

# New confidence bounds for the mean of a Gaussian distribution versus the classical confidence bounds

## Abstract

The aim of the present study is to compare confidence intervals for the mean of a Gaussian distribution. Considering the square error loss function, Approximate Bayesian confidence intervals for the mean of a normal population are derived. Using normal data and SAS software, the obtained approximate Bayesian confidence intervals will then be compared to the ones obtained with the well-known classical method.

Numerical results show that both the classical and the new Approximate Bayesian model perform well and that the Approximate Bayesian model has great coverage accuracy.

**Keywords:** estimation, loss functions, confidence intervals, coverage accuracy, statistical analysis, mean time between failures

Volume 3 Issue 3 - 2017

**Vincent AR Camara**

Research Center for Bayesian Applications Inc, USA

**Correspondence:** Vincent AR Camara, Research Center for Bayesian Applications Inc, Largo, USA, Email gvcamara@ij.net

**Received:** October 20, 2016 | **Published:** June 22, 2017

## Introduction

Bayesian analysis implies the exploitation of suitable prior information and the choice of a loss function in association with Bayes' Theorem.<sup>1,2</sup> It rests on the notion that a parameter within a model is not merely an unknown quantity but rather behaves as a random variable which follows some distribution. In the area of life testing, it is indeed realistic to assume that a life parameter is stochastically dynamic. This assertion is supported by the fact that the complexity of electronic and structural systems is likely to cause undetected component interactions resulting in an unpredictable fluctuation of life parameters. Recently, Drake<sup>3</sup> gave an excellent account for the use of Bayesian statistics in reliability problems. As he points out "He (Bayesian) realizes that his selection of a prior (distribution) to express his present state of knowledge will necessarily be somewhat arbitrary.<sup>4-10</sup> But he greatly appreciates this opportunity to make his entire assumptive structure clear to the world". "Why should an engineer not use his engineering judgment and prior knowledge about the parameters in the statistical distribution he has picked? For example, if it is the mean time between failures (MTBF) of an exponential distribution that must be evaluated from some tests, he undoubtedly has some idea of what the value will turn to be". In the present study, we shall consider a classical and useful underlying model.<sup>11,12</sup> That is, we shall consider the Normal underlying model characterized by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0 \quad (1)$$

Once the underlying model is found to be normally or approximately normally distributed, to construct confidence intervals for a Normal population mean, the well-known classical approach uses the following models that rely on the standard Normal and the student-t statistics:

$$\left( \bar{X} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right) \quad (2)$$

$$\left( \bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \quad (3)$$

## Methodology

In the derivation of our Approximate Bayesian confidence bounds for the mean of a normal distribution, the square error loss function has been used. The square error loss function places a small weight on estimates near the true value and proportionately more weight on extreme deviation from the true value of the parameter. Its popularity is due to its analytical tractability in Bayesian modeling. The square error loss is defined as follows:

$$L_{SE}(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2 \quad (4)$$

Considering the Square Error Loss function, the following Approximate Bayesian confidence bounds for the variance of a Normal distribution<sup>13</sup> have been derived:

$$L_{\sigma^2(SE)} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n-2-2\ln(\alpha/2)} \quad (5)$$

$$U_{\sigma^2(SE)} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n-2-2\ln(1-\alpha/2)} \quad (6)$$

Using the equation

$$\sigma^2 = E(X^2) - \mu^2 \quad (7)$$

Along with equations (5) and (6), the following Approximate Bayesian confidence bounds for a positive mean of a normal distribution<sup>14</sup> have been easily derived:

$$U_{\mu(SE)} = \left( \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} + \bar{x}^2 - \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-2-2\ln(\alpha/2)} \right)^{0.5} \quad (7)$$

$$L_{\mu(SE)} = \left( \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} + \bar{x}^2 - \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-2-2\ln(1-\alpha/2)} \right)^{0.5} \quad (8)$$

