Thailand Statistician
July 2020; 18(3): 340-353
http://statassoc.or.th
Contributed paper

Interval Estimation for the Common Coefficient of Variation of Gamma Distributions

Patarawan Sangnawakij*[a] and Sa-Aat Niwitpong [b]
[a] Department of Mathematics and Statistics, Thammasat University, Rangsit Campus, Pathumthani, Thailand.
[b] Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand.
*Corresponding author; e-mail: patarawan.s@gmail.com

Received: 28 January 2019
Revised: 15 July 2019
Accepted: 18 July 2019

Abstract

The method of variance of estimate recovery (MOVER) has been widely used to construct the confidence interval for the function of parameters. This method is based on estimating the variance of the related estimators and recovering them to the confidence interval for the parameter of interest. However, only confidence intervals for one- or two-sample in the gamma distribution have been reported, whereas in many research areas, more than two samples must be studied. In this paper, we therefore introduce three confidence intervals for the common coefficient of variation (CV) of multiple gamma distributions. The traditional MOVER is extended as the adjusted MOVER. The first two proposed confidence intervals are constructed using the adjusted MOVER with existing confidence limits obtained from the score and Wald methods. The third is formulated using normal approximation. The performance of these estimators was investigated using simulations. The results showed that the confidence interval derived by the adjusted MOVER with the Wald method satisfied the criteria for coverage probability in the general cases. Two real-world datasets were analyzed to confirm the practical application of these estimators.

Keywords: Adjusted method of variance of estimate recovery, asymptotic confidence interval, several populations.

1. Introduction

The coefficient of variation (CV) is widely used in many fields of research to measure the dispersion of a variable of interest. It is expressed by the ratio of the standard deviation (SD) to the mean. Since the CV is a unit-free measure, it can be used to compare several variables in terms of homogeneity expressed at different scales (Rice 2006). This is an important reason for preferring the CV over the variance or SD, as, when the latter is used to report the variation, all variables must have the same means and be expressed in the same units (Banik and Kibria 2011). In one application, the CV is used to measure uncertainty about the amount of a certain chemical in the urine or blood of a patient (Chadban et al. 2003). It is elsewhere used to monitor the physical properties of products in quality control (Castagliola et al. 2013). In environmental studies, the CV has been used to measure the spatial and temporal correlation of solar radiation at two sites in Guadeloupe, and to index the air pollution level in Rome (Battista et al. 2016, Calif and Soubdhan 2016). In medical studies, the CV has been used to investigate the expression level of proteins in mRNA (Popadin et al. 2014). More recent applications have appeared in papers, including Shiina et al. (2012), Kenig et al. (2014), Zou et al. (2015), and Bakowski et al. (2017).

The CV has therefore been widely studied in statistical inference, especially in interval estimation of a single sample. For example, confidence intervals for the CV in the normal and non-normal distributions have been investigated by Wong and Wu (2002), Verrill (2003), Mahmoudvand and Hassani (2009), Niwitpong and Wongkhao (2015) and Sangnawakij and Niwitpong (2016). In the lognormal and delta-lognormal distributions, confidence intervals have been addressed by Koopmans et al. (1964), Chen and Zhou (2006) and Niwitpong (2015). From the previous reviews, many studies have investigated confidence intervals for the single parameter in the distribution which are related to the normal distribution. However, skewed distributions are often encountered. The application of statistical inference in non-symmetric distributions is therefore an important question. Sangnawakij and Niwitpong (2017) introduced confidence intervals for the CV in a gamma distribution. They used two methods based on the likelihood function: the score and Wald methods. In practice, however, parameter estimation often has to be applied to samples drawn from multiple populations, especially in experimental studies. To estimate the parameter of interest, many studies have derived the confidence interval for the common parameter across several distributions. Established approaches are as follows. Lin and Lee (2005) constructed the confidence interval for the common mean of several normal populations using the concept of generalized variables, as did Krishnamoorthy and Lu (2003). Using the same method, Ng (2014) and Tian (2005) proposed confidence intervals for the common CV in lognormal and normal populations. Tian and Wu (2007) introduced confidence intervals for the common mean of several lognormal populations, as did Krishnamoorthy and Mathew (2003), Behboodian and Jafari (2006), and Krishnamoorthy and Oral (2017). Furthermore, the confidence intervals for several populations in two-parameter exponential distributions were presented in Thangjai and Niwitpong (2017).

In this study, we focus on the gamma distribution. This is a right-skewed continuous probability model often used in reliability studies and life experiments. Suppose that X be a random sample from a gamma distribution, denoted as Gamma(α, β), the probability density function of X is given by

$$f_X(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} e^{-x/\beta} x^{\alpha-1}, \tag{1}$$

where $x > 0, \alpha > 0$ and $\beta > 0.$ The mean and variance of X are $\alpha\beta$ and $\alpha\beta^2,$ respectively. This model plays a role of importance in representing the time to failure or point at which the α^{th} event occurs (Gupta and Guttman 2013). It is widely used in applications that take positive values, especially in engineering, science, business, and environmental pollutant studies (Husak et al. 2007, Rohan et al. 2015, Xie and Wu 2017). Note that the CV in the gamma distribution depends on only a single parameter $\alpha,$ in contrast with the variance, which depends on two unknown parameters. The confidence intervals for a single CV have reviewed above. However, when it is known that k independent gamma populations, for $k \geq 2,$ have equal coefficients of variation, a key question arises: how to summarize the statistics of all k samples. No previous studies have investigated the confidence interval for the common CV of gamma distributions. This problem is therefore addressed here.

