5-1-2015

Comparison of Bayesian Credible Intervals to Frequentist Confidence Intervals

Kathy Gray
California State University-Chico, klgray@csuchico.edu

Brittany Hampton
California State University-Chico, brittloveez@gmail.com

Tony Silveti-Falls
California State University-Chico, Tonys.falls@gmail.com

Allison McConnell
California State University-Chico, amcconnell4@mail.csuchico.edu

Casey Bausell
Oregon State University, cbausell@charter.net

Comparison of Bayesian Credible Intervals to Frequentist Confidence Intervals

Comparison of Bayesian Credible Intervals to Frequentist Confidence Intervals

Kathy Gray
California State University
Chico, CA

Brittany Hampton
California State University
Chico, CA

Tony Silveti-Falls
California State University
Chico, CA

Allison McConnell
California State University
Chico, CA

Casey Bausell
Oregon State University
Corvallis, OR

Frequentist confidence intervals were compared with Bayesian credible intervals under a variety of scenarios to determine when Bayesian credible intervals outperform frequentist confidence intervals. Results indicated that Bayesian interval estimation frequently produces results with precision greater than or equal to the frequentist method.

Keywords: Confidence intervals, credible intervals, Bayesian statistics, mean square error

Introduction

Although mathematicians introduced the field of Bayesian statistics in the 1700s, Bayesian methods gained most of its popularity in practice fairly recently (McCarthy & Parris., 2004; Smyth, 2004; Stoyan & Penttinan, 2000). Researchers have used frequentist methods for statistical analysis until technological advances and the introduction of certain algorithms, such as Markov chain Monte Carlo, gave way to increased computational power that enabled complex calculations to be done using Bayesian procedures (Little, 2006). This resulted in an increase in the interest of Bayesian statistics and sparked much controversy and debate regarding which method should be used by researchers (Little, 2006).

The frequentist approach relies on properties based on repeated sampling and takes only sample data into account to estimate the population parameter. Bayesian statistics, however, adds the component of a prior distribution based on prior knowledge and/or expert opinion of the subject. Using the prior information and the observed data, Bayesian methods calculate a refined estimate of the

population parameter. Some claim that this subjective prior is key to most accurately estimating the population parameter while others claim that the lack of objectivity of Bayesian statistics interferes with the results (Choy et al., 2009).

The goal of this study was to compare Bayesian credible intervals to frequentist confidence intervals under a variety of scenarios to determine when Bayesian credible intervals outperform frequentist confidence intervals. The Central Limit Theorem (CLT) states that when a large enough random sample is taken the distribution of the sample means will be approximately normal. This theorem has been widely researched and it is generally accepted that as long as the sample size is around 25 we can rely on the CLT when performing inference on the population mean when the population is not normal (Stonehouse & Forrester, 1998).

Although not as well studied as the CLT, there exists a Bayesian Central Limit Theorem (BCLT) which states that under certain conditions the posterior probability distribution is approximately normal for large enough sample sizes (Walker, 1969). For Bayesian credible intervals, if the data are assumed to follow a normal distribution and if the prior distribution is also assumed to be normal then the calculations are straightforward because the posterior distribution for the population mean will also follow a normal distribution (Kruschke, 2010). If the data are not normal and transformations of the data do not achieve normality, then a more appropriate distribution could be used to model the data, however, this leads to a more complicated analysis. Furthermore, there is no guarantee that an appropriate distribution can be found that models the data. The goal of our study is to examine the robustness of Bayesian credible intervals when the assumption of normally distributed data is violated and to determine under what scenarios Bayesian credible intervals outperform frequentist confidence intervals.

Methodology

In order to investigate the BCLT we generated populations from three different distributions: 1) Standard normal distribution; 2) Beta distribution with parameters $\alpha = 2$ and $\beta = 5$ (moderately skewed distribution); 3) Exponential distribution with parameter $\lambda = 0.5$ (strongly skewed distribution). We repeatedly and randomly sampled from each population for various sample sizes ($n = 10, 15, 20, 25, 30, 40, 50,$ and 75). For each scenario we calculated Bayesian credible intervals and frequentist confidence intervals. The frequentist confidence interval was calculated using the following formula:

$$\bar{x} \pm t_{n-1} \frac{s}{\sqrt{n}} \qquad (1)$$

where t_{n-1} is the critical value for a 95% confidence interval with $n-1$ degrees of freedom. Bayesian confidence intervals were calculated as follows:

$$\mu_1 \pm t_{n-1}\sigma_1 \qquad (2)$$

where

$$\mu_1 = \frac{1/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2}\mu_0 + \frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_0^2}\bar{x} \qquad (3)$$

and

$$\sigma_1^2 = \frac{\sigma^2\sigma_0^2}{\sigma^2 + n\sigma_0} \qquad (4)$$

where σ^2 is the population variance, μ_0 is the prior mean and σ_0^2 is the prior variance.

