Research Article

A Simulation Comparison of New Confidence Intervals for the Coefficient of Variation of a Poisson Distribution

Wararit Panichkitkosolkul*

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Phathumthani, Thailand

**Corresponding author. E-mail address: wararit@mathstat.sci.tu.ac.th*

Received July 6, 2010; Accepted December 13, 2010

Abstract

This paper proposes four new confidence intervals for the coefficient of variation of a Poisson distribution based on obtaining confidence intervals for the Poisson mean. The following confidence intervals are considered: confidence intervals for the coefficient of variation of a Poisson distribution based on Wald (W), Wald with continuity correction (WCC), Scores (S) and Variance stabilizing (VS) confidence interval. Using Monte Carlo simulations, the coverage probabilities and lengths of these confidence intervals are compared. Simulation results have shown that the confidence interval based on WCC has desired closeness coverage probabilities of 0.95 and 0.90. Additionally, the lengths of newly proposed confidence intervals are slightly different. Therefore, the confidence interval based on WCC is more suitable than the other three confidence intervals in terms of the coverage probability.

Key Words: Coefficient of variation; Confidence interval; Coverage probability; Length; Poisson distribution

Introduction

The coefficient of variation is a dimensionless number that quantifies the degree of variability relative to the mean (Kelley, 2007). The population coefficient of variation is defined as

$$\kappa = \frac{\sigma}{\mu} , \qquad (1)$$

where σ is the population standard deviation and μ is the population mean. The typical sample estimate of κ is given as

$$\hat{\kappa} = \frac{S}{\bar{X}} , \qquad (2)$$

where S is the sample standard deviation, the square root of the unbiased estimator of the variance, and $\bar{X}$ is the sample mean.

The coefficient of variation has long been a widely used descriptive and inferential quantity in many applications of science, economics and others. In chemical experiments, the coefficient of variation is often used as a yardstick of precision for measurements. For example, two measurement methods may be used to compare precision on the basis of their respective coefficients of variation. The coefficient of variation can be used to measure relative risks (Miller and Karson, 1977) in finance and actuarial science. Furthermore, testing the equality of the coefficients of variation for two stocks can help determine whether the two stocks possess the same risk or not. In physiology, the coefficient of variation can also be applied to assess homogeneity of bone test samples (Hamer et al., 1995). In the field of safety engineering, the coefficient of variation is used as a tool in the uncertainty of fault trees analysis (Ahn, 1995). Additionally, the coefficient of variation is also employed in assessing the strength of ceramics (Gong and Li, 1999).

Although the point estimator of coefficient of variation can be a useful measure, the greatest use of it is to construct a confidence interval of coefficient of variation for the quantity of interest. (Mahmoudvand and Hassani, 2009), since a confidence interval provides much more information about the population value of the quantity of interest than does a point estimate (e.g., Smithson, 2001; Thompson, 2002; Steiger, 2004).

An approximate $(1-\alpha)100\%$ confidence interval for the coefficient of variation (see e.g. Vangel, 1996) is given by

$$CI = \left\{ \frac{\hat{\kappa}}{\sqrt{t_1(\theta_1\hat{\kappa}^2+1)-\hat{\kappa}^2}}, \frac{\hat{\kappa}}{\sqrt{t_2(\theta_2\hat{\kappa}^2+1)-\hat{\kappa}^2}} \right\}, \quad (3)$$

where $\nu = n-1$, $t_1 \equiv \chi^2_{\nu,1-\alpha/2}/\nu$, $t_2 \equiv \chi^2_{\nu,\alpha/2}/\nu$ and $\theta = \theta(\nu,\alpha)$ is a known function selected so that a random variable $W_\nu \equiv Y_\nu/\nu$, where Y_ν has a χ^2_ν distribution, has approximately the same distribution as a pivotal quantity $Q \equiv \frac{\hat{\kappa}^2(1+\kappa^2)}{(1+\theta\hat{\kappa}^2)\kappa^2}$. This pivotal quantity can be used to construct hypothesis tests and a confidence interval for κ.

