

## *Review Article*

# **A Note on the Comparison of the Bayesian and Frequentist Approaches to Estimation**

**Andrew A. Neath and Natalie Langenfeld**

*Department of Mathematics and Statistics, Southern Illinois University Edwardsville,  
Edwardsville, IL 62026-1653, USA*

Correspondence should be addressed to Andrew A. Neath, [aneath@siue.edu](mailto:aneath@siue.edu)

Received 10 April 2012; Accepted 24 September 2012

Academic Editor: C. D. Lai

Copyright © 2012 A. A. Neath and N. Langenfeld. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Samaniego and Reneau presented a landmark study on the comparison of Bayesian and frequentist point estimators. Their findings indicate that Bayesian point estimators work well in more situations than were previously suspected. In particular, their comparison reveals how a Bayesian point estimator can improve upon a frequentist point estimator even in situations where sharp prior knowledge is not necessarily available. In the current paper, we show that similar results hold when comparing Bayesian and frequentist interval estimators. Furthermore, the development of an appropriate interval estimator comparison offers some further insight into the estimation problem.

## **1. Introduction**

Samaniego and Reneau [1], hereafter referred to as SR, presented a landmark study on the comparison of the Bayesian and frequentist approaches to point estimation. Traditionally, disagreements between the two schools of thought are philosophical in nature. Particular conflict exists as to whether the use of subjective information, quantified in the form of a prior distribution, is scientifically appropriate or not. The work by SR is set apart by its practicality. Point estimates simply are judged by their closeness to the truth. In comparing two estimates, the better estimate is the one that is closest to the desired target. All philosophical arguments are pushed aside for the sake of the comparison.

In the current note, we look back at the SR paper, offering our own views and comments. A grander retrospective of the comparison between the Bayesian and frequentist approaches to point estimation is provided in Samaniego [2]. We will then push the study onward into a comparison of Bayesian and frequentist approaches to the problem of interval estimation. A general theme of Samaniego's work on comparisons between point estimators

is how the Bayesian has a much greater opportunity for improvement on a frequentist estimator than had been previously established. As in the point estimation problem, we show in a comparison of interval estimators that the Bayesian has a generous opportunity for improvement on a frequentist. In the light of Samaniego's history of work, the results we achieve are not unexpected. Nevertheless, the development of an extension of point estimator comparisons to the problem of interval estimator comparisons offers some further insight into the estimation problem.

Since the publication of the SR paper in 1994, several authors have instigated a comparison between the Bayesian and frequentist approaches to estimation. Barnett [3] looks at the conceptual and methodological differences between Bayesian and frequentist methods. Robert [4] uses decision theory to argue the superiority of the Bayesian viewpoint. Berger [5] writes on how objective rather than subjective Bayesian analysis is the appropriate tactic in a scientific undertaking. Samaniego and Neath [6] consider a comparison of estimators in an empirical Bayes framework, leading to the conclusion that the use of prior information, no matter how diffuse, is beneficial. Vestrup and Samaniego [7] compare Bayesian and frequentist shrinkage estimators in a multivariate problem.

## 2. Comparison of Point Estimators

This section provides a review of SR's work on the comparison of Bayesian and frequentist point estimators. We follow the presentation from Samaniego [2], but with slight changes in notation as seen fit. The reader interested in more detailed arguments supporting the rules for comparison should refer to the original sources.

Let  $\hat{\theta}_F$  denote the frequentist entry into the competition. The framework for the study is one for which the "best frequentist estimator" is unambiguous, such as when  $\hat{\theta}_F$  represents a sample mean, a sample proportion, or is unbiased and a function of a complete, sufficient statistic. Let  $\hat{\theta}_B = a\hat{\theta}_F + (1-a)\theta_o$  denote the Bayesian estimator. The framework for the study is one for which the Bayesian estimator is a weighted average of the frequentist estimator  $\hat{\theta}_F$  and a prior estimate  $\theta_o$ , where  $a$  defines the prior weight on the data information. It is required that  $0 \leq a < 1$ , so the Bayesian entry in the competition is distinguished from the frequentist by the use of at least some prior information ( $a < 1$ ), although we do allow for the Bayesian to use no information at all from the data ( $a = 0$ ). In this framework, the choice of a prior reduces to the choice of the pair  $(\theta_o, a)$ .

The estimators  $\hat{\theta}_F$  and  $\hat{\theta}_B$  will be compared under decision theoretic principles. Consider a squared error loss function

$$L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2. \quad (2.1)$$

The corresponding risk function is the mean squared error of estimation, a reasonable criterion for judging estimators. Let  $R(\hat{\theta}_F, \theta)$  and  $R(\hat{\theta}_B, \theta)$  denote the risk functions for the frequentist estimator and Bayesian estimator, respectively. It is easy to derive algebraic forms. Define  $\sigma^2 = \text{Var}(\hat{\theta}_F)$ . Then  $R(\hat{\theta}_F, \theta) = \sigma^2$  and  $R(\hat{\theta}_B, \theta) = a^2\sigma^2 + (1-a)^2(\theta - \theta_o)^2$ . Figure 1 displays a graph of the two risk functions in a representative case. We took  $\sigma^2 = 5$ ,  $\theta_o = 15$ , and  $a = 5/6$  in creating the display.

