual level allocation, α_m^*'s.

3.5 Data Application: Known σ_m's

In-season prediction of batting averages has been studied by several researchers for a couple of reasons, e.g., Efron and Morris [15] and Brown [6]. First, while the small portion of data at the beginning of season is used for the prediction, the batting average throughout the rest of the season can be considered as a success probability, p, which reflects player's true batting capability for the season. Therefore, one can set up a binomial model with the p along with another parameter, n, the total number of at-bats. Second, when the arcsine transformation is utilized to convert the binomial distribution into an approximate normal distribution, the standard deviation is derived solely from the number of at-bats. Therefore, the individual problem can be set up with a normal distribution with a known standard deviation, so that we can utilize the plain version of BMIE Thres procedure to make an inference.

The previous studies on the in-season prediction of batting averages utilized the parametric empirical Bayes approach to predict the true batting ability by using the beginning portion of the data. For example, Efron and Morris [15] and Efron and Morris [16] used the batting average of the first 45 at-bats in the 1970 MLB season and Brown [6] used the batting average of the first three months' data in

the 2005 MLB season for the predictions. These previous studies concentrate on the point estimations evaluated by the total squared error: $\sum(x_m - \theta_m)^2$ where θ_m is the batting average of the mth player and x_m is the corresponding estimate for θ_m. In this study, we focus on constructing a BMIE Thres and evaluating its performance by ascertaining the BREL and BFWCR. We utilize the 2005 MLB season data which is also used in Brown [6], as it involves a larger number of players than the data in Efron and Morris [15].

3.5.1 Problem Setup and Assumptions

Let H_m and N_m be the number of hits and at-bats for the mth player over the whole 2005 season. As we choose the first j month(s) for the prediction period, H_m^j and N_m^j are set as the number of hits and at-bats of the first j month(s) period for $j = 1$, 2, or, 3. Once j has been determined, the notation for $j = 4$ is reserved for the number of hits and at-bats over the rest of the season. In this application, we exclude pitchers as well as batters who have fewer than 11 at-bats either inside or outside of the prediction period. Therefore, each prediction period contains a different number of players, M_j. Given any index j for the prediction period, we first set up the following binomial model:

$$H_m \sim Binom(N_m, p_m), \text{ for } m = 1, 2, \ldots, M$$

Then the arcsine transformation is utilized to obtain the following approximated normal model, and a normal prior distribution is added:

$$X_m = \arcsin\sqrt{\frac{H_m + 1/4}{N_m + 1/2}} \overset{\text{approx.}}{\sim} \mathcal{N}(\mu_m, \sigma_m^2); \;\; \mu_m \sim \mathcal{N}(\eta, \tau^2)$$

where $\mu_m = \arcsin\sqrt{p_m}$ and $\sigma_m^2 = \frac{1}{4N_m}$. In addition, we assume that $\mu_m^j = \mu_m$ for $j = 1$, 2, 3, and 4, meaning the true batting average of the mth player does not change throughout the season. Therefore, we set X_m^4 to be the target parameter, the true batting average $\mu_m = \mu_m^4$.

3.5.2 Hyper-parameter Estimation

In order to implement the BMIE Thres, the hyper-parameters, η and τ, must be estimated. We follow the ML-II approach in Good [23]. This estimation procedure considers the expression of the marginal density of X_m^j, $\mathcal{N}(\eta, (\sigma_m^j)^2 + \tau^2)$, as a likelihood function of the hyper-parameters and seeks the values which maximize the likelihood. The resulting estimators can be obtained by solving the following two equations:

$$\hat{\eta} = \frac{\sum_{m=1}^{M_j} X_m^j / ({\sigma_m^j}^2 + \hat{\tau}^2)}{\sum_{m=1}^{M_j} 1/({\sigma_m^j}^2 + \hat{\tau}^2)} \quad \& \quad \sum_{m=1}^{M_j} \frac{(X_m^j - \hat{\eta})^2}{({\sigma_m^j}^2 + \hat{\tau}^2)^2} = \sum_{m=1}^{M_j} \frac{1}{{\sigma_m^j}^2 + \hat{\tau}^2}.$$

As the estimation is quite sensitive to the initial values of $\hat{\eta}$ and $\hat{\tau}$, we started from the following initial values:

$$\hat{\eta}_{\text{ini}} = \frac{1}{M_j} \sum_{m=1}^{M_j} X_m^j \quad \& \quad \sum_{m=1}^{M_j} \frac{(X_m^j - \hat{\eta}_{\text{ini}})^2}{({\sigma_m^j}^2 + \hat{\tau}_{\text{ini}}^2)^2} = \sum_{m=1}^{M_j} \frac{1}{{\sigma_m^j}^2 + \hat{\tau}_{\text{ini}}^2}.$$

3.5.3 Optimal C^* and α_m^*'s

The optimal C^* and α_m^*'s can be obtained through the optimization procedure in the previous section. Note that the procedure is performed based on the σ_m's and the estimated hyper-parameters $\hat{\eta}$ and $\hat{\tau}$. In addition, the tuning parameter, β, in the optimization is fixed to be 1000 as in chapter 2. With the optimized values, the form of the BIE Thres for the mth player with the jth prediction period becomes as follows:

$$\Gamma_m(X_m^j; \alpha_m^*) = \left(X_m^j - z_{\alpha_m^*/2}\sigma_m^j I\left\{X_m^j > \hat{\eta}^j - C^*\hat{\tau}^j\right\}, X_m^j + z_{\alpha_m^*/2}\sigma_m^j I\left\{X_m^j < \hat{\eta}^j + C^*\hat{\tau}^j\right\}\right).$$

3.5.4 Performance of BMIE Thres on Batting Average Prediction

We implement the procedure for different prediction periods: April, April-May, and April-June. In each case, the target parameter μ_m's are obtained from the rest of the season, respectively: May-October, June-October, and July-October. The global

level is set to be $1-q=0.9$. The application results are summarized in <Table 3.1>. First, the BFWCRs are consistently higher than the global level $1-q$ in all the pre-

Table 3.1: Result for Batting Averages Prediction

Prediction Period	April (j=1)	April~May (j=2)	April~June (j=3)
$\hat{\eta}$	0.5425	0.5438	0.5468
$\hat{\tau}$	0.008	0.0123	0.0151
C^*	2.86	3.04	3.195
BFWCR	92.24% (100%)	95.37% (98.78%)	93.79% (97.70%)
BREL	62.58%	76.46%	82.95%
BTR	73.39%	42.68%	28.05%
M_j	387	410	435

diction periods. Note that the corresponding BFWCRs of the classical z-based MIEs are in the parentheses. The values are quite larger than the global level, implying the z-based MIEs are unnecessarily conservative. Now, note that the estimated τ increases as the prediction periods become wider. This prior information determines the optimal value C^*. As a result, we have the smallest C^* when $j=1$, so the BTR becomes high and the corresponding BREL, 62.58%, shows a considerable reduction. When $j=2$, we have a larger C^*, and this results in a higher BREL of 76.45%. Lastly, we have the highest C^* when $j=3$, so that the highest BREL, 82.95%, is obtained. In conclusion, the BMIE Thres consistently shows reasonably higher-than-nominal-level BFWCRs; at the same time, we can achieve the meaningful reductions on the BRELs. However, the amount of reduction depends on the estimated value of τ which governs the prior information of the target parameter.

3.6 Data Application: Unknown σ_m's

The leukemia data appeared in Efron and Hastie [14] as a type of gene expression data from high-density oligonucleotide microarrays. It consists of $n=72$ patients with $n^1=45$ of ALL (acute lymphoblastic leukemia) group and $n^2=27$ of AML (acute myeloid leukemia) group, which has a worse prognosis. Efron provides small

and large data sets which contain $M_1 = 3571$ and $M_2 = 7128$ genes, respectively. To eliminate response disparities among the M microarrays as well as some outliers, the raw expression levels for the mth gene on the kth patient, X_{mk}, were transformed to a normal score value, $x_{mk} = \Phi^{-1}\left(\frac{rank(X_{mk})-0.5}{M}\right)$. Efron's investigation of the data was about a multiple testing procedure based on the local-FDR in Efron [13]. However, our goal here is to construct the BMIE Thres for the mean difference between the ALL and AML groups and to compare this result with the classical t-based MIE. In general, interval estimation becomes effective when we equally compare two group means because *zero* can be utilized as a criterion to evaluate the coverage of the interval estimation. That is, zero works as if it is the true parameter of the interval estimation in evaluating the empirical coverage probability.

3.6.1 Problem Setup

The data has the form of a $M \times n$ matrix where M is either 3571 or 7128 and $n = n^1 + n^2$ is $72 = 45 + 27$. Next, x_{mk} is the expression level for the mth gene of the kth patient, where $m = 1, 2, \ldots, M$, $k = 1, 2, \ldots, n^1$, for the AML group and $k = 46, 47, \ldots, n$, for the ALL group. This is the case of unknown standard deviations with one of the sample sizes less than 30, so that we could apply a plug-in procedure with the corresponding sample standard deviations. Thus, the basic individual procedure follows the two-sample t-based interval estimation under the equal variances assumption. (Lehmann and Romano [37])

3.6.2 Hyper-Parameter Estimation and Optimal Threshold

For the hyper-parameter estimation, we follow the same procedures as in the case of known σ_m's in the previous section. However, the optimization becomes difficult because the large M causes a problem in the Newton-Raphson method. Therefore, we instead pursue a graphical search for the optimal C^* as in <Figure 3.5>. Note that

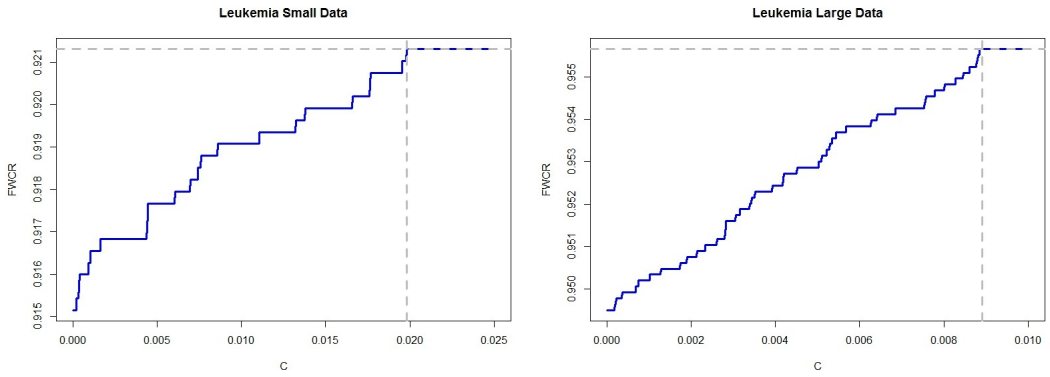


Figure 3.5: Graphical Search for C^* based on BFWCRs

given a global level, $1-q=0.9$, the BFWCRs of the classical t-based MIE for small and large leukemia data sets are 0.921 and 0.956, respectively. Since the BFWCRs should converge to the above values as C increases, we choose the optimal C^* at the joining points: $C^*=0.0198$ for the small data and $C^*=0.0098$ for the large data. These values are the smallest values which have the same BFWCRs as those of the classical MIEs. Then the form of the mth BIE Thres is as follows:

$$\Gamma_m(\boldsymbol{X}_m;\alpha_S)=\Big(\bar{X}_m^1-\bar{X}_m^2-t_{\alpha_S/2,n_m^1-n_m^2-2}S_m^p\sqrt{1/n_m^1+1/n_m^2}\,I\left\{\bar{X}_m>\hat{\eta}-C^*\hat{\tau}\right\},$$
$$\bar{X}_m^1-\bar{X}_m^2+t_{\alpha_S/2,n_m^1-n_m^2-2}S_m^p\sqrt{1/n_m^1+1/n_m^2}\,I\left\{\bar{X}_m<\hat{\eta}+C^*\hat{\tau}\right\}\Big)$$

where
$$\begin{cases} t_{\alpha_S/2,n_m^1+n_m^2-1}=F^{-1}_{T_{n_m^1+n_m^2-2}}(1-\alpha_S/2)\,; \\ S_m^p=\sqrt{\frac{\sum_{k=1}^{n_m^1}(x_{mk}-\bar{x}_m^1)^2+\sum_{k=n_m^1+1}^{n_m}(x_{mk}-\bar{x}_m^2)^2}{n_m^1+n_m^2-2}}; \\ \alpha_S=1-(1-q)^{1/M}, \text{ the Sidak adjustment}; \\ F_T \text{ is the distribution function of } t(n_m^1+n_m^2-2). \end{cases}$$

As the formal optimization procedure is not adopted here, we lose the opportunity to assign the optimal levels into the individual IEs; instead, the constant Sidak adjusted level was plugged in. However, this will not cause an issue in this particular illustration, as we will see in the next subsection.

3.6.3 Performance of BMIE Thres on Leukemia Data

In this subsection, we construct the BMIE Thres under the global level $1-q = 0.9$. The result is summrized in <Table 3.2>. By design, the BFWCRs of the BMIE Thres are

Table 3.2: Result for Leukemia Data

Leukemia	Small Data	Large Data
$\hat{\eta}$	0.0108	0.0014
$\hat{\tau}$	0.5336	0.1598
C^*	0.0198	0.0098
BFWCR	92.13%	95.57%
BREL	50.65%	50.46%
BTR	98.74%	98.81%
M	3571	7128

the same as those of the classical t-based MIEs. However, the BRELs of the BMIE Thres are less than 51% for both cases, implying the strong performance to reduce the global expected content. These seemingly *too good* results can be justified by the motivational sketch of the thresholding approach. As mentioned, the thresholding

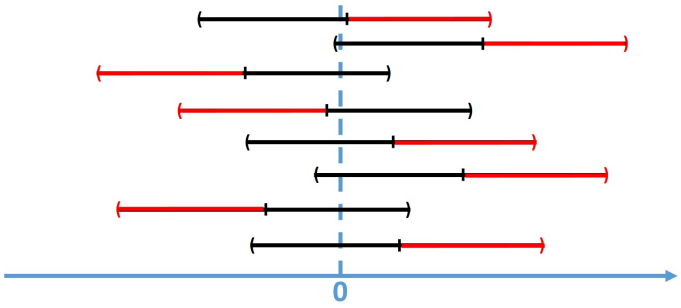


Figure 3.6: Comparing Two Population Means

procedure removes the outer tails of BIEs. However, with regard to comparing two group means, the inner tails cover zeros in most cases as in <Figure 3.6>. In some sense, the thresholding scheme of the BMIE Thres is designed to operate well in this particular application of comparing two population means by discarding a number of redundant outer tails. Because of the significant reductions already achieved, we do not have to exploit the additional reduction from the formal optimization procedure.

This can also be an advantage when we consider the potential applicability of the BMIE Thres to high-throughput data set.

3.7 Simulation

We ascertained the performance of the BMIE Thres by investigating its theoretical global quantities in section 3.4. However, the illustration was based on an ideal situation in which the suggested model perfectly represents the true underlying structure as follows:

$$\bar{X}_m|\mu_m \sim \mathcal{N}\left(\mu_m, \sigma_m^2\right) \;\; \& \;\; \mu_m \sim \mathcal{F}^*(\eta^*, \tau^*) = \mathcal{N}^{\Pi}\left(\eta, \tau^2\right) \text{ for } m = 1, 2, \ldots, M$$

where $\mathcal{F}^*(\eta^*, \tau^*)$ is the true underlying distribution which generates the target parameter μ_m's. Although this setting is effective to show the general behavior of the BMIE Thres, it does not reflect the potential model misspecifications which can frequently occur when a model relies on prior information. This is because no one can guarantee the prior distribution is identical to the true distribution, i.e., it is possible $\mathcal{N}^{\Pi} \neq \mathcal{F}^*$. Moreover, even though the distributions are identical, the parameters can still be distinct from one another, i.e., it is also possible $(\eta^*, \tau^*) \neq (\eta, \tau)$. Therefore, a simulation study is designed in this section to emulate the situations of prior misspecification. That is, we assume the true distribution, $\mathcal{F}^*$, to be normal, uniform, logistic, or exponential, with a fixed true mean, η^*, and standard deviation, τ^*, in order to generate μ_m's. However, the prior distribution is always the normal distribution with the hyper-parameters, η and τ, and we set these to be different from the true parameters.

In addition, although the BMIE Thres is considered to be a frequentist procedure, it is true that the procedure depends on the prior distribution to set up the thresholds. Therefore, it would be reasonable to ascertain the performance in relation to the Bayesian credible MIE as well as the classical z-based MIE. By involving

these comparisons in the simulation, we will consider a total of five MIEs: z-based classical MIE, $\boldsymbol{\Gamma_0(X)}$, BMIE Thres with assumed and estimated hyper-parameters, $\boldsymbol{\Gamma_1(X)}$ and $\boldsymbol{\Gamma_2(X)}$, and Bayesian credible MIE with assumed and estimated hyper-parameters, $\boldsymbol{\Gamma_3(X)}$ and $\boldsymbol{\Gamma_4(X)}$. It is important to note the performances of MIEs, the global coverage probability and global expected content, are measured based on the true target parameters, μ_m's. Therefore, we do not create any bias between the frequentist and Bayesian procedures. The mth individual IEs of the five MIEs are as follows:

$$
\begin{aligned}
\Gamma_0^m(\boldsymbol{X}) &= \left(\bar{X}_m - z_{\alpha_s/2}\sigma_m, \bar{X}_m - z_{\alpha_s/2}\sigma_m\right);\\
\Gamma_1^m(\boldsymbol{X}) &= \left(\bar{X}_m - z_{\alpha_m/2}\sigma_m I\left\{\bar{X}_m > \eta - C^*\tau\right\}, \bar{X}_m + z_{\alpha_m/2}\sigma_m I\left\{\bar{X}_m < \eta + C^*\tau\right\}\right);\\
\Gamma_2^m(\boldsymbol{X}) &= \left(\bar{X}_m - z_{\alpha_m/2}\sigma_m I\left\{\bar{X}_m > \hat{\eta} - C^*\hat{\tau}\right\}, \bar{X}_m + z_{\alpha_m/2}\sigma_m I\left\{\bar{X}_m < \hat{\eta} + C^*\hat{\tau}\right\}\right);\\
\Gamma_3^m(\boldsymbol{X}) &= \left(\beta\bar{X}_m + (1-\beta)\eta - z_{\alpha_s/2}\sigma_m\sqrt{\beta}, \beta\bar{X}_m + (1-\beta)\eta + z_{\alpha_s/2}\sigma_m\sqrt{\beta}\right);\\
\Gamma_4^m(\boldsymbol{X}) &= \left(\hat{\beta}\bar{X}_m + (1-\hat{\beta})\hat{\eta} - z_{\alpha_s/2}\sigma_m\sqrt{\hat{\beta}}, \hat{\beta}\bar{X}_m + (1-\hat{\beta})\hat{\eta} + z_{\alpha_s/2}\sigma_m\sqrt{\hat{\beta}}\right)
\end{aligned}
$$

where $\beta = \frac{\tau^2}{\sigma^2+\tau^2}$, $\hat{\beta} = \frac{\hat{\tau}^2}{\sigma^2+\hat{\tau}^2}$, and $\alpha_s = 1 - (1-q)^{1/M}$, the Sidak adjustment. Note that the last MIE is also called the empirical Bayes MIE under the normal-normal model.

