

Response to the Reviewers

Journal: PeerJ

Manuscript ID: 63996

Title name: Bayesian computation for the common coefficient of variation of delta-lognormal distributions with application to common rainfall dispersion in Thailand

Authors: Noppadon Yosboonruang, Sa-Aat Niwitpong, and Suparat Niwitpong

Dear Reviewers,

We are very grateful for your constructive comments to our manuscript. The manuscript was revised in accordance with your valuable suggestions. Below are our point-by-point replies to your comments.

Lists of correction:

Reviewer 1

1. Minor. Lines 15, 43, etc., describe rainfall as following a delta lognormal distribution. The authors consider daily rainfall. Add “daily” throughout (e.g., replacement at line 81).

Response:

Thank you for your suggestion. We added daily rainfall instead of rainfall according to the context of each sentence throughout the revised manuscript.

2. My sole major concern is that the authors’ Monte-Carlo simulations were made for the delta log-normal. The authors find good performance with the equal-tailed Bayesian based on the independent Jeffreys prior (Figures 1-4). For that analysis type, please add additional simulation with small discrepancy to the probability distribution and evaluate coverage under such conditions.

Response:

We added additional simulation with small discrepancy ($\sigma^2 = 0.1$) to the probability distribution and evaluate coverage under such conditions which presented in Tables 1-3 and Figures 3-8 in a revised manuscript.

Reviewer 2

1. The emphasis of this article is not confirmed, it should be an applied one, just from the title. In this case, it should start with the actual issue (rainfall data) to illustrate the importance of interval estimation.

Response:

Thank you for your recommendation. We have made improvements starting with rainfall data to illustrate the importance of interval estimation in the first paragraph of introduction section on pages 1-2 in a revised manuscript.

2. From the existing literature, the rainfall data follow the delta-lognormal distribution, while the existing confirm interval estimates are not available based on this data, it necessary to study the estimation based on the Lognormal distribution. Then, the next issue is the difficulty of direct estimation?

Response:

Estimating the coefficient of variation of the delta-lognormal distribution, equation (4), is more complicated than the lognormal distribution ($\delta_i = 0$). Therefore, it is difficult to directly estimate the common coefficient of variation of the delta-lognormal distributions because zero values must be estimated from the binomial distributions.

3. The reason to take Bayes?

Response:

We proposed Bayesian method because the highest posterior density interval by this method is the shortest possible interval containing $100(1-\alpha)\%$ of the posterior probability such that the density within the interval have a higher probability than that outside. Moreover, we can give the posterior density function as determined by likelihood function and the several priors' distributions.

4. Conclusion with estimate data from simulation & data. The reasonable illustrate of the research methodologies is not available, just copy & paste.

Response:

We have summarized the results of the actual data in lines 224-227 on page 15 in a revised manuscript as follows:

“The results reveal that the three confidence intervals tested contained the real value of the parameter, thereby reinforcing the conclusions based on the simulation study results. However, the expected length of FGCI was the shortest, thereby making it a good choice for estimating the common CV in the dispersion of precipitation from the three areas in Nan province, Thailand.”

Moreover, we explained further in the conclusion section in lines 244-248 on page 16 in a revised manuscript as follows:

“The results indicate that... This is due to the parameter to be estimated relying on the posterior densities of δ_i^* and σ_i^2 . According to the results of the simulation study and

the real data example, the Bayesian credible interval based on the independent Jeffreys prior is suitable for cases with small variances since it provided the narrowest length of the interval due to it falling in the domain of its posterior density.”

Reviewer 3

1. What does the word "common" mean for this work? Please describe.

Response:

The word “common” for this work means mutually or joint which is used to study more than one delta-lognormal coefficients of variation at one time.

2. What is the method you used to derive the estimated variance in (7)? How?

Response:

We used the delta method to derive the estimated variance in (7) which can be seen from the article by Yosboonruang et al., 2018.

“Yosboonruang, N., Niwitpong, S.-A., and Niwitpong, S. (2018). Confidence intervals for the coefficient of variation of the delta-lognormal distribution. In Anh, L. H., Dong, L. S., Kreinovich, V., and Thach, N. N., editors, *Econometrics for Financial Applications*, volume 760 of *Studies in Computational Intelligence*, pages 327–337. Springer International Publishing, Cham.”

3. Please give a discussion why the MOVER method provides low performance. Especially, for large n , as the method depends on CLT.

Response:

We discussed about the MOVER method provides low performance in the discussion section in lines 233-237 on page 16 in a revised manuscript as follows:

“Although MOVER performed well for cases with a high proportion of non-zero values, it produced coverage probabilities that were lower than the nominal confidence level for cases with small variances. This is probably because the lower and upper bounds of the zero values are used in the confidence interval construction, and the combined effect with the other parameters caused the inadequate coverage probability results.”.

4. Your results for the expected length are not consistent (Figure 4).

Response:

For Figure 4 (changed to Figure 6 in a revised manuscript), we ignored FGCI method since it is too wide of the expected length.

5. It will be useful if you could provide the R code and a set of real data for application, not only the code for simulation.

