Journal of Modern Applied Statistical Methods

Volume 11 | Issue 1 Article 13

5-1-2012

New Approximate Bayesian Confidence Intervals for the Coefficient of Variation of a Gaussian Distribution

Vincent A. R. Camara
Research Center for Bayesian Applications, Inc., Largo, FL, gvcamara@ij.net

Recommended Citation
Camara, Vincent A. R. (2012) "New Approximate Bayesian Confidence Intervals for the Coefficient of Variation of a Gaussian Distribution," *Journal of Modern Applied Statistical Methods*: Vol. 11 : Iss. 1 , Article 13.
DOI: 10.22237/jmasm/1335845520
Available at: http://digitalcommons.wayne.edu/jmasm/vol11/iss1/13

Journal of Modern Applied Statistical Methods
May 2012, Vol. 11, No. 1, 158-166

1538 – 9472/12/$95.00

New Approximate Bayesian Confidence Intervals for the Coefficient of Variation of a Gaussian Distribution

Vincent A. R. Camara
Research Center for Bayesian Applications, Inc.,
Largo, FL

Confidence intervals are constructed for the coefficient of variation of a Gaussian distribution. Considering the square error and the Higgins-Tsokos loss functions, approximate Bayesian models are derived and compared to a published classical model. The models are shown to have great coverage accuracy. The classical model does not always yield the best confidence intervals; the proposed models often perform better.

Key words: Estimation, loss functions, confidence intervals, statistical analysis.

Introduction

A significant amount of research in Bayesian analysis and modeling has been published during the last thirty-five years. Bayesian analysis implies the exploitation of suitable prior information and the choice of a loss function in association with Bayes' Theorem. It is based on the notion that a parameter within a model is not merely an unknown quantity but rather behaves as a random variable that follows some distribution. In the area of life testing, it is 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 the life parameter.

Drake (1966) gave an excellent account of the use of Bayesian statistics in reliability problems. As he pointed out, "He (a Bayesian) realizes…that his selection of a prior (distribution) to express his present state of knowledge will necessarily be somewhat arbitrary. But he greatly appreciates this opportunity to make his entire assumptive structure clear to the world…" (Drake, 1966, p. 315-320).

This study considers a widely used and useful underlying model; that is, the normal underlying model characterized by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2};$$

$$-\infty \prec x \prec \infty,\ -\infty \prec \mu \prec \infty, \sigma \succ 0. \qquad (1)$$

For a given level of significance, the published classical model obtained by Miller (1991) considers the following confidence bounds for the coefficient of variation of a normal distribution.

$$L_M = \frac{s}{\overline{X}} - Z_{\alpha/2}\left[m^{-1}\left(\frac{s}{\overline{X}}\right)^2\left(0.5 + \left(\frac{s}{\overline{X}}\right)^2\right)\right]^{\frac{1}{2}}$$

Vincent A. R. Camara earned a Ph.D. in Mathematics/Statistics. His research interests include the theory and applications of Bayesian and empirical Bayes analyses with emphasis on the computational aspect of modeling. He is featured in Marquis Who's Who in America and in Marquis Who's Who in the World. E-mail him at: gvcamara@ij.net.

and

$$U_M = \frac{s}{\overline{X}} + Z_{\alpha/2}\left[m^{-1}\left(\frac{s}{\overline{X}}\right)^2 \left(0.5 + \left(\frac{s}{\overline{X}}\right)^2\right)\right]^{\frac{1}{2}} \tag{2}$$

with m=n−1, where n is the sample size. This classical approach uses the Z-table to construct confidence intervals for a normal population coefficient of variation. This article relies only on observations under study to construct new approximate Bayesian confidence intervals for the coefficient of variation of a normal population.

Methodology

For model (1), the following results will be used to derive the approximate Bayesian confidence intervals for the coefficient of variation of a normal distribution: approximate Bayesian confidence bounds for a normal population variance (Camara 2003) and approximate Bayesian confidence bounds for a normal population mean (Camara, 2009). Although no specific analytical procedure is available that allows the identification of the appropriate loss function to be used, the most common is the square error loss function. One reason for selecting this loss function is due to its analytical tractability in Bayesian analysis. As this study shows, selecting the square error loss does not always lead to the best approximate Bayesian confidence intervals. However, the obtained approximate Bayesian confidence intervals corresponding to the square error and the Higgins-Tsokos loss functions will be respectively compared to the classical model (2).

