Bayesian estimation of rainfall dispersion in Thailand using gamma distribution with excess zeros

Wansiri Khooriphan, Sa-Aat Niwitpong and Suparat Niwitpong

Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

Corresponding author:
Sa-Aat Niwitpong

Email address: Sa-aat.n@sci.kmutnb.ac.th

ABSTRACT

The gamma distribution is commonly used to model environmental data. However, rainfall data often contain zero observations, which violates the assumption that all observations must be positive in a gamma distribution, and so a gamma model with excess zeros treated as a binary random variable is required. Rainfall dispersion is important and interesting, the confidence intervals for the variance of a gamma distribution with excess zeros help to examine rainfall intensity, which may be high or low risk. Herein, we propose confidence intervals for the variance of a gamma distribution with excess zeros by using fiducial quantities and parametric bootstrapping, as well as Bayesian credible intervals and highest posterior density intervals based on the Jeffrey's, uniform, or normal-gamma-beta prior. The performances of the proposed confidence interval were evaluated by establishing their coverage probabilities and average lengths via Monte Carlo simulations. The fiducial quantity confidence interval performed the best for a small probability of the sample containing zero observations (δ) whereas the Bayesian credible interval based on the normal-gamma-beta prior performed the best for large δ. Rainfall data from the Kiew Lom Dam in Lampang province, Thailand, are used to illustrate the efficacies of the proposed methods in practice.

INTRODUCTION

Thailand is a mainly agrarian country, with the largest agricultural area being in the north of the country due to its cooler climate making it the best place for cultivation. Rainfall is an important factor for cultivation. The rainy season begins in mid-May and ends in mid-October, the southwest monsoon predominate over Thailand to bring abundant annual rainfall. August to September is the wettest period of the year for most of the country, whereas January and December are very dry. Fluctuating rainfall makes it difficult to predict heavy precipitation that could lead to crop loss or damage. Since environmental data, meteorology, climatology and pollution studies are often right-skewed, the gamma distribution is commonly used to model these data. (Piao & Zhi-Sheng, 2015; Pradhan & Kundu, 2011; Son & Oh, 2006; Wang et al., 2019). Many researchers have developed confidence intervals for the parameters of a gamma distribution by using various methods. For example, Krishnamoorthy and León-Novelo (2014) proposed the parametric bootstrap (PB) confidence interval for the mean of a gamma distribution that performed satisfactorily even for small samples. Sangnawakij et al. (2016) proposed the method of variance estimates recovery (MOVER) and Score and Wald intervals to construct confidence intervals for the ratio of the coefficients of variation (CVs) of gamma distributions that performed better than classical estimators in terms of the expected length. Krishnamoorthy and Wang (2016) developed approximate fiducial quantities (FQs) for constructing the confidence interval for the mean of a gamma distribution that performed satisfactorily when the shape parameter was around 0.5 or larger. FQs can be used to establish approximate solutions to many statistical problems and can be readily applied to handle both uncensored and censored samples. Wang et al. (2019) proposed FQs for the differences between the shape parameters, scale parameters, and means of two independent gamma distributions and found that the performances of

the FQ-based confidence intervals were more accurate than other comparable methods.

Rainfall data often contain zero observations at certain times of the year and so this must be taken into account when studying precipitation in Thailand. Aitchison (1995) investigated situations where data contain zero observations with the probability of $0 < \delta < 1$ while the positive observations have a residual probability of $1-\delta$. Aitchison and Brown (1963) introduced the delta-lognormal distribution (a lognormal distribution with an excess of zero observations) for which the number of zero observations comprises a random variable with a binomial distribution and the positive observations comprise a random variable from a lognormal distribution. Many researchers have developed methods to construct confidence intervals for the parameters of a delta-lognormal distribution by using various methods. For example, Yosboonruang et al. (2019) proposed new confidence intervals for the CV of a delta-lognormal distribution by using Bayesian methods based on the independent Jeffreys', Jeffreys' rule, or uniform prior and compared them with the fiducial generalized confidence interval (FGCI); the Bayesian confidence interval based on the independent Jeffreys' prior performed better than the other methods in all situations studied. Maneerat and Niwitpong (2021) proposed confidence intervals for the common mean of several delta-lognormal distributions based on FGCI, the large-sample (LS) approach, MOVER, PB, and highest posterior density intervals (HPD) based on the Jeffreys' rule (HPD-JR) or normal-gamma-beta (HPD-NGB) prior; those based on MOVER and PB outperformed the others in a variety of situations. Several researchers have examined methods for constructing confidence intervals for a gamma distribution with excess zeros. Ren et al. (2021) proposed simultaneous confidence intervals for the difference between the means of multiple zero-inflated gamma distributions by using three fiducial methods and applied them to precipitation data. Muralidharan and Kale (2002) defined a modified gamma distribution with a singularity at zero and produced confidence intervals for the mean of a mixed distribution. Lecomte et al. (2013) provided compound Poisson-gamma and delta-gamma distributions to handle zero-inflated continuous data under variable sampling volume. Kaewprasert et al. (2022) proposed Bayesian estimation for the mean of delta-gamma distributions with application to rainfall data in Thailand.

In statistics, the variance, which gives a measure of the spread or variability of a distribution, is the second central moment, and the positive square root of the variance is the standard deviation (Casella & Berger, 2001). It is one of the most popular parameters of interest for probability and statistical inference.

We are interested to study the confidence interval for the variance of gamma distribution because it is commonly used to model environmental data such as a rainfall dispersion. Rainfall dispersion data can help to examine rainfall intensity, which may be high or low risk. We have studied many research related to constructing the confidence interval for rainfall data, such as Yosboonruang et al. (2019) and Maneerat and Niwitpong (2021). We have found several interesting priors, including: Jeffrey's, uniform, or normal-gamma-beta prior. Therefore, we applied to this study.

Since no publications have yet been forthcoming on constructing confidence intervals for the variance of a gamma distribution with excess zeros, the objective of the present study is to construct the confidence interval for the variance of a gamma distribution with excess zeros based on FQ, PB, and six Bayesian-based methods: three Bayesian confidence intervals based on the Jeffrey's (BAY-J), uniform (BAY-U), or normal-gamma-beta (BAY-NGB) prior and three highest posterior density intervals based on the Jeffrey's (HPD-J), uniform (HPD-U), or normal-gamma-beta (HPD-NGB) prior.