Response:

We provided the R code and a set of real data for application as supplemental file “R-code-raw data.R”.

6. The data used here are interested, but what is the reason to use. I would expect to see the other type of data. So, it will be different from "Measuring the dispersion of rainfall using Bayesian confidence intervals.... (2019)".

Response:

The rainfall data was used because it consists of positive values and true zero values. The positive values have lognormal distribution and true zero values have binomial distribution which corresponds to the delta-lognormal distribution. Moreover, Thailand has flooding in monsoon season due to heavy rain in some areas causing damage. Therefore, we interested to study the dispersion of rainfall in Thailand to apply in planning for preventing with such incidents that may occur in the future. Since this manuscript focused on daily rainfall data. Therefore, the other type of data will be presented here. In addition, we added an example of medical charges data which taken from Zhou and Tu (1999).

“Zhou, X. H. and Tu, W. (1999). Comparison of several independent population means when their samples contain log-normal and possibly zero observations. *Biometrics*, 55:645–651.”

The medical charges were divided into two inpatient groups including control group (non-treatment) and intervention-based treatment group. The summary statistics are $n_1 = 147$, $\hat{\delta}_1 = 0.19$, $\hat{\mu}_1 = 9.03$, $\hat{\sigma}_1^2 = 1.37$ and $n_2 = 119$, $\hat{\delta}_2 = 0.17$, $\hat{\mu}_2 = 9.33$, $\hat{\sigma}_2^2 = 0.41$. The 95% confidence intervals and credible intervals for the common coefficient of variation of medical care costs data as follows:

Methods	Lower	Upper	Lengths
FGCI	2.7465	4.5308	1.7843
E-B.Indj	3.1924	8.3646	5.1722
E-B.U	3.1533	8.2896	5.1363
C-B.Indj	2.9653	7.4937	4.5284
C-B.U	2.8734	7.4031	4.5297
MOVER	2.7328	3.1510	0.4182

This result indicated that MOVER based method outperform the others. Moreover, it might be useful for determining inpatient care costs.

7. Please also fit your data to other distributions which are in the right skew. A measure, for example, AIC should be provided.

Response:

We fitted the daily rainfall data to other distributions which are in the right skew as follows.

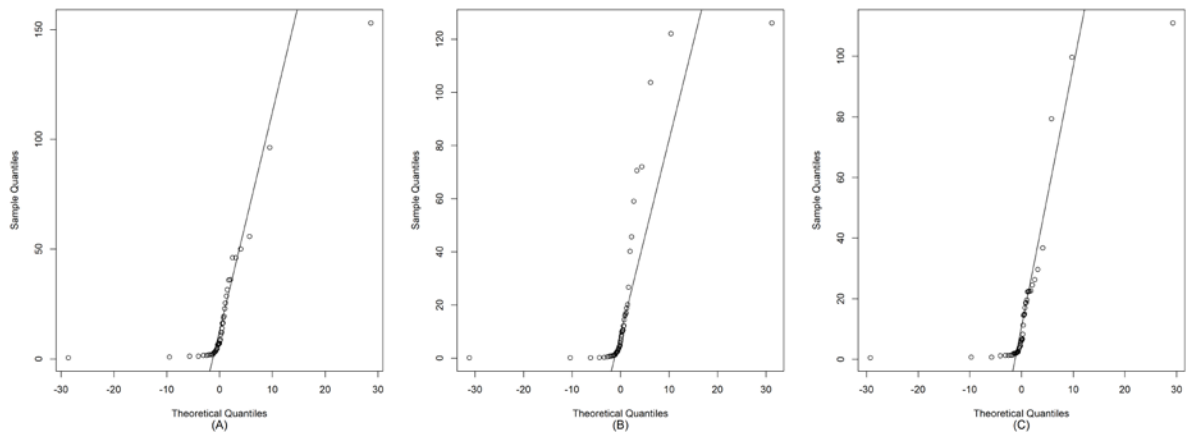


Figure 1. The Cauchy Q-Q plots of the positive daily rainfall data from (A) Chiang Klang (B) Tha Wang Pha, and (C) Pua in Nan, Thailand.

