RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2025.006

ScienceAsia 51 (1): 2025: ID 2025006: **1–13**

On estimation of the population mean in a two-parameter Rayleigh distributed variable with applications to environmental studies

Patarawan Sangnawakij

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani 12120 Thailand

e-mail: patarawan.s@gmail.com

Received 23 Aug 2024, Accepted 17 Dec 2024
Available online 15 Feb 2025

ABSTRACT: This paper considers parameter estimation for the population mean of a two-parameter Rayleigh distribution. We derive a new variance of the mean estimator and provide a confidence interval using the Wald-type method, the large-sample approach, the method of variance estimate recovery, and bootstrap methods. The performance of these interval estimators is conducted through Monte Carlo simulation. According to the studies, the moment estimator has the smallest mean squared error and bias in estimation. The bootstrap-t confidence interval performs very well in all cases in the study. In particular, this method outperforms the compared confidence intervals for small sample sizes as it covers the mean parameter with a given coverage probability. When sample sizes are large, the confidence intervals using maximum likelihood and moment estimators are superior. Three real-world examples of environmental data are used to demonstrate the approaches in applications.

KEYWORDS: bootstrap-t distribution, large-sample approach, mean estimation, particulate matter air pollution, sea surface temperature

MSC2020: 62E17 62F10 62F40

INTRODUCTION

The Rayleigh (RL) distribution is a continuous probability model that has been extensively used for modelling data in areas including reliability, engineering, biological studies, medical science, and the environment. It is generally adequate for data with positively skewed density shapes, such as a Weibull distribution, a chi-square distribution, and an extreme value distribution. In 1880, Lord Rayleigh introduced the RL distribution with a single, unknown shape parameter to address issues in the acoustics and optics fields [1]. Inferential statistics for parameters and functions of parameters in the one-parameter RL distribution were discussed in papers [2–7]. Due to the significance of real-life data, Dey et al [8] established the RL distribution with two parameters, including location (or threshold) and scale parameters. The probability density function of a two-parameter RL variable is given in equation (1). This model is suggested for use in a wider range of applications than the one-parameter RL distribution. Examples are as follows: the study of the time when an electrical appliance is broken after a one-year warranty and the study of wind speed to produce electricity (the wind turbine is likely to begin turning and produce power if the annual average wind speed is at least nine miles per hour). According to these scenarios, the one-parameter RL distribution may not provide a satisfactory result because the lifetimes of the units do not begin at zero.

In theory, we suppose that X is a random variable followed by a two-parameter RL distribution with unknown location μ and scale λ parameters. It is denoted as $X \sim RL(\mu,\lambda)$. The cumulative distribution function of X is given by

$$F(x;\mu,\lambda) = 1-\exp[-\lambda(x-\mu)^2]$$

and the probability density function [8] of X is

$$f(x;\mu,\lambda) = 2\lambda(x-\mu)\exp[-\lambda(x-\mu)^2], \qquad (1)$$

where $x > \mu$, and μ and λ are the real positive values. If $\mu = 0$, model (1) is equivalent to the RL distribution with a shape parameter. Therefore, the two-parameter RL model becomes more flexible and useful when applied to data that is rapidly aging over time. Various shapes of the densities with different values of parameters in the two-parameter RL distributions are displayed in Fig. 1. It can be seen that when λ is increased and μ is fixed, the asymmetry increases. In other words, λ controls the skewness of the distribution. Moreover, when λ is fixed and μ is varied, the location of the density changes, but the distribution has the same skewness. In the literature, several researchers have studied the parameter estimation given in (1). Methods to construct the point estimators for λ and μ are usually based on maximum likelihood (ML) estimation, method of moments (MM), Bayesian method, and

