



Measurement of dispersion of PM 2.5 in Thailand using confidence intervals for the coefficient of variation of an inverse Gaussian distribution

Wasana Chankham, Sa-Aat Niwitpong and Suparat Niwitpong

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

ABSTRACT

Air pollution is a growing concern for the general public in Thailand with PM 2.5 (particulate matter $\leq 2.5 \mu\text{m}$) having the greatest impact on health. The inverse Gaussian (IG) distribution is used for examining the frequency of high concentration events and has often been applied to analyze pollution data, with the coefficient of variation (CV) being used to calculate the quantitative difference in PM 2.5 concentrations. Herein, we propose confidence intervals for the CV of an IG distribution based on the generalized confidence interval (GCI), the adjusted generalized confidence interval (AGCI), the bootstrap percentile confidence interval (BPCI), the fiducial confidence interval (FCI), and the fiducial highest posterior density confidence interval (F-HPDCI). The performance of the proposed confidence intervals was evaluated by using their coverage probabilities and average lengths from various scenarios via Monte Carlo simulations. The simulation results indicate that the coverage probabilities of the AGCI and FCI methods were higher than or close to the nominal level in all of test case scenarios. Moreover, FCI outperformed the others for small sample sizes by achieving the shortest average length. The efficacies of the confidence intervals were demonstrated by using PM 2.5 data from the Din Daeng and Bang Khun Thian districts in Bangkok, Thailand.

Submitted 22 October 2021

Accepted 1 February 2022

Published 17 February 2022

Corresponding author

Sa-Aat Niwitpong,
sa-aat.n@sci.kmutnb.ac.th

Academic editor

Tianfeng Chai

Additional Information and
Declarations can be found on
page 15

DOI 10.7717/peerj.12988

© Copyright
2022 Chankham et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Statistics, Computational Science, Environmental Contamination and Remediation, Environmental Impacts

Keywords Air pollution, Particulate matter, Pollution data, Adjusted generalized confidence interval, Bootstrap percentile confidence interval, Fiducial confidence interval, Fiducial highest posterior

INTRODUCTION

Air pollution is regarded as a serious environmental threat. Inefficient forms of transportation (polluting fuels and cars), coal-fired power plants, and agricultural and garbage burning are all major causes of air pollution, high levels of which have been linked with several negative health effects (*Pope & Dockery, 2006*). Short and long-term exposure to air pollution has been linked to negative health outcomes, and respiratory infections, heart problems, and lung cancer are all increased by it, with people who are already sick being subjected to more severe consequences, while children and the elderly are particularly

vulnerable. One of the most health-damaging pollutants is PM 2.5 (particulate matter ≤ 2.5 μm), which can penetrate deep into the lungs.

The problem of air pollution is a growing concern for the general public in Thailand. Especially, Bangkok's urban environment is a complex mixture of commercial, residential, and industrial buildings. The growing number of vehicles on the road and increased energy consumption constitute a significant source of PM 2.5 emissions, and consequently, PM 2.5 levels (and population exposure) along roadsides are frequently substantially greater than in other areas. Industrial emissions are also of great concern in some areas. [Chuersuwan et al. \(2008\)](#) investigated PM 10 (particulate matter ≤ 10 μm) and PM 2.5 concentrations in the Bangkok Metropolitan Region over the course of a year and uncovered their key sources in the inner city adjacent to a busy road. In the Bangkok Metropolitan Region, [Oanh et al. \(2013\)](#) reported the mass concentrations of PM 2.5 and gaseous pollutants along specific travel routes and at fixed roadside locations. Since the coefficient of variation (CV) was used in these studies to quantitatively characterize the temporal variation of PM 2.5, we are interested in analyzing PM 2.5 data by using the CV of an IG distribution.

The inverse Gaussian (IG) distribution is an excellent choice for modeling positive and right-skewed data, and some of its statistical features were discussed by [Tweedie \(1957\)](#). The IG distribution has been used in hydrology, cardiology, pharmacokinetics, demography, economics, finance, among others. [Schrodinger \(1915\)](#) used it to model the first passage of the time distribution for Brownian motion. [Chhikara & Folks \(1989\)](#) proposed its application to model lifetime and wind energy data. The IG distribution has also been used to investigate insulating fluid failure time, market incidence models, particle cycle time distribution in the blood, health costs, and air pollution. For example, [Kumar & Goel \(2016\)](#) calculated PM 10, PM 2.5, and PM 1 zones of influence under various driving situations and fitted the pdfs of fixed-site data for different PM types at signalized traffic junctions, for which the inverse Gaussian was found to be the most suitable. [Gavril et al. \(2006\)](#) used the pdf in the analysis of the distribution of air pollution in central Athens; they found that the inverse Gaussian distribution provided a better result than the beta, gamma, and Weibull distributions. Therefore, we are interested in studying PM 2.5 pollution by using the IG distribution.

Numerous investigations on the parameters and confidence intervals for IG distributions have previously been conducted. For example, [Arefi, Borzadaran & Vaghei \(2008\)](#) proposed the likelihood ratio interval and the Wald interval for calculating the mean of an IG distribution. [Ye, Ma & Wang \(2010\)](#) employed the generalized confidence interval (GCI) approach for hypothesis testing and interval estimation for the common mean of several IG populations. [Tian & Wilding \(2005\)](#) presented confidence intervals of the ratio of means of two independent IG populations using modified directed likelihood ratio statistics. [Krishnamoorthy & Tian \(2008\)](#) proposed a method for constructing confidence intervals for IG based on the generalized variable approach and a modified likelihood ratio test. In another study, [Ismail & Auda \(2007\)](#) used Bayesian and fiducial inference via the Gibbs sampling process for IG distributions with Type-II censored data. [Jayalath & Chhikara \(2020\)](#) developed a thorough survival analysis for an IG distribution via the

Gibbs sampling method by employing Bayesian and fiducial approaches that require a Monte Carlo Markov Chain (MCMC) method.

The CV, which is the ratio of the standard deviation to the mean (it is useful to measure data dispersion with different units), has been used in a variety of fields, including agriculture, biology, medicine, and finance, with many researchers having developed CV parameters and confidence intervals. [Banik & Kibria \(2011\)](#) presented confidence intervals for estimating the CVs of symmetric and positively skewed distributions. [Wongkhao, Niwitpong & Niwitpong \(2015\)](#) suggested confidence intervals for the CV of a normal distribution. [Sangnawakij & Niwitpong \(2017\)](#) used three approaches: the method of variance of estimates recovery (MOVER), GCI, and the asymptotic confidence interval to establish confidence intervals for the CV of a two-parameter exponential distribution. [Niwitpong \(2013\)](#) created a new confidence interval for the CV of a lognormal distribution with restricted parameters. [Yosboonruang, Niwitpong & Niwitpong \(2019\)](#) provided the confidence intervals for the CV of a lognormal distribution by utilizing Bayesian and fiducial GCI methods. Confidence intervals for the difference between the CVs of Weibull distributions were proposed by [La-ongkaew, Niwitpong & Niwitpong \(2021\)](#). [Chankham, Niwitpong & Niwitpong \(2019\)](#) used GCI and bootstrap percentile confidence interval (BPCI) methods to calculate confidence intervals for the CV of an IG distribution.