Our approach is based on an adjustment to the method of variance of estimate recovery (MOVER) introduced by Zou and Donner (2008). It is applied in this study to construct the confidence interval for the common CV across two or more gamma distributions. Since this method is derived from the confidence limits for each individual CV, the confidence intervals obtained from the score and Wald methods, which have a good performance in terms of coverage probability and expected length are used (Sangnawakij and Niwitpong 2017). The methodologies for deriving the confidence intervals for the common CV of several gamma populations are described in Section 2. That section also includes a review of the existing confidence intervals for a single CV. Simulations are used to evaluate the performance of the proposed confidence intervals in terms of coverage probability and expected length. The results are presented in Section 3. Section 4 illustrates the use of these confidence intervals using two real data examples: the time in hours of successive failures of the air conditioning system of a military plane, and air pollution incidents in Bangkok. Finally, Section 5 reports our conclusions.

2. Statistical Methodologies

Let $(X_{11}, X_{12}, ..., X_{1n_1})$, $(X_{21}, X_{22}, ..., X_{2n_2})$, $\ldots$, $(X_{k1}, X_{k2}, ..., X_{kn_k})$ be k random samples, each drawn from population i which has a gamma distribution, denoted as $X_{ij} \sim \text{Gamma}(\alpha_i, \beta_i)$, for $i = 1, 2, ..., k$ and $j = 1, 2, ..., n_{ik}$. The mean and variance of X_{ij} are given by $\alpha_i \beta_i$ and $\alpha_i \beta_i^2$, respectively. Assume that $\tau_i = 1/\sqrt{\alpha_i}$ is the coefficient of variation (CV) of the i^{th} sample. In the case where $\tau_i = \tau$, for all i, corresponds to the common CV of the gamma distributions.

From the i^{th} sample, the maximum likelihood estimator for parameter τ_i is approximated by $\hat{\tau}_i = 1/\sqrt{\hat{\alpha}_i}$, where $\hat{\alpha}_i = 1 \Big/ \left(2\left(\ln \bar{X}_i - \sum_{j=1}^{n_i} \ln X_{ij} / n_i \right) \right)$ and $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij} / n_i$. It can be seen that $\hat{\tau}_i$ is in the complex function. The approximate approach, based on Taylor series expansion, is known as the delta method (Casella and Berger 2002) and is applied to estimate the variance of $\hat{\tau}_i$. Since we derive that the approximate variance of $\hat{\alpha}_i$ is given by $Var(\hat{\alpha}_i) = 2\alpha_i^2 / n_i$, the variance estimate of $\hat{\tau}_i$ based on the delta method is given as

$$\overline{Var}(\hat{\tau}_i) \cong \left[\frac{\partial}{\partial \hat{\alpha}_i} \left(\frac{1}{\sqrt{\hat{\alpha}_i}} \right) \right]^2 Var(\hat{\alpha}_i) \Bigg|_{\alpha_i = \hat{\alpha}_i} = \frac{1}{2 n_i \hat{\alpha}_i}. \tag{2}$$

In Sangnawakij and Niwitpong (2017), interval estimators for a single CV, τ_i, were introduced based on the score and Wald methods, denoted as CI_{si} and CI_{wi}, respectively. They derived these confidence intervals for τ_i as follows:

$$CI_{si} = (l_{si}, u_{si}) = \left(\sqrt{\frac{2}{n_i}\left(T_{1i} - z_{\gamma/2}\sqrt{\frac{n_i}{2\hat{\alpha}_i^2}} \right)}, \sqrt{\frac{2}{n_i}\left(T_{1i} + z_{\gamma/2}\sqrt{\frac{n_i}{2\hat{\alpha}_i^2}} \right)} \right),$$

and

$$CI_{wi} = \left(l_{wi}, u_{wi}\right) = \left(\frac{1}{\sqrt{\hat{\alpha}_i + z_{\gamma/2}\sqrt{2\hat{\alpha}_i^2 / n_i}}}, \frac{1}{\sqrt{\hat{\alpha}_i - z_{\gamma/2}\sqrt{2\hat{\alpha}_i^2 / n_i}}} \right),$$

where $T_{1i} = n_i \ln \bar{X}_i - \sum_{j=1}^{n_i} \ln X_{ij}$ and $z_{\gamma/2}$ is the $(\gamma/2)100^{\text{th}}$ percentile of the standard normal distribution. Note that this is used for a single sample in a gamma distribution only. Our novel methods for deriving the confidence interval for the common CV of several gamma populations are presented in the following sections.

2.1. Approximate confidence interval

We first consider the point estimator for the overall or common CV of gamma distributions,The estimator is derived using the weighted average method to minimize the variance (Finkelstein and Levin 2001, Hartung et al. 2008). Here, the weight of the individual CV is given by

$$v_i = \frac{1/\overline{Var}\left(\hat{\tau}_i\right)}{\sum_{i=1}^{k} 1/\overline{Var}\left(\hat{\tau}_i\right)}.$$

Therefore, the estimated common CV for τ, which is the parameter of interest, is of the form

$$\hat{\tau} = \sum_{i=1}^{k} \hat{\tau}_i v_i = \frac{\sum_{i=1}^{k} \hat{\tau}_i / \overline{Var}(\hat{\tau}_i)}{\sum_{i=1}^{k} 1/\overline{Var}(\hat{\tau}_i)}, \tag{3}$$

with mean $E(\hat{\tau}) = \tau$ and variance $\overline{Var}(\hat{\tau}) = 1/\sum_{i=1}^{k} 1/\overline{Var}(\hat{\tau}_i)$. Note that this point estimator corresponds to the CV and variance estimates for each sample i. Moreover, $\hat{\tau}$ is the unbiased estimator for τ.

