Comparing Ordinary Least Squares Confidence Intervals and Bayes Regression Credible Intervals

Marieke Sorge

April 2022

1 Introduction

The contrast between Frequentist and Bayesian methods of analysis in statistics has been widely debated in the recent history of the field, particularly estimates of uncertainty in analysis. In this project, I will examine and discuss the contrast between Bayesian linear Regression and the frequentist Ordinary Least Squares linear regression. While they are similar in both objective and theory, there are key difference between the Bayes linear regression and Ordinary Least squares. Some differences of note include that Bayesian Regression includes previous knowledge about the parameter distributions (prior) and makes inference on a (posterior) distribution rather than a point estimate, as Ordinary Least Squares does. The benefits of Bayesian regression include the ability to incorporate previous knowledge into parameter estimation via the prior and greater knowledge about the uncertainty of the inferred prediction as it originates from a distribution that can be characterized. Characterizing the uncertainty about an estimate is particularly relevant to the examination of confidence and credible intervals as this is their primary conceptual purpose. However, Bayesian regression can be computationally expensive since sampling from the posterior distribution is necessary for inference. The Bayesian method of defining credible intervals finds a high probability range for a parameter estimate on a distribution called the "a posteriori probability" or more commonly the "posterior" using both the prior and the observed data. Whereas the frequentist confidence intervals provide a high probability range of parameter values assuming the data originates from a t-distribution with mean and standard deviation equivalent to the mean and standard deviation of the sample.

Figures 1 and 2 show a sample of posterior distributions for two simulated coefficients on the covariates with the priors $\beta_1\ N(100,1), \beta_1\ N(3,1), \beta_2\ N(2,1)$ with $X_1\ N(50,9)$ and $X_2\quad N(200,40)$, sample size=100, and the true values of the coefficients corresponding to $\beta_0 = 150, \beta_1 = 4, \beta_2 = 2.5$.

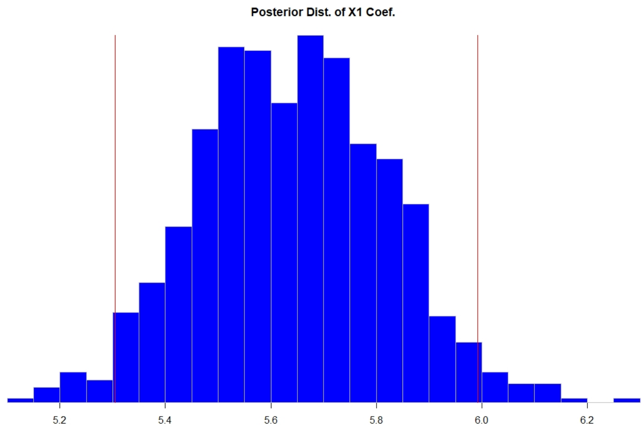

Figure 1: Posterior Distribution of β_1 the coefficient on X1 with red cut-off line at the end of the 95 percent CI

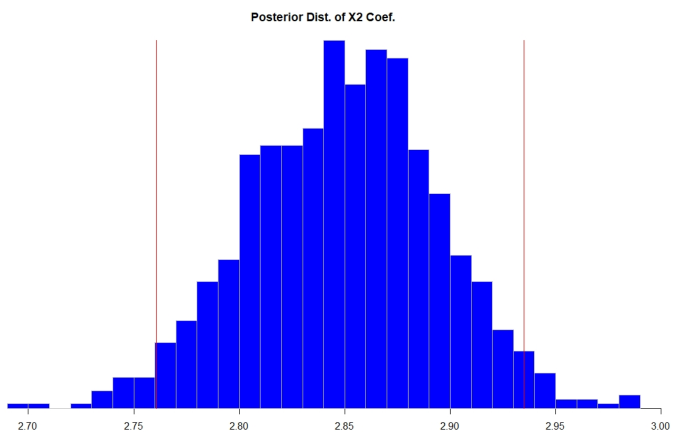

Figure 2: Posterior Distribution of β_2 the coefficient on X2 with red cut-off line at the end of the 95 percent CI

2 Methods

For this exploration data was simulated with linear relationship between two "x" covariates plus standard normal noise and a y response variable, yielding the model:

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \epsilon$$

with ϵ distributed N(0,1).

The true intercept and coefficients of the relationship between the x variables and the y variable were set as constants: $\beta_0 = 150, \beta_1 = 4, \beta_2 = 2.5$ so that performance of the respective models can be evaluated under different conditions, in particular focusing on the coverage and length of the respective CI for each method. The coverage of the ordinary least squares confidence intervals and the Bayes regression credible intervals were examined under four conditions: exact prior on the Bayesian regression, small sample size, uninformative prior on the Bayesian regression, and a mis-informative prior where the Bayesian prior does not correctly reflect the distribution of the population. All priors were normally distributed but the mean and standard deviation of the priors were adjusted to reflect the simulation situation.