The population variance was always estimated with the sample variance, s^2. For each population distribution and sample size we calculated Bayesian credible intervals using a prior mean that wasn't biased, a prior that had a low bias, and a prior that had a large bias. We use the term bias to represent how far off the prior mean is from the population mean. A bias of 0.25 times the standard deviation was considered as a small bias in the prior mean and a bias of 0.50 times the standard deviation as a large bias in the prior mean. For the normal distribution, the bias was added to the prior before running the simulations. For the skewed distributions, we looked at both positive (prior mean > population mean) and negative biases (prior mean < population mean). The prior variance can be thought of as how confident one is in the prior mean. For instance, if there is a lot of confidence in the prior mean then the prior variance would be small since the researcher has honed in on the population mean. If there is little confidence in their prior mean then the prior variance would be large to reflect this. A confidence in the precision was considered to be equal to a value that would be equivalent to a sample size of about 12. In other words, about as much confidence

was placed in the prior as would be if there was a sample of 12 from the population. This value was somewhat arbitrary; however, it represents the typical confidence in a prior mean. Thus, the prior variance was calculated as

$$\sigma_0^2 = \frac{\sigma^2}{12} \tag{5}$$

For each scenario (combination of sample size, bias, and population shape) we computed capture rate as the percent of the intervals that contained the true population mean. Additionally, the mean squared error (MSE) was calculated for each scenario. The MSE combines both the bias of an estimator as well as the variance. The MSE was calculated as

$$\mathrm{MSE}\left(\hat{\mu}\right) = \mathrm{bias}^2 + \mathrm{var}\left(\hat{\mu}\right) \tag{6}$$

The bias is the difference between the estimated value and the true mean of the population. For the frequentist method it can be shown that the bias of the sample mean is 0, therefore, the MSE is var($\overline{y}$) for frequentist methods. All statistical analyses were performed using R (R Development Core Team, 2007).

Results

The capture rates of the frequentist and Bayesian intervals are shown in Figure 1 for various scenarios and sample sizes. As expected, the frequentist intervals have a 95% capture rate when the population distribution is normal. The frequentist method does quite well for the moderately skewed population where a sample size of 30 is needed to obtain a 95% capture rate. For the strongly skewed population, the frequentist intervals do not capture the parameter at the stated 95% level, however, when the sample size is 75 the capture rate remains at about 94%. For all scenarios, the no bias and positive, low bias scenarios performed best for small sample sizes with capture rates above 95% for both the normal population and the moderately skewed population. These capture rates decreased when sample size increased since the credible intervals were weighted more heavily by the data rather than the prior information and thus conform to frequentist properties.

For the strongly skewed data, the scenario that performed the best was the positive, low bias prior. This scenario captured the mean about 95% of the time for all sample sizes. A negative bias increased the capture rate when compared to

a positive bias. As expected, the high bias scenarios performed the worst with respect to capture rates. For all sample sizes, the high bias scenarios gave worse results than the frequentist intervals, indicating that sample sizes need to be larger than 75 to dilute the bias in the prior. The results indicate that as long as the bias in the prior distribution is not too large then one can have results better than frequentist's methods even for strongly skewed distributions.

The MSE is given in Figure 2 for each sample size. The MSE accounts for both bias and variance of an estimator and, therefore, the smaller the MSE the better the performance of the statistic in estimating the parameter. Since the sample mean is an unbiased estimate of the population mean, the frequentist confidence interval has a bias equal to zero and the MSE is only based on the variance of the estimator. The bias for the Bayesian credible intervals varied from no bias to a bias of half of a standard deviation. For all three population distributions, there were no significant differences between MSE when the sample size reached 75. Similar results were obtained for the normal and moderately skewed population.