The purpose of the current study is to establish new confidence intervals for the CV of an IG distribution by using generalized confidence interval (GCI), adjusted generalized confidence interval (AGCI), bootstrap percentile confidence interval (BPCI), fiducial confidence interval (FCI), and fiducial highest posterior density confidence interval (F-HPDCI) methods. The paper is organized as follows. The ideas of the proposed methods are clarified in Methods section. The simulation results are presented in Results section, and then the proposed methods are applied to the real world datasets, as detailed in An empirical application. The last two sections contain discussion and conclusions on the study.

METHODS

Let $X = (X_1, X_2, \dots, X_n)$ be a random sample from an IG distribution denoted as $IG(\mu, \lambda)$, where μ and λ are the mean and scale parameters of X , respectively. Subsequently, the probability density function of X is given by

$$f(x, \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\lambda(x - \mu)^2}{2\mu^2 x} \right\}, \quad x > 0, \mu > 0, \lambda > 0, \quad (1)$$

The maximum likelihood estimates (MLEs) of parameters μ and λ are $\mu = \bar{X} = \frac{1}{n} \sum_{i=1}^{n_i} X_i$ and $\lambda_i^{-1} = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij}^{-1} - \bar{X}_i^{-1})$, respectively. For notation convenience, we use $V_i = \lambda_i^{-1} \cdot X_i$ and V_i are mutually independent random variables with respective distributions

$$\bar{X}_i \sim IG(\mu, n_i \lambda_i) \quad \text{and} \quad n_i \lambda_i V_i \sim \chi_{n_i-1}^2, \quad (2)$$

where $\chi_{n_i-1}^2$ denotes a chi-square distribution with $n_i - 1$ degrees of freedom. Reproducing the exponential property of the IG explains the independence of these two statistics. It is

simple to show that (\bar{X}_i, V_i) are form a set of complete sufficient statistics. The population mean and variance of X can respectively be expressed as

$$E(X) = \mu, \quad (3)$$

and

$$Var(X) = \frac{\mu^3}{\lambda}. \quad (4)$$

Therefore, the CV of X is

$$CV(X) = \theta = \frac{\sqrt{Var(X)}}{E(X)} = \sqrt{\frac{\mu}{\lambda}}. \quad (5)$$

Here, we present the five methods for constructing confidence intervals for θ .

The GCI method

Since the GCI approach was first introduced by [Weerahandi \(1993\)](#), several researchers have used it to provide statistical inferences ([Tsui & Weerahandi, 1989](#); [Weerahandi, 1993](#); [Weerahandi, 1995](#); [Ye & Wang, 2007](#); [Krishnamoorthy & Tian, 2008](#)). After that, [Ye, Ma & Wang \(2010\)](#) presented the generalized pivot quantity (GPQ) criterion for the parameters and the constructed confidence intervals for the common mean of several inverse Gaussian populations. Furthermore, [Chankham, Niwitpong & Niwitpong \(2019\)](#) recommended GCI for constructing confidence intervals for the coefficient of variation of an IG distribution; they found that GCI provide coverage probabilities greater than or equal to the nominal confidence level at 0.95. Therefore, GCI was selected as the baseline for comparison with the proposed methods of this study.

The confidence interval for the CV is calculated using the concept of the GPQ. Let $X = (X_1, X_2, \dots, X_n)$ be random variables from a distribution defined by probability density function $f_X(x; \theta, \delta)$, with θ and δ being the sought after and nuisance parameters, respectively. Meanwhile, GPQ $R(X; x, \theta, \delta)$ satisfies the following two conditions.

- (i) The probability distribution of function $R(X; x, \theta, \delta)$ is independent of the unknown parameters.
- (ii) The observed value of $R(X; x, \theta, \delta)$, $X = x$ does not depend on the nuisance parameters.

If $R(X; x, \theta, \delta)$ satisfies both conditions, then the GCI for the parameter of interest is calculated using the percentiles of the GPQ. Let $[R_{\alpha/2}, R_{1-\alpha/2}]$ be a $100(1-\alpha)\%$ two-sided GCI for the parameter of interest, where R_α and $R_{1-\alpha}$ are denoted as $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ of $R(X; x, \theta, \delta)$, respectively.

[Ye, Ma & Wang \(2010\)](#) proposed the respective GPQs for λ_i and μ_i as follows:

$$R_{\lambda_i} = \frac{n_i \lambda_i V_i}{n_i v_i} \sim \frac{\chi_{n_i-1}^2}{n_i v_i}, \quad i = 1, 2, \dots, k, \quad (6)$$

where $\chi_{n_i-1}^2$ denotes a chi-squared distribution with $n_i - 1$ degrees of freedom. Thus, the GPQ for μ_i is given by

$$R_{\mu_i} = \frac{\bar{x}_i}{\left| 1 + \frac{\sqrt{n_i \lambda_i (\bar{x}_i - \mu)}}{\mu \sqrt{\bar{x}_i}} \sqrt{\frac{\bar{x}_i}{n_i R_{\lambda_i}}} \right|} \stackrel{d}{\sim} \frac{\bar{x}_i}{\left| 1 + Z_i \sqrt{\frac{\bar{x}_i}{n_i R_{\lambda_i}}} \right|}, \quad (7)$$

where $\tilde{\cdot}^d$ is approximately distributed and $Z_i \sim N(0, 1)$. The approximate distribution is derived from the moment matching method of [Chhikara & Folks \(1989\)](#). Note that $\sqrt{n_i \lambda_i}(\bar{X} - \mu_i)/(\mu_i \sqrt{\bar{X}_i})$ is a limiting distribution of $N(0, 1)$. Consequently, the observed value of R_{μ_i} is μ_i . Therefore, the GCI for the CV of an IG distribution is given by

$$R_\theta = \sqrt{\frac{R_\mu}{R_\lambda}}. \quad (8)$$

The $100(1 - \alpha)\%$ two-sided confidence interval for the CV of an IG distribution based on the GCI method is given by

$$CI_{GCI} = [L_{GCI}, U_{GCI}] = [R_\theta(\alpha/2), R_\theta(1 - \alpha/2)], \quad (9)$$

where $R_\theta(\alpha/2)$ and $R_\theta(1 - \alpha/2)$ are the $100(\alpha/2)$ -th and $100(1 - \alpha/2)$ -th percentiles of the distribution of R_θ , respectively.