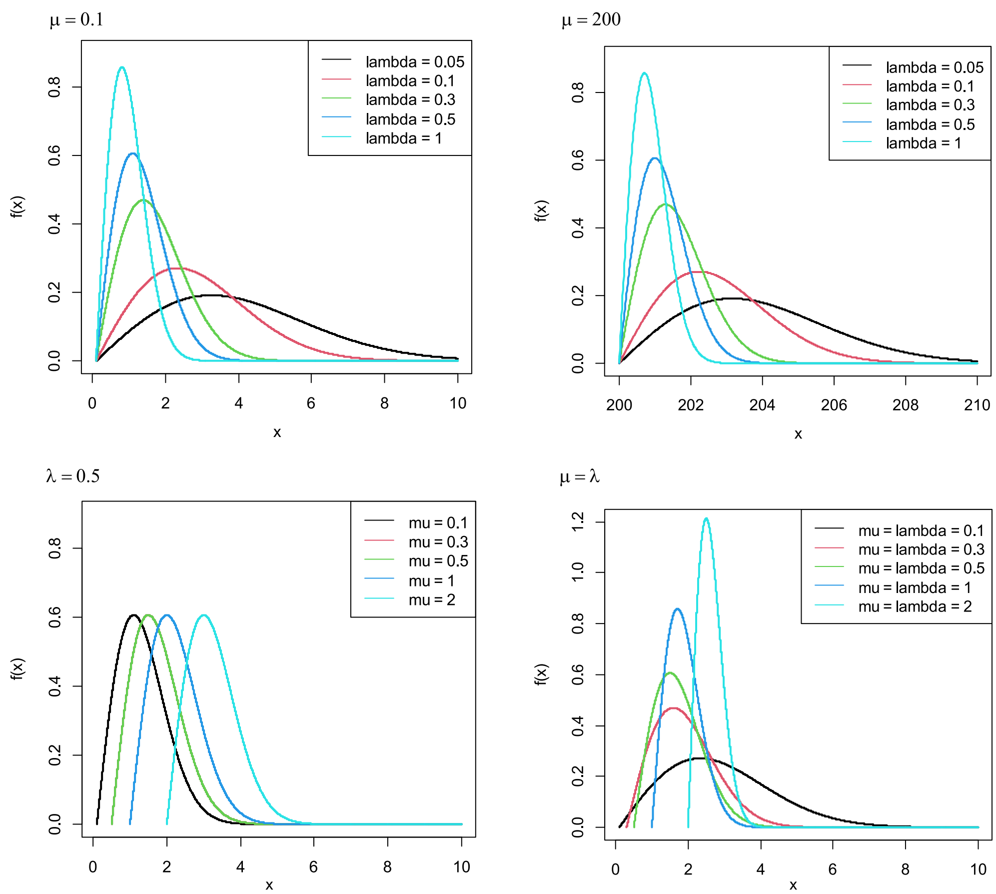

Fig. 1 Density plots of $X \sim RL(\mu, \lambda)$ by various values of scale λ and location μ parameters.

L-moment estimator [8–11]. Moreover, interval estimation is a widely used statistical inference in various applications. It is utilized to estimate the parameter of interest with a guaranteed probability or confidence level [12, 13]. Regarding inference on λ and μ in the two-parameter RL distribution, Asgharzadeh et al [14] introduced the exact confidence limits for parameters under progressive censoring by using the Lagrangian method. Shang and Wenhao [15] used the pivotal quantities from the ML estimators to construct the confidence intervals for λ and μ. Awwad et al [16] proposed the estimators for λ and μ from the Bayesian approach using different loss functions. Krishnamoorthy et al [17] derived confidence intervals for the mean, quantiles, and survival probability using pivotal quantities. However, the probability density function used in the previous two papers is completely different from (1) which is presented in the original paper [8].

The mean, one of the measures of central tendency, plays a crucial role in describing the middle of continuous data or numerical data distribution. It is also widely used in many areas of study and applied research. Unfortunately, Dey et al [8] only focused on point estimation of the location and scale parameters in the continuous two-parameter RL distribution. No research has studied the confidence interval for the population mean of model (1). In the current work, the uncertainty of the mean estimation will be evaluated using the confidence interval. We derive the new interval estimators using parametric and nonparametric approaches. Here are the salient features of this paper:

- It shows the new variance of the mean estimator.
- It presents the new confidence intervals using several methods, including the Wald-type approximation, method of variance estimate recovery, percentile bootstrap, and bootstrap-t.
- It presents a simulation study under several scenarios to investigate the performance of interval estimators.
- It illustrates the concepts with a case study on natural phenomena and environmental pollution.