The largest difference in MSE between the different scenarios occurred for small sample sizes. The MSE was significantly larger for the frequentist confidence intervals than the Bayesian credible intervals until a sample size of 40 for the high biased Bayesian scenario. When comparing the frequentist confidence intervals to the low bias and no bias credible intervals, they are significantly lower until the sample size reaches 75. All Bayesian scenarios performed better than the frequentist intervals until a sample size of 30 was reached. The no bias and low bias continued to perform better than the frequentist interval until a sample size of 75 was reached.

The degree of bias needed before the capture rate drops below 0.95 is investigated in Figure 3. Iterations were performed using samples sizes of 15, 30, 50, and 75. Surprisingly, there was not much difference between the three different population shapes (normal, moderately skewed, strongly skewed) even for smaller sample sizes. In addition, the sample size did not have much effect on capture rate. When the sample size is 15 both the normal distribution and the moderately skewed distribution are above the 95% capture rate until the bias was equal to 0.4, at which point the capture rate dropped very quickly. The strongly skewed population performed only slightly below the 95% capture rate when $n = 15$ until a bias equal to 0.4 at which point it dropped off significantly. The differences between the three distributions were very small for all sample sizes. For sample sizes larger than 15 the capture rate dropped below the 95% level at a

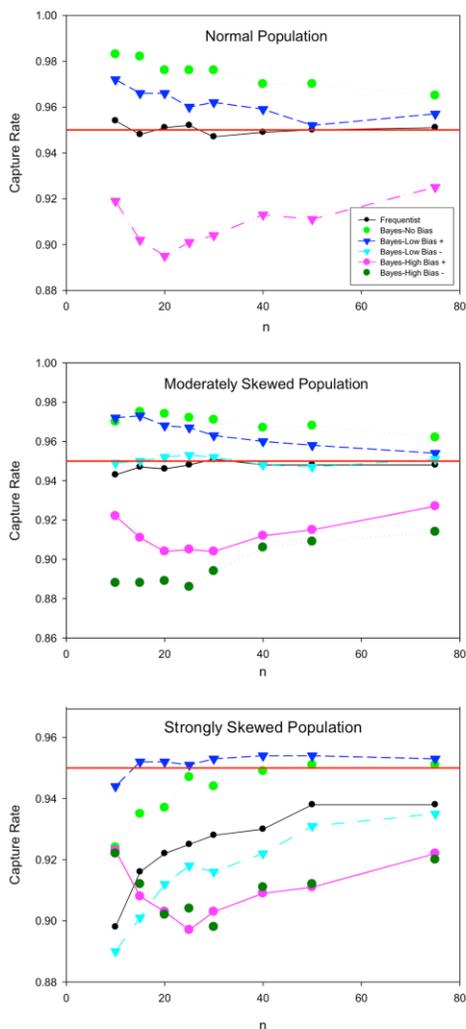0.3 level of bias. Thus, it appears that the effects of the bias are slightly worse for larger sample sizes.

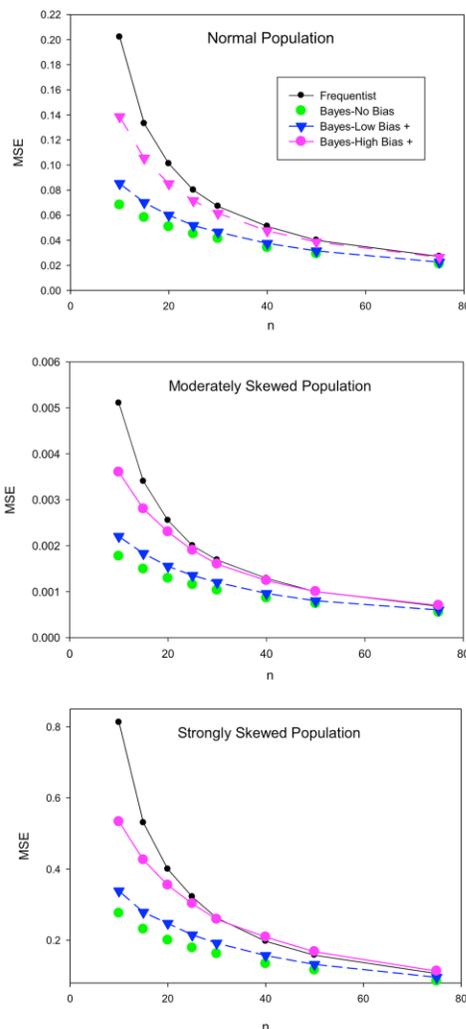**Figure 1.** Capture rates for confidence intervals and Bayesian credible intervals.