The following algorithm was used to construct GCI:

Algorithm 1

- (1) Generate X_1, X_2, \dots, X_n from an IG distribution.
- (2) Compute $\hat{\mu}$ and $\hat{\lambda}$.
- (3) Generate χ_{n-1}^2 from a Chi-square distribution with $n - 1$ degrees of freedom and Z from a standard normal distribution.
- (4) Use [Eqs. \(6\)](#), [\(7\)](#) and [\(8\)](#) to calculate R_λ , R_μ , and R_θ , respectively.
- (5) Repeat Steps 3-4, 5,000 times and obtain an array of R_θ .
- (6) Calculate the 95% confidence intervals for θ by using [Eq. \(9\)](#). If $L \leq \theta \leq U$, then set $cp = 1$; else, set $cp = 0$.
- (7) Repeat Steps 1-6, 15,000 times to compute the coverage probability and the average length.

The AGCI method

According to [Ye, Ma & Wang \(2010\)](#), an approach similar to the GCI method can be utilized for the single coefficient of variation. The GPQ of λ uses the same of GCI method. Subsequently, [Krishnamoorthy & Tian \(2008\)](#) established an estimated GPQ of $\tilde{\mu}_i$ as follows:

$$R_{\tilde{\mu}_i} = \frac{\bar{X}_i}{\max \left\{ 0, 1 + t_{n_i-1} \sqrt{\frac{\bar{X}_i v_i}{n_i - 1}} \right\}}, \quad (10)$$

where t_{n_i-1} denotes a t -distribution with t_{n_i-1} degrees of freedom. However, the denominator can become zero when t_{n_i-1} takes a negative value, and thus $R_{\tilde{\mu}_i}$ is an approximate GPQ.

Therefore, the AGCI for the CV of an IG distribution is given by

$$R_{\tilde{\theta}} = \sqrt{\frac{R_{\tilde{\mu}}}{R_\lambda}}. \quad (11)$$

Subsequently, the $100(1 - \alpha)\%$ two-sided confidence interval for the CV of an IG distribution based on the AGCI method is given by

$$CI_{AGCI} = [L_{AGCI}, U_{AGCI}] = [R_{\tilde{\theta}}(\alpha/2), R_{\tilde{\theta}}(1 - \alpha/2)], \quad (12)$$

where $R_{\hat{\theta}}(\alpha/2)$ and $R_{\hat{\theta}}(1-\alpha/2)$ which are the $100(\alpha/2)$ -th and $100(1-\alpha/2)$ -th percentiles of the distribution of $R_{\hat{\theta}}$, respectively can be obtained from the notion of Algorithm 1.

The BPCI method

When applying this method, the distribution of the bootstrap sample statistic is a direct approximation of the data sample (Efron & Tibshirani, 1986). It is proceeded by re-sampling the data with replacement from the distribution. Chankham, Niwitpong & Niwitpong (2019) reported that BPCI performed more poorly than GCI. However, to provide context, this approach was still included in the comparative analysis. Suppose $X = (X_1, X_2, \dots, X_n)$ is a random sample of size n from an IG distribution. Sampling is replaced by $X^* = (X_1^*, X_2^*, \dots, X_n^*)$, which can be obtained by the bootstrapping the sample B times. Efron & Tibshirani (1986) claimed that a minimum of approximately $B = 1,000$ bootstrap resamples is usually sufficient for obtaining reasonable accurate confidence interval estimates for CV of IG distribution.

The $100(1-\alpha)\%$ two-sided confidence interval for the CV of an IG distribution based on BPCI is defined as

$$CI_{BPCI} = [L_{BPCI}, U_{BPCI}] = [\theta^*(\alpha/2), \theta^*(1-\alpha/2)], \quad (13)$$

where $\theta^*(\alpha/2)$ and $\theta^*(1-\alpha/2)$ are percentiles of the distribution.

The following algorithm is used to construct BPCI:

Algorithm 2

- (1) Generate X_1, X_2, \dots, X_n from an IG distribution.
- (2) Obtain bootstrap sample $X_1^*, X_2^*, \dots, X_n^*$ from Step 1.
- (3) Compute θ^* .
- (4) Repeat Steps 2 and 3, 1,000 times.
- (5) Compute 95% confidence interval based on BPCI according to Eq. (13).

The FCI method

Although fiducial inference proposed by Fisher (1973) is similar to the Bayesian framework, it does not require prior knowledge of the distribution for estimating the parameters involved. Fiducial inference is the only type of inference with frequentist interpretation that uses conditionality on the data. Hence, it allows the implementation of Gibbs sampling (Geman & Geman, 1984), which is an MCMC method commonly used to generate a sample from the posterior distribution for Bayesian inference by sweeping through a variable to sample from its conditional distribution while the remaining variables are fixed at their current values. The sampling distributions of the MLEs of both μ and λ are used for an IG distribution. The fiducial distributions of μ and λ are easily obtained simply by replacing them with their MLEs when they appear in their sample distributions as follows

$$\mu \sim IG(\hat{\mu}, n\hat{\lambda}), \quad (14)$$

and

$$\lambda \sim (\hat{\lambda}/n)\chi_{n-1}^2, \quad (15)$$

where $\hat{\mu}$ and $\hat{\lambda}$ are the MLEs of the μ and λ .

Subsequently, the $100(1 - \alpha)\%$ two-sided confidence interval for the CV of an IG distribution based on fiducial inference is given by

$$CI_{FCI} = [L_{FCI}, U_{FCI}] = [\theta^t(\alpha/2), \theta^t(1 - \alpha/2)], \quad (16)$$

where $\theta^t(\alpha/2)$ and $\theta^t(1 - \alpha/2)$ denotes the $100(\alpha/2)$ -th and $100(1 - \alpha/2)$ -th percentiles of θ^t , respectively.

Algorithm 3 The Gibbs sampling

- (1) Take the initial values (MLEs) of parameters $(\mu^{(0)}, \lambda^{(0)})$
- (2) Generate $\mu^{(t)} \sim IG(\hat{\mu}^{(t-1)}, n\hat{\lambda}^{(t-1)})$
- (3) Generate $\lambda^{(t)} \sim (\hat{\lambda}^{(t-1)}/n)\chi_{n-1}^2$
- (4) Repeat Step 2–3, T times, where T is the number of MCMC replications.
- (5) Burn-in 1,000 samples and compute the parameter of interest.
- (6) Compute the 95% confidence interval based on the fiducial inference [Eq. \(16\)](#).

The F-HPDCI method

Hence, the HPD credible interval of the parameters are obtained by using the MCMC method ([Chen & Shao, 1999](#)). It is assumed that each value inside the interval has a higher posterior density than any of the values outside of it ([Box & Tiao, 2011](#)).