The goal of the simulation is to perform simultaneous interval estimation on $M = 1000$ normal location parameters, μ_m's, given the global level $1 - q = 0.9$. These μ_m's are generated from different true distributions – normal, uniform, logistic, and exponential – with the fixed true mean (η^*) and standard deviation (τ^*). To construct Γ_1 and Γ_3, we directly use the prior mean (η) and standard deviation (τ), and these values will deviate from the true mean and standard deviation. To construct Γ_2 and Γ_4, we use the estimated prior mean ($\hat{\eta}$) and standard deviation ($\hat{\tau}$) from the data to plug these into the procedure. The estimated prior standard deviation is also used to find out the optimal threshold C^* and α_m^*'s. The simulation scheme is summarized in (Algorithm 1).

We first consider the case in which the true generating distribution is normal, $\mathcal{N}^{\Pi} =$

Algorithm 1 Simulation Scheme

```
## M = 1000 and σ_m's are from unif(0.01, 10)
for i = 1 in 1:4 do
  for j = 1 in 1:3 do
    for k = 1 to Nrep = 1000 do
      Data← Generator(η*[i],τ*[j])
      C*←Optimizer(Data,τ̂[j])
      MIE←Constructor(Data,C*,η[i],τ[j],η̂[i],τ̂[j])
      Out[k]←Evaluator(MIE,η*[i],τ*[j])
    end for
    Result[[i]][[j]] ←Summarizer(Out)
  end for
end for
Tabulator(Result)
```

$\mathcal{F}^*$, but the hyper-parameters deviate from the true mean and standard deviation, $(\eta^*, \tau^*) \neq (\eta, \tau)$. That is, the true mean and standard deviation are fixed to be 0 and 2, but η and τ take the values (0, 2, 4, 6) and (1, 2, 3), respectively. When we compare the performances of five MIEs, the global coverage probability is obtained by the ratio of the IEs which cover the true target parameters in MIEs, and the global expected content is measure by the relative average expected length of MIEs, i.e., the average expected lengths of the MIEs in relation to the average expected length of the corresponding z-based MIE.

<Table 3.3> shows the simulation result under the well specified prior up to its distributional level. When the prior mean (η) and prior standard deviation (τ) are equal to the true values, $\eta^* = 0$ and $\tau^* = 2$, all the MIEs perform well, showing the satisfactory global coverage probability and the reductions in the global expected contents. In particular, the global expected contents of the Bayes credible MIEs, Γ_3 and Γ_4, are significantly smaller than those of the z-based MIE and the BMIE Thres. However, as η deviates from the true mean, the coverage of Γ_3 rapidly decreases; still, Γ_4 performs well with the estimated hyper-parameters except for a slight degradation in the global coverage rate. When the prior standard deviation is specified to be

Table 3.3: Prior Misspecification when the True Distribution is Normal

Normal		Global Coverage Probability				Global Expected Content			
$1-q=0.9$	MIEs	$\eta^*=0$	$\eta=2$	$\eta=4$	$\eta=6$	$\eta^*=0$	$\eta=2$	$\eta=4$	$\eta=6$
$\tau=1$	Γ_0	0.896	0.909	0.911	0.873	1.000	1.000	1.000	1.000
	Γ_1	0.904	0.902	0.810	0.082	0.898	0.910	0.895	0.861
	Γ_2	0.904	0.903	0.912	0.882	0.898	0.893	0.902	0.918
	Γ_3	0.000	0.000	0.000	0.000	0.181	0.181	0.181	0.181
	Γ_4	0.878	0.879	0.870	0.878	0.328	0.327	0.327	0.327
$\tau^*=2$	Γ_0	0.909	0.903	0.922	0.883	1.000	1.000	1.000	1.000
	Γ_1	0.909	0.905	0.922	0.888	0.915	0.927	0.951	0.946
	Γ_2	0.909	0.905	0.922	0.888	0.914	0.902	0.917	0.906
	Γ_3	0.894	0.271	0.000	0.000	0.328	0.328	0.328	0.328
	Γ_4	0.861	0.860	0.867	0.860	0.326	0.326	0.326	0.327
$\tau=3$	Γ_0	0.887	0.897	0.907	0.885	1.000	1.000	1.000	1.000
	Γ_1	0.887	0.895	0.907	0.885	0.909	0.925	0.941	0.940
	Γ_2	0.887	0.895	0.907	0.885	0.914	0.879	0.910	0.902
	Γ_3	0.981	0.983	0.832	0.160	0.446	0.446	0.446	0.446
	Γ_4	0.866	0.883	0.888	0.865	0.326	0.327	0.327	0.327

less ($\tau = 1$) than the true standard deviation, the global coverage probabilities are affected by the concentrated prior distribution, resulting in very large degradation. In this situation, Γ_1 is also affected by the concentrated prior, showing a gradual degradation in the global coverage probability as η deviates from the true mean. The global expected contents of MIEs are generally narrower than in the previous case, reflecting the concentrated prior information. Lastly, when the prior standard deviation is specified to be larger ($\tau = 3$) than the true standard deviation, the global level requirements are generally well satisfied due to the effect of the diffused prior, except for Γ_3 which shows rapid decrease as η deviates. To compensate, the global expected contents become larger in general. In terms of the global coverage probability, the MIEs with the estimated parameters, Γ_2 and Γ_4, show the robust result as expected. Among those, Γ_4 performs very well compared to Γ_2 in terms of the global expected content. As a result, when the prior distribution well reflects the true underlying structure up to its distributional level, Γ_4 would be the best choice if one can endure a slight degradation of the global coverage probability; however, if

the global level requirement needs to be strictly satisfied, then Γ_2 would be the choice as it shows a very robust result for any hyper-parameter combinations.

From now on, the simulations reflect the case the prior distribution itself deviates from the true distribution: $\mathcal{N}^{\Pi} \neq \mathcal{F}^*$. First, <Table 3.4> presents the simulation result when the true underlying distribution is a uniform distribution.

Table 3.4: Prior Misspecification when the True Distribution is Uniform

Uniform		Global Coverage Rate				Global Content			
$1-q=0.9$	MIEs	$\eta^*=0$	$\eta=2$	$\eta=4$	$\eta=6$	$\eta^*=0$	$\eta=2$	$\eta=4$	$\eta=6$
$\tau=1$	Γ_0	0.902	0.893	0.884	0.889	1.000	1.000	1.000	1.000
	Γ_1	0.900	0.892	0.884	0.889	0.712	0.789	0.829	0.847
	Γ_2	0.900	0.892	0.884	0.889	0.736	0.746	0.746	0.746
	Γ_3	0.522	0.000	0.000	0.000	0.181	0.181	0.181	0.181
	Γ_4	0.985	0.980	0.979	0.981	0.327	0.327	0.327	0.327
$\tau^*=2$	Γ_1	0.928	0.895	0.902	0.911	1.000	1.000	1.000	1.000
	Γ_1	0.920	0.895	0.902	0.910	0.718	0.793	0.832	0.850
	Γ_2	0.920	0.895	0.902	0.910	0.747	0.746	0.746	0.737
	Γ_3	0.986	0.929	0.001	0.000	0.328	0.328	0.328	0.328
	Γ_4	0.984	0.987	0.981	0.985	0.327	0.327	0.327	0.327
$\tau=3$	Γ_1	0.889	0.890	0.909	0.904	1.000	1.000	1.000	1.000
	Γ_1	0.889	0.888	0.905	0.896	0.717	0.797	0.829	0.850
	Γ_2	0.889	0.888	0.905	0.896	0.746	0.755	0.747	0.746
	Γ_3	0.987	0.981	0.931	0.522	0.446	0.446	0.446	0.446
	Γ_4	0.985	0.987	0.992	0.986	0.327	0.327	0.327	0.327

While the classical z-based MIE always shows a consistent result as it has nothing do to with the prior distribution, the other four MIEs show better performances than the normal prior case. In particular, Γ_1 and Γ_2 provide greater reductions on the global contents compared to the previous normal true distribution case; still, the global level requirements are well satisfied, showing the global coverage probabilities are greater than $1-q=0.9$. Compare to this, Γ_3 and Γ_4 achieve the larger reductions on the global expected content with the higher global coverage probabilities, implying the MIEs can reach the same reductions with less efforts. Still, Γ_3 are very sensitive to the location- or scale-wise deviations of the hyper-parameters.

The shapes of logistic and exponential distributions are quite far from the normal

prior distribution. Thus, we would expect worse performances of the MIEs compared to the previous cases. For this reason, we reset the true standard deviation of the generating distribution to be $\tau = 1$ to have better comparisons among the MIEs.

Table 3.5: Prior Misspecification when the True Distribution is Logistic

logistic		Global Coverage Rate				Global Content			
$1-q=0.9$	MIEs	$\eta^*=0$	$\eta=2$	$\eta=4$	$\eta=6$	$\eta^*=0$	$\eta=2$	$\eta=4$	$\eta=6$
$\tau=0.5$	Γ_0	0.897	0.920	0.895	0.894	1.000	1.000	1.000	1.000
	Γ_1	0.877	0.637	0.000	0.000	0.792	0.779	0.505	0.504
	Γ_2	0.896	0.917	0.895	0.902	0.863	0.840	0.905	0.851
	Γ_3	0.000	0.000	0.000	0.000	0.095	0.095	0.095	0.095
	Γ_4	0.267	0.244	0.246	0.231	0.179	0.179	0.179	0.179
$\tau^*=1$	Γ_1	0.892	0.904	0.905	0.910	1.000	1.000	1.000	1.000
	Γ_1	0.887	0.904	0.904	0.869	0.857	0.873	0.891	0.868
	Γ_2	0.887	0.904	0.905	0.907	0.846	0.845	0.842	0.872
	Γ_3	0.241	0.000	0.000	0.000	0.181	0.181	0.181	0.181
	Γ_4	0.236	0.246	0.255	0.271	0.179	0.179	0.179	0.179
$\tau=2$	Γ_1	0.909	0.896	0.897	0.896	1.000	1.000	1.000	1.000
	Γ_1	0.914	0.890	0.897	0.901	0.848	0.873	0.902	0.912
	Γ_2	0.914	0.890	0.897	0.901	0.830	0.837	0.868	0.861
	Γ_3	0.990	0.961	0.401	0.000	0.328	0.328	0.328	0.328
	Γ_4	0.253	0.275	0.263	0.229	0.179	0.181	0.179	0.180

<Table 3.5> provides the simulation result when the true underlying distribution is a logistic distribution. Note that the global coverage probabilities of Bayes Credible MIEs, Γ_3 and Γ_4, are lower than the nominal global level in any hyper-parameter combinations except for the Γ_3 with the less deviated prior mean ($\eta = 0, 2$) and diffused prior standard deviation ($\tau = 2$). This is because the center value of the credible interval, the posterior mean, is off the target as it is derived based on the normal-normal model assumption. Note that the BMIE Thres with estimated hyper-parameters, Γ_2, still works well under this logistic true distribution. This implies that the hyper-parameter estimation procedure itself is still viable. Lastly, Γ_1 also shows quite reasonable result except for some cases with the concentrated prior. The results provide a example which re-enlighten the notion that the Bayesian approach can be totally off when prior distribution is misspecified.

Table 3.6: Prior Misspecification when the True Distribution is Exponential

exponential		Global Coverage Rate				Global Content			
$1-q=0.9$	MIEs	$\eta^*=0$	$\eta=2$	$\eta=4$	$\eta=6$	$\eta^*=0$	$\eta=2$	$\eta=4$	$\eta=6$
$\tau=0.5$	Γ_0	0.889	0.901	0.888	0.874	1.000	1.000	1.000	1.000
	Γ_1	0.675	0.877	0.883	0.000	0.792	0.779	0.741	0.504
	Γ_2	0.896	0.900	0.888	0.872	0.916	0.919	0.922	0.922
	Γ_3	0.000	0.000	0.000	0.000	0.095	0.095	0.095	0.095
	Γ_4	0.010	0.015	0.011	0.013	0.177	0.178	0.178	0.178
$\tau^*=1$	Γ_0	0.896	0.917	0.896	0.896	1.000	1.000	1.000	1.000
	Γ_1	0.896	0.917	0.896	0.896	0.928	0.900	0.795	0.769
	Γ_2	0.895	0.917	0.896	0.895	0.920	0.922	0.921	0.921
	Γ_3	0.003	0.341	0.000	0.000	0.181	0.181	0.181	0.181
	Γ_4	0.013	0.008	0.022	0.022	0.178	0.178	0.178	0.178
$\tau=2$	Γ_0	0.889	0.897	0.873	0.904	1.000	1.000	1.000	1.000
	Γ_1	0.889	0.892	0.873	0.904	0.921	0.850	0.820	0.770
	Γ_2	0.889	0.890	0.873	0.904	0.905	0.914	0.913	0.918
	Γ_3	0.916	0.978	0.864	0.001	0.328	0.328	0.328	0.328
	Γ_4	0.018	0.006	0.015	0.014	0.178	0.178	0.177	0.177

Lastly, <Table 3.6> shows the simulation result when the true underlying distribution is exponential. It shows quite unexpected behaviors; that is, the highest global coverage probabilities are achieved not with $\eta = 0$ but with $\eta = 2$, which is slightly deviated value from the truth. This would be due to the right-skewness of the true generating distribution. Still, the general behaviors are similar to those of logistic case. While Γ_3 and Γ_4 suffer from the off the target issue, Γ_1 and Γ_2 show a consistent performance except for the Γ_1 with concentrated prior distribution.

In conclusion, Γ_3 cannot be beaten when the prior is correctly specified in terms of the distribution as well as the corresponding hyper-parameters. It provides significant reductions on the global expected contents while maintaining the global coverage probability at least the global level. However, when the hyper-parameters deviate from the truth, it becomes hard to satisfy the global level requirement; still, Γ_4 is robust for these misspecified hyper-parameters. Now, when the prior distribution itself deviates from the true generating distribution, then the both Γ_3 and Γ_4 suffer from the low global coverage probability. When all these happen, Γ_1 and Γ_2 generally sat-

isfy the global level requirement except for some extreme cases, providing reasonable reductions on the global expected contents. Therefore, we can conclude that BMIE Thres provides satisfactory robustness against the prior misspecification.

3.8 Summary

When a confidence interval is introduced, one faces the temptation to remove one side of the interval to reach a shorter expected length which implies a better precision of the interval estimation. Our procedure was motivated by this intuitive idea and it is designed to realize the removal process. However, the procedure relies on the additional prior information. Because of this, the individual estimator is no longer within the class of classical confidence intervals but within the wider class, the class of Bayes interval estimators which can be defined and evaluated by the integrals of the performance quantities with respect to the prior distribution. We call our procedure a Bayes MIE Thres as the integration process resembles the derivation of the Bayes risk in a general statistical decision problem.

Under the independence assumption, the resulting BMIE Thres can be considered as a hyper-rectangular region estimator in the M dimensional parameter space. As we know, this would create a larger volume compared to an elliptical region estimator, even after we discard one side of some intervals. However, in many situations, researchers also want to inspect individual IEs which cannot be tracked from the elliptical region estimator. In this context, the indepdent setting for the BMIE Thres still has its own advantage despite its restriction. In fact, the real data applications provide meaningful performances compared to the traditional z or t-based MIE.

While we utilize prior information, it is important to realize that the prior distribution, which we assumed for the modeling, can be different from the true underlying distribution, which actually generates the target parameters. The simulation setting in the previous section takes this potential disagreement into consideration, generat-

ing the location parameters, μ_m's, from the distinct true distributions. Coupled with the comparison to the Bayes Credible MIEs, the simulation results flash a warning signal in the use of unjustified prior information, i.e., the issue of prior misspecification. Although a model works well in terms of the formula, it is the case under the correctly specified prior distribution and the reality may be quite different from this ideal situation. This reveals the common disadvantage of the general Bayesian procedures, implying we cannot simply enjoy the benefit of the prior information. Consequently, any Bayesian procedures should be used with discretion. One of the remedies would be the use of non-informative prior and it is closely related to the topic in next chapter.

Chapter 4

Equivariant MIE with Haar measures

4.1 Overview

Let us consider a simple decision theoretic structure. The first component is a parameter space, Θ, which consists of the true states controlled by Mother Nature. The second component is an action space, $\mathcal{A}$, which consists of the actions made by a statistician. The statistician's action, a, is then evaluated through a loss function:

$$L(\theta, a) \in \mathbb{R} \text{ for } \theta \in \Theta,\ a \in \mathcal{A}$$

Since the true state is unknown, the statistician designs an experiment to collect sample data, X. The third component consists of the possible data and is denoted by a sample (data) space, $\mathcal{X}$. Therefore, based on the observed sample data, the statistician's action is determined based on the observed sample data via a (non-randomized) decision, i.e., a function, $\delta(x) \in \mathcal{D}$, from the sample space to the action space. The decision is then evaluated by a risk function:

$$R(\theta, \delta) = E_\theta[L(\theta, \delta(X))] \in \mathbb{R} \text{ for } \theta \in \Theta,\ \delta \in \mathcal{D}$$

assuming the expectation exists. Now the goal of the statistician is to investigate a decision which minimizes the risk function, $R(\theta, \delta)$. As global approaches, we could use a minimax procedure – $\max_\theta R(\theta, \delta^M) = \min_\delta \max_\theta R(\theta, \delta)$ – to obtain the minimax decision (δ^M) or a Bayes procedure – $r(\delta^B; \Pi) = \min_\delta \int_\Theta R(\theta, \delta) d\Pi(\theta)$ – to obtain the Bayes decision (δ^B) under the presence of the prior distribution, Π, on Θ. On the other hand, we would restrict the class of decisions and attempt to reach

the best decision within the restricted class. One of the most widely used classes is the collection of unbiased decisions. Another useful class is the collection of invariant decisions, our main interest in this study.

The idea of the Invariance Principle starts from the groups of transformations. When a statistician is able to impose group structures on the three fundamental components – the parameter space, the action space, and the sample space – the dimension of the corresponding risk function for an equivariant decision can be substantially reduced through the symmetry of the group structures. Moreover, in the case of the transitive parameter space, i.e., when all the elements of the parameter space can be reached through the transformations in the group, we reach a constant risk function, so that the investigation of the optimal decision becomes much more accessible.

