Maejo Int. J. Sci. Technol. **2010**, *4*(01), 1-7

Maejo International Journal of Science and Technology

ISSN 1905-7873
Available online at www.mijst.mju.ac.th

Full Paper

Asymptotic confidence interval for the coefficient of variation of Poisson distribution: a simulation study

Wararit Panichkitkosolkul

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Phathum Thani, 12121, Thailand
E-mail: wararit@mathstat.sci.tu.ac.th

Received: 5 October 2009 / Accepted: 13 January 2010 / Published: 15 January 2010

Abstract: A new asymptotic confidence interval constructed by using a confidence interval for the Poisson mean is proposed for the coefficient of variation of the Poisson distribution. The following confidence intervals are considered: McKay's confidence interval, Vangel's confidence interval and the proposed confidence interval. Using Monte Carlo simulations, the coverage probabilities and expected lengths of these confidence intervals are compared. Simulation results show that all scenarios of the new asymptotic confidence interval have desired minimum coverage probabilities of 0.95 and 0.90. In addition, the newly proposed confidence interval is better than the existing ones in terms of coverage probability and expected length for all sample sizes and parameter values considered in this paper.

Keywords: coefficient of variation, confidence interval, coverage probability, expected length, Poisson distribution

---

Introduction

The coefficient of variation is a dimensionless number that quantifies the degree of variability relative to the mean [1]. The population coefficient of variation is defined as:

$$\kappa = \frac{\sigma}{\mu}, \tag{1}$$

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

$$\hat{\kappa} \;=\; \frac{S}{\overline{X}}, \tag{2}$$

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

The coefficient of variation has long been widely used as a descriptive and inferential quantity in many applications of science, economics and other areas. In chemical experiments, it is often used as a yardstick of precision of measurements; two measurement methods, for example, may be compared on the basis of their respective coefficients of variation. Relative risks in finance and actuarial science can be measured using the coefficient of variation [2]. The test for equality of the coefficients of variation of two stocks can also help determine whether the two stocks possess the same risk. In physiological science, the coefficient of variation can be applied to assess the homogeneity of bone samples [3]. It has been used as a tool in uncertainty analysis of fault trees [4] and in assessing the strength of ceramics [5].

Though useful as a point estimate, perhaps for the best use of the estimated coefficient of variation it is necessary to construct a confidence interval for the population quantity [1]. A confidence interval provides much more information about the population value of the quantity of interest than does a point estimate [e.g. 6-8].

An approximate $(1-\alpha)100\%$ confidence interval for the coefficient of variation for a normal distribution is given [e.g. 9] by:

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

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

McKay [10] has proposed using $\theta = \dfrac{\nu}{\nu+1}$ as a good approximation for the confidence interval in equation (3) but was unable to investigate the small-sample distribution of Q. McKay's approximate confidence interval is

$$CI_1 \;=\; \left\{ \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,1-\alpha/2}}{\nu+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{\nu,1-\alpha/2}}{\nu}\right]^{-1/2}, \hat{\kappa}\left[\left(\frac{\chi^2_{\nu,\alpha/2}}{\nu+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{\nu,\alpha/2}}{\nu}\right]^{-1/2} \right\} \tag{4}$$

where $\nu = n-1$, the degrees of freedom of a χ^2 distribution. Several authors have carried out numerical investigations of the accuracy of McKay's confidence interval. For instance, Iglewicz and Myers [11] compared McKay's confidence interval with the exact one based on the noncentral t distribution and found that McKay's confidence interval is efficient for $n \geq 10$ and $0 < \kappa < 0.3$.