McKay (1932) proposed that the choice $\theta = \frac{\nu}{\nu+1}$ gives a good approximation for the confidence interval in equation (3), but he was unable to investigate the small-sample distribution of Q. McKay's approximate confidence interval is

$$CI_{01} = \left\{ \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,1-\alpha/2}}{\nu+1}-1\right)\hat{\kappa}^2 + \frac{\chi^2_{\nu,1-\alpha/2}}{\nu}\right]^{-1/2}, \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,\alpha/2}}{\nu+1}-1\right)\hat{\kappa}^2 + \frac{\chi^2_{\nu,\alpha/2}}{\nu}\right]^{-1/2} \right\}, \quad (4)$$

where $\nu = n-1$ is the degrees of freedom of the χ^2 distribution. Several authors have carried out numerical investigations of the accuracy of McKay's confidence interval. For instance, Iglewicz and Myers (1970) had compared McKay's confidence interval with the exact confidence interval based on the noncentral t distribution and they found that McKay's confidence interval is efficient for $n \geq 10$ and $0 < \kappa < 0.3$.

Vangel (1996) proposed a new confidence interval for the coefficient of variation which he called the modified McKay's confidence interval. He proposed the use of the function θ where $\theta = \frac{\nu}{\nu+1}\left[\frac{2}{\chi^2_{\nu,\alpha}}+1\right]$. He also suggested that the modified McKay method gave confidence intervals for the coefficient of variation that are closely related to the McKay's confidence interval but they are usually more accurate. The modified McKay's confidence interval for a coefficient of variation is given by

$$CI_{02} = \left\{ \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,1-\alpha/2}+2}{\nu+1}-1\right)\hat{\kappa}^2 + \frac{\chi^2_{\nu,1-\alpha/2}}{\nu}\right]^{-1/2}, \right.$$

$$\hat{\kappa}\left[\left(\frac{\chi^2_{\nu,\alpha/2}+2}{\nu+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{\nu,\alpha/2}}{\nu}\right]^{-1/2}\Bigg\} . \quad (5)$$

When data are normally distributed, McKay's confidence interval and the modified McKay's confidence interval, CI_{01} and CI_{02}, can be used very well in terms of coverage probability and length. However, for non-normal data, these confidence intervals cannot be used practically. The aim in this paper is to construct the new confidence intervals for the coefficient of variation of the Poisson distribution. The modified confidence intervals for the coefficient of variation are obtained from applying confidence intervals for the Poisson mean. Additionally, the coverage probabilities and the lengths of new confidence intervals for a coefficient of variation are compared through a Monte Carlo simulation study.

The paper is organized as follows. In the next section, new confidence intervals for the coefficient of variation of a Poisson distribution are presented. Simulation results obtained from the Monte Carlo simulation and discussions are shown in the third section. The conclusions are presented in the final section.

New Confidence Intervals for the Coefficient of Variation of a Poisson Distribution

In this section the new confidence intervals for the coefficient of variation of a Poisson distribution are presented. Newly proposed confidence intervals are based on confidence intervals for the Poisson mean. Suppose $X_i \sim Poi(\lambda)$, $i=1,2,...,n$. Hence, the population coefficient of variation for a Poisson distribution is given by

$$\kappa = \frac{\sigma}{\mu} = \frac{\sqrt{\lambda}}{\lambda} = \frac{1}{\sqrt{\lambda}} .$$

In order to construct new confidence intervals, there are first mentioned confidence intervals for the Poisson mean. These confidence intervals considered are: (Barker, 2002)

(1) Wald (W) confidence interval. The W confidence interval is derived from the asymptotic standard normal distribution of $(\bar{X}-\lambda)/\sqrt{\bar{X}/n}$. This quantity can be inverted to provide the confidence interval

$$\left(\bar{X}-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}}\ ,\ \bar{X}+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}}\right). \quad (6)$$

(2) Wald with continuity correction (WCC) confidence interval. The W confidence interval uses a continuous distribution (normal) to approximate a discrete distribution (Poisson). A continuity correction might make this approximation more accurate. The WCC confidence interval is given by

$$\left(\bar{X}-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}\ ,\ \bar{X}+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}\right). \quad (7)$$