Hence, the $100(1 - \alpha)\%$ two-sided confidence interval for the CV of an IG distribution based on the F-HPDCI is given by

$$CI_{F-HPDCI} = [L_{F-HPDCI}, U_{F-HPDCI}]. \quad (17)$$

The following algorithm is used to construct confidence interval using the F-HPDCI method for the CV of an IG distribution:

Algorithm 4

- (1) Take the initial values (MLEs) of parameters $(\mu^{(0)}, \lambda^{(0)})$
- (2) Use algorithm 3 to calculate the parameter of interest.
- (3) Compute the 95% confidence interval based on F-HPDCI according to [Eq. \(17\)](#).

RESULTS

A Monte Carlo simulation study using the R statistical programming language was conducted to evaluate the performances of the confidence intervals based on GCI, AGCI, BPCI, FCI, and F-HPDCI for the CV of an IG distribution. The sample size was set as $n = 5, 10, 30, 50, 100$, and 200 ; μ as 0.5 and 1 , and λ as $1, 2, 5$, and 10 . We used $15,000$ replications for each parameter combination. Furthermore, $5,000$ repetitions were used for the GCI and AGCI method, $1,000$ bootstrap samples for the BPCI method, and $20,000$ realizations of MCMC using the Gibbs algorithm with a burn-in of $1,000$ for the fiducial methods. Assessing the performances of the confidence intervals of the five methods was achieved in terms of their coverage probabilities and the average lengths respectively calculated as

$$CP = \frac{c(L \leq \theta \leq U)}{M}, \quad (18)$$

and

$$AL = \frac{\sum_{i=1}^M (U - L)}{M}. \quad (19)$$

where $c(L \leq \theta \leq U)$ is the numbers of simulation replications for θ that lie within the confidence interval. The best-performing confidence interval in each case had a coverage probability is greater than or equal to the nominal confidence level of 0.95 and the shortest average length. M is the number of simulation replications. The computational steps to estimate the coverage probabilities and average length performances of all of the methods were computed by using Algorithm 5.

Algorithm 5.

- (1) Set the values M, m, n, μ , and λ .
- (2) Generate X_1, X_2, \dots, X_n from an IG distribution.
- (3) Use Algorithm 1 to construct GCI $[L_{GCI}, U_{GCI}]$ and AGCI $[L_{AGCI}, U_{AGCI}]$.
- (4) Use Algorithm 2 to construct BPCI $[L_{BPCI}, U_{BPCI}]$.
- (5) Use Algorithm 3 to construct FCI $[L_{FCI}, U_{FCI}]$.
- (6) Use Algorithm 4 to construct F-HPDCI $[L_{F-HPDCI}, U_{F-HPDCI}]$.
- (7) If $(L \leq \theta \leq U)$, set $P = 1$; else set $P = 0$.
- (8) Calculate $(U-L)$.
- (9) Repeat Steps 2–8, M times.
- (10) Compute the coverage probabilities and the average lengths.

The results in Table 1, Figs. 1 and 2 show that GCI, AGCI, and FCI provided coverage probabilities greater than or close to the nominal coverage level of 0.95 in almost all cases. However, those of F-HPDCI were less than the nominal confidence level for some cases ($n = 5, \mu = 0.5, 1, \lambda = 1, 2$; $n = 10, \mu = 0.5, 1, \lambda = 1$; and $n = 30, \mu = 1, \lambda = 1$) and those of BPCI in all cases. Meanwhile, the average lengths of all methods became narrower when the sample size was increased, with FCI and F-HPDCI methods providing the shortest ones for $n = 5, 10$, or 30 and AGCI for $n = 50, 100$, or 200.

AN EMPIRICAL APPLICATION

Example 1

PM 2.5 from the Din Daeng district of Bangkok were collected by the Pollution Control Department, Thailand (<http://air4thai.pcd.go.th/webV2/download.php>) due to this area having a high traffic volume (Table 2). Figure 3 exhibits a Q–Q plot of the PM 2.5 data, indicating that an IG distribution is suitable for this dataset. Before computing the confidence intervals, the minimum Akaike information criterion (AIC) and Bayesian information criterion (BIC) were first used to test the best-fitting distribution for these data. These two criteria are respectively defined as

$$AIC = -2\ln L + 2k, \quad (20)$$

and

$$BIC = -2\ln L + 2k\ln(n), \quad (21)$$

Table 1 The coverage probabilities (CP) and average length (AL) of 95% two-sided confidence intervals for the coefficient of variation of inverse Gaussian distribution.