Using this approach and the normal approximation, the pivotal quantity of τ is derived by

$$Z = \frac{\hat{\tau} - E(\hat{\tau})}{\sqrt{\overline{Var}(\hat{\tau})}} = \frac{\hat{\tau} - \tau}{\sqrt{1/\sum_{i=1}^{k} 1/\overline{Var}(\hat{\tau}_i)}},$$

where $\overline{Var}(\hat{\tau}_i)$ is obtained from (2). The pivot Z converges to the standard normal distribution, and can be used to build the confidence interval for the parametric function. From the general probability statement

$$1 - \gamma = P(-z_{\gamma/2} \le Z \le z_{\gamma/2}),$$

we obtain

$$1 - \gamma = P\left(\hat{\tau} - z_{\gamma/2}\sqrt{\frac{1}{\sum_{i=1}^{k} 1/\overline{Var}(\hat{\tau}_i)}} \le \tau \le \hat{\tau} + z_{\gamma/2}\sqrt{\frac{1}{\sum_{i=1}^{k} 1/\overline{Var}(\hat{\tau}_i)}} \right),$$

where $1 - \gamma$ is the confidence level probability. Thus, the $(1-\gamma)100\%$ two-sided approximate confidence interval for τ is given by

$$CI_A = \left(\hat{\tau} - z_{\gamma/2}\sqrt{\frac{1}{2\sum_{i=1}^{k} n_i\hat{\alpha}_i}}, \hat{\tau} + z_{\gamma/2}\sqrt{\frac{1}{2\sum_{i=1}^{k} n_i\hat{\alpha}_i}} \right), \tag{4}$$

where $\hat{\tau}$ is obtained from (3).

2.2. Confidence intervals based on the adjusted MOVER

The method based on the recovering variance or method of variance of estimate recovery (MOVER) (Zou and Donner 2008, Zou et al. 2009, Donner and Zou 2010) is used to construct the confidence interval for the common CV. Our approach extends from this method. The original MOVER is based on finding the confidence intervals for two single parameters, recovering variance estimates from those confidence intervals, then deriving the confidence interval for the function of the parameter of interest based on the central limit theorem (CLT). Since the common CV shown in (3) is the sum of product between the CV of sample i and the inverse of its variance estimate, variance estimation from a part of the MOVER is applied for the new interval estimator.

The motivation for the MOVER is now given for $k=2$ samples. Following Zou et al. (2009), let θ_1 and θ_2 be two generic parameters with point estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively. Suppose that the parameter of interest is in terms of $\theta_1+\theta_2$. From the CLT, we have

$$(L,U) = \left((\hat{\theta}_1+\hat{\theta}_2) \pm z_{\gamma/2}\sqrt{V(\hat{\theta}_1)+V(\hat{\theta}_2)} \right), \tag{5}$$

where $V(\hat{\theta}_1)$ and $V(\hat{\theta}_2)$ are the unknown variances of $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively. This form assumes the independence of the two estimators. The confidence intervals for parameters θ_1 and θ_2 are respectively given by

$$(l_1',u_1') = \left(\hat{\theta}_1 \pm z_{\gamma/2}\sqrt{V(\hat{\theta}_1)} \right) \text{ and } (l_2',u_2') = \left(\hat{\theta}_2 \pm z_{\gamma/2}\sqrt{V(\hat{\theta}_2)} \right).$$

Zou et al. (2009) showed that $l_1'+l_2'$ is closer to the lower limit L than is $\hat{\theta}_1+\hat{\theta}_2$, and that $u_1'+u_2'$ is closer to the upper limit U than is $\hat{\theta}_1+\hat{\theta}_2$. By assuming $\theta_1=l_1'$ and $\theta_2=l_2'$, the variances used to compute L are estimated as

$$\hat{V}_L(\hat{\theta}_1) = \frac{(\hat{\theta}_1-l_1')^2}{z_{\gamma/2}^2} \text{ and } \hat{V}_L(\hat{\theta}_2) = \frac{(\hat{\theta}_2-l_2')^2}{z_{\gamma/2}^2}.$$

Similarly, if $\theta_1=u_1'$ and $\theta_2=u_2'$, the variance estimates used to compute U are given by

$$\hat{V}_U(\hat{\theta}_1) = \frac{(u_1'-\hat{\theta}_1)^2}{z_{\gamma/2}^2} \text{ and } \hat{V}_U(\hat{\theta}_2) = \frac{(u_2'-\hat{\theta}_2)^2}{z_{\gamma/2}^2}.$$

The variances used to compute L and U are then substituted into (5) to obtain the $(1-\gamma)100\%$ confidence interval for $\theta_1+\theta_2$. A discussion of the other parameter functions, including $\theta_1-\theta_2$ and θ_1/θ_2, can be found in Donner and Zou (2011).

To consider k random samples, where $k \geq 2$, the adjusted MOVER is applied. Here, the variance estimates of $\hat{\theta}_i$ at $\theta_i=l_i'$ and $\theta_i=u_i'$, for $i=1,2,...,k$, are respectively given by

$$\hat{V}_L(\hat{\theta}_i) = \frac{(\hat{\theta}_i - l_i')^2}{z_{\gamma/2}^2} \quad \text{and} \quad \hat{V}_U(\hat{\theta}_i) = \frac{(u_i' - \hat{\theta}_i)^2}{z_{\gamma/2}^2}. \tag{6}$$

The variance estimate of $\hat{\theta}_i$ is then computed by

$$\overline{Var}(\hat{\theta}_i) = \frac{1}{2}\left(\frac{(\hat{\theta}_i - l_i')^2}{z_{\gamma/2}^2} + \frac{(u_i' - \hat{\theta}_i)^2}{z_{\gamma/2}^2} \right).$$

Since we have the score confidence interval for τ_i, $CI_{si} = (l_{si}, u_{si})$, the estimators $\hat{\tau}_i$, l_{si}, and u_{si} are substituted into $\hat{\theta}_i$, l_i', and u_i', respectively, of (6). It follows that the variance estimate of $\hat{\tau}_i$ is given by

$$\overline{Var}_s(\hat{\tau}_i) = \frac{1}{2}\left(\hat{V}_L(\hat{\tau}_i) + \hat{V}_U(\hat{\tau}_i) \right) = \frac{1}{2}\left(\frac{(\hat{\tau}_i - l_{si})^2}{z_{\gamma/2}^2} + \frac{(u_{si} - \hat{\tau}_i)^2}{z_{\gamma/2}^2} \right). \tag{7}$$