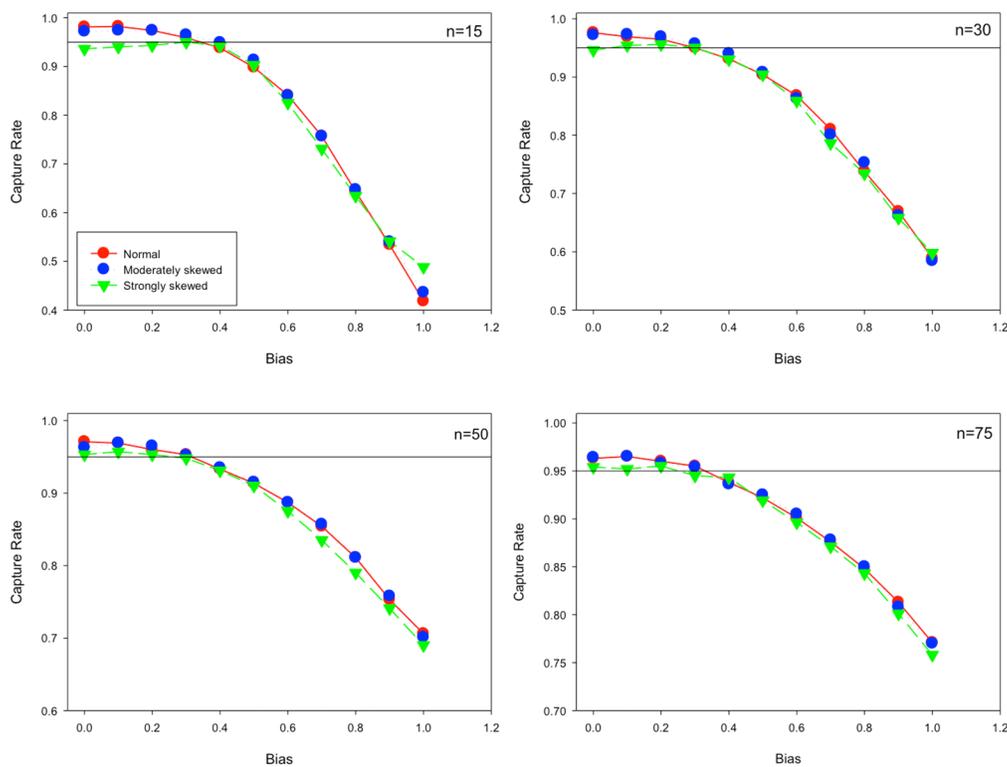**Figure 2.** Mean squared error (MSE) for each scenario.

Figure 3. Capture rates for different sample sizes and different degrees of bias. The bias is calculated by the number of standard deviations above the prior's mean.