In addition, the Invariance Principle provides further useful results in combination with Haar measures. The Haar measures are the mesaures for group structures and invariant for the left- and right-hand side transformations. Under certain conditions, these Haar measures can be exploited as the non-informative prior distributions for the parameter space. The derived non-informative prior distributions are legitimate because it leads to a reconciliation between the frequentist and Bayesian approaches combined with equivariant decisions. Not to mention its theoretical importance, it is also a practical result since determining the best equivariant decision through the frequentist approach is not an easy task. By using the reconciliation, we can instead search for a generalized Bayes equivariant decision based on the non-informative prior distribution.

When the scope of the statistical decisions is reduced to IEs, the reconciliation becomes the equivalence between the frequentist coverage probability and the Bayesian coverage probability with respect to the posterior distribution derived from the right Haar measure. Furthermore, by using the left Haar measure, we can invariantly mea-

sure the expected length of the equivariant IE. These results will be used to establish a coherent framework to handle a set of heterogeneous parameters in multiple interval estimations. That is, we categorize the different parameters in terms of their group structures and reduce the dimension of the problem by using the equivariant IEs along with the left and right Haar measures. Combined with a optimization procedure, the resulting MIE is called an equivariant multiple interval estimator with Haar measures (EMIE Haar).

A brief review of related studies is provided in section 2. Next, the invariance structure and the corresponding left and right Haar measures are presented in section 3. Section 4 contains the main idea of equivariant multiple interval estimator with Haar measures. An optimization procedure modified for the EMIE Haar and a corresponding application is provided in section 5. A paricular application with two different group structures under the normal distribution is illustrated in section 6. Lastly, a summary is presented in section 7.

4.2 Related Studies

The idea of invariance in statistical inference has existed for almost one hundred years. For example, Pitman [44] derived general forms of the best invariant estimator under location and scale transformations. In the frequentist perspective, the Invariance Principle, like the Sufficiency Principle, allows us to achieve dimension reductions in statistical problems. However, since the idea is very general, i.e., essentially related to the use of symmetry structure, research has been developed from many other perspectives. For example, Fraser [21] showed the relation between the invariance and Fisher's fiducial inference (Fisher [18]), and the idea was further developed in his structural inference (Fraser [20]). In addition, with regard to the Bayesian approach, the invariance was studied further in relation to non-informative priors. (Robert [46]) The choice of prior is the issue in Bayesian statistics even in non-informative

cases. Given a group structure, the right Haar measure can be utilized as a non-informative prior which provides a compatible structure. Furthermore, the invariance is considered to be a very useful tool not just in the field of statistics, but also in fields such as quantum mechanics (Helland [25]).

The invariance structure is integrally related to the multivariate statistical setting because it allows us to reduce the dimension of problems as in Eaton [12] and Giri [22]. In particular, Eaton further studied the invariance procedure in-depth in Eaton [11]. Wijsman [55] introduced the cross-section of orbits to derive the distribution of maximal invariant statistics. He also further extended his research on invariance problems in Wijsman [56]. Hooper applied his advisor Wisjman's cross-section idea to confidence sets in Hooper [31] and Hooper [30]. Zidek also worked on invariance problems with his advisees as in Zidek [57] and Brewster, Zidek, et al. [5]. Helland [26] summarized quite well the relation between confidence sets and credible sets under transitive assumptions.

Research has also been conducted for non-transitive cases. One of the most famous results is the Hunt-Stein theorem, which states that the minimax rule within the class of invariant rules is overall minimax. It was first introduced in Stein's unpublished paper, but the content can be found in Lehmann and Romano [37]. The result was extended to general cases in Kiefer et al. [35] and in Kudo [36].

The general idea of the Invariance Principle appears in many classical textbooks as one of the data reduction approaches along with the Sufficiency Principle, e.g., Lehmann and Romano [37] or Ferguson [17]. Its Bayesian interpretation is also summarized in some major textbooks, especially in relation to the non-informative prior distribution, e.g., Berger [3] or Schervish [49]. As mentioned, the compatible non-informative prior distribution turns out to be from the Haar measure. However, the presentation remains a one dimensional case with limited group structures mainly based on the location group, the scale group, and the location-scale group.

4.3 Invariance Structure with Haar Measure

In this section, we introduce a general invariance structure and corresponding left and right Haar measures.

4.3.1 Invariance Group Structure

Recall the three fundamental components with corresponding sigma algebras: the sample space, $(\mathcal{X}, \mathcal{F})$, the parameter space, $(\Theta, \mathcal{T})$, and the action space, $(\mathcal{A}, \mathscr{A})$. Then define $\mathcal{G}$ to be a group of measurable transformations on $\mathcal{X}$ which satisfies the following axioms with a binary (composition) operation, $\cdot$:

i) $g' \cdot g'' \in \mathcal{G}$ for any $g' \in \mathcal{G}$ and $g'' \in \mathcal{G}$

ii) $g \cdot (g' \cdot g'') = (g \cdot g') \cdot g''$ for any g, g', and $g'' \in \mathcal{G}$

iii) There exists $e \in \mathcal{G}$ such that $e \cdot g = g \cdot e = g$ for every $g \in \mathcal{G}$

iv) For every $g \in \mathcal{G}$, there exists $g^{-1} \in \mathcal{G}$ such that $g^{-1} \cdot g = g \cdot g^{-1} = e$

Now we say a family of probability distributions, $\mathcal{F} = \{P_\theta : \theta \in \Theta\}$, is invariant under $\mathcal{G}$ if for any $g \in \mathcal{G}$ and $\theta \in \Theta$, there exists a unique $\theta' \in \Theta$ such that

$$P_\theta[X \in A] = P_{\theta'}[g(X) \in A]$$

for every $A \in \mathcal{F}$. We define $\bar{g}(\theta) = \theta'$ in relation to $g \in \mathcal{G}$. Then $\bar{\mathcal{G}} = \{\bar{g} : g \in \mathcal{G}\}$ forms a group of transformations on Θ. Note that the above equality is equivalent to $P_\theta[g(X) \in A] = P_{\bar{g}(\theta)}[X \in A]$, meaning the probability resulting from the transformation of x via g is equivalent to the probability resulting from the transformation of θ via $\bar{g}$. Throughout the study, we consider parametric settings with continuous distributions. In addition, we say a loss function $L(\theta, a)$ is invariant under $\mathcal{G}$ if for every $g \in \mathcal{G}$ and $a \in \mathscr{A}$ there exists a unique $a' \in \mathscr{A}$ such that

$$L(\theta, a) = L(\bar{g}(\theta), a')$$

for all $\theta \in \Theta$. We also define $\tilde{g}(a) = a'$ in relation to $g \in \mathcal{G}$. Then $\tilde{\mathcal{G}} = \{\tilde{g} : g \in \mathcal{G}\}$ forms a group of transformations on $\mathscr{A}$. The above equality means the loss remains the same if we take the transformations of $\bar{g}(\theta)$ and $\tilde{g}(a)$ concurrently.

Now suppose $\mathcal{D}$ is the class of non-randomized decisions. Then we say a decision function, $\delta \in \mathcal{D}$, is equivariant if the following equality holds true for every $x \in \mathcal{X}$ and $g \in \mathcal{G}$:

$$\delta(g(x)) = \tilde{g}(\delta(x)).$$

This equality implies that the pre-transformed decision function, $\delta(g(x))$, is the same as the post-transformed decision, $\tilde{g}(\delta(x))$. Given the settings, we introduce an important result in the following proposition:

Proposition 4.1. *Suppose $\delta \in \mathcal{D}$ is an equivariant decision. Then the following equality holds true for every $\theta \in \Theta$ and $g \in \mathcal{G}$:*

$$R(\theta, \delta) = R(\bar{g}(\theta), \delta).$$

In general, if θ and θ' are connected to each other via a transformation in $\bar{\mathcal{G}}$, e.g., there exists $\bar{g}$ such that $\theta' = \bar{g}(\theta)$, we say θ and θ' are in an equivalent relation. This equivalent relation partitions the parameter space, Θ, into equivalent classes or orbits. Therefore, the above proposition implies the important Invariance Principle that an equivariant decision function yields a constant risk on each orbit. In addition, if $\bar{\mathcal{G}}$ generates only one orbit on Θ, it is called transitive. As a result, the risk becomes a constant for an equivariant decision in the transitive case.

4.3.2 Right and Left Invariant Haar Measures

In addition to the classical invariance structure in the previous section, we add one more layer, a prior distribution. In particular, the prior distribution should be compatible with the corresponding invariance structure. To achieve this, we first introduce the concept of invariant Haar measures and link it to the prior distribution. Note

that further in-depth material regarding the Haar measures can be found in Nachbin [40]. We first define a left and right transformations.

Definition 4.2. Given $\bar{g}_0 \in \bar{\mathcal{G}}$, we define the right and left transformations as

$$\bar{g} \to \bar{g}\bar{g}_0 \;\; \& \;\; \bar{g} \to \bar{g}_0\bar{g}.$$

Similarly, we define the classes of the right and left transformations as

$$A\bar{g}_0 = \{\bar{g}\bar{g}_0 : \bar{g} \in A\} \;\; \& \;\; \bar{g}_0 A = \{\bar{g}_0\bar{g} : \bar{g} \in A\}.$$

Now the right and left Haar measures are defined in the following ways:

Definition 4.3. Let $\bar{\mathcal{G}}$ be a group with the σ-field of subsets, $\bar{\Lambda}$. Suppose $A \in \bar{\Lambda}$, $\bar{g}A \in \bar{\Lambda}$ for $\bar{g} \in \bar{\mathcal{G}}$, and $A^{-1} = \{\bar{g}^{-1} : \bar{g} \in A\} \in \bar{\Lambda}$. The measures H^R and H^L on $\bar{\Lambda}$ are called right and left Haar measures (RHM & LHM) if

$$H^R(A\bar{g}) = H^R(A) \;\; \& \;\; H^L(\bar{g}A) = H^L(A)$$

for every $\bar{g} \in \bar{\mathcal{G}}$ and every $A \in \bar{\Lambda}$. In addition, the right and left Haar densities on $\bar{\Lambda}$, $h^R(\bar{g})$ and $h^L(\bar{g})$, are defined to be the Radon-Nikodym derivatives of RHM and LHM with respect to a proper measure, so that

$$\int_{A\bar{g}_0} h^R(y)dy = \int_A h^R(y)dy \;\; \& \;\; \int_{\bar{g}_0 A} h^L(y)dy = \int_A h^L(y)dy$$

Intuitively speaking, the measures remain the same with respect to the right and left transformations, respectively.

Property 4.4. Some useful results for the Haar densities are

i) $h^R(\bar{g})$ and $h^L(\bar{g})$ exist and are unique up to a multiplicative constant

ii) If $\bar{\mathcal{G}}$ is symmetric (Abelian), then $h^R(\bar{g})$ and $h^L(\bar{g})$ are equivalent.

The first property is intuitive from the definition of the Haar measures, and the formal proof is provided in Weil [52] and Cartan [7]. The second property is evident from the realization that the right and left transformations are equivalent for the symmetric group.

For actual derivations, we use change-of-variable techniques with respect to the group of transformation. Consider $x = y\bar{g}_0^{-1}$ on the left-hand side of the equality in the definition of the right Haar density. Then,

$$\int_{A\bar{g}_0} h^R(y)dy = \int_A h^R(x\bar{g}_0)J^R_{\bar{g}_0}(x)dx$$

where $J^R_{\bar{g}_0}(x)$ is the Jacobian of the right transformation: $x \to x\bar{g}_0$. Note that the transformation can be multi-dimensional, so that the Jacobian becomes a square matrix with the corresponding dimension. Because the integrand should be equivalent to the right-hand side of the equality in the definition for any $A \in \bar{\Lambda}$, $\bar{g}$ and x in $\bar{\mathcal{G}}$, we have the result $h^R(x\bar{g})J^R_{\bar{g}}(x) = h^R(x)$. Replacing x by the identity element, $\bar{e}$, of $\bar{\mathcal{G}}$, we have the following expression:

$$h^R(\bar{g}) = \frac{h^R(\bar{e})}{J^R_{\bar{g}}(\bar{e})} := \frac{1}{J^R_{\bar{g}}(\bar{e})}$$

The second equality for the definition is due to the property (i) because $h^R(\bar{e})$ is a constant. Likewise, we can derive the form of the left invariant Haar density as follows:

$$h^L(\bar{g}) = \frac{h^L(\bar{e})}{J^L_{\bar{g}}(\bar{e})} := \frac{1}{J^L_{\bar{g}}(\bar{e})}$$

Now examples follow for location, scale, and location-scale groups.

Example 4.5. (Location Group) The first example concerns a location group: $\bar{\mathcal{G}} = \{g_a : g_a(x) = x + a, a \in \mathbb{R}\} = \{(a) : a \in \mathbb{R}\}$. Note that the group is Abelian. So

$$\bar{g}\bar{g}_0 = a + a_0 = a_0 + a = \bar{g}_0\bar{g}$$

In this case, the Jacobian is just 1. Now the Haar densities are identical: $h^L_{\bar{g}_0}(\bar{g}) = h^R_{\bar{g}_0}(\bar{g}) = 1$.

Example 4.6. (Scale Group) The second example concerns a scale group: $\bar{\mathcal{G}} = \{g_b : g_b(x) = bx, b > 0\} = \{(b) : b > 0\}$. This group is also Abelian. So

$$\bar{g}\bar{g}_0 = bb_0 = b_0 b = \bar{g}_0\bar{g}$$

The Jacobian becomes b_0. Then the Haar densities are identical: $h^L_{\bar{g}_0}(\bar{g}) = h^R_{\bar{g}_0}(\bar{g}) = \frac{1}{b_0}$.

Example 4.7. (Location-Scale Group) The last example concerns a location-scale group: $\bar{\mathcal{G}} = \{g_{(a,b)} : g_{(a,b)}(x, y) = (a + bx, by), (a, b) \in \mathbb{R} \times \mathbb{R}^+\} = \{(a, b) : (a, b) \in \mathbb{R} \times \mathbb{R}^+\}$. For any given $\bar{g}_0 = (a_0, b_0) \in \bar{\mathcal{G}}$, the right and left transformations are as follows:

$$\bar{g} = (a, b) \to \bar{g}\bar{g}_0 = (a, b)(a_0, b_0) = (a + ba_0, bb_0)$$

$$\bar{g} = (a, b) \to \bar{g}_0\bar{g} = (a_0, b_0)(a, b) = (a_0 + ba, bb_0)$$

Based on these transformations, we can derive the right and left differentials and corresponding Jacobians as follow:

$$D^R_{\bar{g}_0}(\bar{g}) = \begin{bmatrix} 1 & a_0 \\ 0 & b_0 \end{bmatrix};\ J^R_{\bar{g}_0}(\bar{g}) = b_0 \ \ \& \ \ D^L_{\bar{g}_0}(\bar{g}) = \begin{bmatrix} b_0 & 0 \\ 0 & b_0 \end{bmatrix};\ J^L_{\bar{g}_0}(\bar{g}) = b_0^2$$

As a result, the right and left invariant Haar densities are $h^R_{\bar{g}_0}(\bar{g}) = \frac{1}{b_0}$ and $h^L_{\bar{g}_0}(\bar{g}) = \frac{1}{b_0^2}$, respectively.

4.3.3 Relation to Prior Distribution

We can relate the Haar densities to prior densities via the isomorphism between $\bar{\mathcal{G}}$ and Θ. This means the two spaces are essentially the same in the sense that there exists a bijection between two spaces, preserving the composition structure. In this study, we add further assumptions as in <Table 4.1>. The curly equalities represent the isomorphisms between the spaces. Note that the usual relation between $\mathcal{G}$ and $\bar{\mathcal{G}}$ is the homomorphism, i.e., the injection which preserves the composition structure. This is reasonable because the dimension of the sample space is higher than that of the parameter space in general. However, we impose somewhat stronger assumptions

Table 4.1: Assumptions for Invariance Structure

$$\begin{array}{ccccc} (\mathcal{X},\mathcal{F}) & \approx & (\Theta,\mathcal{T}) & \approx & (\mathscr{A},\mathcal{A}) \\ \wr\wr & & \wr\wr & & \wr\wr \\ \mathcal{G} & \approx & \bar{\mathcal{G}} & \approx & \tilde{\mathcal{G}} \end{array}$$

which are the isomorphisms among the spaces. In the applications, we will resort to the Sufficiency Principle to satisfy the assumptions.

Given the settings, the right and left invariant Haar densities on Θ can be denoted by $\pi^R(\theta)$ and $\pi^L(\theta)$ and called the right and left invariant prior densities. Note that the densities turn out to be non-informative priors which have no information about the parameter θ. Next we revisit the previous examples for location, scale, and location-scale groups.

Example 4.8. (Location Group) Suppose $X_1,\ X_2,\ \ldots, X_n$ is a random sample from a $\mathcal{N}(\mu,\sigma^2)$ with an unknown μ. Then the sampling distribution of $\bar{X}$ follows $\mathcal{N}(\mu,\sigma^2/n)$. Note that $\mathcal{X} = \{\bar{x} : \bar{x} \in \mathbb{R}\}$ and $\Theta = \{\mu : \mu \in \mathbb{R}\}$. This implies $\mathcal{X} \approx \mathcal{G} \approx \bar{\mathcal{G}} \approx \Theta \approx \mathbb{R}$. Therefore, the Haar densities from the previous example can be used for the common prior density as follows:

$$\pi^R(\mu) = \pi^L(\mu) = 1.$$

Note that this is the conventional non-informative prior for a location parameter.

Example 4.9. (Scale Group) Suppose $X_1,\ X_2,\ \ldots, X_n$ is a random sample from a $\mathcal{N}(\mu,\sigma^2)$ with an unknown σ^2. Then the sampling distribution of $(n-1)S^2/\sigma^2$ follows $\chi^2(n-1)$. Note that $\mathcal{X} = \{s^2 : s^2 \in \mathbb{R}^+\}$ and $\Theta = \{\sigma^2 : \sigma^2 \in \mathbb{R}^+\}$. Thus, $\mathcal{X} \approx \mathcal{G} \approx \bar{\mathcal{G}} \approx \Theta \approx \mathbb{R}^+$. As a result, the Haar densities from the previous example can be used for the common prior density as follows:

$$\pi^R(\sigma^2) = \pi^L(\sigma^2) = \frac{1}{\sigma^2}.$$

Note that this is also a non-informative prior for a scale parameter.

Example 4.10. (Location-Scale Group) Suppose $X_1,\ X_2,\ \ldots, X_n$ is a random sample from a $\mathcal{N}(\mu, \sigma^2)$ with an unknown μ and σ^2. The sufficient statistic is $(\bar{X}, S^2)$. Then $\mathcal{X} \approx \mathcal{G} \approx \bar{\mathcal{G}} \approx \Theta \approx (\mathbb{R}, \mathbb{R}^+)$. Therefore, we can use the Haar densities from the previous example for the prior densities:

$$\pi^R((\mu, \sigma^2)) = \frac{1}{\sigma^2} \ \ \& \ \ \pi^L((\mu, \sigma^2)) = \frac{1}{\sigma^4}.$$

4.4 Equivariant Multiple Interval Estimator with Haar Measures

In this section, we define an equivariant MIE with Haar measures (EMIE Haar) and investigate its property. To do so, we start from a single dimensional case with respect to the corresponding group structures.