To perform interval estimation for τ, the large sample approach is used. Hence, the $(1-\gamma)100\%$ two-sided confidence interval for τ using the adjusted MOVER with the score confidence limits is derived as follows

$$CI_{AdMS} = \left(\hat{\tau} - z_{\gamma/2}\sqrt{\frac{1}{\sum_{i=1}^{k} 1/\hat{V}_L(\hat{\tau}_i)}},\ \hat{\tau} + z_{\gamma/2}\sqrt{\frac{1}{\sum_{i=1}^{k} 1/\hat{V}_U(\hat{\tau}_i)}} \right),$$

$$= \left(\hat{\tau} - \sqrt{\frac{1}{\sum_{i=1}^{k} 1/\left(\hat{\tau}_i - l_{si}\right)^2}},\ \hat{\tau} + \sqrt{\frac{1}{\sum_{i=1}^{k} 1/\left(u_{si} - \hat{\tau}_i\right)^2}} \right), \tag{8}$$

where $\hat{\tau}$ is the overall CV estimate obtained from (3) and $\overline{Var}(\hat{\tau}_i)$ derived from (7).

Similarly, we have the Wald confidence interval for τ_i, $CI_{wi} = (l_{wi}, u_{wi})$. Again, we substitute $\hat{\tau}_i$, l_{wi} and u_{wi} into $\hat{\theta}_i$, l_i', and u_i', respectively, of (6). The average variance of $\hat{\tau}_i$ is given by

$$\overline{Var}_w\left(\hat{\tau}_i\right) = \frac{1}{2}\left(\frac{\left(\hat{\tau}_i - l_{wi}\right)^2}{z_{\gamma/2}^2} + \frac{\left(u_{wi} - \hat{\tau}_i\right)^2}{z_{\gamma/2}^2} \right). \tag{9}$$

The $(1-\gamma)100\%$ two-sided confidence interval for τ using the adjusted MOVER with the Wald confidence limits is therefore given as

$$CI_{AdMW} = \left(\hat{\tau} - \sqrt{\frac{1}{\sum_{i=1}^{k} 1/\left(\hat{\tau}_i - l_{wi}\right)^2}},\ \hat{\tau} + \sqrt{\frac{1}{\sum_{i=1}^{k} 1/\left(u_{wi} - \hat{\tau}_i\right)^2}} \right), \tag{10}$$

where $\hat{\tau}$ is the overall CV estimate with the its corresponded variance in (9).

3. Simulation Studies

The proposed confidence intervals, CI_A, CI_{AdMS} and CI_{AdMW}, were evaluated through simulations using the R statistical package (R Core Team 2018). The performance was scored in terms of the coverage probability (CP) and expected length (EL). These were computed as

$$\text{CP} = \frac{c(L \le \tau \le U)}{M} \quad \text{and} \quad \text{EL} = \frac{\sum_{h=1}^{M}(U_h - L_h)}{M},$$

where $c(L \le \tau \le U)$ is the number of simulation runs for a parameter τ that lies within the confidence interval and M is total runs. In the simulations, the data were randomly sampled from independent gamma distributions, where $\beta_i = 2$ and $\alpha_i = 1/\tau^2$, for $i = 1, 2, ..., k$. The true common CVs were set as τ = 0.05, 0.10, 0.20, 0.33, and 0.50. We considered cases in which the number of groups k = 3, 4 and 5 with equal sample sizes $n_i = n = 30$, 50, 100, and 200. We performed $M = 10{,}000$ simulation runs on the generated dataset, and then calculated the 95% confidence intervals for τ based on all combinations of the different settings. From the simulation, a preferred 95% confidence interval of τ would have the coverage probability in the interval $c_0 \pm z_{\gamma/2}\sqrt{\frac{c_0(1-c_0)}{M}}$, with the short expected length. Here, c_0 is the value of the given confidence level, which is given as 0.95. Thus, the 95% confidence interval for the nominal coverage level is [0.9457, 0.9543]. This criteria confirms that the coverage probability is similar to the given confidence level, or a good confidence interval is obtained. The full simulation results are presented in Tables 1-3, but the key findings were given as follows.

As can be seen in Table 1, for all sample sizes, CI_A and CI_{AdMS} produced coverage probabilities below the 0.95 threshold. The coverage probabilities of these two confidence intervals did not meet the criteria. In contrast, the coverage probabilities of CI_{AdMW} satisfied the nominal coverage level in many cases, especially when $n \ge 50$. Only the case $n = 200$ and τ = 0.50, coverage probabilities of CI_{AdMW} were slightly lower than the criteria. The expected lengths of CI_A and CI_{AdMS} were slightly smaller than those of CI_{AdMW}, and those of CI_A smaller than CI_{AdMS}. However, when $n \ge 100$, no significant differences were found in the value of the expected length. In all cases, the expected length decreased as the sample size increased.

In summary, the simulation results suggest that CI_{AdMW} is generally superior to CI_{AdMS} and CI_A, reflecting its stable coverage probability across sample sizes and desirable the nominal coverage level with acceptable expected length in many situations. The results are shown graphically in Figure 1. Furthermore, varying k had little effect on the results (see Tables 2 and 3). We therefore conclude that the method of variance of estimate recovery with the Wald confidence limits proposed to derive the confidence intervals in this work are appropriate for estimating the common CV of multiple gamma populations.