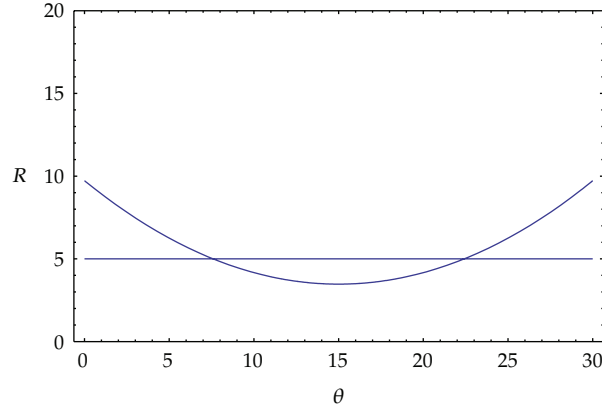


Figure 1: Bayes and frequentist risk functions.

Let  $\theta_*$  denote the true value of the parameter. The preferred estimator is the one with the smallest risk function at the true value  $\theta_*$ . The graph in Figure 1 indicates a first thought: a good prior estimate ( $\theta_*$  is close to  $\theta_o$ ) leads to a situation where the Bayesian estimator  $\hat{\theta}_B$  is better, while a poor prior estimate ( $\theta_*$  is away from  $\theta_o$ ) leads to a situation where  $\hat{\theta}_F$  is better. Consider the following quote from Diaconis and Freedman [8]: “A statistician who has *sharp* prior knowledge of these parameters should use it.”

The above thought process derails our journey towards an answer to the question of which approach is better in an estimation problem. The above only tells us that  $\hat{\theta}_B$  is better when the truth  $\theta_*$  is close to prior estimate  $\theta_o$ . What do we know about whether the truth is close to the prior estimate or not? The Bayesian believes what is specified in the prior distribution. The frequentist, however, is not trusting of such information. The two sides then retreat back to their respective camps and the issue remains unresolved. The SR approach to comparing point estimators is unique in how it turns the problem around. Rather than thinking about how close the truth is to a prior parameterization, instead we think about which prior specifications  $(\theta_o, a)$  lead to a Bayesian estimator which is superior to the frequentist estimator with respect to a truth  $\theta_*$ . Specifically, the problem posed by SR is to determine which choices of  $(\theta_o, a)$  lead to a Bayesian estimator  $\hat{\theta}_B$  which outperforms the frequentist estimator  $\hat{\theta}_F$  as judged by the risk function  $R(\hat{\theta}, \theta_*)$ . One result in particular stands out.

**Theorem 2.1.** Let  $\Delta = (\theta_o - \theta_*)^2 / \sigma^2$  denote the scaled, squared distance between the prior estimate and the true value. Then  $R(\hat{\theta}_B, \theta_*) < R(\hat{\theta}_F, \theta_*)$  if and only if

$$\Delta < \frac{1+a}{1-a}. \quad (2.2)$$

Figure 2 presents the threshold separating the regions of superiority as a graph in  $(a, \Delta)$ . The importance of the inequality in (2.2) is demonstrated in the following arguments.

(1) Since  $a \geq 0$ , then  $(1+a)/(1-a) \geq 1$ . So regardless of the choice of the weight  $a$ , any choice of a prior estimate  $\theta_o$  such that  $\Delta < 1$  leads to a Bayesian estimator which beats the frequentist estimator. Theorem 2.1 then quantifies what is reasonably believed: one should be a Bayesian when one is able to provide a good prior estimate of the true value.

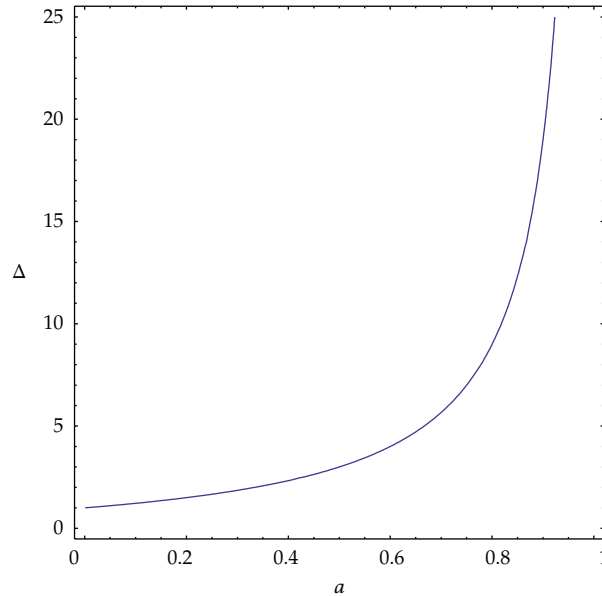


Figure 2: Threshold for point estimator comparison.