1 METHODS

Let X_i be a random variable following gamma(α, β) distribution with shape parameter α and scale parameter β. The probability density function can be derived as follows

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\Gamma(a)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & ; \quad x > 0, \\ 0 & ; \quad \text{otherwise.} \end{cases} \quad (1)$$

Suppose that the population of interest contains both zero and non-zero observations; the zero observations follow a binomial distribution while the non-zero observations follow a gamma distribution. The numbers of zero and non-zero observations are defined as $n_{(0)}$ and $n_{(1)}$ respectively, where $n = n_{(0)} + n_{(1)}$. Let $X = (X_1, X_2, ..., X_n)$ be a random sample from a gamma distribution with excess zeros denoted as $\Delta(\delta, \alpha, \beta)$. The distribution function for the confidence interval can be derived as

$$G(x_i; \delta, \alpha, \beta) = \begin{cases} \delta & ; \quad x = 0, \\ \delta + (1-\delta)F(x; \alpha, \beta) & ; \quad x > 0 \end{cases} \quad (2)$$

where $F(x; \alpha, \beta)$ is the gamma cumulative distribution function.

The maximum likelihood estimator of δ is $\widehat{\delta} = n_{(0)}/n$ The population mean and variance of X are respectively given by

$$E(X) = (1-\delta) \cdot (\alpha\beta) \tag{3}$$

$$Var(X) = \tau = (1-\delta) \cdot (\alpha\beta^2) + \delta(1-\delta) \cdot (\alpha\beta)^2 \tag{4}$$

The approches used to construct the confidence intervals are in the following subsections.

1.1 The FQ confidence interval

Krishnamoorthy, Mathew & Mukherjee (2008) suggested that a gamma distribution can be approximated by applying the cubic transformation of a Gaussian distribution. Let $Y_1, ..., Y_n$ be a sample from a gamma(α, β) distribution. When $X_i = Y_i^{\frac{1}{3}}, i = 1, ..., n$ then X_i are approximately normally distributed with mean μ and variance σ^2 respectively given by

$$\mu = (ba)^{\frac{1}{3}} \left(1 - \frac{1}{9a}\right) \quad \text{and} \quad \sigma^2 = \frac{b^{\frac{2}{3}}}{9a^{\frac{1}{3}}} \tag{5}$$

where shape parameter a and scale parameter b. The FQs for μ and σ^2 are, respectively,

$$Q_\mu = \bar{x} + \frac{Z\sqrt{n-1}}{\sqrt{\chi^2_{n-1}}} \cdot \frac{s}{\sqrt{n}} \quad \text{and} \quad Q_{\sigma^2} = \frac{(n-1)s^2}{\chi^2_{n-1}} \tag{6}$$

where $\bar{x}$ and s are the observed values of $\bar{X}$ and S, respectively; Z and χ^2_{n-1} represent independent random variable of standard normal and chi-squared distribution, respectively; and n is the sample size. The FQs for the parameters of a gamma distribution can thus be derived as

$$Q_a = \frac{1}{9}\left\{\left(1 + 0.5\frac{Q_\mu^2}{Q_{\sigma^2}}\right) + \left[\left(1 + 0.5\frac{Q_\mu^2}{Q_{\sigma^2}}\right)^2 - 1\right]^{\frac{1}{2}}\right\} \tag{7}$$

$$Q_b = 27(Q_a)^{\frac{1}{2}}(Q_{\sigma^2})^{\frac{3}{2}} \tag{8}$$

Krishnamoorthy & Wang (2016) proposed the FQs for the mean of gamma distribution as follows:

$$Q_M = \left\{\frac{Q_\mu}{2} + \sqrt{\left(\frac{Q_\mu}{2}\right)^2 + Q_{\sigma^2}}\right\}^3 \tag{9}$$

where Q_μ and Q_{σ^2} are defined in Equation 7. Li., Zhou & Tian (2013) proposed the FQ for δ as

$$Q_\delta \sim \frac{1}{2}\text{Beta}(n_{(1)}, n_{(0)} + 1) + \frac{1}{2}\text{Beta}(n_{(1)} + 1, n_{(0)}) \tag{10}$$

We can express the FQ for the variance as follows:

If $V = ab^2$, then we can write Equation 5 as

$$\mu = V^{\frac{1}{3}}b^{-\frac{1}{3}}\left(1 - \frac{b^2}{9V}\right) \quad \text{and} \quad \sigma^2 = \frac{b^{\frac{4}{3}}}{9V^{\frac{1}{3}}} \tag{11}$$

By solving the above equations for V, we obtain $V = ((\mu + \sqrt{\mu^2 + 4\sigma^2})/(2(9^{-1/4})(\sigma^2)^{-1/4}))^4$. Thus, the FQ for gamma variance can be obtained as

$$Q_V = \left\{\frac{Q_\mu + \sqrt{Q_\mu^2 + 4Q_{\sigma^2}}}{2(9^{-1/4})(Q_{\sigma^2})^{-1/4}}\right\}^4 \tag{12}$$

where Q_μ and Q_{σ^2} are defined in Equation 7. Thus, the FQ for τ is in the form

$$Q_\tau = (1 - Q_\delta) \cdot Q_V + Q_\delta (1 - Q_\delta) \cdot Q_M^2 \tag{13}$$

Therefore, the $100(1 - \alpha)\%$ confidence interval for τ is

$$CI_{FQ} = [Q_\tau(\alpha/2), Q_\tau(1 - \alpha/2)] \tag{14}$$

where $Q_\tau(\alpha/2)$ and $Q_\tau(1 - \alpha/2)$ are the $(\alpha/2)-$th and $(1 - \alpha/2)-$th percentiles of Q_τ, respectively.

The confidence intervals for τ can be obtained by using Algorithm 1.