Table 1 The results of the 95% confidence intervals for the common CV of gamma distributions (3 sample cases)

k	n	τ	Coverage Probability			Expected Length		
			CI_A	CI_{AdMS}	CI_{AdMW}	CI_A	CI_{AdMS}	CI_{AdMW}
3	30	0.05	0.8450	0.8078	0.9615	0.0140	0.0145	0.0168
		0.10	0.8530	0.8179	0.9591	0.0280	0.0290	0.0337
		0.20	0.8601	0.8243	0.9550	0.0562	0.0582	0.0675
		0.33	0.8672	0.8346	0.9615	0.0932	0.0965	0.1119
		0.50	0.8929	0.8702	0.9522	0.1428	0.1480	0.1716
	50	0.05	0.8909	0.8687	0.9539	0.0110	0.0113	0.0122
		0.10	0.8963	0.8729	0.9538	0.0221	0.0225	0.0245
		0.20	0.8975	0.8788	0.9518	0.0443	0.0452	0.0491
		0.33	0.9096	0.8931	0.9528	0.0734	0.0800	0.0815
		0.50	0.9333	0.9260	0.9501	0.1126	0.1149	0.1249
	100	0.05	0.9241	0.9111	0.9534	0.0079	0.0080	0.0083
		0.10	0.9211	0.9086	0.9503	0.0158	0.0160	0.0166
		0.20	0.9249	0.9149	0.9495	0.0317	0.0320	0.0333
		0.33	0.9378	0.9314	0.9493	0.0526	0.0531	0.0553
		0.50	0.9446	0.9474	0.9468	0.0806	0.0814	0.0847
	200	0.05	0.9338	0.9274	0.9504	0.0056	0.0057	0.0058
		0.10	0.9358	0.9317	0.9508	0.0113	0.0113	0.0115
		0.20	0.9447	0.9420	0.9503	0.0226	0.0227	0.0231
		0.33	0.9462	0.9460	0.9481	0.0375	0.0377	0.0384
		0.50	0.9302	0.9376	0.9438	0.0574	0.0576	0.0588

Table 2 The results of the 95% confidence intervals for the common CV of gamma distributions (4 sample cases)

k	n	τ	Coverage Probability			Expected Length		
			CI_A	CI_{AdMS}	CI_{AdMW}	CI_A	CI_{AdMS}	CI_{AdMW}
4	30	0.05	0.8241	0.7831	0.9571	0.0121	0.0125	0.0145
		0.10	0.8266	0.7811	0.9578	0.0242	0.0251	0.0291
		0.20	0.8369	0.8006	0.9527	0.0486	0.0503	0.0584
		0.33	0.8504	0.8125	0.9595	0.0805	0.0835	0.0968
		0.50	0.8781	0.8491	0.9577	0.1234	0.1279	0.1483
	50	0.05	0.8751	0.8505	0.9525	0.0095	0.0097	0.0106
		0.10	0.8717	0.8457	0.9517	0.0191	0.0195	0.0212
		0.20	0.8864	0.8613	0.9543	0.0383	0.0391	0.0425
		0.33	0.8999	0.8795	0.9543	0.0635	0.0649	0.0705
		0.50	0.9214	0.9090	0.9570	0.0973	0.0993	0.1080
	100	0.05	0.9106	0.8961	0.9502	0.0068	0.0069	0.0072
		0.10	0.9156	0.8998	0.9530	0.0137	0.0138	0.0144
		0.20	0.9299	0.9209	0.9571	0.0275	0.0277	0.0289
		0.33	0.9302	0.9219	0.9505	0.0455	0.0460	0.0478
		0.50	0.9395	0.9444	0.9494	0.0698	0.0705	0.0734
	200	0.05	0.9340	0.9236	0.9510	0.0049	0.0049	0.0050
		0.10	0.9353	0.9309	0.9522	0.0097	0.0098	0.0100
		0.20	0.9397	0.9348	0.9507	0.0195	0.0196	0.0200
		0.33	0.9470	0.9471	0.9437	0.0324	0.0326	0.0332
		0.50	0.9274	0.9381	0.9441	0.0497	0.0499	0.0509

Table 3 The results of the 95% confidence intervals for the common CV of gamma distributions (5 sample cases)

k	n	τ	Coverage Probability			Expected Length		
			CI_A	CI_{AdMS}	CI_{AdMW}	CI_A	CI_{AdMS}	CI_{AdMW}
5	30	0.05	0.8660	0.7410	0.9596	0.0108	0.0112	0.0130
		0.10	0.8116	0.7581	0.9554	0.0216	0.0224	0.0260
		0.20	0.8153	0.7579	0.9576	0.0433	0.0449	0.0520
		0.33	0.8332	0.7910	0.9599	0.0720	0.0746	0.0865
		0.50	0.8691	0.8371	0.9560	0.1103	0.1143	0.1325
	50	0.05	0.8579	0.8276	0.9587	0.0085	0.0087	0.0095
		0.10	0.8575	0.8261	0.9498	0.0171	0.0174	0.0189
		0.20	0.8724	0.8457	0.9530	0.0342	0.0349	0.0380
		0.33	0.8895	0.8669	0.9545	0.0568	0.0580	0.0630
		0.50	0.9219	0.9102	0.9527	0.0871	0.0888	0.0966
	100	0.05	0.9021	0.8875	0.9534	0.0061	0.0062	0.0064
		0.10	0.9076	0.8903	0.9524	0.0122	0.0124	0.0129
		0.20	0.9196	0.9066	0.9504	0.0245	0.0248	0.0258
		0.33	0.9312	0.9233	0.9505	0.0407	0.0411	0.0428
		0.50	0.9433	0.9488	0.9479	0.0624	0.0630	0.0656
	200	0.05	0.9301	0.9216	0.9516	0.0044	0.0044	0.0045
		0.10	0.9317	0.9229	0.9505	0.0087	0.0088	0.0089
		0.20	0.9359	0.9296	0.9442	0.0175	0.0176	0.0179
		0.33	0.9487	0.9478	0.9436	0.0290	0.0291	0.0297
		0.50	0.9193	0.9332	0.9312	0.0444	0.0446	0.0455