For the simulations, the x variables were distributed $X_1 \;\; N(50, 9)$ and X_2 $N(200, 40)$ and the error term on y ϵ was distributed N(0,1). For simulations with a large sample size 100 simulated data points were used. For the small sample simulation the sample size was 10. In each iteration of the simulations the data was randomly sampled, the regression was performed and the 95 percent CI was calculated. Then the condition of whether the true coefficient value was captured in the CI for each coefficient was evaluated and stored by assigning a 1 for a CI that includes the coefficient and a 0 for the failure of the CI to capture the true coefficient value. The percentage of CIs that captured the true coefficient values over 10000 trials was then calculated. Since the 95 percent confidence intervals and credible interval were evaluated we expect these coverage percentages to be approximately 95 percent in a large number of trials. The size of the confidence or credible interval (upper bound - lower bound) was also recorded as a means to evaluate the CI. Using the Monte Carlo error in empirical coverage based on the normal approximation to the binomial: $M = 1.96 * \sqrt{(.05 * .95)/1000} \approx .01351$ as a margin of error on the expected coverage for the credible intervals and the confidence intervals, we can observe whether the intervals perform relative to the expected 95 percent coverage. With this margin of error we would consider coverage that is less than .95- M $\approx$ .9365 as under- performance and coverage that is greater than .95+ M $\approx$ .9635 as over performance.

2.1 Out of Sample Regression Performance Analysis

Before evaluating the confidence and credible intervals around the point estimates the ability of each method to find a reasonable point estimate was evaluated. To assess the relative performance of the Bayes Regression vs. the Ordinary Least Squares Regression estimation of the parameters of the model, the out-of-sample performance metrics: Mean Absolute Error and Mean Squared Error were calculated for a 100 fold cross validation on a 2/3, 1/3 training-testing set split. The data was simulated normal with according to the above specifications with true values of the model parameters $\beta_0 = 150, \beta_1 = 4, \beta_2 = 2.5$.

For evaluation of out-of-sample performance the prior on β_0 is distributed $N(150, 1)$. The prior on β_1 is distributed $N(3, 1)$ and the prior on β_2 is distributed $N(2, 1)$.

Mean absolute error was calculated with the following equation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$$

Mean squared Error was calculated with the following equation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

The results are reported in a table below:

MSE	MSE Bayes	MAE OLS	MAE Bayes
25682.7711	25682.7732	129.3145	129.3145

The out of sample performance for each method is comparable. Thus one does not more accurately attain point estimates for the parameters of the linear model.

3 Results

3.1 Large Sample Exact Prior Simulation

During this simulation the prior on β_0 was distributed $N(150, 1)$. The prior on β_1 was distributed $N(4, 1)$ and the prior on β_2 was distributed $N(2.5, 1)$. These means reflect the exact values of the true parameters with small variance thus producing an exact prior.

Table 1: Coverage for Exact Prior Simulation

	(Intercept)	β_1	β_2
OLS	0.9477	0.9467	0.9468
Bayes	0.9473	0.9469	0.9460

Table 2: Length for Exact Prior Simulation

	(Intercept)	β_1	β_2
OLS	2.99675	0.04418	0.00994
Bayes	2.988624477	0.044042123	0.009901338

When the prior means for the parameters are set to their true values the credible interval captures the true values with the expected coverage when compared the expected 95 percent coverage + the Monte Carlo error (M) 0.9473, 0.9469, 0.9460 which are all within thee .9365-.9636 expected range. The OLS confidence interval performs as expected with coverage between .9365 and .9635 for all the parameters. The credible interval is slightly smaller than the confidence interval. Since the credible interval is estimated from a combination of information from the data and the prior the shorter length in the Bayesian credible intervals occurs. Including a prior that is centered at the true values from which the data is sampled is equivalent to generating a CI on a larger sample size.

3.2 Small Sample Normally Distributed X Values Simulation

For this simulation the prior on β_0 is $N(150, 1)$. The prior on is set to β_1 is $N(3, 1)$ and the prior on β_2 is $N(2, 1)$. These priors are somewhat informative though not as near the true values as those in the previous simulation.

Table 3: Coverage for Small Sample Size Simulation

	(Intercept)	β_1	β_2
OLS	0.9508	0.9501	0.9508
Bayes	0.9315	0.9345	0.9310

Table 4: Length for Small Sample Size Simulation

	(Intercept)	β_1	β_2
OLS	2.9991	0.04420	0.00995
Bayes	2.87326	0.04233	0.00952

For the small sample size the Bayes credible interval under performs across the parameters with coverage on $\beta_0, \beta_1, \beta_2 = 0.9315, 0.9345, 0.9310$ respectively all less than the criteria .9365 set by the expected .95 coverage minus the Monte Carlo error. The OLS confidence interval performs as expected with coverage between .9365 and .9635 for all the parameters. The credible interval is slightly numerically smaller than the confidence interval, but this does not necessarily imply a statistically significant difference between them. Further analysis would need to be done to assess for statistical differences.

3.3 Large Sample Mis-informative Prior Simulation

In this simulation, the prior on β_0 was $N(0, 1)$ distributed, the prior on β_1 was $N(100, 1)$, and the prior on β_2 was distributed $N(200, 1)$. These values are distant from the true values of 4, 2.5, and 150 respectively, creating a mis-informative prior.