Vangel [9] proposed a new confidence interval called the modified McKay's confidence interval for the coefficient of variation and also proposed the use of the function for θ as $\theta = \dfrac{\nu}{\nu+1}\left[\dfrac{2}{\chi^2_{\nu,\alpha}}+1\right]$, claiming that the modified McKay's method gives for the coefficient of variation the confidence intervals that are closely related to McKay's confidence interval, albeit usually more accurate. The modified McKay's confidence interval for the coefficient of variation is given by:

$$CI_2 = \left\{ \hat{\kappa}\left[\left(\frac{\chi^2_{v,1-\alpha/2}+2}{v+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{v,1-\alpha/2}}{v}\right]^{-1/2}, \hat{\kappa}\left[\left(\frac{\chi^2_{v,\alpha/2}+2}{v+1}-1\right)\hat{\kappa}^2+\frac{\chi^2_{v,\alpha/2}}{v}\right]^{-1/2} \right\}. \quad (5)$$

When data are normally distributed, McKay's confidence interval (CI_1) and modified McKay's confidence interval (CI_2) can be used very well in terms of coverage probability and expected length. However, for non-normal data, these confidence intervals cannot be used in practice. The aim of this paper is to construct a new asymptotic confidence interval for the coefficient of variation of the Poisson distribution. The modified asymptotic confidence interval is obtained by applying a confidence interval for the Poisson mean. Additionally, the coverage probabilities and the expected lengths of the new and existing confidence intervals for the coefficient of variation are compared through a Monte Carlo simulation study.

New Asymptotic Confidence Interval for the Coefficient of Variation of the Poisson Distribution

In this section, a new asymptotic confidence interval for the coefficient of variation of the Poisson distribution is presented. The newly-proposed confidence interval is based on a confidence interval for the Poisson mean. 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 a new asymptotic confidence interval, first, a confidence interval for the Poisson mean is used. One $(1-\alpha)100\%$ confidence interval for the Poisson mean with continuity correction [12] is defined as:

$$\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}} \ , \ \bar{X} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}} \right), \quad (6)$$

where $X_i \sim Poi(\lambda), i = 1,2,...,n$, $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, and $Z_{1-\frac{\alpha}{2}}$ is a $\left(1-\frac{\alpha}{2}\right)$th quantile of the standard normal distribution. From Equation (6), a confidence interval for the coefficient of variation of the Poisson distribution can be derived as follows:

$$P\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}} < \lambda < \bar{X} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}} \right) = 1-\alpha$$

$$1-\alpha = P\left(\sqrt{\bar{X} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}} < \sqrt{\lambda} < \sqrt{\bar{X} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}} \right)$$

Maejo Int. J. Sci. Technol. **2010**, *4*(01), 1-7

$$= \quad P\left(\frac{1}{\sqrt{\bar{X} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}}} < \frac{1}{\sqrt{\lambda}} < \frac{1}{\sqrt{\bar{X} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}}} \right)$$

$$= \quad P\left(\frac{1}{\sqrt{\bar{X} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}}} < \kappa < \frac{1}{\sqrt{\bar{X} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}}} \right). \tag{7}$$

Thence, the new $(1-\alpha)100\%$ asymptotic confidence interval for the coefficient of variation of the Poisson distribution is obtained, which is:

$$CI_3 \quad = \quad \left\{ \left(\sqrt{\bar{X} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}} \right)^{-1}, \left(\sqrt{\bar{X} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}+0.5}{n}}} \right)^{-1} \right\}. \tag{8}$$