(2) As  $a \rightarrow 1$ , the ratio  $(1+a)/(1-a) \rightarrow \infty$ . So regardless of the accuracy of the prior estimate as measured by  $\Delta$ , there exists a lower bound  $A$  where any weight  $a$  such that  $A < a < 1$  leads to a Bayesian estimator which beats the frequentist rule. Theorem 2.1 provides

then a truly surprising result. No matter how bad a prior estimate one provides, there exists a choice of prior weight for which the Bayesian estimator improves upon the frequentist estimator. For example, the point  $(a = .8, \Delta = 9)$  lies on the threshold separating Bayesian and frequentist superiority, so even a prior estimate 3 standard deviations away from the truth  $\theta_*$  ( $\Delta = 9$ ) leads to a Bayesian estimator which improves upon the frequentist estimator when the data weight  $a$  is greater than  $A = .8$ .

In case 1, the Bayesian can beat the frequentist by providing an appropriate assessment on the true value of the parameter. In case 2, the Bayesian can beat the frequentist by providing an appropriate assessment on the weight of the prior belief. The findings of SR indicate that Bayesian point estimators work well in more situations than were previously suspected. Unless one is both misguided (poor prior estimate) *and* stubborn (undue weight on the prior estimate), the Bayesian point estimator outperforms the frequentist point estimator.

### 3. Some Comments and Extensions

To gain a better understanding of the competition between Bayesian and frequentist point estimators, and to prepare for an extension of the study to a competition between interval estimators, we define a sampling distribution on the estimator  $\hat{\theta}_F$  as

$$\hat{\theta}_F | \theta \sim N(\theta, \sigma^2). \quad (3.1)$$

Such an assumption aids in mathematical tractability and may be justified based on asymptotic arguments. Suppose the Bayesian's prior information can be modeled using the conjugate prior  $\theta \sim N(\theta_o, \sigma_o^2)$ . The posterior distribution on  $\theta$  based on observing  $\hat{\theta}_F$  can be derived to be

$$\theta \mid \hat{\theta}_F \sim N(\hat{\theta}_B, a\sigma^2) \quad (3.2)$$

by reparameterizing so that  $a = \sigma_o^2 / (\sigma^2 + \sigma_o^2)$ .

Let us explore further the decision theory framework used in Section 2 for comparing point estimators. The posterior risk for a decision  $\delta$  under squared error loss is

$$r(\delta) = E_{\theta \mid \hat{\theta}_F}(\delta - \theta)^2, \quad (3.3)$$

where  $E_{\theta \mid \hat{\theta}_F}$  denotes the expectation with respect to the conditional distribution in (3.2). The Bayes rule found by minimizing  $r(\delta)$  is the conditional expectation  $E(\theta \mid \hat{\theta}_F) = \hat{\theta}_B$ . So, under the decision theoretic framework of the competition,  $\hat{\theta}_B$  is the "best" decision rule when prior information  $(\theta_o, a)$  is used. In a similar fashion, we would like to think of  $\hat{\theta}_F$  as the "best" decision rule when no prior information is used. The Bayes rule when full weight is given to the data ( $a = 1$ ) is indeed the frequentist estimator  $\hat{\theta}_F$ . Both the Bayesian and the frequentist are putting up their best competitors in the point estimation comparison. We will require that the same be true in the interval estimation comparison.

The comparison between point estimators, as set forth in SR, defines a Bayesian estimator as one that necessarily uses some prior information. From this definition, estimators which some may refer to as noninformative Bayesian estimators fall under the frequentist umbrella. If one prefers, the comparison described in Section 2 may be thought of as a competition between an informative Bayesian and a noninformative Bayesian, where we are able to determine when one benefits from the use of prior information.

The frequentist estimator  $\hat{\theta}_F$  can be defined as a limit of Bayes rules as  $a \rightarrow 1$ . Thus,  $\hat{\theta}_F$  is a minimax estimator under the distributions in (3.1) and (3.2). The comparison between point estimators may then also be thought of as a competition between a proper Bayes rule and a minimax rule. Following this line of reasoning, the comparison is able to determine when one benefits by taking an approach other than protection against a worst case scenario.

Finally, if one prefers, the comparison may be looked at as a decision between an unbiased estimator, represented by  $\hat{\theta}_F$ , and a shrinkage estimator, represented by  $\hat{\theta}_B = a\hat{\theta}_F + (1 - a)\theta_o$ . In this light, the comparison is able to determine when the reduction in variance associated with a shrinkage estimator is enough to offset the increase in bias.