Square Loss Error Function

The popular square error loss function places a small weight on estimates near the true value and proportionately more weight on extreme deviations 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:

$$L_{SE}(\hat{\theta},\theta) = \left(\hat{\theta} - \theta\right)^2 \tag{3}$$

Higgins-Tsokos Loss Function

The Higgins-Tsokos loss function places a heavy penalty on extreme over- or under-estimation. That is, it places an exponential weight on extreme errors. The Higgins-Tsokos loss function is defined as:

$$L_{HT}(\hat{\theta},\theta) = \frac{f_1 e^{f_2(\hat{\theta}-\theta)} + f_2 e^{-f_1(\hat{\theta}-\theta)}}{f_1 + f_2} - 1,$$

$$f_1, f_2 \succ 0 \tag{4}$$

The Pareto prior was selected due to its mathematical tractability. The Pareto prior defined as follows:

$$f_1(\theta) = \frac{a}{b}\left(\frac{b}{\theta}\right)^{a+1}; \quad \theta \geq b \succ 0, \quad a \succ 0. \tag{5}$$

where $\theta = 1/\sigma^2$.

The use of good approximations of the Pareto prior (5) along with the square error loss and the Higgins-Tsokos loss leads to the approximate Bayesian confidence bounds for a normal population variance and positive mean (Camara, 2003). For the square error loss function

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

and

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

For the Higgins-Tsokos loss function

$$L_{\sigma^2((HT)} = \frac{1}{\dfrac{n-1-2Ln(\alpha/2)}{\sum_{i=1}^{n}(x_i-\bar{x})^2} - \dfrac{1}{f_1+f_2} Ln\left(\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2+f_2}{\sum_{i=1}^{n}(x_i-\bar{x})^2-f_1}\right)}$$

and

$$U_{\sigma^2(HT)} = \frac{1}{\dfrac{n-1-2Ln(1-\alpha/2)}{\sum_{i=1}^{n}(x_i-\bar{x})^2} - \dfrac{1}{f_1+f_2} Ln\left(\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2+f_2}{\sum_{i=1}^{n}(x_i-\bar{x})^2-f_1}\right)} \qquad (7)$$

The approximate Bayesian confidence interval for a positive normal population mean corresponding to the square error loss (Camara, 2009) is:

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

and

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

The approximate Bayesian confidence bounds for a positive normal population mean corresponding to the Higgins-Tsokos loss function (Camara, 2009) is:

$$L_{\mu(HT)} = \left(s^2-\bar{x}^2-\frac{1}{\dfrac{n-1-2Ln(1-\frac{\alpha}{2})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} - \dfrac{1}{f_1+f_2} Ln\left(\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2+f_2}{\sum_{i=1}^{n}(x_i-\bar{x})^2-f_1}\right)}\right)^{0.5}$$

and

$$U_{\mu(HT)} = \left(s^2+\bar{x}^2-\frac{1}{\dfrac{n-1-2Ln(1-\frac{\alpha}{2})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} - \dfrac{1}{f_1+f_2} Ln\left(\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2+f_2}{\sum_{i=1}^{n}(x_i-\bar{x})^2-f_1}\right)}\right)^{0.5} \qquad (9)$$

Using Equations (6), (7), (8) and (9), the approximate Bayesian confidence intervals are obtained for the coefficient of variation of a normal population corresponding to the square error loss function:

$$L_{Cv}(SE) = \left(\frac{n-1}{\left(1+\left(\frac{\bar{x}}{s}\right)^2\right)\left(n-2-2Ln(\alpha/2)\right)-n+1}\right)^{0.5}$$

and

$$U_{Cv}(SE) = \left(\frac{n-1}{\left(1+\left(\frac{\bar{x}}{s}\right)^2\right)\left(n-2-2Ln(1-\alpha/2)\right)-n+1}\right)^{0.5} \qquad (10)$$