(3) Scores (S) confidence interval. The S confidence interval is derived from the asymptotic standard normality of $(\bar{X}-\lambda)/\sqrt{\lambda/n}$. This quantity can be inverted to provide the S confidence interval

$$\left(\bar{X}+\frac{(Z_{1-\alpha/2})^2}{2n}-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\bar{X}+\frac{(Z_{1-\alpha/2})^2}{n}}{4n}},\right.$$

$$\left.\bar{X}+\frac{(Z_{1-\alpha/2})^2}{2n}+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\bar{X}+\frac{(Z_{1-\alpha/2})^2}{n}}{4n}}\right). \quad (8)$$

(4) Variance stabilizing (VS) confidence interval. The quantity $(\sqrt{\bar{X}}-\sqrt{\lambda})/\sqrt{1/4n}$ is the asymptotically standard normal. This can be inverted into the confidence interval

$$\left(\bar{X} + \frac{(Z_{1-\alpha/2})^2}{4n} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}},\right.$$

$$\left.\bar{X} + \frac{(Z_{1-\alpha/2})^2}{4n} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}}\right), \tag{9}$$

where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $Z_{1-\frac{\alpha}{2}}$ is a $\left(1-\frac{\alpha}{2}\right)th$ quantile of the standard normal distribution. From Equations (6)-(9), we therefore can derive the confidence intervals for a coefficient of variation of a Poisson distribution based on the above confidence intervals for the Poisson mean as follows:

$$1-\alpha = P\left(L_i < \lambda < U_i\right)$$

$$= P\left(\sqrt{L_i} < \sqrt{\lambda} < \sqrt{U_i}\right)$$

$$= P\left(\frac{1}{\sqrt{U_i}} < \frac{1}{\sqrt{\lambda}} < \frac{1}{\sqrt{L_i}}\right)$$

$$= P\left(\frac{1}{\sqrt{U_i}} < \kappa < \frac{1}{\sqrt{L_i}}\right). \tag{10}$$

where L_i and U_i, $i = 1,2,3,4$ denote the lower and upper limit of confidence intervals for the Poisson mean based on W, WCC, S, and VS, respectively.

Hence, we obtain $(1-\alpha)100\%$ four new confidence intervals for the coefficient of variation of a Poisson distribution which are

$$CI_1 = \left\{\left(\sqrt{\bar{X} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}}}\right)^{-1}, \left(\sqrt{\bar{X} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}}}\right)^{-1}\right\}, \tag{11}$$

$$CI_2 = \left\{\left(\sqrt{\bar{X} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}}\right)^{-1}, \left(\sqrt{\bar{X} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}}\right)^{-1}\right\}, \tag{12}$$

$$CI_3 = \left\{\left(\sqrt{\bar{X} + \frac{(Z_{1-\alpha/2})^2}{2n} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\bar{X} + \frac{(Z_{1-\alpha/2})^2}{n}}{4n}}}\right)^{-1}, \left(\sqrt{\bar{X} + \frac{(Z_{1-\alpha/2})^2}{2n} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\bar{X} + \frac{(Z_{1-\alpha/2})^2}{n}}{4n}}}\right)^{-1}\right\}, \tag{13}$$

$$CI_4 = \left\{\left(\sqrt{\bar{X} + \frac{(Z_{1-\alpha/2})^2}{4n} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}}}\right)^{-1}, \left(\sqrt{\bar{X} + \frac{(Z_{1-\alpha/2})^2}{4n} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}}}\right)^{-1}\right\}, \tag{14}$$

where $CI_i, i = 1,2,3,4$ denote the confidence intervals for the coefficient of variation of a Poisson distribution based on W, WCC, S, and VS confidence interval, respectively.

To study the different confidence intervals, we consider their coverage probability and length. For each of the methods considered, we obtain a $(1-\alpha)100\%$ confidence interval denoted by

(L, U) (based on M replicates) and estimated the coverage probability and the length, respectively, by