n	μ	λ	CP					AL				
			GCI	AGCI	BPCI	FCI	F-HPDCI	GCI	AGCI	BPCI	FCI	F-HPDCI
5	0.5	1	0.9750	0.9466	0.5351	0.9647	0.9373	2.622	1.514	0.635	1.564	1.322
		2	0.9638	0.9489	0.5342	0.9625	0.9478	1.526	1.071	0.443	1.092	0.921
		5	0.9568	0.9471	0.5447	0.9523	0.9474	0.782	0.676	0.276	0.681	0.573
		10	0.9534	0.9482	0.5506	0.9520	0.9470	0.515	0.481	0.195	0.483	0.406
	1	1	0.9809	0.9483	0.5057	0.9579	0.9235	4.234	2.127	0.913	2.261	1.909
		2	0.9741	0.9478	0.5335	0.9640	0.9382	2.609	1.511	0.634	1.566	1.323
		5	0.9690	0.9536	0.5348	0.9649	0.9524	1.281	0.957	0.392	0.970	0.817
		10	0.9623	0.9547	0.5480	0.9592	0.9521	0.786	0.679	0.277	0.684	0.576
10	0.5	1	0.9781	0.9490	0.7076	0.9591	0.9432	1.118	0.783	0.534	0.841	0.780
		2	0.9696	0.9492	0.7203	0.9592	0.9502	0.660	0.555	0.367	0.577	0.535
		5	0.9598	0.9501	0.7248	0.9529	0.9519	0.377	0.352	0.229	0.357	0.331
		10	0.9568	0.9525	0.7210	0.9538	0.9528	0.256	0.248	0.159	0.250	0.231
	1	1	0.9892	0.9531	0.6986	0.9566	0.9306	2.208	1.105	0.792	1.256	1.166
		2	0.9806	0.9486	0.7103	0.9614	0.9477	1.121	0.784	0.537	0.843	0.783
		5	0.9673	0.9509	0.7235	0.9571	0.9541	0.571	0.497	0.329	0.513	0.476
		10	0.9588	0.9488	0.7221	0.9545	0.9510	0.375	0.350	0.227	0.356	0.330
30	0.5	1	0.9789	0.9478	0.8652	0.9559	0.9480	0.484	0.384	0.361	0.425	0.414
		2	0.9704	0.9498	0.8711	0.9552	0.9545	0.308	0.272	0.244	0.286	0.279
		5	0.9606	0.9520	0.8649	0.9530	0.9506	0.181	0.172	0.150	0.176	0.171
		10	0.9562	0.9502	0.8700	0.9520	0.9500	0.125	0.120	0.105	0.123	0.119
	1	1	0.9934	0.9486	0.8555	0.9500	0.9423	0.832	0.545	0.549	0.652	0.637
		2	0.9812	0.9487	0.8618	0.9551	0.9492	0.484	0.384	0.360	0.424	0.414
		5	0.9652	0.9471	0.8617	0.9503	0.9479	0.125	0.243	0.216	0.253	0.247
		10	0.9591	0.9500	0.8637	0.9503	0.9479	0.181	0.172	0.150	0.175	0.171
50	0.5	1	0.9820	0.9501	0.8964	0.9505	0.9495	0.359	0.289	0.290	0.321	0.316
		2	0.9730	0.9517	0.8948	0.9505	0.9495	0.229	0.204	0.195	0.216	0.213
		5	0.9621	0.9520	0.8965	0.9531	0.9517	0.136	0.129	0.120	0.132	0.130
		10	0.9559	0.9505	0.8922	0.9501	0.9498	0.094	0.091	0.084	0.092	0.091
	1	1	0.9930	0.9530	0.8899	0.9549	0.9480	0.598	0.409	0.443	0.493	0.485
		2	0.9837	0.9552	0.8956	0.9575	0.9521	0.359	0.289	0.289	0.320	0.315
		5	0.9677	0.9514	0.8895	0.9515	0.9470	0.201	0.182	0.172	0.190	0.187
		10	0.9602	0.9504	0.8973	0.9531	0.9529	0.136	0.129	0.120	0.132	0.130
100	0.5	1	0.9856	0.9547	0.9236	0.9521	0.9500	0.247	0.200	0.212	0.223	0.221
		2	0.9718	0.9519	0.9233	0.9516	0.9509	0.159	0.142	0.142	0.149	0.148
		5	0.9634	0.9526	0.9246	0.9523	0.9519	0.094	0.089	0.087	0.092	0.091
		10	0.9557	0.9506	0.9246	0.9523	0.9519	0.065	0.063	0.061	0.064	0.063
	1	1	0.9931	0.9516	0.9129	0.9496	0.9460	0.407	0.283	0.324	0.344	0.341
		2	0.9839	0.9531	0.9209	0.9528	0.9488	0.247	0.200	0.212	0.223	0.221
		5	0.9696	0.9526	0.9222	0.9526	0.9509	0.139	0.127	0.126	0.133	0.131
		10	0.9633	0.9521	0.9225	0.9530	0.9519	0.094	0.084	0.087	0.092	0.091

(continued on next page)

Table 1 (continued)

n	μ	λ	CP					AL				
			GCI	AGCI	BPCI	FCI	F-HPDCI	GCI	AGCI	BPCI	FCI	F-HPDCI
200	0.5	1	0.9830	0.9521	0.9398	0.9551	0.9540	0.172	0.140	0.152	0.156	0.155
		2	0.9736	0.9517	0.9325	0.9526	0.9506	0.111	0.099	0.102	0.105	0.104
		5	0.9618	0.9519	0.9394	0.9535	0.9518	0.066	0.063	0.063	0.064	0.064
		10	0.9557	0.9508	0.9373	0.9503	0.9483	0.045	0.044	0.044	0.045	0.045
	1	1	0.9944	0.9535	0.9344	0.9521	0.9510	0.282	0.198	0.235	0.241	0.240
		2	0.9847	0.9503	0.9346	0.9520	0.9500	0.172	0.140	0.152	0.156	0.156
		5	0.9678	0.9522	0.9365	0.9505	0.9504	0.097	0.089	0.091	0.093	0.092
		10	0.9565	0.9509	0.9348	0.9505	0.9499	0.066	0.063	0.063	0.064	0.064

Notes.
GCI, generalized confidence interval; AGCI, adjusted generalized confidence interval; BPCI, bootstrap percentile confidence interval; FCI, fiducial confidence interval; F-HPDCI, fiducial highest posterior density confidence interval.
Bold denoted as the coverage probability ≥ 0.95 and the shortest average length.

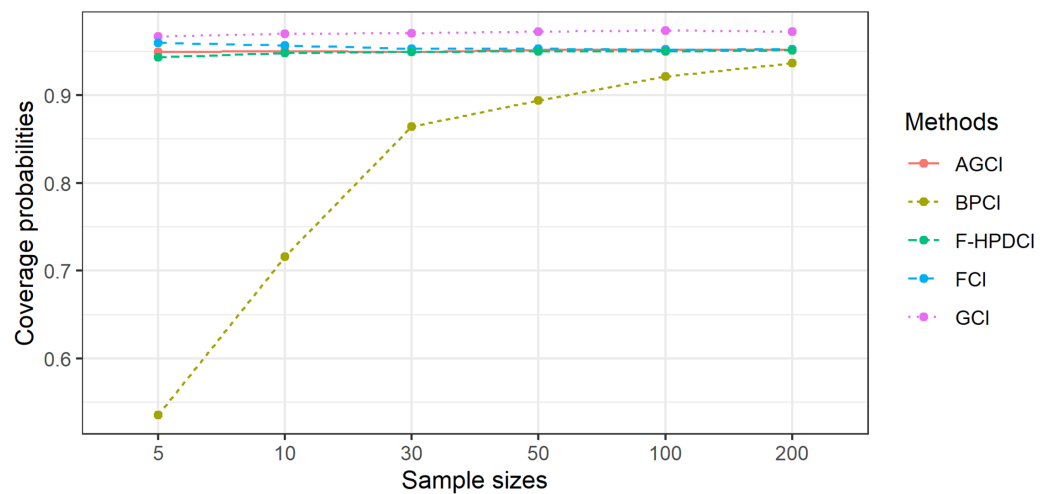


Figure 1 Comparison of the performance of the proposed methods with sample size in terms of coverage probability.

Full-size [DOI: 10.7717/peerj.12988/fig-1](https://doi.org/10.7717/peerj.12988/fig-1)

where L is a likelihood function, k is the number of parameters, and n is the number of recorded measurements.

It was found that the PM 2.5 data fit an IG distribution because the AIC and BIC values for this distribution were smaller than the other tested distributions (normal, lognormal, Cauchy, exponential, and Weibull) (Table 3). The summary statistics were computed: $n = 31$, $\hat{\mu} = 53.1229$, $\hat{\lambda} = 337.9519$, and $CV = 0.3965$. The 95% confidence intervals for the CV of the PM 2.5 dataset by using the five methods are reported in Table 4. These results agree with the simulation results for $n = 30$ in that the F-HPDCI method performed the best in terms of coverage probability and average length, which supports the simulation results.