#### 4. Development of a Risk Function for Comparing Interval Estimators

We now extend the results from Section 2 into a comparison between Bayesian and frequentist interval estimators. The entries into the competition will be the interval estimators considered most standard in the statistical literature; namely, 95% confidence intervals developed within the respective paradigms. Denote an interval estimator as  $\delta = [\delta_L, \delta_U]$ . From the sampling distribution in (3.1), the frequentist entry into the interval estimation competition is

$\hat{\theta}_F \pm 1.96\sigma$ , which we will denote as  $\delta_F$ . From the posterior distribution in (3.2), the Bayesian interval estimator is  $\hat{\theta}_B \pm 1.96\sigma\sqrt{a}$ , which we will denote as  $\delta_B$ . The Bayesian interval estimator features not only a shift in the midpoint from  $\hat{\theta}_F$  to  $\hat{\theta}_B = a\hat{\theta}_F + (1-a)\theta_o$  but also a reduction in the interval half length from  $1.96\sigma$  to  $1.96\sigma\sqrt{a}$ . The weight on the data can vary over  $0 \leq a < 1$ , so conceivably  $\delta_B$  could correspond to a single point when no data weight is employed.

The comparison between Bayesian and frequentist point estimators has a squared error loss function at its core. In this section, we consider the problem of selecting an appropriate loss function for judging the interval estimators. Per the discussion in Section 3, we are looking to determine a loss function for which  $\delta_B$  is the Bayes rule with respect to a proper prior distribution ( $0 \leq a < 1$ ), while  $\delta_F$  is the Bayes rule with respect to a limit of proper priors ( $a \rightarrow 1$ ), or equivalently an improper prior placing full weight on the data ( $a = 1$ ). In this way, the competition will be judged under a loss function for which  $\delta_B$  and  $\delta_F$  are the "best" approaches put forth by their respective camps.

A good starting point for choosing an interval estimate loss function is the form

$$L(\delta, \theta) = c(\delta_L - \theta) + c(\theta - \delta_U) + c_o \times [\delta_U - \delta_L], \quad (4.1)$$

where  $c(t)$  is a nondecreasing function with  $c(t) = 0$  for  $t \leq 0$ . The function  $c(t)$  defines a cost for an incorrect interval. A length penalty is dictated by the constant  $c_o$  multiplied by the length of the interval. The loss function in (4.1) mimics the thought process behind the development of an interval estimate. Of course, a goal is to have an interval that covers the true parameter; failure to do so results in a cost. In contrast, a goal is for an interval with small length, so the cost of an interval increases with its length.

Consider the cost function  $c(t) = I\{t > 0\}$ , where  $I(A)$  denotes the indicator function on set  $A$ . A penalty of 1 cost unit is placed on an incorrect interval. For easy reference, we will name (4.1) under this cost function as the 0-1 loss function. The corresponding risk function, for an interval estimator of the form  $\hat{\theta} \pm m$ , can be derived to be

$$R(\hat{\theta}, m; \theta) = P_{\hat{\theta}|\theta}(\hat{\theta} - m > \theta) + P_{\hat{\theta}|\theta}(\hat{\theta} + m < \theta) + c_o \times [2m]. \quad (4.2)$$

The risk function in (4.2) seems to provide a reasonable means for judging the competing interval estimators, as the risk function returns the probability of an incorrect interval plus a cost for the length of the interval. However, let us look further to see if the condition requiring that  $\delta_B$  and  $\delta_F$  are Bayes rules can be met. For the frequentist interval  $\delta_F : \hat{\theta}_F \pm 1.96\sigma$ , we get  $R(\hat{\theta}_F, m_F; \theta) = .05 + c_o \times 2(1.96)$  where  $c_o$  is to be chosen so that  $\delta_F$  is the Bayes rule under the noninformative prior. Let  $g$  denote the probability density function for the posterior distribution on  $\theta$ . It can be shown that Bayes rule with respect to a 0-1 loss function is the interval  $\delta = [\delta_L, \delta_U]$  found by solving  $g(\delta_L) = g(\delta_U) = c_o$ . So for the posterior based on the noninformative prior ((3.2) with  $a = 1$ ), Bayes rule is found by solving

$$\begin{aligned} \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left\{\frac{-(\delta_L - \hat{\theta}_F)^2}{2\sigma^2}\right\} &= c_o, \\ \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left\{\frac{-(\delta_U - \hat{\theta}_F)^2}{2\sigma^2}\right\} &= c_o. \end{aligned} \quad (4.3)$$

Thus, the frequentist interval  $\delta_F$  is the Bayes rule under the 0-1 loss function with respect to the noninformative prior when  $c_o$  is taken to be

$$c_o = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(1.96)^2}{2\sigma^2}\right\}. \quad (4.4)$$