Hence, for a normal random variable  $X$  with a mean that is smaller or equal to zero, we can infer the following Approximate Bayesian confidence bounds<sup>15-17</sup> for the population mean:

$$U_{\mu(SE)} = \left( \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} + \bar{y}^2 - \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-2-2\ln(\alpha/2)} \right)^{0.5} - a \quad (9)$$

$$L_{\mu(SE)} = \left( \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} + \bar{y}^2 - \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-2-2\ln(1-\alpha/2)} \right)^{0.5} - a \quad (10)$$

Where  $y = x + a$  and “a” is a constant such that  $x + a > 0$

## Numerical and results

For the numerical results, we will use samples that have been obtained from normally distributed populations<sup>18-20</sup> (Examples 1, 2, 3, 4, 7) and approximately normal populations (Examples 5, 6). SAS software is used to obtain the normal population mean  $\mu$  and standard deviation  $\sigma$  corresponding to each of the Normal and approximately Normal data sets that are given below. The lengths of the classical and Approximate Bayesian confidence intervals are respectively denoted by  $W_C$  and  $W_{SE}$ .

**Example 1:** Data obtained from Prem S Mann<sup>21</sup> 24, 28, 22, 25, 24, 22, 29, 26, 25, 28, 19, 29. Normal population distribution obtained with SAS:

$$N(\mu = 25.083, \sigma = 3.1176)$$

$$\bar{x} = 25.08333$$

The corresponding (Table 1) sample mean and sample variance is

$$s^2 = 9.719696$$

**Example 2:** Data obtained from Prem S Mann<sup>21</sup> 13, 11, 9, 12, 8, 10, 5, 10, 9, 12, 13. Normal population distribution obtained with SAS:

$$N(\mu = 10.182, \sigma = 2.4008)$$

$$\bar{x} = 10.181812$$

The corresponding (Table 2) sample mean and sample variance is

$$s^2 = 5.763636$$

**Table 1** Classical and approximate Bayesian confidence intervals for the population mean corresponding to the first example of data set

C.L. %	Approx. Bayesian bounds (SE)	Classical bounds	WC WSE
80	25.0683-25.1311	23.85665-26.31001	39.87
90	25.0661-25.1437	23.46696-26.69971	41.66
95	25.0650-25.1543	23.10246-27.06420	44.36
99	25.0641-25.1734	22.28798-27.87869	51.15

**Table 2** Classical and approximate Bayesian confidence intervals for the population mean corresponding to the second example of data set

C.L. %	Approximate Bayesian bounds (SE)	Classical bounds	WC WSE
0	10.1575-10.2565	9.18869-11.17495	20.06
90	10.1538-10.2756	8.87019-11.49344	21.54
95	10.1520-10.2914	8.56907-11.79457	23.14
99	10.1506-10.3194	7.88792-12.47572	27.18

**Example 3:** Data obtained from Prem S Mann<sup>21</sup> 16, 14, 11, 19, 14, 17, 13, 16, 17, 18, 19, 12. Normal population distribution obtained with SAS:

$$N(\mu = 15.5, \sigma = 2.6799)$$

The corresponding (Table 3) sample mean and sample variance is

$$\bar{x} = 15.5$$

$$s^2 = 7.181818$$

**Table 3** Classical and approximate Bayesian confidence intervals for the population mean corresponding to the third example of data set

C.L. %	Approx. Bayesian bounds (SE)	Classical bounds	WC WSE
80	15.4820-15.5570	14.44556-16.55440	28.12
90	15.4794-15.5721	14.11058-16.88942	29.98
95	15.4781-15.5847	13.79727-17.20273	31.95
99	15.4770-15.6075	13.09714-17.90286	36.83

**Example 4:** Data obtained from Prem S Mann<sup>21</sup> 27, 31, 25, 33, 21, 35, 30, 26, 25, 31, 33, 30, 28. Normal population distribution obtained with SAS:

$$N(\mu = 28.846, \sigma = 3.9549)$$

The corresponding (Table 4) sample mean and sample variance is

$$s^2 = 15.641025$$

$$\bar{x} = 28.846153$$

**Table 4** Classical and approximate Bayesian confidence intervals for the population mean corresponding to the fourth example of data set