The MSE for increasing levels of bias in the prior mean in shown in Figure 4. The three distributions are shown on separate graphs and within each graph are three separate sample sizes. The solid line represents the MSE for frequentist methods for each sample size. For the normal distribution, when the Bayesian methods reach a bias of 0.6 the MSE is about equal to the frequentist methods with the same sample size. After a bias of 0.6 the Bayesian methods perform worse than frequentist methods with respect to the MSE. The differences were minor when comparing distributions. For strongly skewed distributions a smaller biased is required to perform better than frequentist methods. Surprisingly, for all distributions there was not much difference between the bias cutoff for different sample sizes. For the strongly skewed distribution, after a bias of 0.4 the frequentist methods performs better for $n = 50$ compared to a bias of 0.6 for a sample size of $n = 15$.
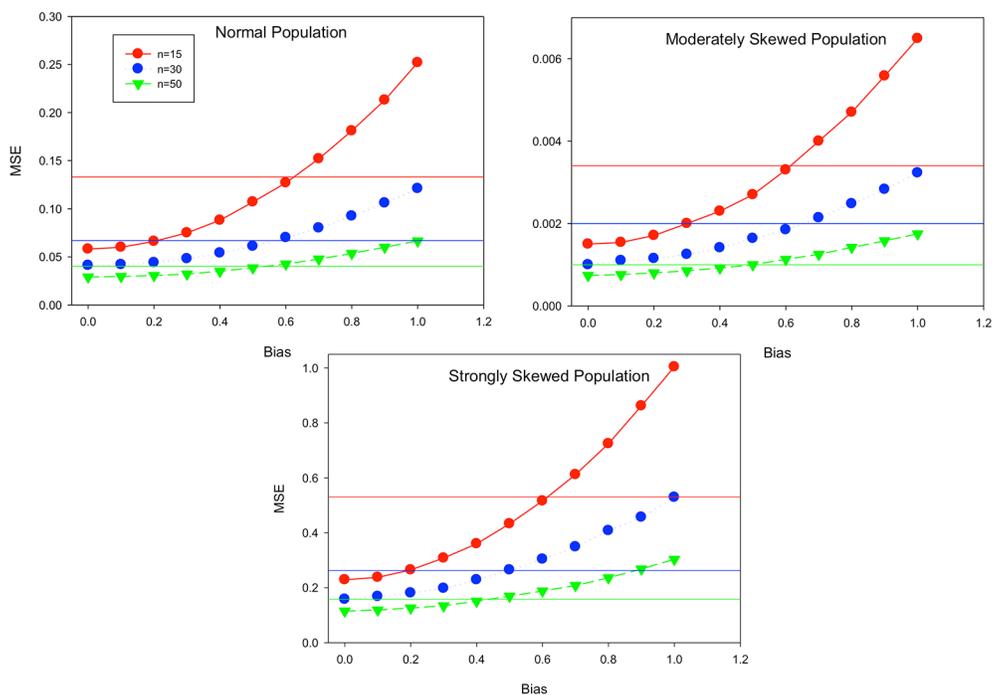
49

Figure 4. MSE calculated for different degrees of bias. The horizontal line reflects the MSE for frequentist methods of the same sample size.

Conclusion

These results indicate that when a prior mean is less than 0.4 to 0.6 standard deviations from the population mean then Bayesian credible intervals outperform frequentist confidence intervals with respect to MSE and capture rate for most scenarios that we looked at. For larger biases, frequentist confidence intervals will perform better with respect to MSE. Additionally, the distribution of the data did not have a large effect on the results even though the methods used assumed that the data came from a normal distribution. For strongly skewed data, neither frequentist nor Bayesian intervals performed at the optimal 95% capture rate with the exception being the Bayesian scenario with small, positive bias. Thus, for strongly skewed data it is suggested to seek a transformation for the data no matter which technique is used. In conclusion, this research demonstrates that Bayesian credible interval can have desirable properties for small sample sizes when the bias can be kept within about 0.5 standard deviations of the mean.

Because researchers will never know the bias of the prior mean they should only use Bayesian techniques when they have good information about the subject being researched.

Acknowledgements

References

Choy, S. L., O'Leary, R., & Mengersen, K. (2009). Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology*, *90*(1), 265-277. doi:10.1890/07-1886.1

Kruschke, J. (2010). *Doing Bayesian data analysis: A tutorial introduction with R*. Waltham, MA: Academic Press.

Little, R. J. (2006). Calibrated Bayes: a Bayes/frequentist roadmap. *The American Statistician*, *60*(3), 213-223. doi:10.1198/000313006X117837

McCarthy, M. A. & Parris, K. M. (2004). Clarifying the effect of toe clipping on frogs with Bayesian statistics. *Journal of Applied Ecology*, *41*, 780–786. doi:10.1111/j.0021-8901.2004.00919.x

R Development Core Team (2007). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at http://www.R-project.org.

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, *3*(1), 1-25. doi:10.2202/1544-6115.1027

Stonehouse, J. M., & Forrester, G. J. (1998). Robustness of the t and U tests under combined assumption violations. *Journal of Applied Statistics*, *25*(1), 63-74. doi:10.1080/02664769823304

Stoyan, D. & Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science 15*(1), 61-78. doi:10.1214/ss/1009212674. http://projecteuclid.org/euclid.ss/1009212674.

Walker, A. M. (1969). On the asymptotic behaviour of posterior distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, *31*(1). 80-88.