Algorithm 1 FQ

1: Generate **x** from a gamma distribution with excess zeros, compute $\bar{x}$, and s^2 of the cube root transformed sample.
2: Generate a standard normal variate Z and a chi-square variate χ^2_{n-1}.
3: Generate $\text{Beta}(n_{(1)}, n_{(0)} + 1)$ and $\text{Beta}(n_{(1)} + 1, n_{(0)})$.
4: Compute Q_μ, Q_{σ^2} and Q_δ from Equation (6) and (10).
5: Compute the FQs for mean (Q_M) and variance (Q_V) of gamma distribution from Equation (9) and (12).
6: Compute Q_τ from Equation (6).
7: Repeat Steps 2–6 5,000 times and obtain an array of Q_τ.
8: Compute the 95% confidence intervals for τ from Equation (14).
9: Repeat Steps 1–8 10,000 times to compute the coverage probabilities (CPs) and the average lengths (ALs).

1.2 The PB confidence interval

The log-likelihood function for the vector of shape α and scale β parameters in gamma distribution is given by Saulo et al. (2018).

$$L(\alpha, \beta) = n\{\alpha \, \log(\beta) - \log[\Gamma(\alpha)]\} + (\alpha - 1) \sum_{i=1}^{n} \log(X_i) - \beta \sum_{i=1}^{n} X_i \tag{15}$$

Then, the maximum likelihood estimators (MLE) of α and β can be derived as

$$\widehat{\alpha} = \frac{0.5}{\log \, \bar{x} - \overline{\log \, x}} \tag{16}$$

$$\widehat{\beta} = \frac{\widehat{\alpha}}{\bar{x}} \tag{17}$$

The PB for variance of gamma distribution with excess zeros can be written as

$$\widehat{\tau}^* = (1 - \widehat{\delta}^*) \cdot \left(\frac{\widehat{\alpha}^*}{\widehat{\beta}^{*2}} \right) + \widehat{\delta}^* (1 - \widehat{\delta}^*) \cdot \left(\frac{\widehat{\alpha}^*}{\widehat{\beta}^*} \right)^2 \tag{18}$$

The $100(1 - \alpha)\%$ confidence interval for τ is

$$CI_{PB} = [\widehat{\tau}^*(\alpha/2), \widehat{\tau}^*(1 - \alpha/2)] \tag{19}$$

1.3 The Bayesian confidence intervals

For this study, let $Y_1, ..., Y_n$ be a sample from a $\text{gamma}(\alpha, \beta)$ distribution, then for $X_i = Y_i^{\frac{1}{3}}, i = 1, ..., n$ then X_i are approximately normally distributed with mean μ and variance σ^2 (Krishnamoorthy, Mathew

---

Algorithm 2 PB

1: Generate **x** from a gamma distribution with excess zeros, compute $\bar{x}$, $\widehat{\delta}$, $\widehat{\alpha}$ and $\widehat{\beta}$.
2: Generate x^* from x.
3: Compute $\bar{x}^*$, $\widehat{\delta}^*$, $\widehat{\alpha}^*$ and $\widehat{\beta}^*$.
4: Compute $\widehat{\tau}^*$ from Equation (18).
5: Repeat Steps 2–4 5,000 times and obtain an array of $\widehat{\tau}^*$.
6: Compute the 95% confidence intervals for $\widehat{\tau}^*$ from Equation (19).
7: Repeat Steps 1–6 10,000 times to compute the CPs and ALs.

---

97 & Mukherjee, 2008). From the law of large numbers, we know that $\mu \sim N(\bar{x}, \sigma^2/n)$ (Casella & Berger,
98 2001). Thus, the marginal posterior distribution of μ is $\mu|\sigma^2, x \sim N(\bar{x}, \sigma^2/n_{(1)})$
99   HPD intervals are constructed from the posterior distribution based on the Bayesian approach. The
100 HPD consists of the values of the parameter for which the posterior density is highest (Casella & Berger,
101 2001), while the HPD interval is the narrowest possible interval for the parameter of interest at probability
102 $100(1-\alpha)\%$ (Maneerat, Niwitpong & Niwitpong, 2020).
103   In this section, the Bayesian confidence interval is constructed upon the Jeffrey's priors, uniform
104 priors and normal-gamma-beta prior.

105 **1.3.1 The BAY-J and HPD-J intervals**

The Jeffrey's prior for δ in a binomial distribution is $p(\delta) \propto (\delta)^{-\frac{1}{2}}(1-\delta)^{\frac{1}{2}}$ (Bolstad & Curran, 2016).
This leads to obtaining the marginal posterior distribution of δ as

$$\delta_{jef}|x \sim Beta\left(n_{(0)} + \frac{1}{2}, n_{(1)} + \frac{3}{2}\right) \tag{20}$$

Jeffrey's prior for σ^2 in a lognormal distribution is $p(\sigma^2) \propto \sigma^{-2}$. Therefore, the marginal posterior
distribution of σ^2 becomes

$$\sigma^2_{jef}|x \sim IG\left(\frac{n_{(1)}}{2}, \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2}\right) \tag{21}$$

The marginal posterior distribution of μ is

$$\mu_{jef}|\sigma^2, x \sim N(\bar{x}, \sigma^2_{jef}/n_{(1)}) \tag{22}$$

We compute the mean and variance of gamma by using $\mu_{jef}|\sigma^2, x$ and $\sigma^2_{jef}|x$ as follows:

$$M_{BAY-J} = \left\{\frac{\mu_{jef}}{2} + \sqrt{\left(\frac{\mu_{jef}}{2}\right)^2 + \sigma^2_{jef}}\right\}^3 \tag{23}$$

$$V_{BAY-J} = \left\{\frac{\mu_{jef} + \sqrt{\mu^2_{jef} + 4\sigma^2_{jef}}}{2(9^{-1/4})(\sigma^2_{jef})^{-1/4}}\right\}^4 \tag{24}$$

So that

$$\widehat{\tau}_{BAY-J} = (1 - \delta_{jef}) \cdot V_{BAY-J} + \delta_{jef}(1 - \delta_{jef}) \cdot M^2_{BAY-J} \tag{25}$$

The confidence interval and HPD interval of τ based on the Jeffrey's prior are obtained as

$$CI_{BAY-J} = [\hat{\tau}_{BAY-J}(\alpha/2), \hat{\tau}_{BAY-J}(1 - \alpha/2)] \tag{26}$$