The approximate Bayesian confidence interval for the coefficient of variation of a normal population corresponding to the Higgins-Stokes Loss functions are:

$$L_{Cvr}(HT) = \frac{1}{\left(\left(s^2 + \bar{x}^2\right) \left(\frac{n-1-2Ln\left(\frac{\alpha}{2}\right)}{\sum_{i=1}^{n}(x-\bar{x})^2} - \frac{1}{f_1 - f_2} Ln\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 + f_2}{\sum_{i=1}^{n}(x_i-\bar{x})^2 + f_1} \right) \right) - 1 \right)^{0.5}}$$

and

$$U_{Cvr}(HT) = \frac{1}{\left(\left(s^2 + \bar{x}^2\right) \left(\frac{n-1-2Ln\left(1-\frac{\alpha}{2}\right)}{\sum_{i=1}^{n}(x-\bar{x})^2} - \frac{1}{f_1 - f_2} Ln\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 + f_2}{\sum_{i=1}^{n}(x_i-\bar{x})^2 + f_1} \right) \right) - 1 \right)^{0.5}} \quad (11)$$

Results

To assess the proposed models, numerical results were obtained using SAS along with samples from normally distributed populations (Examples 1, 2, 3, .4, 7) as well as approximately normal populations (Examples 5, 6). Results were then compared to those obtained with the classical approach (2); for the Higgins-Tsokos loss function, f1 = 1 and f2 = 1 were considered. WC corresponds to the widths of the classical confidence intervals, and WSE and WHT denote the widths of the approximate Bayesian confidence intervals corresponding to the square error and the Higgins-Tsokos loss functions.

Example 1

Data: 24, 28, 22, 25, 24, 22, 29, 26, 25, 28, 19, 29 (Mann, 1998, p. 504).

Normal population distribution obtained with SAS:

Table 1: Classical (2) and Approximate Bayesian Confidence Intervals for the Population Coefficient of Variation Corresponding To Data Set 1

C.L.%	Approximate Bayesian Bounds (SE)	Approximate Bayesian Bounds (HT)	Classical Bounds I
80	0.1077-0.1291	0.1132-0.1276	0.0899-0.1587
90	0.1028-0.1298	0.1077-0.1291	0.0799-0.1687
95	0.0986-0.1301	0.1028-0.1298	0.0716-0.1770
99	0.0905-0.1303	0.0937-0.1303	0.0549-0.1937

C.L.%	Lengths of Approximate Bayesian Intervals (SE)	Lengths of Approximate Bayesian Intervals (HT)	WC / WSE	WC / WHT
80	0.0214	0.0144	3.215	4.778
90	0.0270	0.0214	3.289	4.150
95	0.0315	0.0270	3.346	3.904
99	0.0398	0.0366	3.487	3.792

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

The corresponding coefficient of variation is:

$$\bar{C}_v = \frac{\sigma}{\mu} = 0.124$$

Example 2

Data: 13, 11, 9, 12, 8, 10, 5, 10, 9, 12, 13 (Mann, 1998, p. 504).

Normal population distribution obtained with SAS:

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

The corresponding coefficient of variation is:

$$\bar{C}_v = \frac{\sigma}{\mu} = 0.236$$

Example 3

Data: 16, 14, 11, 19, 14, 17, 13, 16, 17, 18, 19, 12 (Mann, 1998, p. 504).

Normal population distribution obtained with SAS:

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

The corresponding coefficient of variation is:

$$\bar{C}_v = \frac{\sigma}{\mu} = 0.173$$

Example 4

Data: 27, 31, 25, 33, 21, 35, 30, 26, 25, 31, 33, 30, 28 (Mann, 1998, p. 504).

Normal population distribution obtained with SAS:

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

The corresponding coefficient of variation is:

$$\bar{C}_v = \frac{\sigma}{\mu} = 0.137$$

Example 5

Data: 52, 33, 42, 44, 41, 50, 44, 51, 45, 38,37,40,44, 50, 43 (McClave & Sincich, 1997, p. 301).

Normal population distribution obtained with SAS:

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

The corresponding coefficient of variation is:

$$\bar{C}_v = \frac{\sigma}{\mu} = 0.126$$

Example 6

Data: 52, 43, 47, 56, 62, 53, 61, 50, 56, 52, 53, 60, 50, 48, 60, 55 (McClave & Sincich, 1997, p. 301).