The remainder of the paper is structured as follows. In the second section, we briefly explain the maximum likelihood and method of moments estimation to estimate the parameters in the two-parameter

RL distribution. Then, the new variance of the mean estimator are derived. In the third section, the confidence intervals are introduced using the large-sample approximation and bootstrap approaches. We investigate the performance of the proposed estimators using simulations in the next section. Then, three real data sets are used to assess the proposed methods. Finally, we give some conclusions and discussions.

POINT ESTIMATION FOR THE POPULATION MEAN

Suppose that a random variable of size n, denoted as $X = (X_1, X_2, \dots, X_n)$, follows a two-parameter RL distribution with the density in (1). The log-likelihood function of μ and λ when $X_i = x_i$, for $i = 1, 2, \dots, n$, is

$$\log L(\mu, \lambda; x_i) = \sum_{i=1}^{n} \log(x_i - \mu) - \lambda \sum_{i=1}^{n} (x_i - \mu)^2 + n\log(\lambda) + n\log(2). \quad (2)$$

The expected value of X, or population mean of the two-parameter RL variable, is defined by

$$E(X) = \theta = \frac{\Gamma(3/2)}{\sqrt{\lambda}} + \mu. \quad (3)$$

Herein, θ is the parameter of interest in this work. The variance of X is $Var(X) = [1 - \Gamma^2(3/2)]/\lambda$, where $\Gamma(a) = \int_0^\infty \exp(-x)x^{a-1}\,dx$ denotes the gamma function and a is a positive value, so $\Gamma(3/2) \approx 0.8862$. Since μ and λ are unknown values, this section focuses on point estimation for these parameters. It is based on the two classical methods: maximum likelihood (ML) estimation and method of moment (MM) estimation.

We first find the ML estimator for μ, which is obtained by maximizing (2) under the assumption that $x > \mu$. The explicit solution for μ is given as

$$\hat{\mu}_{\mathrm{ml}} = \min\{X_1, X_2, \dots, X_n\} = X_{(1)}, \quad (4)$$

or the first-order statistic. The expected value of $\hat{\mu}_{\mathrm{ml}}$ is $E(\hat{\mu}_{\mathrm{ml}}) = c/\sqrt{n\lambda} + \mu$ and the variance of $\hat{\mu}_{\mathrm{ml}}$ is $Var(\hat{\mu}_{\mathrm{ml}}) = (1 - c^2)/n\lambda$, where $c = \Gamma(3/2)$. It is simple to show that $\lim_{n\to\infty} E(\hat{\mu}_{\mathrm{ml}}) = \mu$ and $\lim_{n\to\infty} Var(\hat{\mu}_{\mathrm{ml}}) = 0$. Therefore, $\hat{\mu}_{\mathrm{ml}}$ is the consistent estimator. To find the ML estimator for λ, we take the first-partial derivative of (2) with respect to λ and set it to zero. This yields the normal equation:

$$\frac{n}{\lambda} - \sum_{i=1}^{n} (x_i - \mu)^2 = 0.$$

Solving the above equation for λ, the estimator is established as

$$\hat{\lambda}_{\mathrm{ml}} = \frac{n}{\sum_{i=1}^{n} (X_i - \hat{\mu}_{\mathrm{ml}})^2}. \quad (5)$$

The estimators given in (4) and (5) are independent [8]. Then, we consider the expected value of $1/\hat{\lambda}_{\mathrm{ml}}$, namely, $E(1/\hat{\lambda}_{\mathrm{ml}})$. It is easy to verify that $E(1/\hat{\lambda}_{\mathrm{ml}}) = (1 + 1/n - 2c^2/\sqrt{n})/\lambda$. Based on the asymptotic properties of the ML estimator, the variance of $\hat{\lambda}_{\mathrm{ml}}$ can be approximated using the inverse of Fisher information. The Fisher information is the negative of the expected value of the second-partial derivative of the log-likelihood function. It can be written as $I(\nu) = -E\left(\frac{\partial^2}{\partial \nu^2} \log L(\nu; x_i)\right)$, where ν is a generic parameter. In our case, we can derive that $Var(\hat{\lambda}_{\mathrm{ml}}) = \lambda^2/n$. The proof is shown in the Appendix section. According to the large-sample theory, $\hat{\lambda}_{\mathrm{ml}}$ is assumed to have an approximate normal distribution. We denote it as $\hat{\lambda}_{\mathrm{ml}} \sim N(\lambda, \lambda^2/n)$. When μ and λ in (3) are replaced by $\hat{\mu}_{\mathrm{ml}}$ and $\hat{\lambda}_{\mathrm{ml}}$, respectively, the estimated mean for θ is of the form