$$\widehat{1-\alpha} = \frac{\#(L \le \kappa \le U)}{M},$$

and $$\widehat{Length} = \frac{\sum_{j=1}^{M}(U_j - L_j)}{M}.$$

Results and Discussions

In this section, the performance of the estimated coverage probabilities of the new asymptotic confidence intervals (11), (12), (13) and (14) and their lengths was examined via Monte Carlo simulations. Data are generated from the Poisson distribution with κ = 0.1, 0.2 and 0.3 sample sizes; n = 10, 15, 25, 50 and 100. All simulations were performed using programs written in the R statistical software (The R Development Core Team, 2009a, 2009b) with the number of simulation runs, M = 50,000 at the level of significance $\alpha = 0.05$ and 0.10. The simulation results are shown in Tables 1 and 2. The following information is presented here: the estimated coverage probabilities of the confidence intervals, CI_1, CI_2, CI_3 and CI_4, and their lengths for a Poisson distribution at α = 0.05 and 0.10, respectively. As can be seen from Tables 1 and 2, the confidence interval based on WCC, CI_2, has a closeness coverage probability of $1-\alpha$ for all sample sizes and values of κ except when $n = 10$, $\kappa = 0.2$ and $\alpha = 0.10$. The other three confidence

Table 1 The estimated coverage probabilities and lengths of a 95% confidence interval in (11), (12), (13) and (14) for a Poisson distribution.

n	κ	Coverage probabilities				Lengths			
		CI_1	CI_2	CI_3	CI_4	CI_1	CI_2	CI_3	CI_4
10	0.1	0.9506	0.9521	0.9488	0.9508	0.0062	0.0062	0.0062	0.0062
	0.2	0.9502	0.9528	0.9468	0.9510	0.0251	0.0254	0.0249	0.0250
	0.3	0.9444	0.9564	0.9481	0.9471	0.0576	0.0590	0.0563	0.0568
15	0.1	0.9505	0.9521	0.9491	0.9491	0.0051	0.0051	0.0051	0.0051
	0.2	0.9506	0.9529	0.9491	0.9481	0.0204	0.0206	0.0203	0.0204
	0.3	0.9473	0.9559	0.9484	0.9488	0.0465	0.0476	0.0458	0.0461
25	0.1	0.9494	0.9506	0.9481	0.9493	0.0039	0.0039	0.0039	0.0039
	0.2	0.9499	0.9518	0.9471	0.9495	0.0158	0.0159	0.0157	0.0157
	0.3	0.9512	0.9537	0.9475	0.9478	0.0357	0.0365	0.0354	0.0355
50	0.1	0.9491	0.9499	0.9490	0.9490	0.0028	0.0028	0.0028	0.0028
	0.2	0.9511	0.9530	0.9512	0.9514	0.0111	0.0112	0.0111	0.0111
	0.3	0.9470	0.9515	0.9478	0.9470	0.0251	0.0257	0.0250	0.0250
100	0.1	0.9503	0.9508	0.9494	0.9501	0.0020	0.0020	0.0020	0.0020
	0.2	0.9507	0.9533	0.9498	0.9510	0.0079	0.0079	0.0078	0.0078
	0.3	0.9515	0.9546	0.9511	0.9512	0.0177	0.0181	0.0177	0.0177

intervals, CI_1, CI_3, and CI_4, give slightly lower coverage probabilities than $1-\alpha$. The estimated coverage probabilities of the CI_1 and CI_2 increase as the values of κ get larger (i.e. for CI_2, n=10 and $\alpha = 0.05$, 0.9521 for κ = 0.1; 0.9528 for κ = 0.2; 0.9564 for κ = 0.3). The lengths of all confidence intervals are slightly different. Further, the lengths increase as the values of κ get larger (i.e. for CI_2, n=10 and $\alpha = 0.05$, 0.0062 for κ = 0.1; 0.0254 for κ = 0.2; 0.0590 for κ = 0.3). Moreover, when the sample sizes increase, the lengths are shorter (i.e. for CI_2, κ=0.1 and $\alpha = 0.05$, 0.0062 for n=10; 0.0051 for n=15; 0.0039 for n=25; 0.0028 for n =50; 0.0020 for n=100).

Conclusions

Four new confidence intervals for the coefficient of variation of the Poisson distribution have been developed. The proposed confidence intervals are compared through a Monte Carlo simulation study. The new confidence intervals are based on a confidence interval for the Poisson mean. The confidence interval based on WCC has closeness coverage probabilities $1-\alpha$. In addition, the lengths of all of the confidence intervals are slightly different. Therefore, if a confidence interval with a closeness coverage probability equal to a pre-specified value is preferred, the confidence interval based on WCC is preferable to the other three confidence intervals.