Normal population distribution obtained with SAS:

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

The corresponding coefficient of variation is:

$$\bar{C}_v = \frac{\sigma}{\mu} = 0.101$$

Example 7

The following observations were obtained from the collection of SAS data sets: 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 SAS:

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

The corresponding coefficient of variation is:

$$\bar{C}_v = \frac{\sigma}{\mu} = 0.401$$

Table 2 Classical (2) and Approximate Bayesian Confidence Intervals for the Population Coefficient of Variation Corresponding To Data Set 2

C.L.%	Approximate Bayesian Bounds (SE)	Approximate Bayesian Bounds (HT)	Classical Bounds I
80	0.2007-0.2463	0.2122-0.2430	0.1647-0.3069
90	0.1908-0.2478	0.2007-0.2463	0.1441-0.3275
95	0.1823-0.2486	0.1908-0.2478	0.1269-0.3447
99	0.1662-0.2492	0.1724-0.2490	0.0924-0.3792

C.L.%	Lengths of Approximate Bayesian Intervals (SE)	Lengths of Approximate Bayesian Intervals (HT)	WC / WSE	WC / WHT
80	0.0456	0.0308	3.118	4.617
90	0.0570	0.0456	3.218	4.022
95	0.0663	0.0570	3.285	3.821
99	0.0830	0.0766	3.455	3.744

Table 3: Classical (2) and Approximate Bayesian Confidence Intervals for the Population Coefficient of Variation Corresponding To Data Set 3

C.L.%	Approximate Bayesian Bounds (SE)	Approximate Bayesian Bounds (HT)	Classical Bounds I
80	0.1495-0.1797	0.1573-0.1776	0.1243-0.2215
90	0.1427-0.1807	0.1495-0.1797	0.1103-0.2354
95	0.1368-0.1811	0.14273-0.1807	0.0985-0.2473
99	0.1255-0.1815	0.1300-0.1842	0.0750-0.2708

C.L.%	Lengths of Approximate Bayesian Intervals (SE)	Lengths of Approximate Bayesian Intervals (HT)	WC / WSE	WC / WHT
80	0.0302	0.0203	3.219	4.788
90	0.0380	0.0302	3.292	4.142
95	0.0443	0.0380	3.359	3.916
99	0.0560	0.0542	3.496	3.613

Table 4: Classical (2) and Approximate Bayesian Confidence Intervals for the Population Coefficient of Variation Corresponding To Data Set 4

C.L.%	Approximate Bayesian Bounds (SE)	Approximate Bayesian Bounds (HT)	Classical Bounds I
80	0.1200-0.1419	0.1258-0.1404	0.1006-0.1736
90	0.1150-0.1426	0.1200-0.1419	0.0901-0.18414
95	0.1104-0.1430	0.1149-0.1426	0.0812-0.1930
99	0.1018-0.1433	0.1052-0.1432	0.0636-0.2107

C.L.%	Lengths of Approximate Bayesian Intervals (SE)	Lengths of Approximate Bayesian Intervals (HT)	WC / WSE	WC / WHT
80	0.0219	0.0146	3.333	5.000
90	0.0276	0.0219	3.406	4.292
95	0.0326	0.0277	3.429	4.036
99	0.0415	0.0380	3.545	3.871

Table 5: Classical (2) and Approximate Bayesian Confidence Intervals for the Population Coefficient of Variation Corresponding To Data Set 5

C.L.%	Approximate Bayesian Bounds (SE)	Approximate Bayesian Bounds (HT)	Classical Bounds I
80	0.1118-0.1293	0.1166-0.1282	0.09472-0.1564
90	0.1076-0.1299	0.1118-0.1293	0.08580-0.1653
95	0.1039-0.1301	0.1076-0.1299	0.0783-0.1728
99	0.0964-0.1303	0.0994-0.1303	0.0634-0.1877

C.L.%	Lengths of Approximate Bayesian Intervals (SE)	Lengths of Approximate Bayesian Intervals (HT)	WC / WSE	WC / WHT
80	0.0175	0.0116	3.526	5.314
90	0.0223	0.0175	3.565	4.543
95	0.0262	0.0223	3.607	4.238
99	0.0339	0.0309	3.667	4.023

Table 6: Classical (2) and Approximate Bayesian Confidence Intervals for the Population Coefficient of Variation Corresponding To Data Set 6

C.L.%	Approximate Bayesian Bounds (SE)	Approximate Bayesian Bounds (HT)	Classical Bounds I
80	0.0906-0.1038	0.0942-0.1029	0.0771-0.1248
90	0.0874-0.1042	0.0906-0.1038	0.0702-0.1317
95	0.0845-0.1044	0.0874-0.1042	0.0645-0.1375
99	0.0787-0.1045	0.0810-0.1045	0.0529-0.1490