4.4.1 The Single Dimensional Case

We start this subsection by providing a classical definition of a confidence interval.

Definition 4.11. Given $\alpha \in (0, 1)$, a $100 \times (1 - \alpha)\%$ confidence interval for $\theta \in \Theta$ is defined by $\Gamma : \mathcal{X} \to \mathcal{T}$, such that

$$P_\theta[\theta \in \Gamma(X)] \geq 1 - \alpha \ \text{ for every } \ \theta \in \Theta.$$

Suppose $\mathcal{G}$ and $\bar{\mathcal{G}}$ are the group structures for $(\mathcal{X}, \mathcal{F})$ and $(\Theta, \mathcal{T})$, respectively. Assume the transitivity of Θ via $\bar{\mathcal{G}}$. Then Γ is called a equivariant IE if it satisfies the following equation:

$$P_\theta[\theta \in \Gamma(X)] = P_{\bar{g}(\theta)}[\bar{g}(\theta) \in \Gamma(X)] \text{ for } \theta \in \Theta,\ X \in \mathcal{X}, \text{ and } g \in \mathcal{G}.$$

Observe that this result is immediate from the (Proposition 4.1) by using $L(\theta, \Gamma) = I\{\theta \in \Gamma\}$. As mentioned previously, determining the best equivariant decision is not an easy task, and the situation is the same for the best equivariant interval estimator (EIE). However, given the assumptions in <Table 4.1>, the equlity between the classical coverage probability and the Bayesian coverage probability holds true, so

the best equivariant IE can be derived as the generalized Bayes decision based on the posterior distribution with respect to the right invariant Haar measure. (Proposition 4.12) provides the result as follows:

Proposition 4.12. *Suppose $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\Theta}$, $\mathcal{G}$, and $\bar{\mathcal{G}}$ are all isomorphic. Then the following equation holds for an equivariant interval estimator, $\Gamma(\cdot)$:*

$$P^{\pi^R(\theta|e)}[\theta \in \Gamma(e)] = P^{\pi^R(\theta|x)}[\theta \in \Gamma(x)] = P_\theta[\theta \in \Gamma(X)] = P_{\bar{e}}[\bar{e} \in \Gamma(X)].$$

As reminder, the prior distribution is the right Haar distribution. Although the result is based on quite strong assumptions, it is a surprising result as it implies a reconciliation between the frequentist and Bayesian procedures. In fact, this is one definition of the non-informative prior distribution, the probability matching prior. Observe that the posterior coverage probability is the same as the classical coverage probability even after updating its information via the prior distribution. This intuively implies the prior distribution has no information about the parameter θ. Furthermore, the coverage probabilities can be obtained with respect to the identity element of the group structure due to the Invariance Principle. Note that the result holds for general equivariant statistical decisions, as in (Lemma 4.13).

Lemma 4.13. *Suppose $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\Theta}$, $\mathcal{G}$, and $\bar{\mathcal{G}}$ are all isomorphic. Then for an equivariant decision rule $\delta(x) = \tilde{x}(\delta(e))$,*

$$E^{\pi^R(\theta|x)}[L(\theta, \tilde{x}(\delta(e)))] = R(\theta, \delta) = \int_{\boldsymbol{\mathcal{X}}} L(\bar{e}, \tilde{y}(\delta(e)))f(y|\bar{e})dy$$

where $\pi^R(\theta|x)$ is the posterior distribution of θ given x with respect to the right Haar prior distribution.

Next, we consider the expected length of an EIE. Note that the length of IEs is positively unbounded, and it is difficult to compare the perfomance of IEs with different group structures. This is quite different from hypothesis testing where the

power of the test can be used to compare the performance of the different procedures. However, Pratt [45] provides a different point of view on the expected length

$$E_\theta[\nu(\Gamma(X))] = \int_\Theta P_\theta[\theta' \in \Gamma(X)]d\theta'$$

where ν is the content measure of the interval estimator. The integrand on the right-hand side is called a false coverage probability, and the equality implies that its integral is equivalent to the expected length of the IE. Note that this relation is proved for the location group case when $\nu(\cdot)$ is the Lebesgue measure. Thus, it must be further generalized to handle different group structures. Here, we utilize the idea based on the left Haar measure as it appeared in Kiefer [34].

Definition 4.14. Suppose $\Gamma(\cdot)$ is an equivariant interval estimator. Then the expected length of the estimator is defined in the following manner:

$$E_\theta[\nu^L(\Gamma(X))] = \int_{\boldsymbol{\mathcal{X}}} \int_{\Gamma(x)} \pi^L(\theta')d\theta' f(x|\theta)dx$$

where $\nu^L(\Gamma(x)) = \int_{\Gamma(x)} \pi^L(\theta)d\theta$.

The suggested expected length can be simplified further through the invariance property of the left Haar measure. Moreover, it can be elegantly interpreted using the false coverage probability. These results are summarized in (Proposition 4.15) as follows:

Proposition 4.15. *Suppose* $\Gamma(\cdot)$ *is equivariant and* $E_\theta[\nu^L(\Gamma(X))]$ *is the expected content defined in (Definition 4.14). Then the following arguments hold true:*

$$E_\theta[\nu^L(\Gamma(X))] = \int_{\boldsymbol{\Theta}} Pr_{\bar{e}}[\theta \in \Gamma(X)]\pi^L(\theta)d\theta = \nu^L(\Gamma(e)).$$

From the first equality, the expected length is the same as the false coverage probability averaged with respect to the left Haar distribution. Therefore, this can be useful to compare the expected lengths among different group structures by considering them to be the averages of the false coverage probabilities. Note that these

averages are *generalized* in the sense that the left Haar density is a non-informative prior. Furthermore, the second equality says the expected length is a constant for any x; therefore, we can replace x with the identity element e. Later, in (Proposition 4.17), this simplicity provides a relatively easy way to derive an optimized EIE.

Finally, we formally define an optimal equivariant interval estimator with the Haar measures (EIE Haar):

Definition 4.16. Given $\alpha \in (0,1)$, a $100 \times (1-\alpha)\%$ interval estimator, $\Gamma(\cdot)$, is the best equivariant if it minimizes $E[\nu^L(\Gamma(X))]$ subject to

$$P^{\pi^R(\theta|x)}[\theta \in \Gamma(x)] = P_\theta[\theta \in \Gamma(X)] \geq 1-\alpha.$$

Observe that this definition encompasses both the frequentist and Bayesian perspectives at the same time. Furthermore, the invariant property simplfies the coverage probability and expected length as in (Proposition 4.12) and (Proposition 4.15). However, it may be difficult to derive the best equivariant interval estimator (BEIE) directly from the definition. For the actual derivation, we can use the following proposition:

Proposition 4.17. *The best equivariant interval estimator (BEIE)* $\Gamma^\star$ *satisfies* $\Gamma^\star(x) = \tilde{x}\Gamma^\star(e)$ *where*

$$\Gamma^\star(e) = \{\theta \in \Theta : \pi^R(\theta|e) \geq C\pi^L(\theta)\}$$

where C *can be chosen by*

$$P^{\pi^R(\theta|e)}[\theta \in \Gamma^\star(e)] = 1-\alpha.$$

As illustrations, we provide a few examples under a normal distribution.

Example 4.18. (Location Group) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $\mathcal{N}(\mu, \sigma^2)$ with an unknown μ and a known σ^2. The goal is to construct the BEIE

for μ. We know this problem has the location group structure and the resulting left and right Haar densities are 1 from (example 4.5). In addition, in (example 4.8), we confirmed the isomorphic structure was represented by $\mathbb{R}$, so that the non-informative prior density becomes $\pi^R(\mu) = \pi^L(\mu) = 1$. Then

$$\pi^R(\mu|\bar{x}) = \frac{f(\bar{x}|\mu)\pi^R(\mu)}{\int f(\bar{x}|\mu)\pi^R(\mu)d\mu} = f(\bar{x}|\mu)$$

Note that the identity element $e = 0$ for the location group and set σ_0^2 to be σ^2/n. In this case, by (Proposition 4.17),

$$\begin{aligned}\Gamma(0) &= \left\{\mu \in \mathbb{R} : \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{\mu^2}{2\sigma_0^2}\right) > C\right\} \\ &= \left\{\mu \in \mathbb{R} : \frac{1}{\sigma_0}\phi\left(\frac{\mu}{\sigma_0}\right) > C\right\}\end{aligned}$$

where C can be chosen by

$$\begin{aligned}1-\alpha =& P^{\pi^R(\mu|0)}[\mu \in \Gamma(0)] = \int I\{L(0) \le \mu \le U(0)\}\pi^R(\mu|0)d\mu \\ =& \int_{L(0)}^{U(0)} \frac{1}{\sigma_0}\phi\left(\frac{\mu}{\sigma_0}\right)d\mu = \Phi\left(\frac{U(0)}{\sigma_0}\right) - \Phi\left(\frac{L(0)}{\sigma_0}\right)\end{aligned}$$

Therefore, $\Gamma(0)$ should have the following form

$$\Gamma(0) = [L(0), U(0)] = [-z_{\alpha/2}\sigma_0, z_{\alpha/2}\sigma_0] = \left[-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$.

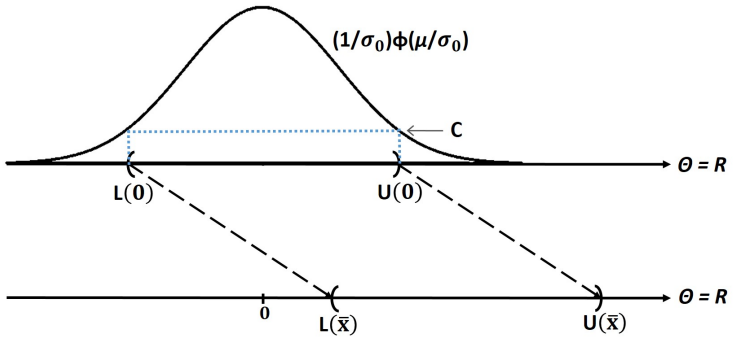


Figure 4.1: BEIE for location group under a normal distribution

Lastly,

$$\Gamma^*(\bar{x}) = \left[\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Note that the resulting BEIE is identical to the classical z-based confidence interval for μ. Just as this known σ^2 case generates the form of a z-based confidence interval, an unknown σ^2 case generates the form of a t-based confidence interval.

Example 4.19. (Scale Group) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $\mathcal{N}(\mu, \sigma^2)$. This time we assume both μ and σ^2 are unknown. The goal is to construct BEIE for σ^2. We know this problem has the scale group structure and the resulting left and right Haar densities are $1/b$ from (example 4.6). In addition, in (example 4.9), we confirmed the isomorphic structure was represented by $\mathbb{R}^+$, so that the non-informative prior density becomes $\pi^R(\sigma^2) = \pi^L(\sigma^2) = 1/\sigma^2$. Suppose f_{n-1} denotes the density of the $\chi^2(n-1)$ distribution. Then

$$\pi^R(\sigma^2|s^2) = \frac{\frac{(n-1)}{\sigma^2} f_{n-1}\left(\frac{(n-1)s^2}{\sigma^2}|\sigma^2\right)\pi^R(\sigma^2)}{\int \frac{(n-1)}{\sigma^2} f_{n-1}\left(\frac{(n-1)s^2}{\sigma^2}|\sigma^2\right)\pi^R(\sigma^2)d\sigma^2} = K\frac{(n-1)}{\sigma^4} f_{n-1}\left(\frac{(n-1)s^2}{\sigma^2}|\sigma^2\right)$$

By adjusting the identity element $e = 1$ to the posterior distribution, the constant K becomes 1. Therefore, by (Proposition 4.17),

$$\begin{aligned}
\Gamma(1) =& [L(1), U(1)] = \left\{\sigma^2 \in \mathbb{R}^+ : \frac{(n-1)}{\sigma^4} f_{n-1}\left(\frac{(n-1)}{\sigma^2}|\sigma^2\right) \geq \frac{C}{\sigma^2}\right\} \\
=& \left\{\sigma^2 \in \mathbb{R}^+ : \frac{(n-1)}{\sigma^2} f_{n-1}\left(\frac{(n-1)}{\sigma^2}|\sigma^2\right) \geq C\right\} \\
=& \left\{\sigma^2 \in \mathbb{R}^+ : \left(\frac{n-1}{\sigma^2}\right)^{\frac{n-1}{2}} \exp\left(-\frac{n-1}{2\sigma^2}\right) > C'\right\} \\
=& \left\{\sigma^2 \in \mathbb{R}^+ : a < \frac{n-1}{\sigma^2} < b\right\} \\
=& \left\{\sigma^2 \in \mathbb{R}^+ : \frac{n-1}{b} < \sigma^2 < \frac{n-1}{a}\right\} = \left[\frac{n-1}{b}, \frac{n-1}{a}\right]
\end{aligned}$$

Note that $t^{\frac{n-1}{2}} \exp(-t/2)$ is unimodal as the kernel of f_{n+1}. Thus, we have the following first equation: $f_{n+1}(a) = f_{n+1}(b)$. In addition, we obtain the second equation from the coverage condition:

$$1 - \alpha = P^{\pi^R(\sigma^2|1)}[\sigma^2 \in \Gamma(1)] = \int I\{L(1) \leq \sigma^2 \leq U(1)\}\pi^R(\sigma^2|1)d\sigma^2$$

$$
\begin{aligned}
&= \int_{L(1)}^{U(1)} \left(\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}\right)^{-1} \left(\frac{n-1}{\sigma^2}\right)^{\frac{n-1}{2}} \exp\left(-\frac{n-1}{2\sigma^2}\right)\frac{1}{\sigma^2}d\sigma^2 \\
&= \int_{a}^{b} \left(\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}\right)^{-1} (t)^{\frac{n-1}{2}-1} \exp\left(-\frac{t}{2}\right)dt \\
&= \int_{a}^{b} f_{n-1}(t)dt
\end{aligned}
$$

By solving the two equations numerically, we have the form, $\Gamma^\star(1) = \left[\frac{(n-1)}{b}, \frac{(n-1)}{a}\right]$.

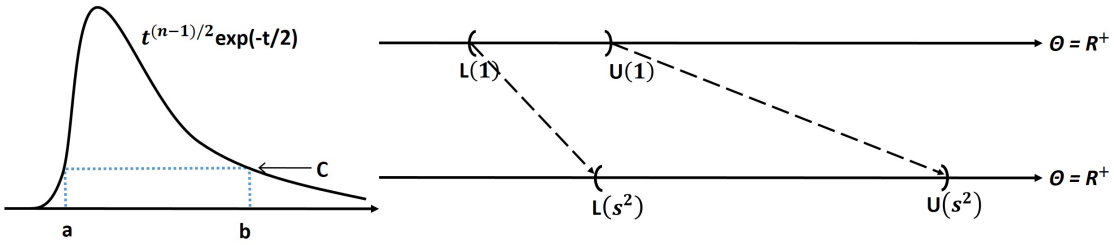


Figure 4.2: BEIE for scale group under a normal distribution

Lastly, the BEIE has the following form:

$$\Gamma^\star(S^2) = \left[\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}\right]$$

Note that the resulting BEIE is identical to the shortest unbiased interval, which minimizes the expected false coverage probability among all unbiased intervals. There are other intervals, the likelihood ratio based interval and the minimum length interval, which satisfy the first equation $f_{n+2}(a) = f_{n+2}(b)$ and $f_{n+3}(a) = f_{n+3}(b)$, respectively, as in Tate and Klett [51]. The length comparison among the IEs is summarized in <Figure 4.3>. IE1 is the BEIE, IE2 is the likelihood ratio based interval, IE2 is the minimum length interval, and IE4 is the usual equi-tailed interval which assigns the same probability to both tails. It is important to note that the interval lengths are determined by the measure based on the left Haar measure.

Example 4.20. (Location-Scale Group) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $\mathcal{N}(\mu, \sigma^2)$ with unknown μ and σ^2. Since the problem has the location-scale group structure, the resulting right and left Haar densities have the form of $1/b$ and $1/b^2$ from (example 4.7). In addition, in (example 4.10), we confirmed the isomorphic

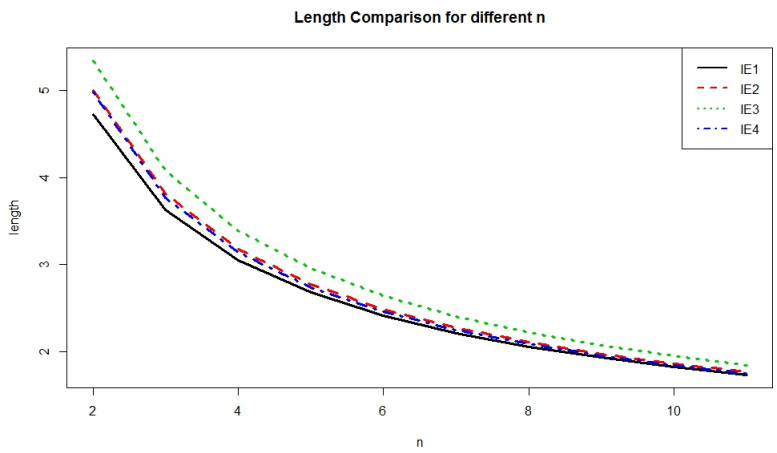


Figure 4.3: Length Comparison between IEs based for different n

structure was represented by $\mathbb{R}\times\mathbb{R}^+$, so that the non-informative prior density becomes $\pi^R((\mu,\sigma^2)) = 1/\sigma^2$ and $\pi^L((\mu,\sigma^2)) = 1/\sigma^4$. Then

$$\begin{aligned}\pi^R((\mu,\sigma^2)|(\bar{x},s^2)) =& Kf((\bar{x},s^2)|(\mu,\sigma^2))\pi^R((\mu,\sigma^2)) \\ =& K'\frac{(s^2)^{\frac{n-3}{2}}}{(\sigma^2)^{\frac{n}{2}+1}}\exp\left(-\frac{(\bar{x}-\mu)^2}{2\sigma^2/n}\right)\exp\left(\frac{-(n-1)s^2}{2\sigma^2}\right)\end{aligned}$$

Note that the identity element $e = (0,1)$ for the location-scale group. Thus, by (Proposition 4.17),

$$\begin{aligned}\Gamma((0,1)) =& \left\{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}^+ : K'\frac{1}{(\sigma^2)^{\frac{n}{2}+1}}\exp\left(-\frac{n\mu^2}{2\sigma^2}\right)\exp\left(\frac{-(n-1)}{2\sigma^2}\right) > C\frac{1}{\sigma^4}\right\} \\ =& \left\{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}^+ : (\sigma^2)^{-(n/2-1)}\exp\left(-\frac{n\mu^2+(n-1)}{2\sigma^2}\right) > C'\right\}\end{aligned}$$

where C' is chosen by $P^{\pi^R((\mu,\sigma^2)|(0,1))}[(\mu,\sigma^2)\in\Gamma(0,1)] = 1-\alpha$.