Monte Carlo Simulation Results

The performance of the estimated coverage probabilities of the confidence intervals (4), (5) and (8) and their expected lengths were examined via Monte Carlo simulations, with particular emphasis on comparison between the new and existing approaches. Data were generated from Poisson distribution with κ = 0.1, 0.2 and 0.3, and sample size n = 10, 15, 25, 50 and 100. All simulations were performed using programs written in the R statistical software [13-15] with the number of simulation runs M = 50,000 at level of significance $\alpha = 0.05$ and 0.10. The simulation results are shown in Tables 1-2, in which the following information, viz. the estimated coverage probabilities of the confidence intervals in (4), (5) and (8) (CI_1, CI_2 and CI_3) and their expected lengths for the coefficient of variation of a Poisson distribution at α = 0.05 and 0.10, is presented respectively. As can be seen the Tables, all of the confidence intervals, CI_1, CI_2 and CI_3, have minimum coverage probability of $1-\alpha$ for all sample sizes and values of κ. The estimated coverage probability increases as the value of κ gets larger (e.g. for CI_3, n=10 and $\alpha = 0.05$; 0.9506 for κ = 0.1; 0.9511 for κ = 0.2; and 0.9559 for κ = 0.3). In addition, the empirical coverage probabilities of the proposed confidence interval, CI_3, are closer to the nominal value of $1-\alpha$ than those of CI_1 and CI_2. Furthermore, the expected lengths of the proposed confidence interval, CI_3, are much shorter than those of CI_1 and CI_2 in all conditions. The expected length increases as the value κ gets larger (e.g. for CI_3, n=10 and $\alpha = 0.05$; 0.0062 for κ = 0.1; 0.0254 for κ = 0.2; and 0.0590 for κ = 0.3). Moreover, when the sample size increases, the expected length is shorter (e.g. for CI_3, 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; and 0.0020 for n=100).

Maejo Int. J. Sci. Technol. **2010**, *4*(01), 1-7

Table1. Estimated coverage probabilities and expected lengths of 95% confidence intervals in (4), (5) and (8) for a Poisson distribution

n	κ	Coverage probability			Expected length		
		CI_1	CI_2	CI_3	CI_1	CI_2	CI_3
10	0.1	0.9506	0.9508	0.9506	0.1133	0.1126	0.0062
	0.2	0.9556	0.9565	0.9511	0.2449	0.2387	0.0254
	0.3	0.9618	0.9631	0.9559	0.4248	0.3958	0.0590
15	0.1	0.9517	0.9518	0.9514	0.0845	0.0843	0.0051
	0.2	0.9556	0.9561	0.9520	0.1786	0.1764	0.0206
	0.3	0.9625	0.9628	0.9540	0.2940	0.2850	0.0476
25	0.1	0.9522	0.9524	0.9521	0.0612	0.0611	0.0039
	0.2	0.9575	0.9577	0.9520	0.1278	0.1270	0.0159
	0.3	0.9632	0.9635	0.9539	0.2057	0.2027	0.0365
50	0.1	0.9519	0.9518	0.9515	0.0414	0.0413	0.0028
	0.2	0.9568	0.9569	0.9514	0.0857	0.0855	0.0112
	0.3	0.9632	0.9634	0.9536	0.1362	0.1354	0.0257
100	0.1	0.9516	0.9516	0.9512	0.0286	0.0286	0.0020
	0.2	0.9560	0.9560	0.9517	0.0591	0.0590	0.0079
	0.3	0.9632	0.9635	0.9541	0.0933	0.0930	0.0181

Table2. Estimated coverage probabilities and expected lengths of 90% confidence intervals in (4), (5) and (8) for a Poisson distribution

n	κ	Coverage probability			Expected length		
		CI_1	CI_2	CI_3	CI_1	CI_2	CI_3
10	0.1	0.9022	0.9025	0.9021	0.0909	0.0905	0.0052
	0.2	0.9045	0.9053	0.9023	0.1936	0.1895	0.0212
	0.3	0.9193	0.9206	0.9100	0.3255	0.3081	0.0491
15	0.1	0.9031	0.9033	0.9020	0.0690	0.0688	0.0043
	0.2	0.9114	0.9118	0.9009	0.1449	0.1432	0.0173
	0.3	0.9188	0.9197	0.9014	0.2354	0.2291	0.0398
25	0.1	0.9034	0.9035	0.9005	0.0506	0.0505	0.0033
	0.2	0.9084	0.9094	0.9023	0.1053	0.1047	0.0134
	0.3	0.9209	0.9215	0.9059	0.1686	0.1664	0.0306
50	0.1	0.9029	0.9030	0.9006	0.0344	0.0344	0.0023
	0.2	0.9106	0.9109	0.9045	0.0713	0.0711	0.0094
	0.3	0.9179	0.9188	0.9035	0.1130	0.1123	0.0215
100	0.1	0.9016	0.9017	0.9004	0.0239	0.0239	0.0016
	0.2	0.9093	0.9096	0.9044	0.0494	0.0493	0.0067
	0.3	0.9206	0.9208	0.9072	0.0779	0.0777	0.0152