C.L.%	Lengths of Approximate Bayesian Intervals (SE)	Lengths of Approximate Bayesian Intervals (HT)	$\frac{WC}{WSE}$	$\frac{WC}{WHT}$
80	0.0132	0.0087	3.614	5.483
90	0.0168	0.0132	3.661	4.659
95	0.0199	0.0168	3.668	4.345
99	0.0258	0.0235	3.725	4.089

Table 7: Classical (2) and Approximate Bayesian Confidence Intervals for the Population Coefficient of Variation Corresponding To Data Set 7

C.L.%	Approximate Bayesian Bounds (SE)	Approximate Bayesian Bounds (HT)	Classical Bounds I
80	0.3790-0.4063	0.3870-0.4047	0.3305-0.4715
90	0.3714-0.4071	0.3790-0.4063	0.3101-0.4919
95	0.3643-0.4075	0.3714-0.4071	0.2930-0.5090
99	0.3492-0.4077	0.3598-0.4077	0.2588-0.5431

C.L.%	Lengths of Approximate Bayesian Intervals (SE)	Lengths of Approximate Bayesian Intervals (HT)	$\frac{WC}{WSE}$	$\frac{WC}{WHT}$
80	0.0273	0.0177	5.165	7.966
90	0.0357	0.0273	5.092	6.660
95	0.0432	0.0357	5.000	6.050
99	0.0585	0.0479	4.860	5.935

Each of the seven randomly selected samples shows that the proposed approximate Bayesian confidence intervals contain the population coefficient of variation. The models are also strictly included in their counterparts obtained with classical approach (2). For each of the seven examples, the proposed approximate Bayesian approach has greater coverage accuracy than its classical counterpart (2).

Conclusion

New approximate Bayesian confidence intervals for the coefficient of variation of a normal population under two different loss functions were derived. The loss functions employed were the square error and the Higgins-Tsokos. Based on the above numerical results the following conclusions are put forth:

- The proposed approach and models rely only on the observations under study, contrary to the classical approach (2) that uses the standard normal distribution, The classical approach (2) used to constructing confidence intervals for the coefficient of variation of a normal population does not always perform better than the approximate Bayesian approach. In fact, in each of the examples provided, the obtained approximate Bayesian confidence intervals had greater coverage accuracy than those obtained with the classical approach (2). Each of the obtained approximate Bayesian confidence intervals contains the population coefficient of variation, and is strictly included in its classical counterpart obtained with the classical approach (2).

- With the proposed approach, approximate Bayesian confidence intervals for a normal population coefficient of variation may be obtained for any level of significance and any sample size.

- The approximate Bayesian approach under the square error loss function does not always yield the best approximate Bayesian results. In fact, in the examples provided, the Higgins-Tsokos loss function performs better.

- Bayesian analysis contributes to reinforcing well-known statistical theories such as the estimation theory.

References

Miller, S. G. (1991), Asymptotic test statistics for Coefficient of variation. *Communications in Statistics – Theory and Methods*, *20*(*10*), 3351-3363.

Camara, V. A. R. (2003). Approximate Bayesian confidence intervals for the variance of a Gaussian distribution. *Journal of Modern Applied Statistical Methods*, *2*(*2*), 350-358.

Higgins, J. J., & Tsokos, C. P. (1976). Comparison of Bayes estimates of failure intensity for fitted priors of life data. In *Proceedings of the Conference on the Theory an Applications of Reliability with Emphasis on Bayesian and Nonparametric Methods*. New York, NY: Academic.

Mann, P. S. (1998). Introductory Statistics, 3rd Edition. New York, NY: Wiley.

McClave, J. T., & Sincich, T. A. (1997). First course in Statistics, 6th Ed. New York, NY: Prentice Hall.

Camara, V. A. R. (2009). Approximate Bayesian confidence intervals for the mean of a Gaussian distribution versus Bayesian models. *Journal of Modern Applied Statistical Methods*, *8*(*2*), 526-533.

Camara, V. A. R., & Tsokos, C. P. (1998). *Bayesian reliability modeling with applications.* Michigan: UMI Publishing Company.

Drake, A. W. (1966). Bayesian statistics for the reliability eng. proc. *Annual Symposium on Reliability*, 315-320.

Camara, V. A. R. (2011). New approximate Bayesian confidence intervals for the coefficient of variation of a Gaussian distribution. In *JSM Proceedings*, *Statistical Computing Section*, 1121-1131. Alexandria, VA: American Statistical Association.