The optimal estimator can be numerically derived, so that

$$\Gamma^*((\bar{x},s^2)) = \overline{(\bar{x},s^2)}\Gamma((0,1)) = \{(\bar{x}+s^2v, s^2w) : (v,w)\in\Gamma((0,1))\}.$$

4.4.2 The Multi-Dimensional Extension

In this subsection, we extend the single EIE Haar to an EMIE Haar. While the MIE still attempts to minimize the global expected content under the global coverage

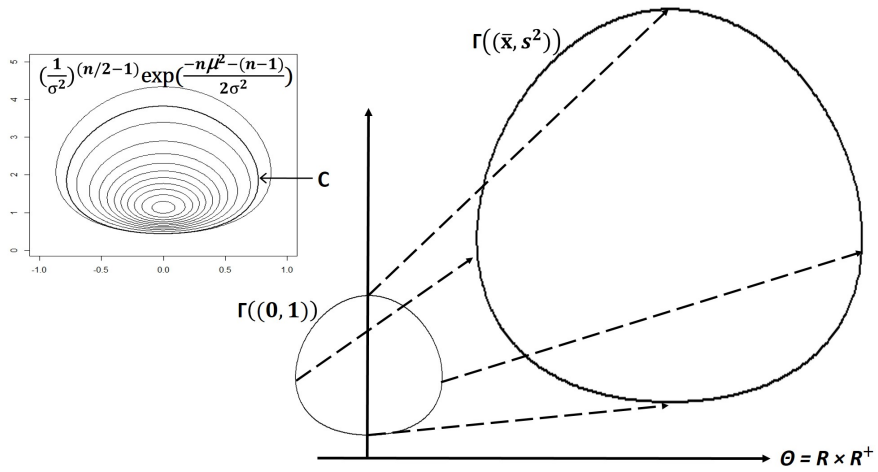
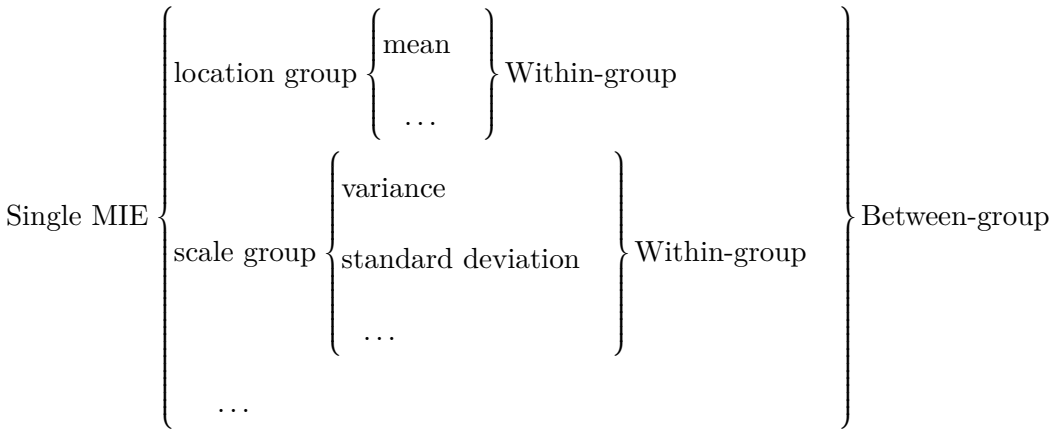


Figure 4.4: BEIE for location-scale group under a normal distribution

constraint, it has an additional goal by contrast to the BMIE Thres. Note that the BMIE Thres handles M *homogeneous* parameters simultaneously. These are homogeneous in the sense that the parameters are of a single particular type, e.g., the population means, from identical probability distributions, e.g., normal distributions. On the other hand, the EMIE Haar is designed to handle M *heterogeneous* parameters which can be different types from distinct probability distributions. To do so, we categorize the different parameters with respect to corresponding group structures so that the Invariance Principle can be utilized for the parameter groups, respectively. Through the process, we raise the unit of estimation from the individual parameter level to the individual group level. Still, this is not a simple task because we have to deal with the both within- and between-group structures as follows:



As with the BMIE Thres, we first introduce the mathematical setting for the EMIE Haar. Consider a statistical model $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{P}})$ where $\boldsymbol{\mathcal{X}} = \times_{m=1}^{M} \mathcal{X}_m$, $\boldsymbol{\mathcal{F}} = \sigma\left(\times_{m=1}^{M} \mathcal{F}_m\right)$, and $\boldsymbol{\mathcal{P}}$ is the collection of probability measures on the product space. In addition, $(\boldsymbol{\Theta}, \boldsymbol{\mathcal{T}})$ is a measurable space with $\boldsymbol{\Theta} = \times_{m=1}^{M} \Theta_m$ and $\boldsymbol{\mathcal{T}} = \sigma\left(\times_{m=1}^{M} \mathcal{T}_m\right)$, where Θ_m and $\mathcal{T}_m$ are the mth parameter space and its σ-field. Suppose $\boldsymbol{P_\theta}$ is the family of continuous probability distribution functions with respect to $\boldsymbol{\mathcal{P}}$ on $\boldsymbol{\mathcal{X}}$. From the model, a random quantity $\boldsymbol{X}$ is observed from $\boldsymbol{P_\theta}$. Assume the conditional independence of $X_m|\theta_m$'s throughout the index $m = 1, 2, \ldots, M$, so that $\boldsymbol{P_\theta}(\boldsymbol{X}) = \prod_{m=1}^{M} P_{\theta_m}(X_m)$. We also allow the individual distribution functions to be from different probability distributions. Lastly, for each fixed $\theta_m \in \Theta_m$, let f_{θ_m} be the density of P_{θ_m} and assume the mapping $(\theta_m, x_m) \mapsto f_{\theta_m}(x_m) = f(x_m|\theta_m)$ is product-measurable.

Define $(\boldsymbol{\mathcal{G}}, \bar{\boldsymbol{\mathcal{G}}})$ to be the families of groups for the sample and parameter spaces. That is, both $\boldsymbol{\mathcal{G}}$ and $\bar{\boldsymbol{\mathcal{G}}}$ contain K number of heterogeneous group structures, such as location, scale, location-scale, general linear, etc. Therefore, we assume that M_k number of individual estimations are assigned to $\mathcal{G}_k$ for $k = 1, 2, \ldots, K$ with $\sum_{i=1}^{K} M_k = M$. Then the forms of Haar measures depend on the specified group structures. Let $\mathcal{H}_k^L$ and $\mathcal{H}_k^R$ be the kth left and right Haar measures for $\mathcal{G}_k$ with the corresponding distribution and density functions, $H_k^L(\cdot)$, $H_k^R(\cdot)$, $h_k^L(\cdot)$, and $h_k^R(\cdot)$. In this study, we further assume all $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\Theta}$, $\boldsymbol{\mathcal{G}}$ and $\bar{\boldsymbol{\mathcal{G}}}$ satisfy the additional assumption in <Table

4.1> along with the transitivities. Under these conditions, the kth Haar distributions and densities will be shared as the prior distribution and densities for any individual estimations which have the kth group structures. In this study, we mainly restrict ourselves to the location-scale group structures, but further generalizations are possible involving a larger class of groups. As mentioned, once all the M dimensional random quantities are observed, then they can be rearranged so that proper group structures can be assigned to the individual estimations.

$$
\left.\begin{array}{l}
X^1_1, X^1_{12}, \ldots, X^1_{1n_1} \sim f^1_1(\cdot\,|\theta^1_1) \\
\qquad\vdots \\
X^1_{m1}, X^1_{m2}, \ldots, X^1_{mn_m} \sim f^1_J(\cdot\,|\theta^1_1) \\
\qquad\vdots \\
X^1_{M_1 1}, X^1_{M_1 2}, \ldots, X^1_{M_1 n_{M_1}} \sim f^1_J(\cdot\,|\theta^1_{M_1})
\end{array}\right\} \theta^1_1, \theta^1_2, \ldots, \theta^1_{M_k} \sim \pi^L_1(\cdot)\ \&\ \pi^R_1(\cdot)
$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$
\left.\begin{array}{l}
\vdots \\
X^k_{m1}, X^k_{m2}, \ldots, X^k_{mn_m} \sim f^k_m(\cdot\,|\theta^k_m) \\
\vdots
\end{array}\right\} \theta^k_1, \theta^k_2, \ldots, \theta^k_{M_k} \sim \pi^L_k(\cdot)\ \&\ \pi^R_k(\cdot)
$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$
\left.\begin{array}{l}
X^K_{11}, X^K_{12}, \ldots, X^K_{1n_1} \sim f^K_1(\cdot\,|\theta^K_1) \\
\qquad\vdots \\
X^K_{M_K 1}, X^K_{M_K 2}, \ldots, X^K_{M_K n_{M_K}} \sim f^K_{M_K}(\cdot\,|\theta^K_{M_K})
\end{array}\right\} \theta^K_1, \theta^K_2, \ldots, \theta^K_{M_K} \sim \pi^L_K(\cdot)\ \&\ \pi^R_K(\cdot)
$$

Suppose a random sample follows the mth probability distribution of the kth group structure. Then, by the Sufficiency Principle, let us say the sample information regarding the target parameter is summarized in T^k_m. Given $\alpha^k_m \in (0,1)$, recall that the corresponding equivariant interval estimator (EIE) satisfies the relation, $\Gamma^k_m(g_k(t^k_m); \alpha^k_m, \mathcal{G}_k) = \bar{g}_k(\Gamma^k_m(t^k_m); \alpha^k_m, \mathcal{G}_k)$. Then (Corollary 4.12) suggests its coverage

probability can be summarized either by the frequentist or Bayesian perspective as follows:

$$P^{\pi_k^R(\theta_m^k|t_m^k)}\left[\theta_m^k \in \Gamma_m^k(t_m^k;\alpha_m^k,\mathcal{G}_k)\right] = P_{\theta_m^k}\left[\theta_m^k \in \Gamma_m^k(T_m^k;\alpha_m^k,\mathcal{G}_k)\right]$$

for $m = 1,2,\ldots,M_k$ and $k = 1,2,\ldots,K$. Note that the posterior probability on the left-hand side is based on the prior distribution with respect to the right Haar distribution, $H_k^R(\theta_m)$. Since the equality holds for any k, we define the global coverage probability of the EMIE Haar as follows.

Definition 4.21. Let $\boldsymbol{\mathcal{G}}$ be the family of group structures with K distinct groups. Suppose the kth group is assigned to M_k number of estimations for $k = 1,2,\ldots,K$. Then, given $q \in (0,1)$, a $100\times(1-q)\%$ equivariant multiple interval estimator with Haar measure for $\boldsymbol{\theta}$ is a map, $\boldsymbol{\Gamma(\cdot)} : \boldsymbol{\mathcal{X}} \longrightarrow \boldsymbol{\mathcal{T}}$, such that

$$\prod_{k=1}^{K}\prod_{m=1}^{M_k} P^{\pi_k^R(\theta_m^k|e_m^k)}[\theta_m^k \in \Gamma_m^k(e_m^k;\alpha_m^k,\mathcal{G}_k)] = \prod_{k=1}^{K}\prod_{m=1}^{M_k} P_{\bar{e}_m^k}[\bar{e}_m^k \in \Gamma_m^k(T_m^k;\alpha_m^k,\mathcal{G}_k)] \geq 1-q.$$

In addition, the left-hand side quantity is called the equivariant family-wise coverage rate of $\boldsymbol{\Gamma(X)}$ and denoted by EFWCR$[\boldsymbol{\alpha};\boldsymbol{\mathcal{G}}]$.

Our goal is to determine the best EMIE Haar. This requires an optimality condition which is based on the global expected content as follows.

Definition 4.22. The best equivariant multiple intervals estimator with Haar measures is the EMIE Haar which minimizes the following global expected content:

$$\boldsymbol{\nu^L(\Gamma(x;\alpha,\mathcal{G}))} = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{\sum_{m=1}^{M_k}\nu_k^L(\Gamma_m^k(e_m^k;\alpha_m^k,\mathcal{G}_k))}{\beta_k + \sum_{m=1}^{M_k}\nu_k^L(\Gamma_m^k(e_m^k;\alpha_m^k,\mathcal{G}_k))}\right]$$

where β_k's are positive constants.

Note that $\nu_k^L(\Gamma_m^k(e_m^k;\alpha_m^k,\mathcal{G}_k))$ is the individual expected length in (Definition 4.14). Here, the β_k's have an important role in handling the within- and between-group structures. First, they are the decision factors for the individual level allocations within the groups through the confidence level investing strategy. In addition,

they decides the weights of the groups as seen in the following reparametrization:

$$\frac{1}{K}\sum_k \frac{\sum_m \nu_k^L(\Gamma_m^k)}{\beta_k + \sum_m \nu_k^L(\Gamma_m^k)} = \frac{1}{K}\sum_k \frac{1/\beta_k \sum_m \nu_k^L(\Gamma_m^k)}{1 + 1/\beta_k \sum_m \nu_k^L(\Gamma_m^k)} = \frac{1}{K}\sum_k \frac{\sum_m \gamma_k \nu_k^L(\Gamma_m^k)}{1 + \sum_m \gamma_k \nu_k^L(\Gamma_m^k)}.$$

4.5 Optimization

In this section, we introduce an optimization procedure which can handle different group structures in a MIE. Note that this procedure is a variation of the general optimization strategy in chapter 2 with the folowing goal:

$$\text{minimize } \boldsymbol{R_0^\beta(\theta,\delta)} \text{ subject to } \boldsymbol{R_1(\theta,\delta)} \leq q$$

where two global risk functions are refined as follows:

$$\begin{aligned}
\boldsymbol{R_0^\beta(\theta,\delta)} =& \frac{1}{K}\sum_{k=1}^{K}\left[\frac{\sum_{m=1}^{M_k} \nu_k^L(\Gamma_m^k(e_m^k;\alpha_m^k,\mathcal{G}_k))}{\beta_k + \sum_{m=1}^{M_k} \nu_k^L(\Gamma_m^k(e_m^k;\alpha_m^k,\mathcal{G}_k))}\right];\\
\boldsymbol{R_1(\theta,\delta)} =& 1 - \prod_{k=1}^{K}\prod_{m=1}^{M_k} P^{\pi_k^R(\theta_m^k|e_m^k)}[\theta_m^k \in \Gamma_m^k(e_m^k;\alpha_m^k,\mathcal{G}_k)].
\end{aligned}$$

A challenging part of the optimization is that it consists of K different components, which reflect the corresponding group structures. If we had information about group-wise level conditions, $1-q_k$ for $k = 1, 2, \ldots, K$ such that $1 - q = \prod_{k=1}^{K}(1-q_k)$, then we could simply perform the K separate optimizations and combine these results. However, this approach has an apparent limitation: it requires information about q_k's; moreover, it does not offer any interactions among the group structures. In this study, we do not assume the information and we let the procedure decide the weights, β_k's. Before proceeding to the actual application, we rewrite the optimization problem in terms of the EFWCR as follows:

$$\begin{aligned}
&\text{minimize } \frac{1}{K}\sum_{k=1}^{K}\left[\frac{\sum_{m=1}^{M_k} \nu_k^L(\Gamma_m^k(e_m^k;\alpha_m^k,\mathcal{G}_k))}{\beta_k + \sum_{m=1}^{M_k} \nu_k^L(\Gamma_m^k(e_m^k;\alpha_m^k,\mathcal{G}_k))}\right]\\
&\text{subject to } \sum_{k=1}^{K}\sum_{m=1}^{M_k} \log\left(P^{\pi_k^R(\theta_m^k|e_m^k)}[\theta_m^k \in \Gamma_m^k(e_m^k;\alpha_m^k,\mathcal{G}_k)]\right) \geq \log(1-q).
\end{aligned}$$

4.6 Application: Mean and Variance of Normal Distribution

In this section, we prepare an application to illustrate the case in which we determine the best EMIE Haar. This particular application is for the means and variances of normal distributions. Note that this is the case for $K = 2$ where G_1 and G_2 are location and scale groups, respectively.