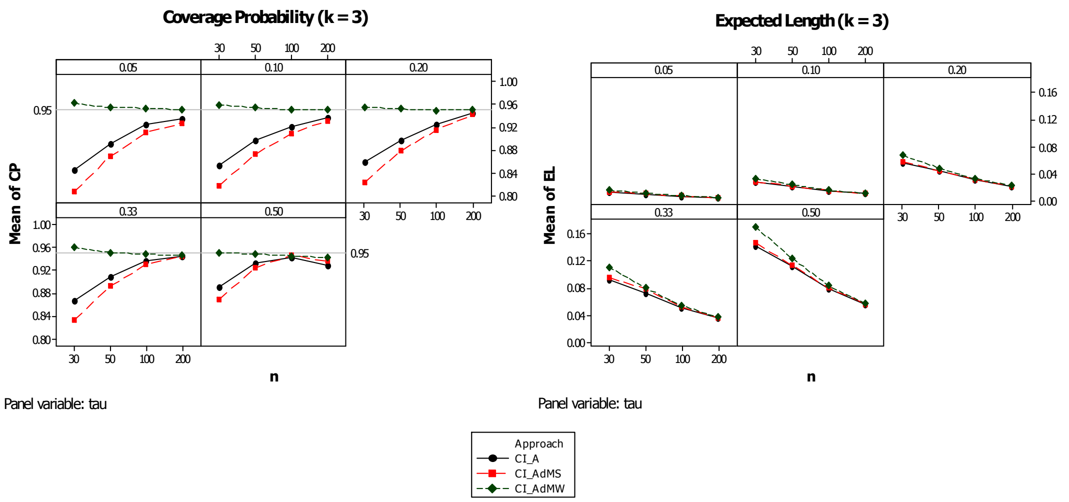

Figure 1 Coverage probability (left) and expected length (right) of the 95% confidence intervals for the common CV for 3 sample cases

4. Applications to Real Data

We illustrated the use of confidence intervals using two real data examples: 1) the time in hours of successive failures of the air conditioning system of a military plane and 2) air pollution incidents in Bangkok.

4.1. Example 1

Proschan (1963) provided the time in hours of successive failures of the air conditioning system of each member of a fleet of 720 Boeing aircraft. We used the data from the four groups with plan numbers 7908, 7909, 7912, and 7913. The sample sizes were reasonably consistent across groups. The distribution of the data was verified using the Anderson-Darling (AD) test and the graphical plot for each group is shown in Figure 2. All samples had a gamma distribution, so that these data were appropriate for computing the confidence intervals in our approach. In computation, the basic statistics of these data are shown in Table 4.

Table 4 Mean and CV for the data on time to failure of the air conditioning system for four planes

Plane number	n_i	$\bar{x}_i$	$\hat{\tau}_i$
7908	23	95.70	1.10
7909	29	83.52	0.81
7912	30	59.60	1.21
7913	27	76.80	1.00

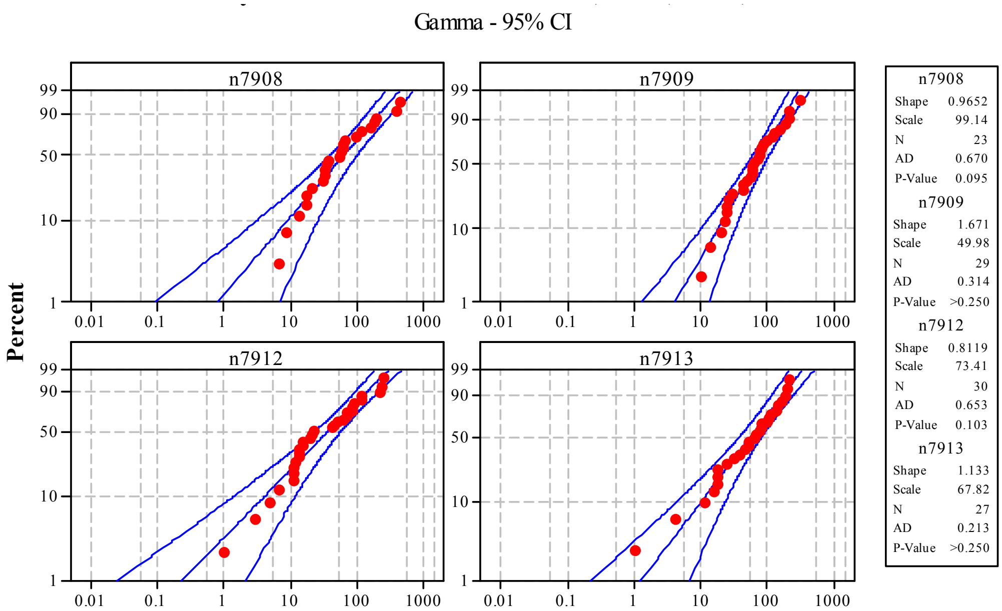

Figure 2 Probability plots of the successive time to failure by group

The estimated common CV of time to failure was 0.98. The 95% confidence intervals for the common CV were then computed. The approximate confidence interval was CI_A = (0.85, 1.11), the adjusted MOVER with the score confidence limits was CI_{AdMS} = (0.82, 1.10), and the adjusted MOVER with the Wald confidence limits was CI_{AdMW} = (0.88, 1.20). The interval lengths of CI_A, CI_{AdMS}, and CI_{AdMW} were 0.26, 0.27, and 0.32, respectively. From the results, little difference was evident in the time to failure of the different groups. CI_A provided the smallest length. The numerical results therefore matched those from the simulations.