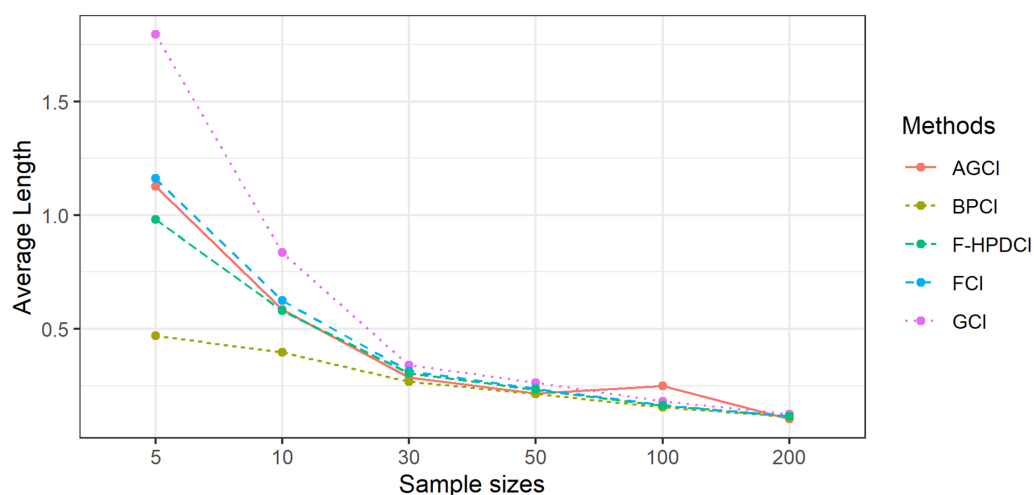


Figure 2 Comparison of the performance of the proposed methods with sample size in terms of average length.

Full-size DOI: [10.7717/peerj.12988/fig-2](https://doi.org/10.7717/peerj.12988/fig-2)

Table 2 Daily PM 2.5 data ($\mu\text{g}/\text{m}^3$) in January 2021 of Din Daeng district, Bangkok, Thailand.

Day	1	2	3	4	5	6	7	8	9	10	11
PM 2.5	27.54	32.58	47.29	54.88	42.25	47.46	37.83	29.88	23.58	32.63	35.75
Day	12	13	14	15	16	17	18	19	20	21	22
PM 2.5	30.00	52.92	77.00	93.46	105.79	60.46	34.63	53.96	77.79	82.08	91.88
Day	23	24	25	26	27	28	29	30	31		
PM 2.5	89.71	47.96	44.83	41.33	49.46	51.00	49.04	51.58	50.26		

Notes.

Source: Air Quality and Noise Management Bureau, Pollution Control Department, Thailand.
<http://air4thai.pcd.go.th/webV2/download.php>.

Example 2

PM 2.5 data from the Bang Khun Thian district, Bangkok, were collected by the pollution Control Department, Thailand (<http://air4thai.pcd.go.th/webV2/download.php>) due to many factors contributing to PM 2.5 pollution in this area (e.g., road construction, car repair shops, and small and large factories) (Table 5). The summary for statistics were $n = 31$, $\hat{\mu} = 56.6161$, $\hat{\lambda} = 225.1443$, and $CV = 0.5015$. A Q–Q plot of these data determining the appropriateness of using an IG distribution is shown in Fig. 4. Furthermore, it was found that the AIC and BIC values for the IG distributions were lower than other tested distributions (Table 6), and thus it provided the best fit for the data. The 95% confidence intervals for the CV of this dataset by using the five methods are reported in Table 7. In agreement with the simulation results for $n = 30$ the F-HPDCI method provided the best confidence interval performance in terms of coverage probability and average length.

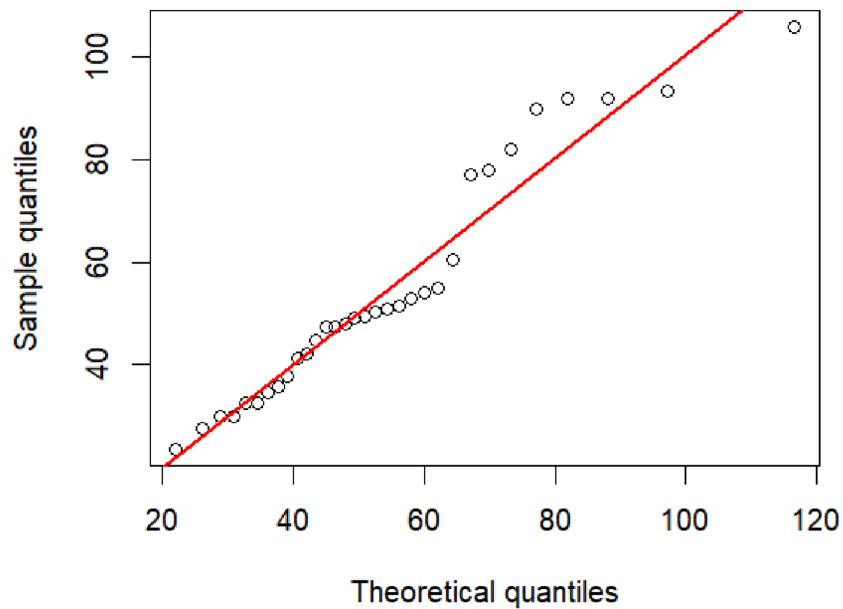


Figure 3 The inverse Gaussian Q–Q plot for the particulate matter data in Din Daeng district, Bangkok Thailand.

Full-size [DOI: 10.7717/peerj.12988/fig-3](https://doi.org/10.7717/peerj.12988/fig-3)

Table 3 AIC and BIC results to check the distribution of PM 2.5 data from Din Daeng district, Bangkok, Thailand.

Densities	Normal	Lognormal	Cauchy	Exponential	Weibull	Inverse Gaussian
AIC	282.9754	275.0629	286.3146	310.5793	280.4643	274.7625
BIC	285.8434	277.9309	289.1826	312.0133	283.3322	277.6305

Notes.

Bold denoted as the minimum AIC and BIC.

Table 4 The 95% confidence intervals for single coefficient of variation of the PM 2.5 data from Din Daeng district, Bangkok, Thailand.

Methods	Confidence intervals for θ		Length of intervals
	Lower	Upper	
GCI	0.3168	0.5470	0.2302
AGCI	0.3231	0.5405	0.2174
BPCI	0.2955	0.4752	0.1797
FCI	0.3180	0.5426	0.2247
F-HPDCI	0.3094	0.5289	0.2195

DISCUSSION

The results show that the AGCI method performed well in all of the scenarios with large sample sizes as its coverage probabilities were consistently greater than or close to the nominal confidence level while its average lengths were the shortest. In addition, FCI and

Table 5 Daily PM 2.5 data ($\mu\text{g}/\text{m}^3$) in January 2021 of Bang Khun Thian district, Bangkok, Thailand.