4.6.1 Application Procedure

Suppose we observe random quantities as follows:

$$\left.\begin{array}{l} X^1_{11}, X^1_{12}, X^1_{13}, \ldots, X^1_{1n_1} \sim \mathcal{N}(\mu^1_1, (\sigma^1_1)^2) \\ \qquad\qquad \vdots \\ X^1_{M_1 1}, X^1_{M_1 2}, \ldots, X^1_{M_1 n_{M_1}} \sim \mathcal{N}(\mu^1_{M_1}, (\sigma^1_{M_1})^2) \end{array}\right\} \mu^1_1, \mu^1_2, \ldots, \mu^1_{M_1} \sim \pi^L_1(\cdot)\ \&\ \pi^R_1(\cdot)$$

$$\left.\begin{array}{l} X^2_{11}, X^2_{12}, X^2_{13}, \ldots, X^2_{1n_1} \sim \mathcal{N}(\mu^2_1, (\sigma^2_1)^2) \\ \qquad\qquad \vdots \\ X^2_{M_2 1}, X^2_{M_2 2}, \ldots, X^2_{M_2 n_{M_2}} \sim \mathcal{N}(\mu^2_{M_2}, (\sigma^2_{M_2})^2) \end{array}\right\} (\sigma^2_1)^2, (\sigma^2_2)^2, \ldots, (\sigma^2_{M_2})^2 \sim \pi^L_2(\cdot)\ \&\ \pi^R_2(\cdot)$$

where $(\sigma^1_1)^2, (\sigma^1_2)^2, \ldots, (\sigma^1_{M_1})^2$ and $\mu^2_1, \mu^2_2, \ldots, \mu^2_{M_2}$ are known. As the location and scale groups are commutative, the corresponding left and right prior densities are identical:

$$\begin{aligned} \pi^L_1(\mu) &= \pi^R_1(\mu) = 1; \\ \pi^L_2(\sigma^2) &= \pi^R_2(\sigma^2) = \frac{1}{\sigma^2}. \end{aligned}$$

The resulting mth individual EIEs based on the identity elements have similar forms in (Example 4.18) and (Example 4.19):

$$\begin{aligned} \Gamma^1_m(e^1_m; \alpha^1_m, \mathcal{G}_1) &= \left[\pm z_{\alpha^1_m/2} \frac{\sigma^1_m}{\sqrt{n^1_m}} \right]; \\ \Gamma^2_m(e^2_m; \alpha^2_m, \mathcal{G}_2) &= \left[\frac{(n^2_m)}{b_m(\alpha^2_m)}, \frac{(n^2_m)}{a_m(\alpha^2_m)} \right]. \end{aligned}$$

where $a_m(\alpha_m^2)$ and $b_m(\alpha_m^2)$ are determined by two equations, $\int_{a_m}^{b_m} f_{n_m}(t)dt = 1 - \alpha_m^2$ and $f_{n_m+2}(a_m) = f_{n_m+2}(b_m)$, with f_{n_m} being the density function of the $\chi^2(n_m)$ distribution. From the indivdual EIEs, the optimization problem becomes as follows:

$$\begin{aligned}&\text{minimize } \frac{\sum_{m=1}^{M_1} 2z_{\alpha_m^1/2}\sigma_m^1}{\beta_1 + \sum_{m=1}^{M_1} 2z_{\alpha_m^1/2}\sigma_m^1} + \frac{\sum_{m=1}^{M_2} \log(b_m(\alpha_m^2)/a_m(\alpha_m^2)}{\beta_2 + \sum_{m=1}^{M_2} \log(b_m(\alpha_m^2)/a_m(\alpha_m^2))}\\ &\text{subject to } \sum_{m=1}^{M_1} \log(1-\alpha_m^1) + \sum_{m=1}^{M_2} \log(1-\alpha_m^2) \geq \log(1-q).\end{aligned}$$

We adjust the reparametrization $\nu_m = \Phi^{-1}(1-\alpha_m^1/2)$, $\nu_m^U = F_{n_m}^{-1}(1-\alpha_m^2/2)$ and $\nu_m^L = F_{n_m}^{-1}(\alpha_m^2/2)$, with the $\chi^2(n)$ distribution function F_n and provide a form we can solve through the Lagrange equation and Newton-Raphson method:

$$\begin{aligned}&\text{minimize } \frac{\sum_{m=1}^{M_1} 2\nu_m\sigma_m^1}{\beta_1 + \sum_{m=1}^{M_1} 2\nu_m\sigma_m^1} + \frac{\sum_{m=1}^{M_2} \log(\nu_m^U/\nu_m^L)}{\beta_2 + \sum_{m=1}^{M_2} \log(\nu_m^U/\nu_m^L)}\\ &\text{subject to } \sum_{m=1}^{M_1} \log(\Phi(2\nu_m - 1)) + \sum_{m=1}^{M_2} \log(F_{n_m^2}(\nu_m^U) - F_{n_m^2}(\nu_m^L)) \geq \log(1-q).\end{aligned}$$

4.6.2 Application Result

We implement the application for four different cases as follows. First, fix $M_2 = 100$ and choose $n_1^2, n_2^2, \ldots, n_{100}^2$ to be a sequence of positive integers from 2 to 101. Then assign four different set of positive numbers for $\sigma_1^1, \sigma_2^1, \ldots, \sigma_{M_1}^1$ so that we can ascertain the results for the different parameter combinations. <Table 4.2> presents the setting of the application. It is natural to have different equivariant global expected lengths (EGEL) from the four cases as we assume different σ_m^1's. Therefore, for each case, we provide the relative EGEL compared to the Sidak adjusted EGEL. In addition, we compare the best β_k's among the cases to see whether they work

Table 4.2: Description of Application

Case	Size of σ_m^1's	σ_m^1's	M_1	n_m^2's	M_2	M
I	Small	0.1, 0.2, ..., 1	10	2, 3, ..., 101	100	110
II	Medium	0.1, 0.31, ..., 3	15	2, 3, ..., 101	100	115
III	Small	0.005, 0.01, ..., 1	200	2, 3, ..., 101	100	300
IV	Large	90, 91, ..., 100	10	2, 3, ..., 101	100	110

Table 4.3: Result of Application

Case	β_1	β_2	Relative EGEL
I	1000	1000	85.01%
II	1000	226.52	94.84%
III	1000	2	99.47%
IV	1000	2	85.36%

properly as weighting devices. The results are summarized in <Table 4.3>. First, let us observe that the relative EGEL is less than 1 for each individual case, implying the optimization procedure helps to reduce the EGEL compared to the EMIE Haar with Sidak adjustment. Notice that the β's vary for different cases. We can ascertain that the procedure chooses appropriate weights, β_1 and β_2, for different parameter combinations. As a result, this process allows us to achieve the minimum relative EGEL. The values are gathered through a graphical search as presented in <Figure 4.5>. Observe that the value 1000 is used in the practical sense for several cases.

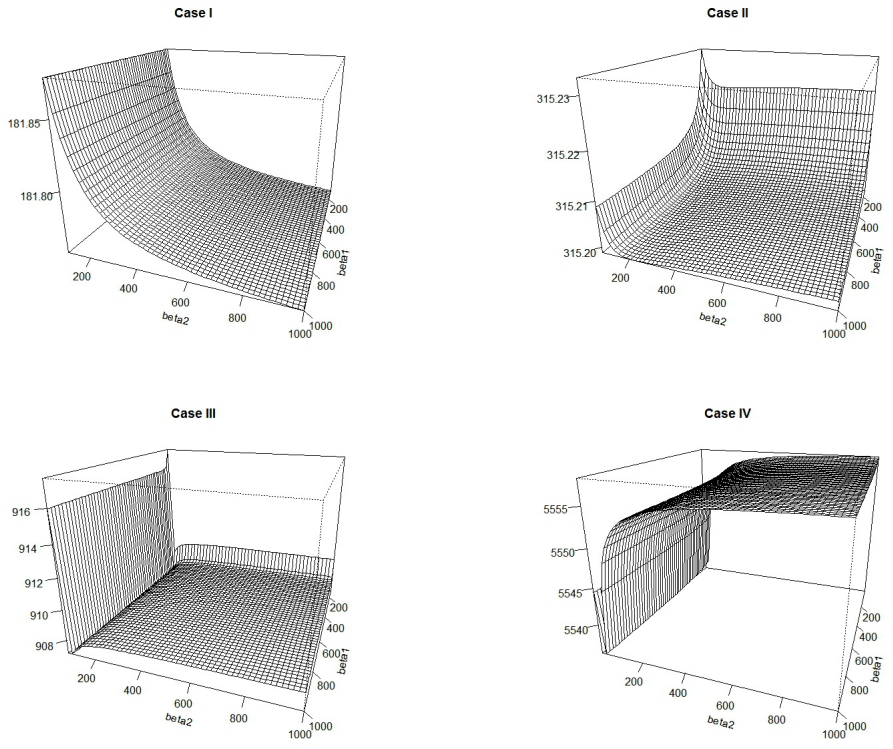


Figure 4.5: Choice of β_1 and β_2 for Different Cases

This is because the EGEL becomes stable for numbers larger than 1000.

4.7 Summary

The philosophical basis of Bayesian statistics is subjectivism. However, in practice, many Bayesian procedures utilize non-informative priors. The *standard* way to choose a non-informative prior distribution is due to Jeffreys. (Kass and Wasserman [33]) The formal rule is proposed in Jeffreys [32] as follows:

$$\pi(\theta) \propto \det(\boldsymbol{I}(\theta))^{1/2}$$

where $\boldsymbol{I}(\theta)_{ij} = E\left[-\frac{\partial^2 l}{\partial\theta_i \partial\theta_j}\right]$, the Fisher information matrix. There are many other ways to derive non-informative priors and our study corresponds to the method based on the Invariance Principle with Haar measures. Intuitively, if a prior distribution is truly non-informative, then the amount of information in the posterior distribution should be equivalent to the amount of information in the likelihood function. (Lemma 4.13) formally represents this reconciliation between the frequentist and Bayesian procedures. That is, given an invariance structure combined with an appropriate non-informative prior, an equivariant decision function results in the same posterior and classical risks. This relation then holds true for IEs as in (Proposition 4.12).

Notice that the method based on the Invariance Principle generates two non-informative priors with respect to the right and left Haar measures. Staticians differ on which of the two priors is most suitable. For example, the prior distribution with respect to the left Haar measure coincides with the Jeffreys' prior for non-location parameters. On the other hand, the prior distribution with respect to the right Haar measure is used in the EIE Haar because it provides the reconciliation between the coverage probabilities. The left Haar measure is also utlized to measure the expected lengtht of the EIE Haar. Due to the left-hand side invariance property, the expected content can also be calculated with respect to the identity element e. Combined

with the invariance property of the coverage probability, this simplifies the way to determine the best EIE Haar in (Proposition 4.17).

It is worth noticing that the result of the reconciliation in (Lemma 4.13) is based on the strong assumptions in <Table 4.1>. For example, the isomorphic relation between the sample space and parameter space may be unrealistic as the dimension of the sample space is generally higher than that of the parameter space. This is the reason we exploit the Sufficiency Principle prior to the actual procedure. The strong assumption can be partially relaxed by using homomorphisms. Under the relaxed assumption, a standard invariance technique which is based on a probability distribution conditioned by maximal invariant statistics ensures the reconciliation result as in Schervish [49]. Note that a famous example of the standard technieque is the Pitman estimator, or minimum risk equivariant estimator, in Pitman [44]. Further relaxations are possible; however, these require considerably deeper knowledge in group theory, e.g., the concept of amenability in Bondar and Milnes [4].

The main focus of the EMIR Haar, the multiple extension of EIE Haar, is how to handle a set of heterogeneous parameters simultaneously. To do so, each parameter is categorized with respect to the corresponding group structure. Then the basic unit of estimation becomes the set of IEs which share the group structure. This makes the IEs in the same group structure comparable. At the same time, the modified optimization procedure handles the between-group structure, assigning proper weights to the groups. This approach is meaningful in the sense that it deals with the heterogeneous parameters among different group structures. Moreover, the optimization excludes the subjective choice on the weights of the groups by letting the procedure decides the best weights to minimize the global expected content.

Chapter 5

Conclusion and Future Works

5.1 Concluding Remarks

The goal of this dissertation was to construct MIEs under the existence of prior information. It started by determining the proper extension of the global performance quantities: the global coverage probability and global expected content. Throughout the manuscript, the family-wise coverage rate was defined and utilized for the global coverage probability, and the average expected length was used for the global expected content. In terms of the prior information, we impose informative or non-informative prior distribution for different goals and purposes. Therefore, the two major topics, the BMIE Thres and EMIE Haar, assume the existence of prior distributions to enhance the performance of the MIEs. However, they are not intended to follow the frequentist or Bayesian paradigm. The BMIE Thres utilized an informative prior distribution to estabilish thresholds; however, the hyper-parameters could later be estimated in a way similar to the parametric empirical Bayes approach. On the other hand, the EMIE Thres utilized a non-informative prior distribution which played a role as a link between two paradigms in statistics under certain assumptions.

As mentioned, it is possible to categorize the BMIE Thres as one of the empirical Bayes MIEs. Several previous studies on empirical Bayes MIEs have investigated hyper-ellipsoid estimators. However, the resulting BMIE Thres is essentially a hyper-rectangular estimator. Although the BMIE Thres will be dominated by the former estimators in terms of the global expected content, it has an advantage in evaluating

the individual IEs' performances which is demanded in actual practices. In addition, the BMIE Thres with estimated hyper-parameters achieves a better level of robustness compared to the regular hyper-rectangular empirical Bayes MIE. This provides its own niche combined with the confirmed reduction of the global expected content via thresholding.

It is a quite ambitious idea to build up a procedure to simultaneously handle heterogeneous parameters which may even come from different probability distributions. The idea of EMIE Haar is to categorize the parameters within the proper groups so that they can be treated in similar fashion. The procedure mainly relies on the Invariance Principle with the corresponding Haar measures to simplify the problem as much as possible. Due to its complexity, the problem may not be completely handled by this analytic procedure. Then, the remaining parts are covered by the optimization procedure which allocates individual levels as well as weights to group structures. Certainly, there exists more room for further improvement to EMIE Haar. However, we believe the procedure delivers a meaningful result in the topic of multiple interval estimation.

Throughout the investigations on MIEs, we can see the dynamics and importance of the topic and realize again that they are worth independently study. One limitation of our study is the use of the FWCR for the global coverage probability. While major studies of MTPs have been reorganized based on FDR for the global type-I error rate, there exists a need for a refined global coverage probability to replace the FWCR. This is not a simple task, as the exact dual procedure of the FDR is impossible to obtain due to the non-existence of alternative hypothesis information in MIEs. If a refined global coverage probability is established, we can reformulate the BMIE Thres and EMIE Haar by replacing the global risk function in the optimization procedure.

5.2 Future Works

5.2.1 Bayes MIE with thresholding

In section 3.6, we used a plug-in procedure in the data application to construct the BMIE Thres under unknown standard deviations. However, we could also have built the procedure upwards from the full prior structure as follows:

$$\begin{aligned} \overline{X}_m|\mu_m,\lambda_m \sim & N\left(\mu_m,(n_m\lambda_m)^{-1}\right) \quad \& \quad (n_m-1)S_m^2|\lambda_m \sim g_1 = Gamma(\tfrac{n_m-1}{2},\tfrac{\lambda_m}{2}) \\ \mu_m|\lambda_m \sim & N\left(\eta,(\kappa_m\lambda_m)^{-1}\right) \quad \& \quad \lambda_m \sim g_2 = Gamma(a,b) \end{aligned}$$

where the normal distributions consist of location and precision parameters and the gamma distributions consist of shape and rate parameters. Given this structure, the form of the mth individual BIE Thres has the following form:

$$\Gamma_m(X_m;\alpha_m) = \left(\overline{X}_m - t_{\alpha_m/2}^{n_m-1}\tfrac{S_m}{\sqrt{n_m}} I_L(\overline{X}_m,\eta), \overline{X}_m + t_{\alpha_m/2}^{n_m-1}\tfrac{S_m}{\sqrt{n_m}} I_R(\overline{X}_m,\eta)\right)$$

$$\text{where } \begin{cases} I_L(\overline{X}_m,\eta) = I\left\{\overline{X}_m - \eta > -C\sqrt{\tfrac{b}{\kappa_m}}\tfrac{\Gamma(a-\frac{1}{2})}{\Gamma(a)}\right\}, \\ I_R(\overline{X}_m,\eta) = I\left\{\overline{X}_m - \eta < C\sqrt{\tfrac{b}{\kappa_m}}\tfrac{\Gamma(a-\frac{1}{2})}{\Gamma(a)}\right\}, \\ t_{\alpha_m/2}^{n_m-1} = F_{t_{n_m-1}}^{-1}(1-\alpha_m/2); \;\; S_m^2 = \tfrac{1}{n_m-1}\sum_{i=1}^{n_m}(X_i-\overline{X}_m)^2. \end{cases}$$

The corresponding Bayes coverage probability and Bayes expected length can be derived, but no closed forms exist. The hardest part is the estimation of the hyper parameters: η, a, b, and κ_m's. If this issue can be resolved, then we will be able to apply the above BMIE Thres to actual data.

Another interesting direction for future study of BMIE Thres is its nonparametric extensions. Note that under the normal-normal model, BMIE Thres shares the same setting as the parametric empirical Bayes. Therefore, a reasonable starting point for the extension would be the nonparametric empirical Bayes setting, and we could derive a threshold based on that nonparametric structure. Another extension would be the discovery of a method which does not rely on prior information, but instead

borrows the information for thresholding from a domain study such as economics, engineering, or biology.

5.2.2 Equivariant MIE with Haar Measures

Relaxation of the assumptions regarding the EMIE Haar was mentioned mainly in terms of the isomorphisms. In addition, the relaxation on the transitivity assumption is an interesting problem. When we lose the transitivity, a risk function has a distinct value on each orbit. Therefore, we must depend on the appropriate minimax estimation with respect to the orbits. The concept of cross-sections introduced in Wijsman [55] can be helpful in relaxing the assumptions. Note that this idea was also used by Hooper [31] and Hooper [30] in studies of invariant set estimation.

As in the BMIE Thres case, nonparametric extension of the EMIE Haar is an interesting area of study. Under the nonparametric setting, the distribution itself may be considered a nuisance parameter, so the current invariant structure should be refined to involve handling the additional structure. A good starting point was arranged in Peña and Kim [42]. Under a single dimensional nonparametric setting, they utilized the relaxed invariance structure to derive the optimal equivariant estimator within a certain class of estimators. Moving forward from their work, the next key step would be to determine the proper form of Haar measures for the extended invariant structure, i.e., the Haar measure of function spaces. It might be a reasonable conjecture that this Haar measure will be a non-informative nonparametric prior distribution.

Bibliography

[1] Yoav Benjamini and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the royal statistical society. Series B (Methodological)* (1995), pp. 289–300.

[2] Yoav Benjamini and Daniel Yekutieli. "False discovery rate–adjusted multiple confidence intervals for selected parameters". In: *Journal of the American Statistical Association* 100.469 (2005), pp. 71–81.

[3] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

[4] James V Bondar and Paul Milnes. "Amenability: A survey for statistical applications of Hunt-Stein and related conditions on groups". In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 57.1 (1981), pp. 103–128.

[5] J-Fa Brewster, JV Zidek, et al. "Improving on equivariant estimators". In: *The Annals of Statistics* 2.1 (1974), pp. 21–38.

[6] Lawrence D Brown. "In-season prediction of batting averages: A field test of empirical Bayes and Bayes methodologies". In: *The Annals of Applied Statistics* (2008), pp. 113–152.

[7] Henri Cartan. "Sur la mesure de Haar". In: *CR Acad. Sci. Paris* 211 (1940), pp. 759–762.

[8] George Casella and Jiunn Tzon Hwang. "Empirical Bayes confidence sets for the mean of a multivariate normal distribution". In: *Journal of the American Statistical Association* 78.383 (1983), pp. 688–698.

[9] George Casella and Jiunn Tzon Hwang. "Evaluating confidence sets using loss functions". In: *Statistica Sinica* (1991), pp. 159–173.

[10] Edgar Dobriban et al. "Optimal multiple testing under a Gaussian prior on the effect sizes". In: *Biometrika* 102.4 (2015), pp. 753–766.

[11] Morris L Eaton. “Group invariance applications in statistics”. In: (1989), pp. i–133.

[12] Morris L Eaton. “Multivariate statistics: a vector space approach.” In: *JOHN WILEY & SONS, INC., 605 THIRD AVE., NEW YORK, NY 10158, USA, 1983, 512* (1983).

[13] Bradley Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Vol. 1. Cambridge University Press, 2012.

[14] Bradley Efron and Trevor Hastie. *Computer age statistical inference*. Vol. 5. Cambridge University Press, 2016.

[15] Bradley Efron and Carl Morris. “Data analysis using Stein’s estimator and its generalizations”. In: *Journal of the American Statistical Association* 70.350 (1975), pp. 311–319.

[16] Bradley Efron and Carl Morris. “Stein’s paradox in statistics”. In: *Scientific American* 236.5 (1977), pp. 119–127.

[17] Thomas S Ferguson. *Mathematical statistics: A decision theoretic approach*. Vol. 1. Academic press, 2014.

[18] Ronald A Fisher. “Statistical methods and scientific inference.” In: (1956).

[19] William Fithian, Dennis Sun, and Jonathan Taylor. “Optimal inference after model selection”. In: *arXiv preprint arXiv:1410.2597* (2014).

[20] Donald Alexander Stuart Fraser. “Structural inference”. In: *Encyclopedia of Statistical Sciences* 13 (2004).