For the risk function in (4.2) to serve appropriately as a judge between the Bayesian and frequentist interval estimators, the Bayesian interval  $\delta_B$  must be the Bayes rule under the 0-1 loss function with respect to a proper prior ((3.2) with  $0 \leq a < 1$ ) for the choice of  $c_o$  in (4.4). But we see that the equations for determining the interval which satisfies Bayes rule

$$\begin{aligned} \frac{1}{\sqrt{2\pi a\sigma^2}} \exp\left\{\frac{-(\delta_L - \hat{\theta}_B)^2}{2a\sigma^2}\right\} &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(1.96)^2}{2\sigma^2}\right\}, \\ \frac{1}{\sqrt{2\pi a\sigma^2}} \exp\left\{\frac{-(\delta_U - \hat{\theta}_B)^2}{2a\sigma^2}\right\} &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(1.96)^2}{2\sigma^2}\right\} \end{aligned} \quad (4.5)$$

yield the solution

$$\hat{\theta}_B \pm \sqrt{(1.96)^2 + \ln\left(\frac{1}{a}\right)}\sigma\sqrt{a}. \quad (4.6)$$

As the interval  $\delta_B : \hat{\theta}_B \pm 1.96\sigma\sqrt{a}$  is not the Bayes rule interval in (4.6), the risk function based on the 0-1 loss function does not meet the conditions we require to judge the competition between the frequentist interval estimator  $\delta_F$  and the Bayesian interval estimator  $\delta_B$ . The Bayes rule (4.6) under the 0-1 loss function is wider than the interval  $\delta_B$ . The 0-1 loss function assigns the same penalty to any incorrect interval, no matter how close an endpoint is to the truth. The "best" interval estimator under a proper prior distribution is wider than the interval estimator  $\delta_B$ , in order to provide better protection against an incorrect interval. This is particularly so for  $a$  near zero, corresponding to a prior which places very little weight on the data.

Instead, let us consider a cost function that penalizes an incorrect interval proportional to the distance between an endpoint and the truth. Take  $c(t) = t \cdot I\{t > 0\}$  in (4.1). For easy reference, we will name (4.1) under this choice of a cost function as the increasing loss function. The risk function for an interval estimator of the form  $\hat{\theta} \pm m$  under the increasing loss function becomes

$$R(\hat{\theta}, m; \theta) = E_{\hat{\theta}|\theta} \left[ (\hat{\theta} - m - \theta) I\{\hat{\theta} - m > \theta\} + (\theta - \hat{\theta} - m) I\{\hat{\theta} + m < \theta\} \right] + c_o \times [2m]. \quad (4.7)$$

Recall that  $c_o$  is to be chosen so that  $\delta_F$  is the Bayes rule under the noninformative prior. Now, let  $G$  denote the cumulative distribution function of the posterior distribution on  $\theta$ . It can be



shown that Bayes rule under the increasing loss function is the interval  $\delta = [\delta_L, \delta_U]$  found by solving  $G(\delta_L) = c_o$  and  $G(\delta_U) = 1 - c_o$ . That is, Bayes rule under the increasing loss function is formed by the lower and upper  $100 \times c_o$ th percentiles. Clearly, we take  $c_o = .025$  for the frequentist interval  $\delta_F$  to be "best" under the noninformative prior. Since our Bayesian interval  $\delta_B$  is formed by the lower and upper percentiles of the posterior distribution in (3.2), interval estimator  $\delta_B$  is "best" under the proper prior.

It has been determined that the risk function in (4.7), the expected value of the loss function in (4.1) with  $c(t) = t \cdot I\{t > 0\}$  and cost  $c_o = .025$ , is an appropriate judge for the competition between the Bayesian and frequentist interval estimators.

## 5. Comparison of Interval Estimators

We are now in a position to carry out a comparison between Bayesian and frequentist interval estimators in a manner analogous to the comparison of point estimators presented in Section 2. Denote the frequentist interval estimator as  $\hat{\theta}_F \pm m_F$  and the Bayesian interval estimator as  $\hat{\theta}_B \pm m_B$ . The goal of our comparison is the same as it was for SR: determine all prior specifications  $(\theta_o, a)$  such that

$$R(\hat{\theta}_B, m_B; \theta_*) < R(\hat{\theta}_F, m_F; \theta_*), \quad (5.1)$$

where  $\theta_*$  denotes the true value of the parameter. The risk function is given by (4.7) with  $c_o = .025$ .