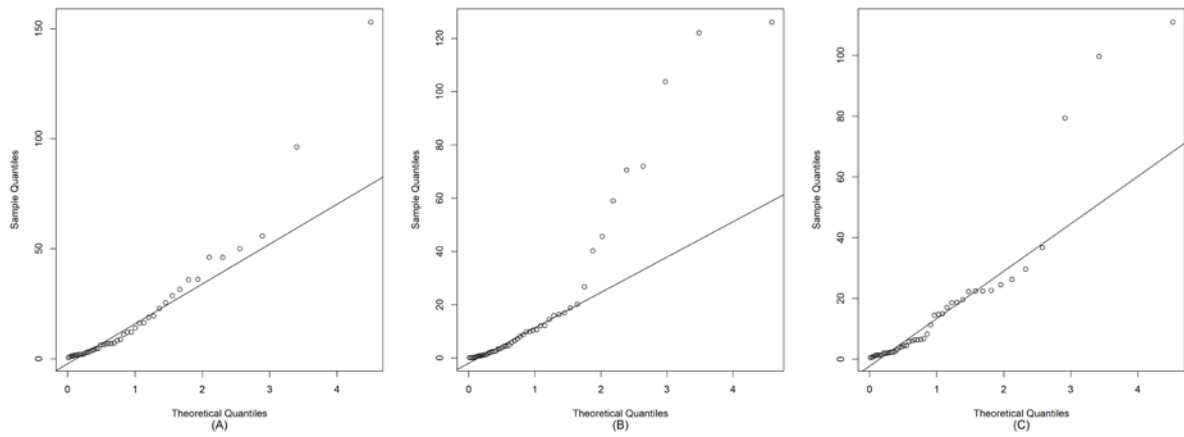


Figure 2. The exponential Q-Q plots of the positive daily rainfall data from (A) Chiang Klang (B) Tha Wang Pha, and (C) Pua in Nan, Thailand.

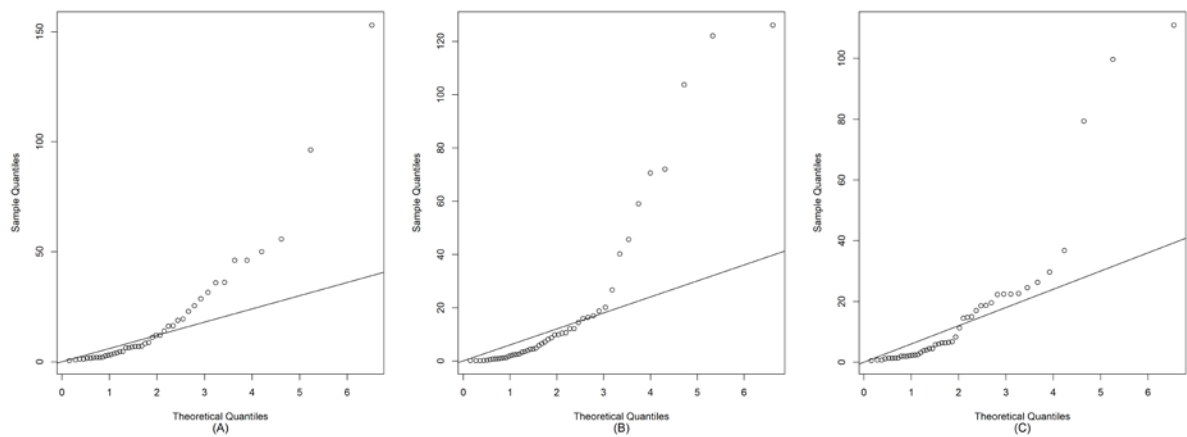


Figure 3. The gamma Q-Q plots of the positive daily rainfall data from (A) Chiang Klang (B) Tha Wang Pha, and (C) Pua in Nan, Thailand.

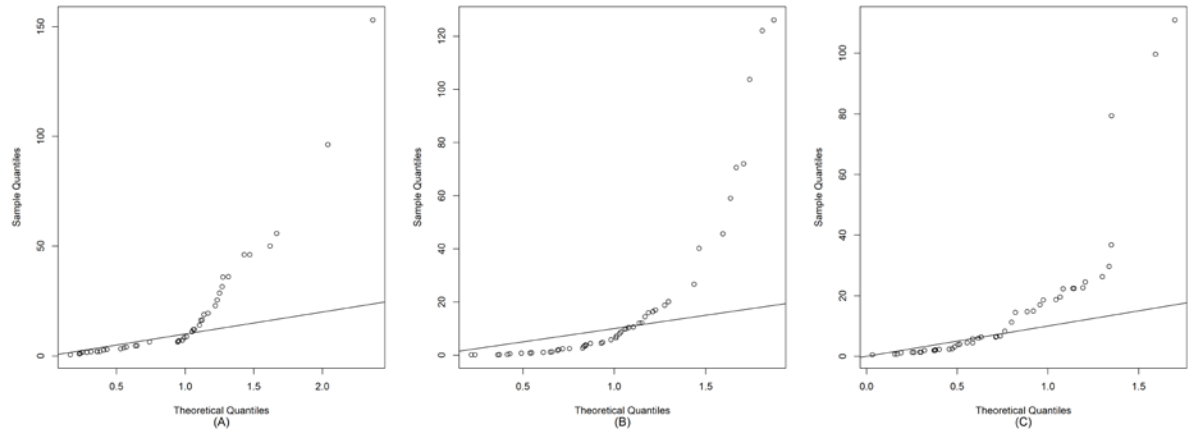


Figure 4. The Weibull Q-Q plots of the positive daily rainfall data from (A) Chiang Klang (B) Tha Wang Pha, and (C) Pua in Nan, Thailand.

The measure, for example, AIC and BIC are

$$AIC = 2k - 2\ln(L)$$

and

$$BIC = k \ln(n) - 2\ln(L),$$

where k is the number of estimated parameters in the model, n is the number of observations, and L is the maximized value of the likelihood function for the model.

Best Regards,

Noppadon Yosboonruang, Sa-Aat Niwitpong, and Suparat Niwitpong

The authors