$$\hat{\theta}_{\mathrm{ml}} = \frac{\Gamma(3/2)}{\sqrt{\hat{\lambda}_{\mathrm{ml}}}} + \hat{\mu}_{\mathrm{ml}}. \quad (6)$$

The property of this estimator will be discussed again in the next section.

The MM estimator is generally derived by equating the first k population moments and their samples for solving the resulting system of simultaneous equations for the parameter of interest. Based on this process, it can be written in the formula as

$$E(X^k) = \frac{1}{n} \sum_{i=1}^{n} X_i^k, \quad (7)$$

where the expressions on the left and right hand sides of the above equation denote the k-th population moment and sample moment, respectively. Applying to our case, let $X = (X_1, X_2, \dots, X_n)$ be a random sample from a $RL(\mu, \lambda)$. For $k = 1$, it follows that $E(X) = \bar{X}$ or $\theta = \bar{X}$. Therefore, the MM estimator for θ is given by

$$\hat{\theta}_{\mathrm{mm}} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i. \quad (8)$$

Clearly, $\bar{X}$ is an unbiased estimator for θ. Furthermore, by the definition we know that $E(X^2) = Var(X) + [E(X)]^2$. From (7) with $k = 2$, it gives $E(X^2) = \sum_{i=1}^{n} X_i^2/n$. This follows that $Var(X) + [E(X)]^2 = \sum_{i=1}^{n} X_i^2/n$ and

$$\frac{1}{\lambda}\left[1 - \Gamma^2(3/2)\right] + \bar{X}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

Solving the above equation, the MM estimator for λ is established as

$$\hat{\lambda}_{\mathrm{mm}} = \frac{1 - \Gamma^2(3/2)}{\sum_{i=1}^{n} (X_i - \bar{X})^2/n}. \quad (9)$$

Note that moment estimation for θ uses a one-stage approach to derive the estimator for θ. It differs from ML estimation, which requires two steps: estimating λ and μ to get the estimator for θ.

The point estimators described in this section will be used to construct the confidence intervals for θ. They are discussed in the next section. Their performances will also be investigated in terms of bias and mean squared error given in the simulation study.

CONFIDENCE INTERVAL FOR THE POPULATION MEAN

In this section, the new confidence intervals for θ are introduced. They are derived based on the Wald method using the ML estimator, the large-sample approximation using the MM estimator, the method of variance estimate recovery (MOVER), and the bootstrap method. The details are provided in the following subsections.

Wald method using maximum likelihood estimator

The Wald method is a popular statistical tool used to derive the pivot quantity and construct the confidence interval for a parameter. If the mean and variance of the estimator related to the parameter exist, the Wald-type confidence interval is obtained on the basis of a normal approximation.

We first consider some properties of the ML estimator for θ in terms of the mean and variance. The expected value of $\hat{\theta}_{\mathrm{ml}}$ is given by

$$E(\hat{\theta}_{\mathrm{ml}}) = E\left(\frac{\Gamma(3/2)}{\sqrt{\hat{\lambda}_{\mathrm{ml}}}} + \hat{\mu}_{\mathrm{ml}}\right) \approx c\sqrt{\frac{1}{\lambda}\left(1+\frac{1}{n}-\frac{2c^2}{\sqrt{n}}\right)} + \frac{c}{\sqrt{n\lambda}} + \mu, \tag{10}$$