Table 2 The estimated coverage probabilities and lengths of a 90% confidence interval in (11), (12), (13) and (14) for a Poisson distribution.

n	κ	Coverage probabilities				Lengths			
		CI_1	CI_2	CI_3	CI_4	CI_1	CI_2	CI_3	CI_4
10	0.1	0.9003	0.9003	0.9037	0.9006	0.0052	0.0052	0.0052	0.0052
	0.2	0.8980	0.8980	0.9052	0.8984	0.0210	0.0213	0.0209	0.0210
	0.3	0.8895	0.9104	0.9041	0.9023	0.0480	0.0491	0.0472	0.0475
15	0.1	0.9007	0.9007	0.8985	0.9013	0.0043	0.0043	0.0042	0.0043
	0.2	0.9009	0.9009	0.8967	0.9018	0.0171	0.0173	0.0170	0.0171
	0.3	0.9033	0.9033	0.8967	0.8958	0.0388	0.0397	0.0384	0.0386
25	0.1	0.9000	0.9024	0.9026	0.9004	0.0033	0.0033	0.0033	0.0033
	0.2	0.8961	0.9001	0.9011	0.8963	0.0132	0.0133	0.0132	0.0132
	0.3	0.8999	0.9048	0.9010	0.8946	0.0299	0.0306	0.0297	0.0298
50	0.1	0.8986	0.9002	0.9007	0.8991	0.0023	0.0023	0.0023	0.0023
	0.2	0.9000	0.9058	0.9030	0.9003	0.0093	0.0094	0.0093	0.0093
	0.3	0.8958	0.9042	0.9007	0.9003	0.0210	0.0215	0.0210	0.0210
100	0.1	0.8992	0.9004	0.8992	0.8992	0.0016	0.0016	0.0016	0.0016
	0.2	0.8981	0.9024	0.9007	0.8987	0.0066	0.0067	0.0066	0.0066
	0.3	0.9004	0.9065	0.8977	0.9008	0.0148	0.0152	0.0148	0.0148

Acknowledgements

The author is very appreciate of valuable insights and comments provided by two anonymous referees, leading to many improvements in the paper. The author also wishes to thank Sheldon Daniels for his careful reading.

References

Ahn, K. (1995) Use of coefficient of variation for uncertainty analysis in fault tree analysis. *Reliability Engineering and System Safety* 47: 229-230.

Barker, L. (2002) A comparison of nine confidence intervals for a Poisson parameter when the expected number of events is ≤ 5. *American Statistician* 56: 85-89

Gong, J. and Li, Y. (1999) Relationship between the estimated Weibull modulus and the coefficient of variation of the measured strength for ceramics. *Journal of the American Ceramic Society* 82: 449-452.

Hamer, A.J., Strachan, J.R., Black, M.M., Ibbotson, C., and Elson, R.A. (1995) A new method of comparative bone strength measurement. *Journal of Medical Engineering and Technology* 19: 1-5.

Iglewicz, B. and Myers, R.H. (1970) Comparisons of approximations to the percentage points of the sample coefficient of variation. *Technometrics* 12: 166-169.

Kelley, K. (2007) Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behavior Research Methods* 39: 755-766.

Mahmoudvand, R. and Hassani, H. (2009) Two new confidence intervals for the coefficient of variation in a normal distribution. *Journal of Applied Statistics* 36: 429-442.

McKay, A.T. (1932) Distribution of the coefficient of variation and the extended t distribution. *Journal of the Royal Statistics Society Series B* 95: 695-698.

Miller, E.G. and Karson, M.J. (1977) Testing the equality of two coefficients of variation. *American Statistical Association: Proceedings of the Business and Economics Section* Part I, 278-283.

Smithson, M. (2001) Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement* 61: 605-632.

Steiger, J.H. (2004) Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods* 9: 164-182.

The R Development Core Team. (2009a) An introduction to R. Vienna: R Foundation for Statistical Computing. [Online://cran.r-project.org/doc/manuals/R-intro.pdf] accessed on October 16, 2009.

The R Development Core Team. (2009b) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. [Online: URL:// cran.r-project.org/doc/manuals/refman.pdf] accessed on October, 16, 2009.

Thompson, B. (2002) What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher* 31: 25-32.

Vangel, M.G. (1996) Confidence intervals for a normal coefficient of variation. *American Statistician* 50: 21-26.