4.2. Example 2

The real data on air pollution used in this section were obtained from Thailand's Pollution Control Department (http://www.aqmthai.com). They reported airborne particulate matter 2.5 (PM 2.5: μg/m^3.) in central Bangkok, Thailand, between 23 December 2018 and 22 January 2019. The data were reported from Pha Ya Thai, Din Dang, and Lat Phrao air monitoring quality stations. The distributions of the data are presented in Figure 3. By the Anderson-Darling (AD) test, PM 2.5 from all stations had a gamma distribution. The statistics are given in Table 5.

Table 5 Mean and CV for the data on air pollution from three stations in Bangkok

Station	n_i	$\bar{x}_i$	$\hat{\tau}_i$
Pha Ya Thai	31	40.10	0.39
Din Dang	31	56.10	0.34
Lat Phrao	31	51.65	0.34

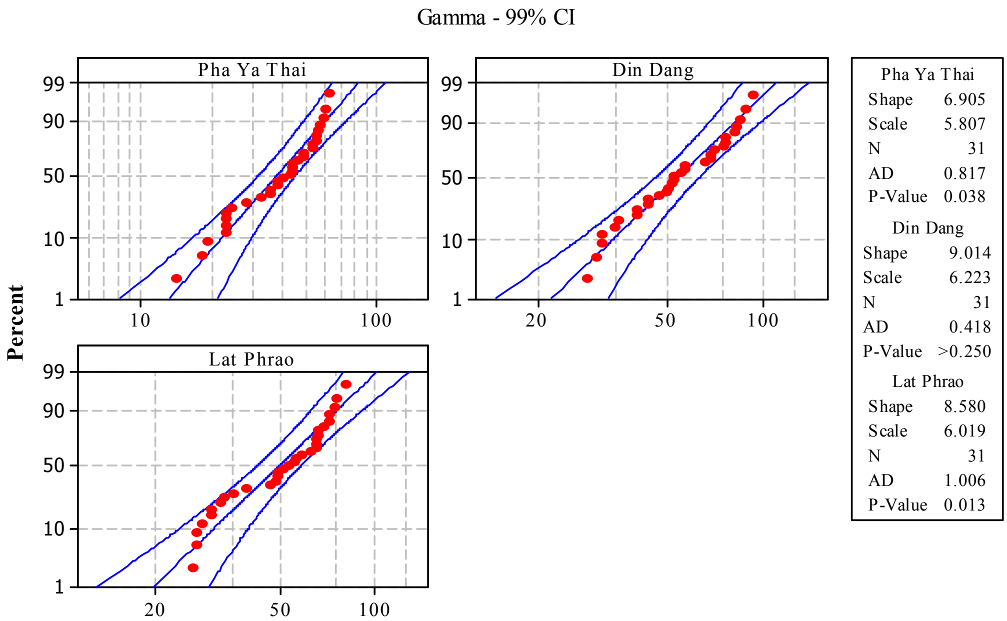

Figure 3 Probability plots of PM 2.5 from three air monitoring quality stations in Bangkok

On average, health effects would be expected from these PM 2.5 levels, as this reported unhealthy. From the dispersion, the estimated common CV in the three areas was 0.35. The 95% confidence intervals for the common CV were $CI_A = (0.30, 0.40)$, $CI_{AdMS} = (0.29, 0.40)$, and $CI_{AdMW} = (0.32, 0.44)$ with interval lengths of 0.10, 0.11, and 0.12, respectively. Again, CI_A gave the smallest interval length, and the real data confirmed the simulation results.

5. Conclusions

A basic approach to constructing the confidence interval for a parameter of interest involves finding the mean and variance of the parameter estimates. In cases where the exact mean and variance are difficult to determine, the delta method or generalized variable approach has been applied to construct the confidence interval. However, in the case of CV in the gamma distribution, the generalized pivotal quantity cannot be derived, because the generalized function depends on the parameters. Since its introduction, the method of variance of estimate recovery (MOVER) has been used to construct the confidence interval. Since the common CV in which we are interested is related to the CV for each sample together with its variances, variance estimation conducted using an extended version of the MOVER is applied. This approach is called the adjusted MOVER.

In this study, the performance of the proposed confidence intervals based on normal approximation (CI_A), the adjusted MOVER with the score method (CI_{AdMS}), and the adjusted MOVER with the Wald method (CI_{AdMW}) was evaluated using simulations. For 3, 4, and 5 populations, CI_{AdMW} performed well in terms of coverage probability. It had coverage probabilities satisfied the nominal coverage level in many cases. The expected lengths of our estimator were found to be small. Our interval estimator also provides the explicit close forms, which are expedient in practical computations. The findings confirm that the adjusted MOVER can be used to estimate the common CV of multiple gamma distributions, and are satisfied for many sample cases. Based on the numerical results, we suggest CI_{AdMW} as an alternative confidence interval for the common coefficient of variation of gamma distributions.

Acknowledgements

The authors thank the reviewers for suggestions which lead to considerable improvements in the manuscript.

References

Bakowski A, Radziszewski L, Zmindak M. Analysis of the coefficient of variation for injection pressure in a compression ignition engine. Procedia Eng. 2017; 177: 297-302.

Banik S, Kibria BMG. Estimating the population coefficient of variation by confidence intervals. Commun Stat Simul Comput. 2011; 40(8): 1236-1261.

Battista G, Pagliaroli T, Mauri L, Basilicata C, Vollaro R. Assessment of the air pollution level in the city of Rome (Italy). Sustainability. 2016; 8(9): 1-15.

Behboodian J, Jafari AA. Generalized inference for the common mean of several lognormal populations. J Stat Theory Appl. 2006; 5(3): 240-259.

Calif R, Soubdhan T. On the use of the coefficient of variation to measure spatial and temporal correlation of global solar radiation. Renew Energ. 2016; 88: 192-199.

Casella G, Berger RL. Statistical inference. California: Duxbury Press; 2002.