Day	1	2	3	4	5	6	7	8	9	10	11
PM 2.5	24.00	29.42	44.08	45.64	67.38	84.54	62.17	25.63	19.04	35.29	37.14
Day	12	13	14	15	16	17	18	19	20	21	22
PM 2.5	35.71	62.58	99.38	112.08	73.70	71.67	33.27	58.79	93.58	100.33	130.67
Day	23	24	25	26	27	28	29	30	31		
PM 2.5	76.83	36.38	36.29	37.63	37.00	42.25	39.17	53.50	49.96		

Notes.

Source: Air Quality and Noise Management Bureau, Pollution Control Department, Thailand.

<http://air4thai.pcd.go.th/webV2/download.php>.

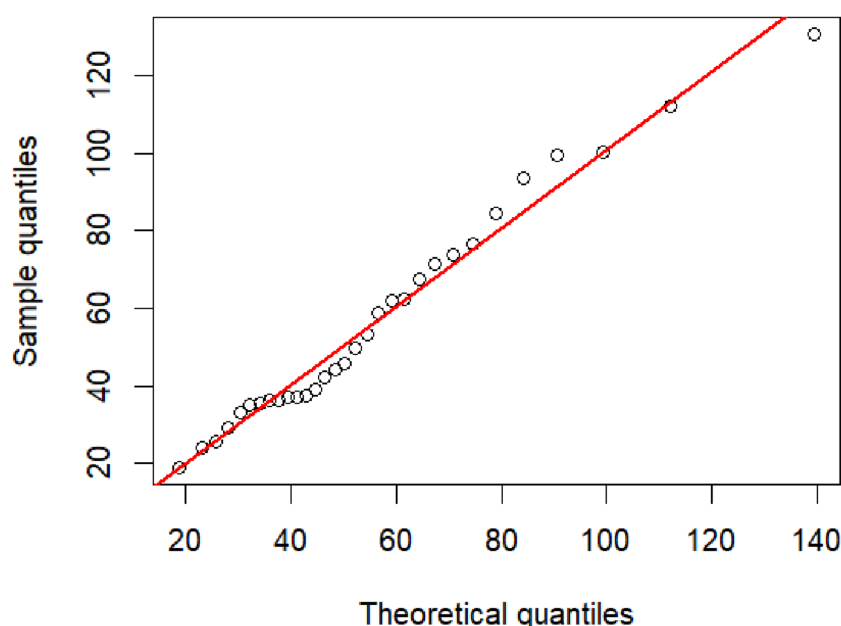


Figure 4 The inverse Gaussian Q-Q plot for the particulate matter data in Bang Khun Thian district, Bangkok Thailand.

Full-size [DOI: 10.7717/peerj.12988/fig-4](https://doi.org/10.7717/peerj.12988/fig-4)

F-HPDCI produced similar results and performed well for small sample sizes. Moreover, when the sample size and λ increased, the average lengths of all of the methods were reduced. The findings of this work are significantly different from previous related studies because we developed a new method for predicting the air pollution level based on fiducial inference derived by using a Gibbs sampler. Moreover, the proposed methods provided the narrowest average lengths, and so can be used to effectively and accurately estimate confidence intervals for various distributions in various fields. Our approach could aid environmentalists and policymakers to monitor air pollution in specific locations and give alarm signals when the air pollution level approaches a dangerous level. Moreover, the proposed method can also be used in air pollution monitoring systems to mitigate the damage caused by climate change and poor air quality. In the same way, the authorities could leverage our research results to control environmental, social, and health impacts

Table 6 AIC and BIC results to check the distribution of PM 2.5 data from Bang Khun Thian district, Bangkok, Thailand.

Densities	Normal	Lognormal	Cauchy	Exponential	Weibull	Inverse Gaussian
AIC	298.1317	289.2471	306.2360	314.2502	293.3355	288.7999
BIC	300.9997	292.1151	309.1040	315.6842	296.2035	291.6679

Notes.

Bold denoted as the minimum AIC and BIC.

Table 7 The 95% confidence intervals for single coefficient of variation of the PM 2.5 data from Bang Khun Thian district, Bangkok, Thailand.

Methods	Confidence intervals for θ		Length of intervals
	Lower	Upper	
GCI	0.3965	0.7022	0.3058
AGCI	0.4068	0.6781	0.2713
BPCI	0.3788	0.6013	0.2225
FCI	0.4001	0.6943	0.2942
F-HPDCI	0.3833	0.6686	0.2853

by promoting laws and regulations to ban vehicles with black smoke emission and or diesel engines, forbid people from burning rubbish, develop integrated urban planning with emission reduction policies, replace delivery trucks with electric vehicles, and collect environmental tax or fees according to the “polluter pays” principle.

CONCLUSIONS

In summary, this research paper aims to propose the construction of confidence intervals for the CV of an IG distribution by using the GCI, AGCI, BPCI, FCI, and F-HPDCI approaches. The performances of these methods were compared using the coverage probability and the average length via simulations studies. The results show that AGCI and FCI performed the best in situations with large ($n = 50, 100$, and 200) and small ($n = 5, 10$ and 30) samples sizes, respectively, and thus, they can be recommended for constructing confidence intervals for the CV of an IG distribution in these two scenarios. Finally, two real pollution datasets were utilized to analyze the performance of the proposed method in real situations. In future work, carbon monoxide, lead, nitrogen dioxide, ozone, sulfur dioxide, and other criteria pollutants should be examined by the CV of IG distribution. Furthermore, new credible intervals based on Bayesian inference for the CV of an IG distribution could be developed.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was funded by King Mongkut's University of Technology North Bangkok, contract number: KMUTNB-PHD-63-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

King Mongkut's University of Technology North Bangkok: KMUTNB-PHD-63-01.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Wasana Chankham conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Sa-Aat Niwitpong conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Suparat Niwitpong performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Data set and R code are available in the [Supplementary Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.12988#supplemental-information>.

REFERENCES

- Arefi M, Borzadaran G, Vaghei Y. 2008.** A note on interval estimation for the mean of inverse Gaussian distribution. *SORT* **32(1)**:49–56.
- Banik S, Kibria BG. 2011.** Estimating the population coefficient of variation by confidence intervals. *Communications in Statistics—Simulation and Computation* **40(8)**:1236–1261 DOI [10.1080/03610918.2011.568151](https://doi.org/10.1080/03610918.2011.568151).
- Box GE, Tiao GC. 2011.** *Bayesian inference in statistical analysis*. Volume 40. New York: John Wiley and Sons.
- Chankham W, Niwitpong S, Niwitpong S. 2019.** Confidence intervals for coefficient of variation of inverse Gaussian distribution. In: *Proceedings of the 3rd international conference on vision, image and signal Processing*. 1–6 DOI [10.1145/3387168.3387254](https://doi.org/10.1145/3387168.3387254).