106 **1.3.2 The BAY-U and HPD-U intervals**

The uniform prior for δ in a binomial distribution is $p(\delta) \propto 1$ (Bolstad & Curran, 2016). This leads to
obtaining the marginal posterior distribution of δ as

$$\delta_{unif}|x \sim Beta(n_{(0)} + 1, n_{(1)} + 1) \tag{27}$$

The uniform prior for σ^2 is $\sigma^2 \propto 1$ (Kalkur & Rao, 2017). Subsequently, the marginal posterior distribution of σ^2 becomes

$$\sigma^2_{unif}|x \sim IG\left(\frac{n_{(1)}-2}{2}, \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}\right) \tag{28}$$

The marginal posterior distribution of μ as

$$\mu_{unif}|\sigma^2,x \sim N(\bar{x}, \sigma^2_{unif}/n_{(1)}) \tag{29}$$

We compute the mean and variance of a gamma distribution using $\mu_{unif}|\sigma^2,x$ and $\sigma^2_{unif}|x$ as follows:

$$M_{BAY-U} = \left\{\frac{\mu_{unif}}{2} + \sqrt{\left(\frac{\mu_{unif}}{2}\right)^2 + \sigma^2_{unif}}\right\}^3 \tag{30}$$

$$V_{BAY-U} = \left\{\frac{\mu_{unif} + \sqrt{\mu^2_{unif} + 4\sigma^2_{unif}}}{2(9^{-1/4})(\sigma^2_{unif})^{-1/4}}\right\}^4 \tag{31}$$

So that

$$\hat{\tau}_{BAY-U} = (1-\delta_{unif})\cdot V_{BAY-U} + \delta_{unif}(1-\delta_{unif})\cdot M^2_{BAY-U} \tag{32}$$

The confidence interval and HPD interval of τ based on the uniform prior are respectively obtained as

$$CI_{BAY-U} = [\hat{\tau}_{BAY-U}(\alpha/2), \hat{\tau}_{BAY-U}(1-\alpha/2)] \tag{33}$$

1.3.3 The BAY-NGB and HPD-NGB intervals

Maneerat & Niwitpong (2021) defined the normal-gamma-beta prior as

$$p(\tau) \propto \lambda^{-1}[(\delta)(1-\delta)]^{-1/2} \tag{34}$$

where $\lambda = \sigma^{-2}$, (μ, λ) follows a normal-gamma distribution and δ follows a beta distribution (Maneerat & Niwitpong, 2021). Thus, the marginal posterior distributions of δ, σ^2 and μ respectively become

$$\delta_{NGB}|x \sim Beta\left(n_{(0)}+\frac{1}{2}, n_{(1)}+\frac{1}{2}\right) \tag{35}$$

$$\sigma^2_{NGB}|x \sim IG\left(\frac{n_{(1)}-1}{2}, \frac{\sum_{i=1}^{n_{(1)}}(x_i-\mu)^2}{2}\right) \tag{36}$$

$$\mu_{NGB}|x \sim t_{2(n_{(1)}-1)}\left(\bar{x}, \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n_{(1)}(n_{(1)}-1)}\right) \tag{37}$$

We compute the mean and variance of a gamma distribution by using $\mu_{NGB}|x$ and $\sigma^2_{NGB}|x$ as follows:

$$M_{BAY-NGB} = \left\{\frac{\mu_{NGB}}{2} + \sqrt{\left(\frac{\mu_{NGB}}{2}\right)^2 + \sigma^2_{NGB}}\right\}^3 \tag{38}$$

$$V_{BAY-NGB} = \left\{\frac{\mu_{NGB} + \sqrt{\mu^2_{NGB} + 4\sigma^2_{NGB}}}{2(9^{-1/4})(\sigma^2_{NGB})^{-1/4}}\right\}^4 \tag{39}$$

So that

$$\hat{\tau}_{BAY-NGB} = (1-\delta_{NGB})\cdot V_{BAY-NGB} + \delta_{NGB}(1-\delta_{NGB})\cdot M^2_{BAY-NGB} \tag{40}$$

The confidence interval and HPD interval of τ based on the normal-gamma-beta prior are respectively obtained as

$$CI_{BAY-NGB} = [\hat{\tau}_{BAY-NGB}(\alpha/2), \hat{\tau}_{BAY-NGB}(1-\alpha/2)] \tag{41}$$

Algorithm 3 Bayesian interval

1: Generate **x** from a gamma distribution with excess zeros, compute $\widehat{\delta}$, $\widehat{\mu}$, and $\widehat{\sigma}^2$.
2: Generate $\delta|x$ from Equation (20), (26) and (33).
3: Generate $\sigma^2|x$ from Equation (21), (27) and (34).
4: Given $\sigma^2|x$ generate $\mu|\sigma^2, x$.
5: Compute mean and variance of gamma distribution from Equation (22), (23), (28), (29), (36) and (37).
6: Compute $\widehat{\tau}$ from Equation (24), (30) and (38).
7: Compute the 95% confidence intervals and HPD for $\widehat{\tau}$ from Equation (25), (31) and (39).
8: Repeat Steps 1–7 10,000 times to compute the CPs and ALs.
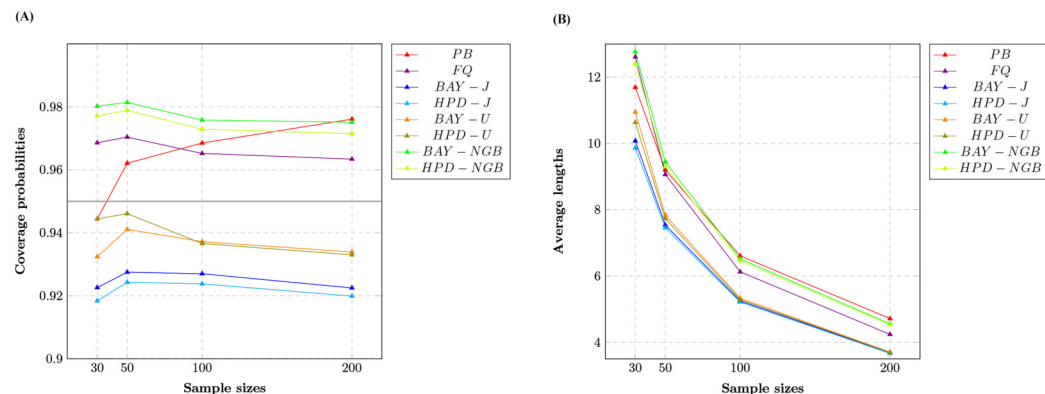
Figure 1. Line graphs of (A) coverage probabilities and (B) average lengths of all methods in case of the different sample sizes.