[21] Donald AS Fraser. “The fiducial method and invariance”. In: *Biometrika* 48.3/4 (1961), pp. 261–280.

[22] Narayan C Giri. *Multivariate statistical inference*. Academic Press, 2014.

[23] Irving John Good. *Good thinking: The foundations of probability and its applications*. U of Minnesota Press, 1983.

[24] Joshua D Habiger and Edsel A Peña. “Compound p-value statistics for multiple testing procedures”. In: *Journal of multivariate analysis* 126 (2014), pp. 153–166.

[25] Inge S Helland. “Quantum mechanics from symmetry and statistical modeling”. In: *International journal of theoretical physics* 38.7 (1999), pp. 1851–1881.

[26] Inge S Helland. "Statistical inference under symmetry". In: *International Statistical Review* 72.3 (2004), pp. 409–422.

[27] Y. Hochberg and A. C. Tamhane. *Multiple Comparison Procedures*. New York: Wiley, 1987.

[28] Yosef Hochberg. "A sharper Bonferroni procedure for multiple tests of significance". In: *Biometrika* 75.4 (1988), pp. 800–802.

[29] Sture Holm. "A simple sequentially rejective multiple test procedure". In: *Scandinavian journal of statistics* (1979), pp. 65–70.

[30] Peter M Hooper. "Invariant confidence sets with smallest expected measure". In: *The Annals of Statistics* (1982), pp. 1283–1294.

[31] Peter M Hooper. "Sufficiency and invariance in confidence set estimation". In: *The Annals of Statistics* (1982), pp. 549–555.

[32] Harold Jeffreys. "An invariant form for the prior probability in estimation problems". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 186.1007 (1946), pp. 453–461.

[33] Robert E Kass and Larry Wasserman. "The selection of prior distributions by formal rules". In: *Journal of the American Statistical Association* 91.435 (1996), pp. 1343–1370.

[34] J Kiefer. "Multivariate optimality results". In: *Multivariate Analysis.* (1966), pp. 255–274.

[35] J Kiefer et al. "Invariance, minimax sequential estimation, and continuous time processes". In: *The Annals of Mathematical Statistics* 28.3 (1957), pp. 573–601.

[36] Hirokichi Kudo. "On minimax invariant estimates of the transformation parameter". In: (1955).

[37] Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

[38] Rupert G Miller Jr. *Simultaneous statistical inference*. McGraw-Hill, 1966.

[39] Carl N Morris. "Parametric empirical Bayes confidence intervals". In: *Scientific inference, data analysis, and robustness*. Elsevier, 1983, pp. 25–50.

[40] Leopoldo Nachbin. *The haar integral*. RE Krieger Pub. Co., 1976.

[41] Edsel A Peña, Joshua D Habiger, and Wensong Wu. "Power-enhanced multiple decision functions controlling family-wise error and false discovery rates". In: *Annals of statistics* 39.1 (2011), p. 556.

[42] Edsel A Peña and Taeho Kim. "Median Confidence Regions in a Nonparametric Model". In: *Electronic Journal of Statistics* (2019).

[43] Edsel A Peña and Taeho Kim. "Optimal Confidence Regions under Group Invariance". In: *Unpublished Manuscript* (2017).

[44] EJG Pitman. "The estimation of the location and scale parameters of a continuous population of any given form". In: *Biometrika* 30.3/4 (1939), pp. 391–421.

[45] John W Pratt. "Length of confidence intervals". In: *Journal of the American Statistical Association* 56.295 (1961), pp. 549–567.

[46] Christian Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.

[47] SN Roy and Raj Chandra Bose. "Simultaneous confidence interval estimation". In: *The Annals of Mathematical Statistics* (1953), pp. 513–536.

[48] Henry Scheffé. "A method for judging all contrasts in the analysis of variance". In: *Biometrika* 40.1-2 (1953), pp. 87–110.

[49] Mark J Schervish. *Theory of statistics*. Springer Science & Business Media, 2012.

[50] Juliet Popper Shaffer. "Multiple hypothesis testing". In: *Annual review of psychology* 46.1 (1995), pp. 561–584.

[51] Robert F Tate and Gerald W Klett. "Optimal confidence intervals for the variance of a normal distribution". In: *Journal of the American statistical Association* 54.287 (1959), pp. 674–682.

[52] André Weil. "L'intégration dans les groupes topologiques et ses applications". In: *Hermann et Cie.* (1940).

[53] Peter H Westfall, Alok Krishen, and S Stanley Young. "Using prior information to allocate significance levels for multiple endpoints". In: *Statistics in medicine* 17.18 (1998), pp. 2107–2119.

[54] Peter H Westfall and Stanley S Young. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. Wiley-Interscience, 1993.

[55] Robert A Wijsman. “Cross-sections of orbits and their application to densities of maximal invariants”. In: 1 (1967), pp. 389–400.

[56] Robert A Wijsman. *Invariant measures on groups and their use in statistics*. 1990.

[57] James V Zidek. “A representation of Bayes invariant procedures in terms of Haar measure”. In: *Annals of the Institute of Statistical Mathematics* 21.1 (1969), pp. 291–308.

Appendix A

Proof of Propositions

Proposition A.1. *The Bayes expected length (BEL) of the mth BIE Thres has the form of*

$$BEL[C, \alpha_m; \sigma_m, \tau] = 2z_{\alpha_m/2}\sigma_m\Phi(C_m)$$

where $C_m = C\tau/\sqrt{\sigma_m^2 + \tau^2}$*, and the BEL approach the expected length of the corresponding z-based IE as* C *goes to infinity.*

Proof. We can first manipulate the original length as follows.

$$\begin{aligned}\nu(\Gamma_m(\boldsymbol{X}_m; \alpha_m)) =& z_{\alpha_m/2}\sigma_m\left[I\left\{\bar{X}_m < \eta + C\tau\right\} + I\left\{\bar{X}_m > \eta - C\tau\right\}\right] \\ =& z_{\alpha_m/2}\sigma_m\left[1 + I\left\{\eta - C\tau < \bar{X}_m < \eta + C\tau\right\}\right]\end{aligned}$$

where ν is the content measure which is Lebesgue in this location parameters case. To derive the Bayes length, we take the expectations as follows:

$$\begin{aligned}&\int_{\Theta_m}\int_{\mathcal{X}_m}\nu(\Gamma_m(x_m; \alpha_m))dP_m(x_m|\mu_m)d\Pi(\mu_m) \\ =& E_{\mu_m}\left[E_{\bar{X}_m|\mu_m}[\nu(\Gamma_m(\boldsymbol{X}_m; \alpha_m))]\right] \\ =& z_{\alpha_m/2}\sigma_m\left(1 + E_{\mu_m}P_{Z_m}\left[\tfrac{\eta - C\tau - \mu_m}{\sigma_m} < Z_m < \tfrac{\eta + C\tau - \mu_m}{\sigma_m}\right]\right) \\ =& z_{\alpha_m/2}\sigma_m\left(1 - E_{\mu_m}\left[\Phi\left(\tfrac{\mu_m - \eta - C\tau}{\sigma_m}\right) - \Phi\left(\tfrac{\mu_m - \eta + C\tau}{\sigma_m}\right)\right]\right) \\ =& z_{\alpha_m/2}\sigma_m\left(1 - E_{Z'_m}\left[\Phi\left(\tfrac{Z'_m - C}{\sigma_m/\tau}\right) - \Phi\left(\tfrac{Z'_m + C}{\sigma_m/\tau}\right)\right]\right) \\ =& z_{\alpha_m/2}\sigma_m\left(1 - \Phi\left(\tfrac{-C\tau}{\sqrt{\sigma_m^2+\tau^2}}\right) + \Phi\left(\tfrac{C\tau}{\sqrt{\sigma_m^2+\tau^2}}\right)\right) = 2z_{\alpha_m/2}\sigma_m\Phi\left(\tfrac{C\tau}{\sqrt{\sigma_m^2+\tau^2}}\right)\end{aligned}$$

where the Z_m and Z'_m follow the standard normal distribution. Once the form is derived, it is easy to observe, as C goes to infinity, this Bayes length approaches the expected length, $2z_{\alpha_m/2}\sigma_m$, of the classical z-based IE. □

Proposition A.2. *The Bayes coverage probability (BCP) of the mth BIE Thres has the form of*

$$BCP[C,\alpha_m;\sigma_m,\tau] = 2\int_{-\infty}^{C_m}\left\{\Phi\left(\tfrac{\sigma_m}{\tau}y+\sqrt{1+\tfrac{\sigma_m^2}{\tau^2}}z_{\alpha_m/2}\right)-\Phi\left(\tfrac{\sigma_m}{\tau}y\right)\right\}d\Phi(y)$$

where $C_m = C\tau/\sqrt{\sigma_m^2+\tau^2}$*, and BCP converges to the coverage probability of the corresponding z-based IE as C goes to infinity.*

Proof. For the derivation, it is better to use the posterior and marginal distributions:

$$\mu_m|\bar{X}_m \sim N\left(\tfrac{\tau^2}{\tau^2+\sigma_m^2}\bar{X}_m+\tfrac{\sigma_m^2}{\tau^2+\sigma_m^2}\eta,\tfrac{\tau^2\sigma_m^2}{\tau^2+\sigma_m^2}\right)\ \ \&\ \ \bar{X}_m\sim N\left(\eta,\sigma_m^2+\tau^2\right).$$

Then, the Bayes coverage probability becomes:

$$\begin{aligned}
&\int_{\Theta_m} P_{\mu_m}[\mu_m\in\Gamma_m(\boldsymbol{X}_m;\alpha_m)]d\Pi(\mu_m)\\
=&E_{\bar{X}_m}E_{\mu_m|\bar{X}_m}I\left\{LB_m\le\mu_m\le UB_m\right\}\\
=&E_{\bar{X}_m}P_{\mu_m|\bar{X}_m}\left[\mu_m<\bar{X}_m+z_{\alpha_m/2}\sigma_m I\left\{\bar{X}_m-\eta<C\tau\right\}\right]\\
-&E_{\bar{X}_m}P_{\mu_m|\bar{X}_m}\left[\mu_m<\bar{X}_m-z_{\alpha_m/2}\sigma_m I\left\{\bar{X}_m-\eta>-C\tau\right\}\right]\\
=&E_{Z'_m}\Phi\left(\tfrac{\sigma_m}{\tau}Z'_m+\tfrac{\sqrt{\tau_m^2+\sigma_m^2}}{\tau}z_{\alpha_m/2}I\left\{Z'_m<\tfrac{C\tau}{\sqrt{\tau^2+\sigma_m^2}}\right\}\right)\\
-&E_{Z'_m}\Phi\left(\tfrac{\sigma_m}{\tau}Z'_m-\tfrac{\sqrt{\tau_m^2+\sigma_m^2}}{\tau}z_{\alpha_m/2}I\left\{Z'_m>\tfrac{-C\tau}{\sqrt{\tau^2+\sigma_m^2}}\right\}\right)\\
=&\int_{-C_m}^{C_m}\left\{\Phi\left(\tfrac{\sigma_m}{\tau}z'_m+\tfrac{\sqrt{\tau_m^2+\sigma_m^2}}{\tau}z_{\alpha_m/2}\right)-\Phi\left(\tfrac{\sigma_m}{\tau}z'_m-\tfrac{\sqrt{\tau_m^2+\sigma_m^2}}{\tau}z_{\alpha_m/2}\right)\right\}\phi(z'_m)dz'_m\\
+&\int_{C_m}^{\infty}\left\{\Phi\left(\tfrac{\sigma_m}{\tau}z'_m\right)-\Phi\left(\tfrac{\sigma_m}{\tau}z'_m-\tfrac{\sqrt{\tau_m^2+\sigma_m^2}}{\tau}z_{\alpha_m/2}\right)\right\}\phi(z'_m)dz'_m\\
+&\int_{-\infty}^{-C_m}\left\{\Phi\left(\tfrac{\sigma_m}{\tau}z'_m+\tfrac{\sqrt{\tau_m^2+\sigma_m^2}}{\tau}z_{\alpha_m/2}\right)-\Phi\left(\tfrac{\sigma_m}{\tau}z'_m\right)\right\}\phi(z'_m)dz'_m\\
=&2\int_{-\infty}^{C_m}\left\{\Phi\left(\tfrac{\sigma_m}{\tau}z'_m+\tfrac{\sqrt{\tau_m^2+\sigma_m^2}}{\tau}z_{\alpha_m/2}\right)-\Phi\left(\tfrac{\sigma_m}{\tau}z'_m\right)\right\}\phi(z'_m)dz'_m
\end{aligned}$$

where $C_m = \frac{C\tau}{\sqrt{\tau^2+\sigma_m^2}}$. Although is has no closed form, it is a smooth increasing function with respect to C and approaches the nominal coverage probability of the classical z-based IE, $1-\alpha_m$. That is,

$$\lim_{C\to\infty}BCP[C,\alpha_m;\sigma_m,\tau] = 2\int_{-\infty}^{\infty}\left\{\Phi\left(\frac{y+\frac{\sqrt{\tau^2+\sigma_m^2}}{\sigma_m}z_{\alpha_m/2}}{\tau/\sigma_m}\right)-\Phi\left(\tfrac{y-0}{\tau/\sigma_m}\right)\right\}\phi(y)dy$$

$$=2\left\{\Phi\left(\Phi^{-1}\left(\alpha_m/2\right)\right)-\tfrac{1}{2}\right\}=1-\alpha_m$$

□

Proposition A.3. *The Bayes threshold ratio (BTR) has the form of*

$$BTR[C;\boldsymbol{\sigma},\tau]=\tfrac{2}{M}\sum_{i=1}^{M}\Phi(-C_m)$$

where $C_m=C\tau/\sqrt{\sigma_m^2+\tau^2}$*. As* C *increases from zero to infinity, the BTR decreases from 1 to 0.*

Proof. Note that $\bar{X}_m$ marginally follows $\mathcal{N}(\eta,\sigma_m^2+\tau^2)$. Then,

$$\begin{aligned}
&E_{\bar{X}_m}\left[\frac{1}{M}\sum_{i=1}^{M}\left[I\{\bar{X}_m>\eta+C\tau\}+I\{\bar{X}_m<\eta-C\tau\}\right]\right]\\
=&\frac{1}{M}\sum_{i=1}^{M}\left[P[\bar{X}_m>\eta+C\tau]+P[\bar{X}_m<\eta-C\tau]\right]\\
=&\frac{1}{M}\sum_{i=1}^{M}\left[P\left[\frac{\bar{X}_m-\eta}{\sqrt{\sigma_m^2+\tau^2}}>\frac{C\tau}{\sqrt{\sigma_m^2+\tau^2}}\right]+P\left[\frac{\bar{X}_m-\eta}{\sqrt{\sigma_m^2+\tau^2}}<\frac{-C\tau}{\sqrt{\sigma_m^2+\tau^2}}\right]\right]\\
=&\frac{1}{M}\sum_{i=1}^{M}\left[1-\Phi\left(\frac{C\tau}{\sqrt{\sigma_m^2+\tau^2}}\right)+\Phi\left(\frac{-C\tau}{\sqrt{\sigma_m^2+\tau^2}}\right)\right]\\
=&\frac{2}{M}\sum_{i=1}^{M}\Phi\left(\frac{-C\tau}{\sqrt{\sigma_m^2+\tau^2}}\right)=\frac{2}{M}\sum_{i=1}^{M}\Phi(-C_m)
\end{aligned}$$

where $C_m=C\tau/\sqrt{\sigma_m^2+\tau^2}$. □

Corollary A.4. *Given* M *parameters and a* $1-q$ *global level, the Bayes relative expected length (BREL), the Bayes family-wise coverage rate (BFWCR), and the Bayes thresholding ratio (BTW) have the following forms:*

$$\begin{aligned}
BREL[C,\boldsymbol{\alpha};\boldsymbol{\sigma},\tau]=&\frac{1}{M}\sum_{i=1}^{M}\left[2z_{\alpha_S}\sigma_m\Phi(C_m)\right]\Big/\frac{1}{M}\sum_{i=1}^{M}\left[2z_{\alpha_S}\sigma_m\right];\\
BFWCR[C,\boldsymbol{\alpha};\boldsymbol{\sigma},\tau]=&\prod_{i=1}^{M}\left[2\int_{-\infty}^{C_m}\left\{\Phi\left(\frac{\sigma_m}{\tau}y+\sqrt{1+\tfrac{\sigma_m^2}{\tau^2}}z_{\alpha_S}\right)-\Phi\left(\tfrac{\sigma_m}{\tau}y\right)\right\}d\Phi(y)\right];\\
BTR[C;\boldsymbol{\sigma},\tau]=&\frac{2}{M}\sum_{i=1}^{M}\Phi(-C_m)
\end{aligned}$$

where $C_m=C\tau/\sqrt{\sigma_m^2+\tau^2}$*, and* α_S *is the Sidak adjusted level,* $1-(1-q)^{1/M}$*.*

Proof. From (Proposition 3.4) and (Definition 3.3), we can obtain the BAELs of BMIE Thres and z-based MIE, respectively. Then $BREL[C, \boldsymbol{\alpha}; \boldsymbol{\sigma}, \tau]$ is the ratio of these two BAELs. In addition, the $BFWCR[C, \boldsymbol{\alpha}; \boldsymbol{\sigma}, \tau]$ can be obtained from (Proposition 3.5) and (Definition 3.2) by multiplying M individual BIEs. Lastly, $BTR[C; \boldsymbol{\sigma}, \tau]$ is an immediate result from (Proposition 3.6). □

Proposition A.5. *Suppose $\delta \in \mathcal{D}$ is an equivariant decision function. Then the following equality holds trus:*

$$R(\theta, \delta) = R(\bar{g}(\theta), \delta)$$

for every $\theta \in \Theta$ and $g \in \mathcal{G}$.

Proof.

$$\begin{aligned} LHS =& E_\theta[L(\theta, \delta(X))] \\ =& E_\theta[L(\bar{g}(\theta), \tilde{g}(\delta(X)))] \\ =& E_\theta[L(\bar{g}(\theta), \delta(g(X)))] \\ =& E_{\bar{g}(\theta)}[L(\bar{g}(\theta), \delta(X))] = RHS \end{aligned}$$

Note that the second equality is due the invariant loss function and the third equality is because δ is equivariant. Lastly, the fourth equality is because of the invariance on the family of probability distributions. □

Lemma A.6. *Suppose $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\Theta}$, $\mathcal{G}$, and $\bar{\mathcal{G}}$ are all isomorphic. Then for an equivariant decision rule $\delta(x) = \tilde{x}(\delta(e))$,*

$$E^{\pi^R(\theta|x)}[L(\theta, \tilde{x}(\delta(e)))] = R(\theta, \delta) = \int_{\boldsymbol{\mathcal{X}}} L(\bar{e}, \tilde{y}(\delta(e)))f(y|\bar{e})dy$$

where $\pi^R(\theta|x)$ is the posterior distribution of θ, given x, with respect to the right invariant Haar prior density.