- Chen M-H, Shao Q-M. 1999.** Monte Carlo estimation of bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics* **8**(1):69–92 DOI [10.1080/10618600.1999.10474802](https://doi.org/10.1080/10618600.1999.10474802).
- Chhikara RS, Folks JL. 1989.** *The inverse Gaussian distribution*. New York: Marcel Dekker.
- Chuersuwan N, Nimrat S, Lekphet S, Kerdumrai T. 2008.** Levels and major sources of PM2.5 and PM10 in Bangkok Metropolitan Region. *Environment International* **34**(5):671–677 DOI [10.1016/j.envint.2007.12.018](https://doi.org/10.1016/j.envint.2007.12.018).
- Efron B, Tibshirani R. 1986.** Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* **1**(1):54–77 DOI [10.1214/ss/1177013815](https://doi.org/10.1214/ss/1177013815).
- Fisher RA. 1973.** *Statistical methods and scientific inference*. New York: Hafner Publishing Co.
- Gavril I, Grivas G, Kassomenos P, Chaloulakou A, Spyrellis N. 2006.** An application of theoretical probability distributions, to the study of pm10 and pm 2.5 time series in Athens, Greece. *EGlobal NEST Journal* **8**(3):241–251.
- Geman S, Geman D. 1984.** Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**:721–741 DOI [10.1109/TPAMI.1984.4767596](https://doi.org/10.1109/TPAMI.1984.4767596).
- Ismail SA, Auda HA. 2007.** Bayesian and fiducial inference for the inverse Gaussian distribution via Gibbs sampler. *Journal of Applied Statistics* **33**(8):787–805 DOI [10.1080/02664760600742268](https://doi.org/10.1080/02664760600742268).
- Jayalath KP, Chhikara RS. 2020.** Survival analysis for the inverse Gaussian distribution with the Gibbs sampler. *Journal of Applied Statistics* Epub ahead of print 2020 3 October DOI [10.1080/02664763.2020.1828314](https://doi.org/10.1080/02664763.2020.1828314).
- Krishnamoorthy K, Tian L. 2008.** Inferences on the difference and ratio of the means of two inverse Gaussian distributions. *Journal of Statistical Planning and Inference* **138**(7):2082–2089 DOI [10.1016/j.jspi.2007.09.005](https://doi.org/10.1016/j.jspi.2007.09.005).
- Kumar P, Goel A. 2016.** Concentration dynamics of coarse and fine particulate matter at and around signalised traffic intersections. *Environmental Science: Processes & Impacts* **18**:1220–1235 DOI [10.1039/c6em00215c](https://doi.org/10.1039/c6em00215c).
- La-ongkaew M, Niwitpong S, Niwitpong S. 2021.** Confidence intervals for the difference between the coefficients of variation of Weibull distributions for analyzing wind speed dispersion. *PeerJ* **9**:e11676 DOI [10.7717/peerj.11676](https://doi.org/10.7717/peerj.11676).
- Niwitpong S-A. 2013.** Confidence intervals for coefficient of variation of lognormal distribution with restricted parameter space. *Applied Mathematical Sciences* **7**(77):3805–3810 DOI [10.12988/ams.2013.35251](https://doi.org/10.12988/ams.2013.35251).
- Oanh NK, Kongpran J, Hang NT, Parkpian P, Hung NTQ, Lee S-B, Bae G-N. 2013.** Characterization of gaseous pollutants and PM2.5 at fixed roadsides and along vehicle traveling routes in Bangkok Metropolitan Region. *Atmospheric Environment* **77**:674–685 DOI [10.1016/j.atmosenv.2013.06.001](https://doi.org/10.1016/j.atmosenv.2013.06.001).

- Pope CA, Dockery DW. 2006.** Health effects of fine particulate air pollution: lines that connect. *Journal of the Air and Waste Management Association* **56**(6):709–742 DOI [10.1080/10473289.2006.10464485](https://doi.org/10.1080/10473289.2006.10464485).
- Sangnawakij P, Niwitpong S-A. 2017.** Confidence intervals for coefficients of variation in two-parameter exponential distributions. *Communications in Statistics—Simulation and Computation* **46**(8):6618–6630 DOI [10.1080/03610918.2016.1208236](https://doi.org/10.1080/03610918.2016.1208236).
- Schrodinger E. 1915.** Zur theorie der fall-und steigversuche an teilchen mit brownscer bewegung. *Physikalische Zeitschrift* **16**:289–295.
- Tian L, Wilding GE. 2005.** Confidence intervals of the ratio of means of two independent inverse Gaussian distributions. *Journal of Statistical Planning and Inference* **133**(2):381–386 DOI [10.1016/j.jspi.2004.01.012](https://doi.org/10.1016/j.jspi.2004.01.012).
- Tsui K-W, Weerahandi S. 1989.** Generalized p -Values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association* **84**(406):602–607.
- Tweedie MCK. 1957.** Statistical properties of inverse Gaussian distributions-I,II. *The Annals of Mathematical Statistics* **28**:362–377 DOI [10.1214/aoms/1177706964](https://doi.org/10.1214/aoms/1177706964).
- Weerahandi S. 1993.** Generalized confidence intervals. *Journal of the American Statistical Association* **88**(423):899–905 DOI [10.1080/01621459.1993.10476355](https://doi.org/10.1080/01621459.1993.10476355).
- Weerahandi S. 1995.** *Exact statistical methods for data analysis*. New York: Springer.
- Wongkhao A, Niwitpong S, Niwitpong S. 2015.** Confidence intervals for the ratio of two independent coefficients of variation of normal distribution. *Far East Journal of Mathematical Sciences* **98**(6):741–757.
- Ye RD, Ma TF, Wang SG. 2010.** Inferences on the common mean of several inverse Gaussian populations. *Computational Statistics & Data Analysis* **54**:906–915 DOI [10.1016/j.csda.2009.09.039](https://doi.org/10.1016/j.csda.2009.09.039).
- Ye RD, Wang SG. 2007.** Generalized p -Values and generalized confidence intervals for variance components in general random effect model with balanced data. *Journal of Systems Science and Complexity* **20**:572–584 DOI [10.1007/s11424-007-9055-9](https://doi.org/10.1007/s11424-007-9055-9).
- Yosboonruang N, Niwitpong S, 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 DOI [10.7717/peerj.7344](https://doi.org/10.7717/peerj.7344).