Without loss of generality, we can take  $\theta_* = 0$  and  $\sigma^2 = 1$ . This will simplify the presentation of the results a good bit. The problem then becomes a comparison of the intervals  $\hat{\theta}_F \pm 1.96$  and  $\hat{\theta}_B \pm 1.96\sqrt{a}$  as estimators of the truth  $\theta_* = 0$ . The squared, scaled distance between the prior estimate and the truth simplifies as  $\Delta = (\theta_o - \theta_*)^2 / \sigma^2 = \theta_o^2$ . The distribution on  $\hat{\theta}_F$  under the truth  $\theta_* = 0$  simplifies to  $\hat{\theta}_F \sim N(0, 1)$ . Let  $\phi$  denote the probability density function for the standard normal distribution. After some further algebraic simplifications, the risk function for the frequentist interval estimator becomes

$$\begin{aligned} R(\hat{\theta}_F, m_F; \theta_*) &= \int_{-\infty}^{-1.96} [0 - (\hat{\theta}_F + 1.96)] \phi(\hat{\theta}_F) d\hat{\theta}_F \\ &+ \int_{1.96}^{\infty} [(\hat{\theta}_F - 1.96) - 0] \phi(\hat{\theta}_F) d\hat{\theta}_F + (.025) \times 2(1.96). \end{aligned} \quad (5.2)$$

The first integral in (5.2) depicts the risk from the interval underestimating the true  $\theta_* = 0$ , while the second integral is the risk from overestimation. The risk function for the Bayesian interval estimator is derived to be

$$\begin{aligned} R(\hat{\theta}_B, m_B; \theta_*) &= \int_{-\infty}^{\theta_l} [0 - (\hat{\theta}_B + m_B)] \phi(\hat{\theta}_F) d\hat{\theta}_F \\ &+ \int_{\theta_u}^{\infty} [(\hat{\theta}_B - m_B) - 0] \phi(\hat{\theta}_F) d\hat{\theta}_F + (.025) \times 2(m_B), \end{aligned} \quad (5.3)$$



where

$$\begin{aligned}
 \hat{\theta}_B &= a\hat{\theta}_F + (1-a)\theta_o, \\
 m_B &= 1.96\sqrt{a}, \\
 \theta_l &= \frac{-1.96}{\sqrt{a}} - \frac{(1-a)\theta_o}{a}, \\
 \theta_u &= \frac{1.96}{\sqrt{a}} - \frac{(1-a)\theta_o}{a}.
 \end{aligned} \tag{5.4}$$

We look to solve the inequality  $R(\hat{\theta}_B, m_B; \theta_*) < R(\hat{\theta}_F, m_F; \theta_*)$  in terms of  $a$  and  $\Delta = \theta_o$ , as was accomplished in Theorem 2.1 for the point estimator comparison. The solution for comparing interval estimators is much more difficult to attain analytically, but we can rely on Mathematica [9], or some other computer algebra system, to solve the inequality for us.

The graph of the threshold dividing Bayesian and frequentist superiority is presented in Figure 3. The solution for interval estimation takes on a similar form to the solution in Figure 2 for point estimation, so the comments made earlier apply here as well. In particular, we see once again the famous result from SR that the Bayesian has room to beat the frequentist with a proper choice of data weight  $a$ , regardless of how far the prior estimate  $\theta_o$  is from the true value  $\theta_*$ .

Let us investigate the frequentist risk (5.2) a bit further. The last term in the sum can be thought of as the component to risk from interval length. The first two terms in the sum can be thought of as components to risk from the chance of an incorrect interval. We can compute the risk due to interval length as  $(.025)2(1.96) = 0.098$ . The integrals in the risk due to incorrectness are calculated to be 0.018 (the risk for the frequentist interval estimator is constant). Note how the risk due to interval length is much greater than the risk due to incorrectness. The framework for our comparison reveals the underlying nature of the frequentist interval estimator as one developed to provide strong protection against putting forth an incorrect interval at the expense of greater length. This is analogous to the frequentist point estimator favoring the property of unbiasedness, at the expense of greater variance in comparison to the Bayesian estimator. There lies an opportunity for the Bayesian to provide improvement in interval estimation. The Bayesian may produce an interval where the increase in risk due to incorrectness is less than the benefit from the reduced length. We will explore this idea in the next section.

It is possible for a Bayesian interval estimator with an accurate prior estimate  $\theta_o$  to universally defeat the frequentist no matter the choice of data weight  $a$ . It can be determined numerically that prior estimates  $\theta_o$  with  $\Delta < 0.0136$  lead to Bayesian interval estimators that beat the frequentist no matter the choice of data weight. Let us examine this idea further by focusing on the case of no weight on the data ( $a = 0$ ). In this case, the Bayesian "interval" estimator is the point  $\theta_o$ . If  $\theta_o \neq \theta_*$ , ( $\Delta > 0$ ), then the Bayesian estimator  $\hat{\theta}_B$  is incorrect with certainty. Such an estimator, if close to the truth, may still be judged by risk function (4.7) to be better than an estimator represented as a proper interval. The threshold for the Bayesian prior estimate  $\theta_o$  alone to beat the frequentist is tighter for interval estimation than for point estimation ( $\Delta < 0.0136$  for interval estimation,  $\Delta < 1$  for point estimation). This is reasonable. Although we can see how a point estimator that does not involve the data at all can be better than a purely data based estimator, the level of accuracy for a point estimator ( $\theta_o$ ) to beat