2 SIMULATION STUDIES AND RESULTS

A Monte Carlo simulation study with 10,000 replications (M) and 5,000 repetitions (m) for FQ and PB, was conducted at a nominal confidence level of 0.95. We set sample size *n* as 30, 50, 100 or 200 and probability of zeros δ as 0.2, 0.5 or 0.8, for which we set shape parameter α as 7.00, 7.50 or 7.75; 2.00, 2.50 or 2.75; and 1.25, 1.50 or 1.75, respectively. We set rate parameter β as 1 for all cases. The performances of the confidence intervals were assessed by comparing their coverage probabilities (CPs) and average lengths (ALs); the best-performing confidence interval for a particular situation was identified as having a CP close or greater than 0.95 and the shortest AL. The confidence intervals for the variance of gamma distribution with excess zeros constructed using FQ, PB, BAY-J, HPD-J, BAY-U, HPD-U, BAY-NGB and HPD-NGB.

We report the coverage probabilities and the average lengths of nominal 95% two-sided confidence intervals for variance of gamma distribution with excess zeros in Table 1 and Figures 1-3.
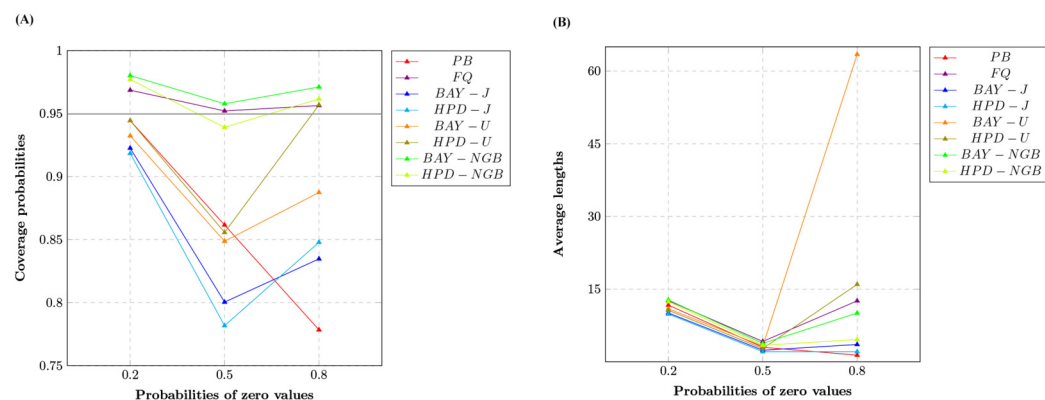
Figure 2. Line graphs of (A) coverage probabilities and (B) average lengths of all methods in case of the different probabilities of zero values.

Table 1. The Coverage probabilities and (Average lengths) of nominal 95% two-sided confidence intervals for variance of gamma distribution with excess zeros

n	δ	α	Coverage probability (Average length)							
			PB	FQ	BAY-J	HPD-J	BAY-U	HPD-U	BAY-NGB	HPD-NGB
30	0.2	7.00	0.9444	**0.9686**	0.9226	0.9184	0.9324	0.9444	**0.9802**	**0.9771**
			(11.6924)	(12.6084)	(10.0751)	(9.8624)	(10.9472)	(10.6335)	(12.7642)	(*12.4067*)
		7.50	0.9480	**0.9728**	0.9317	0.9293	0.9420	**0.9522**	**0.9826**	**0.9789**
			(12.8866)	(13.6348)	(11.0851)	(10.8819)	(11.9674)	(*11.6665*)	(13.9564)	(13.6000)
		7.75	**0.9541**	**0.9731**	0.9378	0.9334	0.9482	**0.9569**	**0.9827**	**0.9807**
			(13.5974)	(14.3134)	(11.7150)	(11.5094)	(12.6155)	(*12.3114*)	(14.6896)	(14.3333)
	0.5	2.00	0.8616	**0.9521**	0.8004	0.7817	0.8487	0.8557	**0.9578**	0.9391
			(2.9978)	(4.1962)	(2.3918)	(2.0896)	(3.2420)	(2.7330)	(*3.8034*)	(3.3989)
		2.50	0.8629	**0.9500**	0.7903	0.7788	0.8354	0.8502	**0.9556**	0.9440
			(3.7780)	(5.3509)	(3.0529)	(2.7002)	(4.0796)	(3.4959)	(*4.8638*)	(4.4099)
		2.75	0.8601	0.9467	0.7850	0.7767	0.8308	0.8433	**0.9543**	0.9454
			(4.1300)	(5.8440)	(3.3308)	(2.9635)	(4.4162)	(3.8167)	(5.3407)	(4.8762)
	0.8	1.25	0.7784	**0.9564**	0.8347	0.8479	0.8874	**0.9569**	**0.9711**	**0.9616**
			(1.3763)	(12.5762)	(3.5742)	(2.0919)	(63.4518)	(15.9813)	(10.0543)	(*4.5952*)
		1.50	0.7932	**0.9615**	0.8403	0.8577	0.8897	**0.9603**	**0.9754**	**0.9671**
			(1.6638)	(13.2999)	(4.0138)	(2.4576)	(63.9289)	(6.6673)	(10.6502)	(*5.1506*)
		1.75	0.8048	**0.9621**	0.8489	0.8647	0.8937	**0.9637**	**0.9793**	**0.9725**
			(1.9395)	(12.9027)	(4.1595)	(2.7093)	(53.5498)	(15.2379)	(10.3024)	(*5.4055*)
50	0.2	7.00	**0.9621**	**0.9704**	0.9275	0.9243	0.9411	0.9461	**0.9814**	**0.9789**
			(9.2009)	(*9.0634*)	(7.5400)	(7.4561)	(7.8353)	(7.7315)	(9.4418)	(9.2934)
		7.50	**0.9625**	**0.9704**	0.9338	0.9296	0.9447	**0.9506**	**0.9807**	**0.9779**
			(10.1651)	(9.9058)	(8.3868)	(8.3065)	(8.6808)	(*8.5812*)	(10.4194)	(10.2715)
		7.75	**0.9655**	**0.9729**	0.9374	0.9367	0.9463	0.9498	**0.9844**	**0.9826**
			(10.6812)	(*10.3530*)	(8.8378)	(8.7599)	(9.1334)	(9.0356)	(10.9368)	(10.7863)
	0.5	2.00	0.9054	0.9478	0.7868	0.7573	0.8238	0.8155	**0.9505**	0.9285
			(2.4797)	(2.6883)	(1.6201)	(1.4938)	(1.8801)	(1.7160)	(2.5473)	(2.3981)
		2.50	0.9010	0.9475	0.7890	0.7693	0.8228	0.8202	**0.9514**	0.9341