Castagliola P, Achouri A, Taleb H, Celano G, Psarakis S. Monitoring the coefficient of variation using a variable sampling interval control chart. Qual Reliab Eng Int. 2013; 29(8): 1135-1149.

Chadban SJ, Briganti EM, Kerr PG, et al. Prevalence of kidney damage in Australian adults: the AusDiab kidney study. Clin J Am Soc Nephrol. 2003; 14: 131-138.

Chen YH, Zhou XH. Interval estimates for the ratio and difference of two lognormal means. Stat Med. 2006; 25(23): 4099-4113.

Donner A, Zou GY. Closed-form confidence intervals for functions of the normal mean and standard deviation. Stat Methods Med Res. 2010; 21(4): 347-359.

Donner A, Zou GY. Estimating simultaneous confidence intervals for multiple contrasts of proportions by the method of variance estimates recovery. Stat Biopharm Res. 2011; 3(2): 320-335.

Finkelstein MO, Levin B. Statistics for lawyers: Statistics for social and behavioral sciences. New York: Springer-Verlag; 2001.

Gupta BC, Guttman I. Statistics and probability with applications for engineers and scientists. New Jersey: John Wiley; 2013.

Hartung J, Knapp G, Sinha BK. Statistical meta-analysis with applications. New Jersey: John Wiley; 2008.

Husak GJ, Michaelsen J, Funk C. Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. Int J Climatol. 2007; 27(7); 935-944.

Kenig B, Patenkovic A, Andelkovic M, Rada SM. Life history variation of drosophila subobscura under lead pollution depends on population history. Genetika. 2014; 46(3): 693-703.

Koopmans LH, Owen DB, Rosenblatt JI. Confidence interval for the coefficient of variation for the normal and lognormal distributions. Biometrika. 1964; 51(1): 25-32.

Krishnamoorthy K, Lu Y. Inferences on the common mean of several normal populations based on the generalized variable method. Biometrics. 2003; 59(2): 237-247.

Krishnamoorthy K, Mathew T. Inferences on the means of lognormal distributions using generalized p-values and generalized confidence interval. J Stat Plan Infer. 2003; 115(1): 103-121.

Krishnamoorthy K, Oral E. Standardized likelihood ratio test for comparing several log-normal means and confidence interval for the common mean. Stat Methods Med Res. 2017; 26(6): 2919-2937.

Lin SH, Lee JC. Generalized inferences on the common mean of several normal populations. J Stat Plan Infer. 2005; 134(2): 568-582.

Mahmoudvand R, Hassani H. Two new confidence intervals for the coefficient of variation in a normal distribution. J Appl Stat. 2009; 36(4): 429-442.

Ng CK. Inference on the common coefficient of variation when populations are lognormal: A simulation-based approach. J Stat Adv Theory Appl. 2014; 11(2): 117-134.

Niwitpong S. Confidence intervals for the difference and the ratio of lognormal means with bounded parameters. Songklanakarin J Sci Technol. 2015; 37(2): 231-240.

Niwitpong S, Wongkhao A. Confidence interval for the inverse of normal mean. Far East J Math Sci. 2015; 98(6): 689-698.

Popadin KY, Gutierrez-Arcelus M, Lappalainen T, et al. Gene age predicts the strength of purifying selection acting on gene expression variation in Humans. Am J Hum Genet. 2014; 95(6): 660-674.

Proschan F. Theoretical explanation of observed decreasing failure rate. Technometrics. 1963; 5(3): 375-383.

R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna: Austria; 2018 [cited 2018 Mar 1]. Available from: https://www.R-project.org.

Rice JA. Mathematical statistics and data analysis. California: Duxbury Press; 2006.

Rohan M, Fairweather A, Grainger N. Using gamma distribution to determine half-life of rotenone applied in freshwater. Sci Total Environ. 2015; 527-528: 246-251.

Sangnawakij P, Niwitpong S. Confidence Intervals for coefficients of variation in two-parameter exponential distributions. Commun Stat Simul Comput. 2016; 46(8): 6618-6630.

Sangnawakij P, Niwitpong S. Confidence intervals for functions of coefficients of variation with bounded parameter spaces in two gamma distributions. Songklanakarin J Sci Technol. 2017; 39(1): 27-39.

Shiina K, Tomiyama H, Takata Y, et al. Overlap syndrome: Additive effects of COPD on the cardiovascular damages in patients with OSA. Respir Med. 2012; 106(9): 1335-1341.

Thangjai W, Niwitpong S. Confidence intervals for the weighted coefficients of variation of two-parameter exponential distributions. Cogent Math. 2017; 4(1): 1-16.

Tian L. Inferences on the common coefficient of variation. Stat Med. 2005; 24(14): 2213-2220.

Tian L, Wu J. Inferences on the common mean of several log-normal populations: The generalized variable approach. Biom J. 2007; 49(6): 944-51.

Verrill S. Confidence bounds for normal and lognormal distribution coefficients of variation. United Stages Department of Agriculture, Forest Service, Forest Products Laboratory. Research Paper 2003; FPL-RP-609: 1-13.

Wong ACM, Wu J. Small sample asymptotic inference for the coefficient of variation: Normal and non-normal models. J Stat Plan Infer. 2002; 104(1): 73-82.

Xie H, Wu X. A conditional autoregressive range model with gamma distribution for financial volatility modelling. Econ Model. 2017; 64: 349-356.

Zou GY, Donner A. Construction of confidence limits about effect measures: A general approach. Stat Med. 2008; 27(10): 1693-1702.

Zou GY, Huang W, Zhang X. A note on confidence interval estimation for a linear function of binomial proportions. Comput Stat Data An. 2009; 53(4): 1080-1085.

Zou J, Dai W, Gong S, Ma Z. Analysis of spatial variations and sources of heavy metals in farmland soils of Beijing suburbs. PLoS One. 2015; 10(2): 1-13.