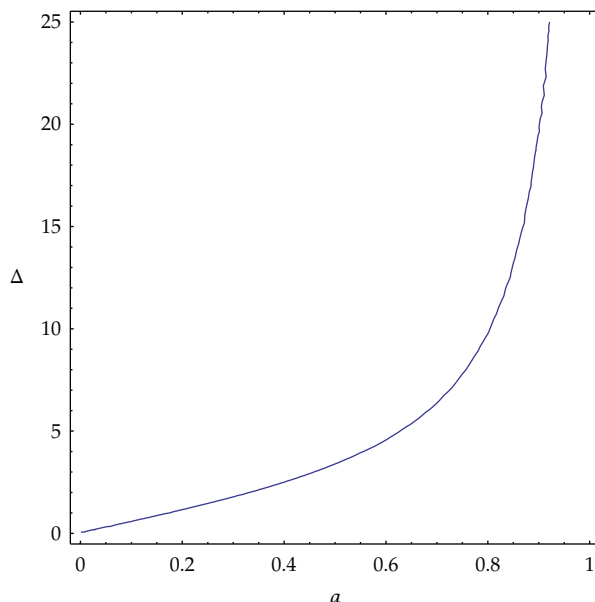


Figure 3: Threshold for interval estimator comparison.

an interval estimator ( $\delta_F$ ) should be greater than for a point estimator to beat another point estimator ( $\hat{\theta}_F$ ). However, there is greater opportunity for Bayesian interval estimators to beat the frequentist interval estimator when the choice of data weight  $a$  is more moderate. This idea also will be explored further in the next section.

## 6. The Word-Length Experiment

Samaniego [2] discusses an experiment in which 99 of his students were asked to construct a prior distribution for an estimation problem involving the first words on the 758 pages of a particular edition of Somerset Maugham's novel *Of Human Bondage*. The parameter of interest  $\theta$  is the proportion of these first words classified as "long" (six or more letters). The data information available for estimating  $\theta$  will be from a random sample of 10 pages. The frequentist point estimator for this problem,  $\hat{\theta}_F$ , will be the sample proportion of long words. Each of the students in the experiment was asked to provide a prior estimate  $\theta_o$  and a weight on the data  $a$ , so that the experiment consists of 99 Bayesian point estimators  $\hat{\theta}_B$  in competition against the frequentist point estimator. A scatterplot of the prior specifications  $\{(\theta_o, a)\}$  is displayed in Figure 4.

The goal of the experiment is to see how many of the 99 Bayesian point estimators are superior to the frequentist point estimator under a competition whose rules were set forth in Section 2. The scatterplot in Figure 4 indicates a diverse set of opinions as to the true proportion of long first words,  $\theta_*$ . For this reason, the experiment provides useful empirical evidence as to the utility of Bayesian estimators. In practice, prior information may be difficult to quantify. An empirical comparison of Bayesian estimation to frequentist estimation is best accomplished across a set of differing opinions. The word length experiment is representative of a situation that is realistic to practitioners. The parameter of interest is familiar enough that

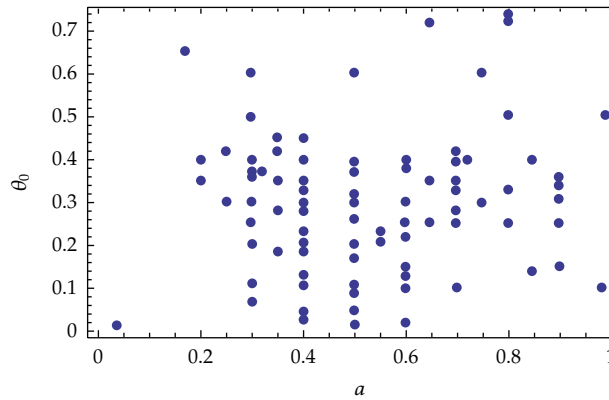


Figure 4: Priors for word-length experiment.

prior information is available, but not so familiar that this prior information can be easily quantified.

The true proportion of long first words turns out to be  $\theta_* = 228/758 = 0.3008$ . The variance of the frequentist estimator is then calculated to be  $\sigma^2 = (.3008)(.6992)/10 = 0.02103$ . The squared scaled distance between prior estimate and truth for the word length experiment becomes  $\Delta = (\theta_o - 0.3008)^2/0.02103$ . From Theorem 2.1, a Bayesian point estimator is superior to the frequentist point estimator when the specified prior is such that  $\Delta < (1 + a)/(1 - a)$ . As reported in Samaniego [2], Bayesian point estimators were superior in 88 of the 99 cases. A prior estimate with  $\Delta < 1$  leads to a Bayesian point estimator which beats the frequentist no matter the choice of data weight. Of the Bayesian priors put forth by the students in the experiment, 66 out of the 99 had prior point estimates accurate enough to beat the frequentist even without the aid of the data.