Table 1. Coverage probability and (Average length)*continued*

			(3.0615)	(3.4346)	(2.0567)	(1.9090)	(2.3755)	(2.1861)	(3.2687)	(3.1047)
		2.75	0.9039	**0.9515**	0.7892	0.7674	0.8223	0.8193	**0.9538**	0.9417
			(3.3850)	(3.8265)	(2.2825)	(2.1243)	(2.6329)	(2.4295)	*(3.6435)*	(3.4714)
	0.8	1.25	0.8435	**0.9559**	0.8337	0.8262	0.8882	0.9116	**0.9688**	0.9476
			(1.1826)	(2.4727)	(1.2830)	(1.0168)	(2.5640)	(1.7317)	*(2.1219)*	(1.6296)
		1.50	0.8550	**0.9569**	0.8384	0.8402	0.8860	0.9161	**0.9699**	**0.9538**
			(1.4185)	(2.8663)	(1.5275)	(1.2448)	(2.8649)	(2.0242)	(2.4800)	*(1.9666)*
		1.75	0.8675	**0.9602**	0.8515	0.8537	0.8930	0.9239	**0.9736**	**0.9622**
			(1.6807)	(3.2832)	(1.7911)	(1.4943)	(3.1990)	(2.3387)	(2.8757)	*(2.3349)*
100	0.2	7.00	**0.9685**	**0.9652**	0.9270	0.9238	0.9372	0.9366	**0.9758**	**0.9729**
			(6.6077)	*(6.1266)*	(5.2494)	(5.2171)	(5.3267)	(5.2916)	(6.5195)	(6.4617)
		7.50	**0.9732**	**0.9682**	0.9321	0.9311	0.9394	0.9407	**0.9801**	**0.9785**
			(7.3264)	*(6.7357)*	(5.8730)	(5.8403)	(5.9473)	(5.9122)	(7.2284)	(7.1679)
		7.75	**0.9760**	**0.9702**	0.9437	0.9426	0.9501	0.9512	**0.9841**	**0.9817**
			(7.6438)	*(7.0104)*	(6.1733)	(6.1407)	(6.2465)	(6.2117)	(7.5695)	(7.5095)
	0.5	2.00	0.9332	0.9292	0.7597	0.7285	0.7931	0.7636	0.9316	0.9120
			(1.8169)	(1.6738)	(1.0400)	(0.9967)	(1.1103)	(1.0616)	(1.6412)	(1.5930)
		2.50	0.9360	0.9420	0.7703	0.7434	0.7995	0.7783	0.9436	0.9306
			(2.2541)	(2.1692)	(1.3337)	(1.2817)	(1.4222)	(1.3641)	(2.1292)	(2.0751)
		2.75	0.9301	0.9392	0.7763	0.7528	0.7995	0.7875	0.9425	0.9295
			(2.4789)	(2.4163)	(1.4761)	(1.4200)	(1.5739)	(1.5114)	(2.3735)	(2.3171)
	0.8	1.25	0.9076	0.9439	0.8161	0.7969	0.8573	0.8475	**0.9550**	0.9302
			(0.9191)	(1.0226)	(0.6335)	(0.5746)	(0.7685)	(0.6809)	(0.9717)	(0.8766)
		1.50	0.9159	**0.9526**	0.8333	0.8141	0.8667	0.8630	**0.9624**	0.9427
			(1.0920)	*(1.2439)*	(0.7821)	(0.7184)	(0.9349)	(0.8409)	(1.1916)	(1.0887)
		1.75	0.9123	**0.9544**	0.8394	0.8267	0.8696	0.8697	**0.9678**	0.9482
			(1.2881)	(1.4819)	(0.9445)	(0.8765)	(1.1158)	(1.0159)	*(1.4312)*	(1.3194)
200	0.2	7.00	**0.9761**	**0.9634**	0.9225	0.9199	0.9339	0.9330	**0.9751**	**0.9715**
			(4.7169)	*(4.2392)*	(3.6845)	(3.6666)	(3.7070)	(3.6888)	(4.5589)	(4.5303)
		7.50	**0.9785**	**0.9665**	0.9317	0.9304	0.9442	0.9428	**0.9775**	**0.9755**
			(5.1932)	*(4.6428)*	(4.1173)	(4.0987)	(4.1390)	(4.1201)	(5.0479)	(5.0179)

Table 1. Coverage probability and (Average length)*continued*

	7.75	**0.9822**	**0.9692**	0.9403	0.9384	0.9483	0.9469	**0.9817**	**0.9799**
		(5.4485)	*(4.8598)*	(4.3497)	(4.3307)	(4.3699)	(4.3503)	(5.3069)	(5.2765)
0.5	2.00	0.9477	0.8978	0.6997	0.6659	0.7285	0.6938	0.9000	0.8774
		(1.3034)	(1.1146)	(0.7016)	(0.6854)	(0.7237)	(0.7066)	(1.1126)	(1.0944)
	2.50	0.9463	0.9201	0.7363	0.7060	0.7590	0.7326	0.9209	0.9059
		(1.6261)	(1.4556)	(0.9051)	(0.8852)	(0.9330)	(0.9121)	(1.4510)	(1.4304)
	2.75	0.9470	0.9297	0.7443	0.7145	0.7667	0.7419	0.9302	0.9162
		(1.7859)	(1.6291)	(1.0051)	(0.9835)	(1.0358)	(1.0131)	(1.6250)	(1.6031)
0.8	1.25	0.9426	0.9268	0.7829	0.7506	0.8168	0.7872	0.9383	0.9111
		(0.6664)	(0.5858)	(0.3827)	(0.3651)	(0.4131)	(0.3922)	(0.5860)	(0.5575)
	1.50	0.9476	0.9450	0.8173	0.7885	0.8451	0.8224	**0.9553**	0.9334
		(0.8008)	(0.7339)	(0.4859)	(0.4665)	(0.5214)	(0.4983)	(0.7377)	(0.7060)
	1.75	0.9462	0.9480	0.8317	0.8124	0.8542	0.8393	**0.9594**	0.9419
		(0.9469)	(0.8890)	(0.5953)	(0.5742)	(0.6363)	(0.6114)	(0.8986)	(0.8637)