C.L.%	Approximate Bayesian bounds (SE)	Classical bounds	WC WSE
80	28.8270-28.9087	27.35878-30.33353	36.41
90	28.8242-28.9256	26.89151-30.80080	38.55
95	28.8228-28.9400	26.45604-31.2362	40.79
99	28.8217-28.9663	25.49517-32.19714	46.35

**Example 5:** Data obtained from James T et al.<sup>22</sup> 52, 33, 42, 44, 41, 50, 44, 51, 45, 38, 37, 40, 44, 50, 43. Normal population distribution obtained with SAS:

$$N(\mu = 43.6, \sigma = 5.4746)$$

The corresponding (Table 5) sample mean and sample variance is

$$\bar{x} = 43.6$$

$$s^2 = 29.971428$$

**Table 5** Classical and approximate Bayesian confidence intervals for the population mean corresponding to the fifth example of data set

C.L.%	Approximate Bayesian bounds (SE)	Classical bounds	WC WSE
80	43.5794-43.6703	41.69879-45.50121	41.83
90	43.5764-43.6902	41.11076-46.08924	43.75
95	43.5749-43.7074	40.56796-46.63204	63.3
99	43.5738-43.7395	39.39189-47.80811	50.79

**Example 6:** Data obtained from James T et al.<sup>22</sup> 52, 43, 47, 56, 62, 53, 61, 50, 56, 52, 53, 60, 50, 48, 60, 55. Normal population distribution obtained with SAS:

$$N(\mu = 53.625, \sigma = 5.4145)$$

The corresponding (Table 6) sample mean and sample variance is

$$\bar{x} = 53.625$$

$$N(\mu = 82.861, \sigma = 33.226)$$

**Example 7:** The following observations have been obtained from the collection of SAS data sets.<sup>8</sup> 50, 65, 100, 45, 111, 32, 45, 28, 60, 66, 114, 134, 150, 120, 77, 108, 112, 113, 80, 77, 69, 91, 116, 122, 37, 51, 53, 131, 49, 69, 66, 46, 131, 103, 84, 78. Normal population distribution obtained with S

$$N(\mu = 82.861, \sigma = 33.226)$$

The corresponding (Table 7) sample mean and sample variance is

$$\bar{x} = 82.8611$$

$$s^2 = 1103.951587$$

All the above tables show that the obtained Approximate Bayesian confidence intervals contain the population mean and are strictly included in their classical counterparts; also, the widths of the classical confidence intervals are more than twenty times greater than the ones corresponding to their Approximate Bayesian counterparts.

**Table 6** Classical and approximate Bayesian confidence intervals for the population mean corresponding to the sixth example of data set

C.L.%	Approximate Bayesian bounds (SE)	Classical bounds	WC WSE
80	53.6098-53.6779	51.80979-55.44021	53.31
90	53.6076-53.6932	51.25210-55.99790	55.44
95	53.6065-53.7064	50.74043-56.50957	57.75
99	53.6056-53.7315	49.63588-57.61412	63.37

**Table 7** Classical and approximate Bayesian confidence intervals for the population mean corresponding to the seventh example of data set

C.L.%	Approximate Bayesian bounds (SE)	Classical bounds	WC WSE
80	82.7072-83.4808	75.6261-90.0959	18.7
90	82.6856-83.6884	73.5052-92.2168	18.66
95	82.6751-83.8815	71.6196-94.1024	18.64
99	82.6669-84.2823	67.7793-97.9427	18.67

## Summary and conclusion

- The classical and our Approximate Bayesian models perform well. The Approximate Bayesian model has great coverage accuracy and performs better.
- The classical method used to constructing confidence intervals for the mean of a Normal population does not always yield the best coverage accuracy. In fact, the above Approximate Bayesian models perform better than their classical counterparts.
- Contrary to the classical method that uses the standard Normal and the student-t statistics, the new Approximate Bayesian approach and confidence bounds rely only on the observations that are under study.
- With the new Approximate Bayesian approach, Approximate Bayesian confidence intervals for a Normal or approximately Normal population mean are easily obtained for any level of significance.
- Bayesian Analysis contributes to reinforcing well-known statistical theories such as the Estimation and Decision-Making theories.