Proof. We prove the theorem based on the following three small claims and prove these claims later.

1. $f(x|\bar{g}) = f(g^{-1}(x)|\bar{e})J^L_{g^{-1}}(x)$

2. There exists $K \in \mathbb{R}$ such that for any integrable function t

   $\int t(x^{-1})h^R(x)dx = K\int t(y)h^L(y)dy$

3. There exists $\nu(g)$ such that for any integrable function t

   $\int t(yg^{-1})h^L(y)dy = \nu(g)\int t(y)h^L(y)dy$

$$\begin{aligned}
\int L(\theta, \tilde{x}(\delta(e)))f(x|\theta)h^R(\theta)d\theta &\overset{2}{=} K\int L(\bar{y}^{-1}, \tilde{x}(\delta(e)))f(x|\bar{y}^{-1})h^L(y)dy \\
&= K\int L(\bar{e}, \tilde{y}\tilde{x}(\delta(e)))f(x|\bar{y}^{-1})h^L(y)dy \\
&\overset{1}{=} K\int L(\bar{e}, \tilde{yx}(\delta(e)))f(yx|\bar{y}^{-1})J^L_y(x)h^L(y)dy \\
&= K\int L(\bar{e}, \tilde{yx}(\delta(e)))f(yx|\bar{y}^{-1})\frac{h^L(x)}{h^L(yx)}h^L(y)dy \\
&\overset{3}{=} Kh^L(x)\nu(x)\int L(\bar{e}, \tilde{yx}(\delta(e)))f(yx|\bar{y}^{-1})dy
\end{aligned}$$

Note that the second equality holds due to the invariant loss function and the fourth equality holds because of the form of the left Haar density. Now set the loss function to be 1 in the first and last terms of the equations. Then, the integral with the density on the right-hand side becomes zeto, so that

$$\int f(x|\theta)H^R(\theta)d\theta = Kh^L(x)\nu(x).$$

Therefore,

$$\Pi^R(\theta|x) = \frac{f(x|\theta)h^R(\theta)}{\int f(x|\theta)h^R(\theta)d\theta} = \frac{f(x|\theta)h^R(\theta)}{Kh^L(x)\nu(x)}$$

But then we verify the first and the third terms are equivalent by moving $Kh^L(x)\nu(x)$ to the other side. Furthermore, the second term is obtained immediately from the third term becuase of transitivity.

proof of 1. Note the following equation holds due to the invariant family of probability distributions.

$$\int_{g^{-1}A} f(x|\theta)dx = P_\theta[X \in g^{-1}A] = P_\theta[gX \in A] = P_{\bar{g}(\theta)}[X \in A] = \int_A f(x|\bar{g}(\theta))dx$$

Then for any $A \in \mathcal{F}$, by the change of variable with $g^{-1}y = x$,

$$LHS = \int_A f(g^{-1}y|\theta) J^L_{g^{-1}}(y)dy = RHS.$$

Thus, $f(g^{-1}y|\theta) = f(y|\bar{g}(\theta))$ for any y, θ, and g. Then the result is immediate by choosing $\theta = \bar{e}$.

proof of 2. Consider $\int_A h^R(x)dx = \int_{Ag} h^R(x)dx$. Set up a change of variable with $x = y^{-1}$.

Then

$$\int_{A^{-1}} h^R(y^{-1})J(y)dy = \int_{g^{-1}A^{-1}} h^R(y^{-1})J(y)dy.$$

Observe that $h^R(y^{-1})J(y)$ is a left Haar density. Thus, $h^R(y^{-1})J(y) = Kh^L(y)$ because the Haar density is unique up to constants. Thus,

$$\int t(x^{-1})h^R(x)dx = \int t(y)h^R(y^{-1})J(y)dy = \int t(y)Kh^L(y)dy.$$

proof of 3. Let $H^L_g(A) = \int_{Ag} h^L(y)dy$. Then it is a left Haar measure because $H^L_g(g_0A) = H^L_g(A)$. Then $H^L_g(A) = \nu(g)H^L(A)$ because the Haar measure is unique up to constants. Therefore, the following holds for any $A \in \mathcal{F}$

$$LHS = \int_{Ag} h^L(y)dy = \int_A h^L(xg)J^R(x)dx;$$
$$RHS = \nu(g) \int_A h^L(y)dy.$$

Therefore, $h^L(xg)J^R_g(x) = \nu(g)H^L(x)$ for all x and g. Lastly,

$$\int t(yg^{-1})h^L(y)dy = \int t(xgg^{-1})h^L(xg)J^R_g(x)dx = \int t(x)\nu(g)h^L(x)dx.$$

Note that this proof is based on a result in Berger [3]. Slightly different approach is provided in Schervish [49]. □

Proposition A.7. *Under the same assumptions, suppose* $\Gamma(\cdot)$ *is an equivariant interval estimator. Then,*

$$P^{\pi^R(\theta|e)}[\theta \in \Gamma(e)] = P^{\pi^R(\theta|x)}[\theta \in \Gamma(x)] = P_\theta[\theta \in \Gamma(X)] = P_{\bar{e}}[\bar{e} \in \Gamma(X)]$$

Proof. The result is immediate from (Lemma 4.13) with the following loss function, $L(\theta, \Gamma(x)) = I\{\theta \in \Gamma(a)\}$. □

Proposition A.8. *Suppose* $\Gamma(\cdot)$ *is equivariant and* $E[\nu^L(\Gamma(X))]$ *is the expected content defined in (Definition 4.14). Then, the following arguments hold true:*

$$E[\nu^L(\Gamma(X))] = \int_{\boldsymbol{\Theta}} Pr_{\bar{e}}[\theta \in \Gamma(X)]\pi^L(\theta)d\theta = \nu^L(\Gamma(e))$$

Proof. The first equality is related to the false coverage probability used for deriving the uniformly most accurate interval estimator (Lehmann and Romano [37]).

$$\begin{aligned}
E[\nu^L(\Gamma(X))] &= \int_{\boldsymbol{\mathcal{X}}} \int_{\Gamma(x)} \pi^L(\theta')d\theta' f(x|\theta)dx \\
&= \int_{\boldsymbol{\mathcal{X}}} \int_{\boldsymbol{\Theta}} I\{\theta' \in \Gamma(x)\}\pi^L(\theta')d\theta' f(x|\theta)dx \\
&= \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{\mathcal{X}}} I\{\theta' \in \Gamma(x)\}f(x|\theta)dx\pi^L(\theta')d\theta' \\
&= \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{\mathcal{X}}} I\{\theta' \in \Gamma(x)\}f(x|\theta)dx\pi^L(\theta')d\theta' \\
&= \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{\mathcal{X}}} I\{\theta' \in \Gamma(x')\}f(x'|\bar{e})dx'\pi^L(\theta')d\theta' = \int_{\boldsymbol{\Theta}} Pr_{\bar{e}}[\theta \in \Gamma(X)]
\end{aligned}$$

The second equality follows straightforwardly from the property of the left Haar measure.

$$\begin{aligned}
E[\nu^L(\Gamma(X))] &= \int_{\boldsymbol{\mathcal{X}}} \int_{\Gamma(x)} \pi^L(\theta')d\theta' f(x|\theta)dx \\
&= \int_{\boldsymbol{\mathcal{X}}} \int_{\bar{x}\Gamma(e)} \pi^L(\theta')d\theta' f(x|\theta)dx \\
&= \int_{\boldsymbol{\mathcal{X}}} \int_{\Gamma(e)} \pi^L(\theta')d\theta' f(x|\theta)dx \\
&= \int_{\Gamma(e)} \pi^L(\theta')d\theta' \int_{\boldsymbol{\mathcal{X}}} f(x|\theta)dx = \nu^L(\Gamma(e))
\end{aligned}$$

□

Proposition A.9. *The best equivariant interval estimator (BEIE)* $\Gamma^\star$ *satisfies* $\Gamma^\star(x) = \tilde{x}\Gamma^\star(e)$ *where*

$$\Gamma^\star(e) = \{\theta \in \Theta : \pi^R(\theta|e) \geq C\pi^L(\theta)\}$$

with C chosen in order that

$$P^{\pi^R(\theta|e)}[\theta \in \Gamma^\star(e)] = 1 - \alpha$$

Proof. First, define $O^*(x) = I\{\pi^R(x|e) \geq C\pi^L(x)\}$. In addition, we let $O(x)$ to be an arbitrary indicator function which satisfy the condition $\int O(x)\pi^R(x|e)dx \geq 1 - \alpha$. Note that $O^*(x) = \begin{cases} 1 \text{ if } \pi^R(x|e) \geq C\pi^L(x) \\ 0 \text{ if } \pi^R(x|e) < C\pi^L(x) \end{cases}$. Therefore, the following inequality always holds true:

$$(O^*(x) - O(x))(\pi^R(x|e) - C\pi^L(x)) \geq 0$$

Integrating both sides,

$$\begin{aligned} &\int (O^*(x) - O(x))(\pi^R(x|e) - C\pi^L(x))dx \geq 0 \\ \Leftrightarrow &\int (O^*(x) - O(x))\pi^R(x|e)dx \geq C \int (O^*(x) - O(x))\pi^L(x)dx. \end{aligned}$$

However, by construction, the left hand side is non-positive, implying the right hand side is also non-positive. Therefore,

$$\int O^*(x)\pi^L(x)dx \leq \int O(x)\pi^L(x)dx.$$

Since $O(x)$ has been chosen arbitrary, this implies $\Gamma^\star(e)$ minimizes the expected content defined in (Definition 4.14). Note that this proof is based on (Theorem 3) in Peña and Kim [43]. □

Appendix B

Solving Optimization Problem

In this appendix, we provide the derivation of the Lagrange equation and corresponding Newton-Raphson method to solve the optimization problem in chapter 2.

B.1 Application: Mean of Normal Distribution

$$\begin{aligned} &\text{Minimize } \frac{1}{M}\sum_{m=1}^{M}\frac{2\nu_m\sigma_m}{\beta+2\nu_m\sigma_m} \\ &\text{subject to } \sum_{m=1}^{M}\log(2\Phi(\nu_m)-1)\geq\log(1-q) \end{aligned}$$

Now, we set up a Lagrange equation:

$$J(\underline{\nu},\lambda)=\sum_{m=1}^{M}\frac{2\nu_m\sigma_m}{\beta+2\nu_m\sigma_m}+\lambda\left\{\sum_{m=1}^{M}\log(2\Phi(\nu_m)-1)-\log(1-q)\right\}$$

The corresponding partial derivatives are:

$$\begin{aligned} \frac{\partial}{\partial\nu_m}J(\underline{\nu},\lambda)=&\frac{2\beta\sigma_m}{(\beta+2\nu_m\sigma_m)^2}+\lambda\frac{2\phi(\nu_m)}{2\Phi(\nu_m)-1}; \\ \frac{\partial}{\partial\lambda}J(\underline{\nu},\lambda)=&\sum_{m=1}^{M}\log(2\Phi(\nu_m)-1)-\log(1-q). \end{aligned}$$

Setting the first partial derivatives to zeros, we obtain m equations by taking log of both sides. In addition, by averaging those m equations and substituting one of the terms using the second partial derivative, we can obtain one more equation.

$$\log(2\Phi(\nu_m)-1)+\log(2\beta\sigma_m)=\log(-\lambda)+\log(2\phi(\nu_m))+2\log(\beta+2\nu_m\sigma_m)$$

$$\frac{\log(1-q)}{M}+\frac{\sum\log(2\beta\sigma_m)}{M}=\log(-\lambda)+\frac{\sum\log(2\phi(\nu_m))}{M}+\frac{2\sum\log(\beta+2\nu_m\sigma_m)}{M}$$

Subtracting one equation from the other equation, we obtain m equations as follows:

$$f_m^{\beta}(\boldsymbol{\nu}) = \log(2\Phi(\nu_m) - 1) + \log(\sigma_m) - \log(\phi(\nu_m)) - 2\log(\beta + 2\nu_m\sigma_m)$$
$$-\tfrac{1}{M}\sum(\log(\sigma_m) - \log(\phi(\nu_m)) - 2\log(\beta + 2\nu_m\sigma_m)) - \tfrac{1}{M}\log(1-q) = 0.$$

After we vertically merge those equations, $\boldsymbol{f^{\beta}(v)} = [f_1^{\beta}(\boldsymbol{v}), f_2^{\beta}(\boldsymbol{v}), \ldots, f_M^{\beta}(\boldsymbol{v})]^T$, we can solve this system of equations using Newton-Raphson method:

$$\begin{bmatrix} \\ \boldsymbol{v_{new}} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{v_{old}} \\ \\ \end{bmatrix} - \begin{bmatrix} \\ \frac{\partial}{\partial \boldsymbol{v}}\boldsymbol{f^{\beta}(v_{old})} \\ \\ \end{bmatrix}^{-1} \times \begin{bmatrix} \\ \boldsymbol{f^{\beta}(v_{old})} \\ \\ \end{bmatrix}$$

where the Jacobian consists of

$$\text{diagonal mth element: } \tfrac{\partial}{\partial\nu_m} f_m(\boldsymbol{v}) = \tfrac{2\phi(\nu_m)}{2\Phi(\nu_m)-1} + \tfrac{M-1}{M}(\nu_m - \tfrac{4\sigma_m}{\beta+2\nu_m\sigma_m})$$
$$\text{off-diagonal (m,n)th element: } \tfrac{\partial}{\partial\nu_n} f_m(\boldsymbol{v}) = -\tfrac{1}{M}(\nu_n - \tfrac{4\sigma_n}{\beta+2\nu_n\sigma_n})$$

B.2 Application: Variance of Normal Distribution

$$\text{Minimize } \frac{1}{M}\sum_{m=1}^{M}\frac{\log(\nu_m^U) - \log(\nu_m^L)}{\beta + \log(\nu_m^U) - \log(\nu_m^L)}$$
$$\text{subject to } \sum_{m=1}^{M}\log(F_{n_m}(\nu_m^U) - F_{n_m}(\nu_m^L)) \geq \log(1-q)$$

where $\nu_m^U = \chi^2_{\alpha_m/2,n_m}$ and $\nu_m^L = \chi^2_{1-\alpha_m/2,n_m}$. By managing the upper and lower quantiles separately, we have twice as many equations compared to the normal mean case. Then, the form of Lagrange multiplier becomes

$$J(\boldsymbol{\nu^U}, \boldsymbol{\nu^L}, \lambda) = \frac{1}{M}\sum_{m=1}^{M}\frac{\log(\nu_m^U) - \log(\nu_m^L)}{\beta + \log(\nu_m^U) - \log(\nu_m^L)} + \lambda\left\{\sum_{m=1}^{M}\log(F_{n_m}(\nu_m^U) - F_{n_m}(\nu_m^L)) - \log(1-q)\right\}$$

The corresponding partial derivatives are:

$$\frac{\partial}{\partial\nu_m^U}J = \frac{1}{M}\frac{\beta/\nu_m^U}{(\beta + \log(\nu_m^U) - \log(\nu_m^L))^2} + \lambda\frac{f_{n_m}(\nu_m^U)}{F_{n_m}(\nu_m^U) - F_{n_m}(\nu_m^L)}$$
$$\frac{\partial}{\partial\nu_m^L}J = -\frac{1}{M}\frac{\beta/\nu_m^L}{(\beta + \log(\nu_m^U) - \log(\nu_m^L))^2} - \lambda\frac{f_{n_m}(\nu_m^L)}{F_{n_m}(\nu_m^U) - F_{n_m}(\nu_m^L)}$$

$$\frac{\partial}{\partial\lambda}J = \sum_{m=1}^{M}\log(F_{n_m}(\nu_m^U) - F_{n_m}(\nu_m^L)) - \log(1-q)$$

The rest of the procedures are similar to the previous normal mean case except that we have $2M$ equations for the ν_m^U's and ν_m^L's.

B.3 Application: Median of Unknown Distribution

$$\begin{aligned}\text{Minimize}\quad & \frac{1}{M}\sum_{m=1}^{M}\frac{\left\{\hat{F}_h^{-1}(1-\alpha_m/2) - \hat{F}_h^{-1}(\alpha_m/2)\right\}\hat{\sigma}_m/\sqrt{n_m}}{\beta + \left\{\hat{F}_h^{-1}(1-\alpha_m/2) - \hat{F}_h^{-1}(\alpha_m/2)\right\}\hat{\sigma}_m/\sqrt{n_m}}\\ \text{subject to}\quad & \sum_{m=1}^{M}\log(1-\alpha_m) \geq \log(1-q)\end{aligned}$$

Similar to the normal case, we do reparametrizations as follows:

$$\nu_m^L = \hat{F}_h^{-1}(\alpha_m/2);\ \ \nu_m^U = \hat{F}_h^{-1}(1-\alpha_m/2));\ \ \sigma_m = \hat{\sigma}_m/\sqrt{n_m}.$$

Then, the problem becomes

$$\begin{aligned}\text{minimize}\quad & \frac{1}{M}\sum_{m=1}^{M}\frac{(\nu_m^U - \nu_m^L)\sigma_m}{\beta + (\nu_m^U - \nu_m^L)\sigma_m}\\ \text{subject to}\quad & \sum_{m=1}^{M}\log\left\{\hat{F}_h(\nu_m^U) - \hat{F}_h(\nu_m^L)\right\} \geq \log(1-q).\end{aligned}$$

By managing the upper and lower quantiles separately, $58 = 29\times 2$ unknown equations for the optimization. Then, the form of Lagrange multiplier becomes

$$J(\boldsymbol{\nu^U}, \boldsymbol{\nu^L}, \lambda) = \frac{1}{M}\sum_{m=1}^{M}\frac{(\nu_m^U - \nu_m^L)\sigma_m}{\beta + (\nu_m^U - \nu_m^L)\sigma_m} + \lambda\left\{\sum_{m=1}^{M}\log\{\hat{F}_h(\nu_m^U) - \hat{F}_h(\nu_m^L)\} - \log(1-q)\right\}$$

The corresponding partials are:

$$\begin{aligned}\frac{\partial}{\partial\nu_m^U}J &= \frac{\beta\sigma_m}{(\beta + (\nu_m^U - \nu_m^L)\sigma_m)^2} + \lambda\frac{2\hat{f}_h(\nu_m^U)}{\hat{F}_h(\nu_m^U) - \hat{F}_h(\nu_m^L)}\\ \frac{\partial}{\partial\nu_m^L}J &= -\frac{\beta\sigma_m}{(\beta + (\nu_m^U - \nu_m^L)\sigma_m)^2} - \lambda\frac{2\hat{f}_h(\nu_m^L)}{\hat{F}_h(\nu_m^U) - \hat{F}_h(\nu_m^L)}\\ \frac{\partial}{\partial\lambda}J &= \sum_{m=1}^{M}\log(\hat{F}_h(\nu_m^U) - \hat{F}_h(\nu_m^L)) - \log(1-q)\end{aligned}$$

The rest of the procedures are similar to the normal variance case.