We will use the same experiment for comparing Bayesian and frequentist interval estimators. We will take  $\sigma^2 = 0.02103$  as fixed and known and treat  $\hat{\theta}_F$  as a normally distributed random variable. Although the underlying binomial properties are being ignored in order for the problem to fit into the framework of our comparison, the experimental results are still valid since the prior information  $(\theta_o, a)$  put forth by the students in the experiment is not tied to the underlying distributional assumptions. The results for comparing interval estimators are even stronger in favor of the Bayesians; 90 out of the 99 cases result in a Bayesian interval estimator superior to the frequentist estimator as judged by the rules for the competition derived in Sections 4 and 5.

It may be of interest to compare the probability of a correct interval for the Bayesian estimators to the .95 probability attained by the frequentist interval. Of the 90 cases where the Bayesian interval estimator was superior to the frequentist interval estimator, 14 of the Bayesian intervals had a coverage probability less than .95. The smallest of these coverage probabilities is .694, for a case with prior parameters  $\theta_o = .110$  and  $a = .30$ . The Bayesian can beat the frequentist in these cases by reducing the risk due to interval length without an undue increase in the risk due to incorrectness. An interval estimator that is incorrect, yet close, does not face much of a penalty under the increasing loss function since the cost of an incorrect interval is based on the distance between the truth and an endpoint.

For a prior estimate alone to dominate the frequentist interval estimator, an accuracy of  $\Delta < 0.01364$  is required. Only 9 out of the 99 prior estimates were accurate enough to beat

the frequentist interval estimator without the aid of data. The overall success of the Bayesian interval estimators in this experiment illustrates how a reasonable choice of data weight  $a$  can lead to an improvement on an interval dependent on the data alone, even with a prior estimate that is away from its target. We present one student in particular as an illustration of the benefits of a reasonable choice of weight. This student chose a prior estimate of  $\theta_o = .730$ ; a rather poor guess at the truth  $\theta_* = .3008$ . This student, however, places weight  $a = .8$  on the data, an appropriate quantification of uncertainty. Despite the poor choice of prior estimate, this student as a Bayesian beats the frequentist in both the point estimation comparison and the interval estimation comparison.

## 7. Concluding Remarks

Berger and Wolpert [10] write that “advancement of a subject usually proceeds by applying to complicated situations truths discovered in simple settings.” Admittedly, the situation considered for the comparisons in the SR paper, as in the current paper, is relatively simple. The lessons learned, however, are interesting and applicable. Efron [11] writes on the use of indirect information as an important trend in statistics. The comparisons initiated by SR reveal how indirect information, quantified in the form of a prior distribution, can lead to a Bayesian estimator that improves upon a frequentist estimator, even in situations where sharp prior knowledge is not necessarily available. The current paper shows that these results hold for interval estimation problems as well.

## Acknowledgments

The authors wish to thank the referees for comments and suggestions which helped to improve the original version of this paper.

## References

- [1] F. Samaniego and D. Reneau, “Toward a reconciliation of the Bayesian and frequentist approaches to point estimation,” *Journal of the American Statistical Association*, vol. 89, pp. 947–957, 1994.
- [2] F. Samaniego, *A Comparison of the Bayesian and Frequentist Approaches to Estimation*, Springer, New York, NY, USA, 2010.
- [3] V. Barnett, *Comparative Statistical Inference*, Wiley, Chichester, UK, 1999.
- [4] C. Robert, *The Bayesian Choice: A Decision Theoretic Motivation*, Chapman and Hall, London, UK, 2001.
- [5] J. Berger, “The case for objective Bayesian analysis (with Discussion),” *Bayesian Analysis*, vol. 1, pp. 385–402, 2006.
- [6] F. J. Samaniego and A. A. Neath, “How to be a better Bayesian,” *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 733–742, 1996.
- [7] E. Vestrup and F. Samaniego, “Bayes versus frequentist shrinkage in multivariate normal problems,” *Sankhya*, vol. 66, pp. 109–139, 2004.
- [8] P. Diaconis and D. Freedman, “On the consistency of Bayes estimates,” *Annals of Statistics*, vol. 14, pp. 1–25, 1986.
- [9] S. Wolfram, *The Mathematica Book*, Wolfram Media, Champaign, Ill, USA, 2003.
- [10] J. Berger and R. Wolpert, *The Likelihood Principle: A Review, Generalizations, and Statistical Implications*, IMS Lecture Notes, Hayward, Calif, USA, 1988.
- [11] B. Efron, “The future of indirect evidence,” *Statistical Science*, vol. 25, no. 2, pp. 145–157, 2010.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