Maejo Int. J. Sci. Technol. **2010**, *4*(01), 1-7

Conclusions

A new asymptotic confidence interval which is based on a confidence interval for the Poisson mean has been developed for the coefficient of variation of the Poisson distribution. The developed confidence interval was compared with those of McKay and Vangel through a Monte Carlo simulation study. All confidence intervals have minimum coverage probabilities $1-\alpha$. The estimated coverage probabilities and the expected lengths of both McKay's and Vangel's confidence intervals are slightly different when sample sizes are small while they are virtually the same when sample sizes are large. The new asymptotic confidence interval proposed in this paper has several advantages over the existing confidence intervals of McKay and Vangel. Firstly, it has the estimated coverage probabilities which are closer to the nominal value of $1-\alpha$. Secondly, it has shorter expected lengths than those of the other two confidence intervals while having reasonable coverage probabilities. Thus, if a confidence interval with minimum coverage probability equal to a pre-specified value and with a shorter expected length is preferred, the newly proposed confidence interval may be the one of choice.

Acknowledgements

The author would like to thank the editor and the three anonymous referees for their suggestions and helpful comments in improving this paper. The author also acknowledges the excellent comments provided by Dr. Gareth Clayton on earlier drafts of this paper.

References

1. K. Kelley, "Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach", *Behav. Res. Meth.*, **2007**, *39*, 755-766.
2. E. G. Miller and M. J. Karson, "Testing the equality of two coefficients of variation", American Statistical Association: Proceedings of the Business and Economics Section, Part I, **1977**, pp. 278-283.
3. A. J. Hamer, J. R. Strachan, M. M. Black, C. Ibbotson and R. A. Elson, "A new method of comparative bone strength measurement", *J. Med. Eng. Technol.*, **1995**, *19*, 1-5.
4. K. Ahn, "On the use of coefficient of variation for uncertainty analysis in fault tree analysis", *Reliab. Eng. Syst. Safe.*, **1995**, *47*, 229-230.
5. J. Gong and Y. Li, "Relationship between the estimated Weibull modulus and the coefficient of variation of the measured strength for ceramics", *J. Am. Ceram. Soc.*, **1999**, *82*, 449-452.
6. M. Smithson, "Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals", *Edu. Psychol. Meas.*, **2001**, *61*, 605-632.
7. B. Thompson, "What future quantitative social science research could look like: Confidence intervals for effect sizes", *Edu. Researcher*, **2002**, *31*, 25-32.

Maejo Int. J. Sci. Technol. **2010**, *4*(01), 1-7

8. J. H. Steiger, "Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis", *Psychol. Meth.*, **2004**, *9*, 164-182.
9. M. G. Vangel, "Confidence intervals for a normal coefficient of variation" *Amer. Statist.*, **1996**, *50*, 21-26.
10. A. T. McKay, "Distribution of the coefficient of variation and the extended t distribution", *J. Roy. Statist. Soc. Ser. B*, **1932**, *95*, 695-698.
11. B. Iglewicz and R. H. Myers, "Comparisons of approximations to the percentage points of the sample coefficient of variation", *Technometrics*, **1970**, *12*, 166-169.
12. L. Barker, "A comparison of nine confidence intervals for a Poisson parameter when the expected number of events is ≤ 5", *Amer. Statist.*, **2002**, *56*, 85-89.
13. The R Development Core Team, "An Introduction to R", R Foundation for Statistical Computing, Vienna, **2008a**.
14. The R Development Core Team, "R: A Language and Environment for Statistical Computing", R Foundation for Statistical Computing, Vienna, **2008b**.
15. R. Ihaka and R. Gentleman, "R: A language for data analysis and graphics", *J. Comput. Graph. Statist.*, **1996**, *5*, 299-314.