*The coverage probabilities greater than the nominal confidence level of 0.95 be in bold and the shortest average lengths be in italic.
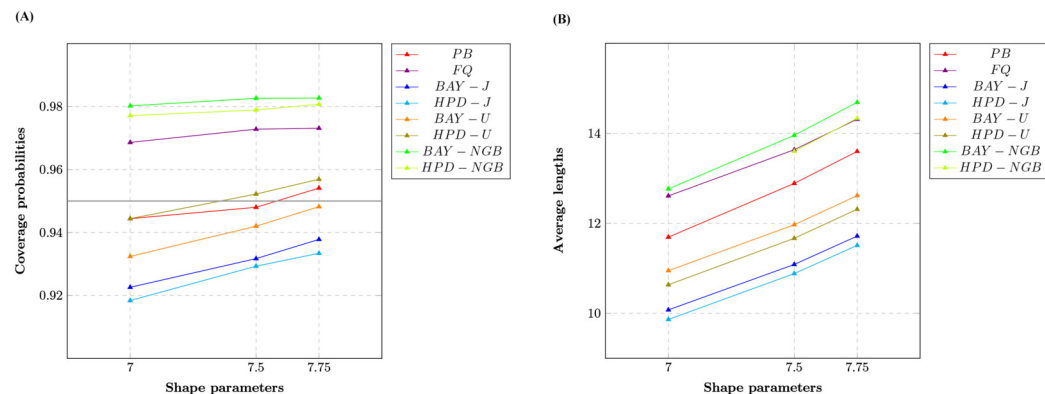
Figure 3. Line graphs of (A) coverage probabilities and (B) average lengths of all methods in case of the different shape parameters.

The CPs of the PB, FQ, HPD-U, BAY-NGB, and HPD-NGB confidence intervals were greater than or close to the nominal confidence level of 0.95 in all situations studied. For a small-to-moderate sample size, FQ and the HPD-U performed well for small δ whereas BAY-NGB and HPD-NGB performed well for large δ. For a large sample size, FQ performed well for small δ whereas BAY-NGB performed well for large δ. Although the expected lengths of the HPD-J were shorter than the other methods, the CPs of BAY-J and HPD-J were lower than the nominal confidence level in all cases.

The findings show that although FQ, HPD-U, BAY-NGB, and HPD-NGB attained acceptable CPs, the ALs of BAY-NGB and the HPD-NGB were shorter than the other methods, and so they can be recommended for constructing the confidence interval for the variance of a gamma distribution with excess zeros. It can be seen that for HPD-NGB developed from the study of Maneerat Niwitpong (2021), the simulation results are similar to these studies. For small-to-large sample size, HPD-NGB performed well. BAY-NGB and HPD-NGB are the best because BAY-NGB and HPD-NGB attained stable CPs and ALs were shorter than the other methods for all sample sizes. A referee suggests to check a validity and robustness of the model for smaller sample sizes with moderate number of zeros. We, therefore, simulated a study with 10,000 replications (M) and 5,000 repetitions (m) for FQ and PB, was conducted at a nominal confidence level of 0.95. We set sample size n as 10 or 20 and probability of zeros δ as 0.2, or 0.5, for which we set shape parameter α as 7.00, 7.50 or 7.75; and 2.00, 2.50 or 2.75, respectively. We set rate parameter β as 1 for all cases. The results (not show here) shown that the CPs of the FQ, HPD-U, BAY-NGB, and HPD-NGB confidence intervals were greater than or close to the nominal confidence level of 0.95 in all situations studied. The findings show that although FQ, HPD-U, BAY-NGB, and HPD-NGB attained acceptable CPs, the ALs of HPD-NGB were shorter than the other methods. Although the sample sizes are small (n=10, n=20), our findings show that BAY-NGB and HPD-NGB can be recommended for constructing the confidence interval for the variance of a gamma distribution with excess zeros.

3 EMPIRICAL APPLICATION OF THE PROPOSED CONFIDENCE INTERVALS

The confidence interval performances were compared by using real-world datasets comprising monthly rainfall data reported by the Upper Northern Region Irrigation Hydrology for January and February 1993 to 2021 at the Kiew Lom Dam, Lampang province, Thailand.

First, the best fit for the positive rainfall data among normal, lognormal, Cauchy, and gamma models was examined by calculating their Akaike information criterion (AIC) and Bayesian information criterion (BIC) values (Table 2). The results show that the lowest AIC and BIC values (207.7139 and 210.2301, respectively) were for the gamma distribution, indicating that it was the best fit for the data.

Table 2. AIC and BIC results of positive rainfall data

Models	Normal	Lognormal	Cauchy	Gamma
AIC	224.9317	216.186	230.4221	207.7139
BIC	227.4479	218.7022	232.9383	210.2301

The summary statistics for the rainfall data in Kiew Lom Dam Lampang province are $\bar{x} = 18.6461, n = 58, n_{(1)} = 26, n_{(0)} = 32$, while the maximum likelihood estimators for δ, α, β and τ are $\hat{\delta} = 0.5517, \hat{\alpha} = 0.7297, \hat{\beta} = 0.0391$ and $\hat{\tau} = 299.5542$, respectively. The calculated two-sided confidence intervals for τ are reported in Table 3.

Table 3. The 95% two-sided confidence intervals for variance of rainfall data in Kiew Lom Dam in Lampang province

Methods	confidence intervals for θ		length of intervals
	Lower	Upper	
PB	115.6468	543.4372	427.7903
FQ	115.3533	974.3039	858.9506
BAY-J	138.7433	764.2119	625.4687
HPD-J	107.4391	613.0399	505.6008
BAY-U	146.4527	1078.71	932.2578
HPD-U	111.2196	809.1257	697.9061
BAY-NGB	135.4990	885.4536	749.9546
HPD-NGB	102.1386	685.1513	583.0128

For $n = 50$ and $\delta = 0.5$, FQ and BAY-NGB obtained CPs close to the nominal confidence level of 0.95, but BAY-NGB bay obtained the shortest length method. Thus, the BAY-NGB method is recommended for constructing the confidence interval for the variance in rainfall data in January and February at the Kiew Lom Dam in Lampang province.