For the mis-informative prior simulation, the Bayes credible interval under performs at an even higher rate than in the small sample simulation. The coverage for $\beta_0, \beta_1, \beta_2 = 0.9283, 0.9260, 0.9481$ respectively all less than the .9365

Table 5: Coverage for Mis-informative Prior Simulation

	(Intercept)	β_1	β_2
OLS	0.9450	0.9446	0.9481
Bayes	0.9283	0.9260	0.9481

Table 6: Length for Mis-informative Prior Simulation

	(Intercept)	β_1	β_2
OLS	2.99879	0.04419	0.00996
Bayes	2.87151	0.042282	0.00953

= .95 - M. The confidence interval performs as expected. The credible interval is once again slightly smaller than the confidence interval.

3.4 Large Sample uninformative prior Simulation

In the uninformative prior simulation the prior mean values were all set to 0 and the prior variance set to 100, making the priors on β_0, β_1, β_2, all $N(0, 100)$ distributed.

Table 7: Coverage for Uninformative Prior Simulation

	(Intercept)	β_1	β_2
OLS	0.9476	0.9497	0.9450
Bayes	0.9300	0.9336	0.9261

Table 8: Length for Uninformative Prior Simulation

	(Intercept)	β_1	β_2
OLS	3.00582	0.04428	0.00995
Bayes	2.88186	0.04245	0.00953

In this simulation OLS performs as expected with coverage between .9365 and .9635 for all the parameters. (nothing has changed), and the Bayes method slightly under-performs for $beta_0$ and β_2 by capturing the true values of in 93.00, 92.61 percent of the simulations, respectively. The coverage for β_2 just marginally exceeds the set cut-off for under-performance. Considering the other two parameters we can conclude the credible interval under-performs in this setting.

3.4.1 Sample Variation with Fewer Trials

To illustrate the fact that the expected 95 percent coverage rate for a 95 percent confidence interval or 95 percent credible interval is an expected value over many trails. The simulation with prior $N(0,1)$ for all three parameters on the Bayes regression was run 5 times with 100 trials in each simulation and the average coverage for each of the 5 runs with 100 trial each was recorded.

		(Intercept)	β_1	β_2
Run 1	OLS	0.97	0.93	0.95
Run 1	Bayes	0.93	0.92	0.96
Run 2	OLS	0.88	0.93	0.91
Run 2	Bayes	0.88	0.93	0.89
Run 3	OLS	0.97	0.95	0.94
Run 3	Bayes	0.96	0.92	0.91
Run 4	OLS	0.96	0.94	0.93
Run 4	Bayes	0.93	0.93	0.94
Run 5	OLS	0.93	0.95	0.95
Run 5	Bayes	0.92	0.93	0.92

Although OLS seems to be mostly centered around that expected 95 percent coverage it varies with minimum coverage of 88 percent and maximum coverage of 97 percent. Similarly the coverage for the credible intervals fluctuates from about 89 percent to 96 percent. Since the simulations were run identically (with randomly selected samples) these discrepancies are due to sample variation.

4 Discussion and Conclusions

In general the Bayes credible intervals are smaller in size but they are subject to failing to capture the mean in less than ideal situations. If the prior on the Bayes Regression is informative Bayes outperforms Ordinary least squares regression in both coverage and length of the credible/confidence interval. In a small sample situation, ordinary least squares out-performs Bayesian regression analysis in terms of coverage although they both perform under the expected 95 percent coverage rate. It is interesting to note that although coverage varies across simulation settings the length of the confidence intervals among the methods (OLS vs Bayes) does not vary much. Additionally, although the settings for the OLS confidence intervals did not change between the the simulations: exact prior, uninformative prior and misinformation prior there is slight fluctuation in the coverage. This again exemplifies changes due to sample variation. If the Bayesian prior is informative and accurate this method has a higher coverage rate and shorter confidence intervals, and thus is the preferred method. However if the prior is uninformative or incorrect the coverage for the Bayesian credible intervals decreases in comparison with the OLS confidence intervals. Although, due to the simplicity of the simulations done in this exploration computational resources were not much of a concern, it is important to note that Bayesian credible intervals require significantly more computational power to generate compared to OLS confidence intervals since Monte Carlo draws from the posterior are used.

5 References

Lu, D., Ye, M., Hill, M. C. (2012). Analysis of regression confidence intervals and Bayesian credible intervals for uncertainty quantification. Water resources research, 48(9). https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2011WR011289

Hespanhol, L., Vallio, C. S., Costa, L. M., Saragiotto, B. T. (2019). Understanding and interpreting confidence and credible intervals around effect estimates. Brazilian journal of physical therapy, 23(4), 290-301. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6630113/

Gray, K., Hampton, B., Silveti-Falls, T., McConnell, A., Bausell, C. (2015). Comparison of Bayesian credible intervals to frequentist confidence intervals. Journal of Modern Applied Statistical Methods, 14(1), 8. https://digitalcommons.wayne.edu/jmasm/vol14/iss1/8/