## Acknowledgements

None.

## Conflict of interest

The author declares no conflict of interest.

## References

1. Camara VAR, Tsokos CP. Sensitivity behavior of bayesian reliability analysis for different loss functions. *International Journal of Applied Mathematics*. 1999.
2. Camara VAR, Tsokos CP. The effect of loss functions on empirical Bayes reliability analysis. *Mathematical Problems in Engineering*. 1999;4(6):539–560.
3. Drake AW. Bayesian statistics for the reliability engineer. *Annual Symposium on Reliability*. USA; 1966. p. 315–320.
4. Britney RR, Winkler RL. *Bayesian III point estimation under various loss functions*. Proc Business and Economic Statistics Section, USA: Amer Statist Assoc; 1968. p. 356–364.
5. Canfield RV. A Bayesian approach to reliability estimation using a loss function. *IEEE Transactions on Reliability*. 1970;R-19(1):13–16.
6. Ray ES, Collins John, Luden ML, et al. *Bayesian reliability demonstration, Phase-I Data for the a priori distribution*. Griffis AFB NY RADC-TR-69-389, USA: Rome Air Development Center; 1970.
7. Schafer RE, Sheffield TS. *Bayesian Reliability Demonstration: Phase II. Development of a Priori Distribution*. Griffis AFR, NY RADC-YR-71-209, USA: Rome Air Development Center; 1971.
8. Camara VAR, Tsokos CP. Bayesian, Reliability modeling with a new loss function. *Statistica*. 1999;61(4).
9. Schafer RE, Sheffield TS, Collins JR. *Bayesian reliability demonstration Phase III- Development of test plans*. Griffis AFB, NY RADC-TR-73-39, USA: Rome Air development Center; 1973.
10. Shafer RE, Feduccia AJ. Prior distribution fitted to observed reliability data. *Prior distribution fitted to observed reliability data*. 1972;R-21(3):148–154.
11. Bhattacharya SK. Bayesian approach to life testing and reliability estimation. *Journal of the American Statistical Association*. 1967;62(317):48–62.
12. Camara VAR, Tsokos CP. Bayesian estimate of a parameter and choice of the loss function. *Nonlinear studies journal*, China; 1999.
13. Camara VAR. Approximate Bayesian confidence intervals for the variance of a gaussian distribution. *JMASM*. 2003;2(2).
14. Camara VAR. Approximate Bayesian confidence intervals for the mean of an exponential distribution versus Fisher Matrix Bounds Models. *JMASM*. 2007;6(1).
15. Camara VAR. A New Approximate Bayesian approach for decision making about the variance of a gaussian distribution versus the classical approach. *JMASM*. 2009;8(1).
16. Camara VAR. Approximate Bayesian confidence intervals for the mean of a gaussian distribution versus bayesian models. *JMASM*. 2009;8(2).
17. Camara VAR. New Approximate Bayesian Confidence. Intervals for the coefficient of variation of a normal distribution. *JMASM*. 2012;11(1):167–178.
18. M Fogel. The statistics problem solver. *Research & Education Association*. USA; 1991. p. 502–505.
19. Miller GE. Asymptotic test statistics for Coefficient of variation. *Communications in Statistics-Theory and Methods*. 1991;20(10):3351–3363.
20. Nelson W. *Applied Life Data Analysis*. New York, USA: John Wiley & Sons; 1982.
21. Prem S Mann. *Introductory Statistics*. 3rd ed. USA; 1998. 504 p.
22. James T, McClave TS. *A first course in Statistics*. 6th ed. USA; 1997. 301 p.