4 CONCLUSIONS

We constructed confidence intervals for the variance of a gamma distribution with excess zeros by using the PB, FQ, BAY-J, HPD-J, BAY-U, HPD-U, BAY-NGB, and HPD-NGB approaches. The CPs and ALs of the methods were assessed by Monte Carlo simulation for various situations and by using real precipitation data following a gamma distribution with excess zeros. Our findings show that BAY-NGB and HPD-NGB can be recommended for constructing the confidence interval for the variance of a gamma distribution with excess zeros. In future research, we will investigate constructing confidence intervals for the difference between the variances of gamma distributions with excess zeros.

REFERENCES

[1] Aitchison, J. (1955). On the Distribution of a Positive Random Variable Having a Discrete Probability Mass at the Origin. *Journal of the American Statistical Association*, 50(271):901-908. https://doi.org/10.2307/2281175.

[2] Aitchison, J., and Brown, J. A. C. (1963). The lognormal distribution: With special reference to its uses in economics London, UK: Cambridge University Press.

[3] Bolstad, W. M., Curran, J. M. (2016). Introduction to Bayesian Statistics (3rd ed.). Wiley.

[4] Casella, G., Berger, R. L. (2001). Statistical Inference (2nd ed.). Cengage Learning.

[5] Kaewprasert, T., Niwitpong, S. A., Niwitpong, S. (2022). Bayesian estimation for the mean of delta-gamma distributions with application to rainfall data in Thailand. *PeerJ*, 10:e13465. https://doi.org/10.7717/peerj.13465.

[6] Kalkur TA, Rao A. (2017). Bayes estimator for coefficient of variation and inverse coefficient of variation for the normal distribution. *International Journal of Statistics and Systems*. 12(4):721–732.

[7] Krishnamoorthy, K., León-Novelo, L. (2014). Small sample inference for gamma parameters: one-sample and two-sample problems. *Environmetrics*, 25(2):107–126. https://doi.org/10.1002/env.2261.

[8] Krishnamoorthy, K., Mathew, T., Mukherjee, S. (2008). Normal-Based Methods for a Gamma Distribution. *Technometrics*, 50(1):69–78. https://doi.org/10.1198/004017007000000353.

[9] Krishnamoorthy, K., Wang, X. (2016). Fiducial confidence limits and prediction limits for a gamma distribution: Censored and uncensored cases. *Environmetrics*, 27(8):479–493. https://doi.org/10.1002/env.2408.

[10] Lecomte, J. B., Benoît, H. P., Ancelet, S., Etienne, M. P., Bel, L., Parent, E. (2013). Compound Poisson-gamma vs. delta-gamma to handle zero-inflated continuous data under a variable sampling volume. In R. B. O'Hara (Ed.), Methods in Ecology and Evolution (pp. 1159–1166). Wiley. https://doi.org/10.1111/2041-210x.12122.

[11] Li, X., Zhou, X., Tian, L. (2013). Interval estimation for the mean of lognormal data with excess zeros. *Statistics Probability Letters*, 83(11): 2447–2453. https://doi.org/10.1016/j.spl.2013.07.004.

[12] Maneerat, P., Niwitpong, S. A. (2021). Estimating the average daily rainfall in Thailand using confidence intervals for the common mean of several delta-lognormal distributions. *PeerJ*, 9:e10758. https://doi.org/10.7717/peerj.10758.

[13] Maneerat, P., Niwitpong, S., Niwitpong, S. (2020). Bayesian confidence intervals for the difference between variances of delta-lognormal distributions. *Biometrical Journal*, 62(7):1769–1790. https://doi.org/10.1002/bimj.201900079.

[14] Muralidharan, K., Kale, B. K. (2002). Modified gamma distributions with singularity at zero. Communications in Statistics - Simulation and Computation, 31(1): 143–158. https://doi.org/10.1081/sac-120002720.

[15] Piao, C., Zhi-Sheng, Y. (2015, December). Tolerance limits for gamma distribution based on generalized fiducial method. 2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). https://doi.org/10.1109/ieem.2015.7385803.

[16] Pradhan, B., Kundu, D. (2011). Bayes estimation and prediction of the two-parameter gamma distribution. *Journal of Statistical Computation and Simulation*, 81(9):1187–1198. https://doi.org/10.1080/00949651003796335.

[17] Ren, P., Liu, G., Pu, X. (2021). Simultaneous confidence intervals for mean differences of multiple zero-inflated gamma distributions with applications to precipitation. Communications in Statistics - Simulation and Computation, 1–12. https://doi.org/10.1080/03610918.2021.1966466.

[18] Sangnawakij, P., Niwitpong, S. A., Niwitpong, S. (2015). Confidence Intervals for the Ratio of Coefficients of Variation of the Gamma Distributions. Lecture Notes in Computer Science, 193–203. https://doi.org/10.1007/978-3-319-25135-6_19.

[19] Saulo, H., Bourguignon, M., Zhu, X., Balakrishnan, N. (2018). Some simple estimators for the two-parameter gamma distribution. Communications in Statistics - Simulation and Computation, 48(8):2425–2437. https://doi.org/10.1080/03610918.2018.1457693.

[20] Son, Y. S., Oh, M. (2006). Bayesian Estimation of the Two-Parameter Gamma Distribution. Communications in Statistics - Simulation and Computation, 35(2):285–293. https://doi.org/10.1080/03610910600591925.

[21] Wang, X., Zou, C., Yi, L., Wang, J., Li, X. (2019). Fiducial inference for gamma distributions: two-sample problems. Communications in Statistics - Simulation and Computation, 50(3):811–821. https://doi.org/10.1080/03610918.2019.1568471.

[22] Yosboonruang, N., Niwitpong, S. A., Niwitpong, S. (2019). Measuring the dispersion of rainfall using Bayesian confidence intervals for coefficient of variation of delta-lognormal distribution: a study from Thailand. *PeerJ*, 7:e7344. https://doi.org/10.7717/peerj.7344.