where $c = \Gamma(3/2)$. Here, $E\left(1/\sqrt{\hat{\lambda}_{\mathrm{ml}}}\right) \approx \sqrt{E(1/\hat{\lambda}_{\mathrm{ml}})}$ by the delta method based on the first-order Taylor series expansion [12]. The proof of (10) is shown in the Appendix section. Referring to Dey et al [8], they showed that $\hat{\mu}_{\mathrm{ml}}$ and $\hat{\lambda}_{\mathrm{ml}}$ are independent. This means that the covariance of these two estimators is zero. Therefore, we can find the variance of $\hat{\theta}_{\mathrm{ml}}$, given as

$$Var(\hat{\theta}_{\mathrm{ml}}) = Var\left(\frac{c}{\sqrt{\hat{\lambda}_{\mathrm{ml}}}} + \hat{\mu}_{\mathrm{ml}}\right) \approx \frac{4-3c^2}{4n\lambda}. \tag{11}$$

Again, the delta method is used to approximate the variance of $\sqrt{1/\hat{\lambda}_{\mathrm{ml}}}$. We have

$$Var\left(\sqrt{1/\hat{\lambda}_{\mathrm{ml}}}\right) \approx \left(\frac{\partial}{\partial\lambda}\sqrt{1/\hat{\lambda}_{\mathrm{ml}}}\right)^2 Var\left(\frac{1}{\hat{\lambda}_{\mathrm{ml}}}\right) = \frac{1}{4n\lambda}.$$

The proof is demonstrated in the Appendix section. From (10) and (11), the mean of $\hat{\theta}_{\mathrm{ml}}$ converges to θ and variance goes to zero for large n, i.e., $\lim_{n\to\infty} E(\hat{\theta}_{\mathrm{ml}}) = \theta$ and $\lim_{n\to\infty} Var(\hat{\theta}_{\mathrm{ml}}) = 0$. Hence, $\hat{\theta}_{\mathrm{ml}}$ is a consistent estimator. According to $E(\hat{\theta}_{\mathrm{ml}})$ for $n \to \infty$, the pivotal quantity for θ can be established as

$$Z_1 = \frac{\hat{\theta}_{\mathrm{ml}} - \theta}{\sqrt{\widehat{Var}(\hat{\theta}_{\mathrm{ml}})}}, \tag{12}$$

where $\widehat{Var}(\hat{\theta}_{\mathrm{ml}}) = (4-3c^2)/4n\hat{\lambda}_{\mathrm{ml}}$ is the estimated variance of $\hat{\theta}_{\mathrm{ml}}$. Equation (12) has an approximate standard normal distribution. Using pivot Z_1, we solve the probability statement

$$1-\alpha = P(-z_{1-\alpha/2} \leqslant Z_1 \leqslant z_{1-\alpha/2}) \tag{13}$$

for the lower and upper limits of θ. Hence, a $(1-\alpha)100\%$ Wald-type confidence interval becomes

$$CI_{\mathrm{ml}} = \left(\hat{\theta}_{\mathrm{ml}} - z_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_{\mathrm{ml}})}, \hat{\theta}_{\mathrm{ml}} + z_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_{\mathrm{ml}})}\right), \tag{14}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha)100$-th percentile of a standard normal distribution and $\alpha \in (0,1)$ is a significance level.

Large-sample estimation using moment estimator

In the previous section, the method of moment provides the sample mean as the estimator for θ. The method is simple, but it is typically used as a starting point for constructing the confidence interval in theory and widely applied in applications, particularly when the sample size gets large. From the previous section, we point out that $E(\hat{\theta}_{\mathrm{mm}}) = \theta$ and $\hat{\theta}_{\mathrm{mm}}$ is an unbiased estimator for θ. Its variance is exactly equal to $Var(\hat{\theta}_{\mathrm{mm}}) = (1-c^2)/n\lambda$. Using the central limit theorem and properties of the MM estimator, the pivot function for θ is derived by

$$Z_2 = \frac{\hat{\theta}_{\mathrm{mm}} - \theta}{\sqrt{\widehat{Var}(\hat{\theta}_{\mathrm{mm}})}} = \frac{\bar{X} - \theta}{\sqrt{\widehat{Var}(\hat{\theta}_{\mathrm{mm}})}}.$$

Z_2 has a limiting distribution of a standard normal distribution. Again, the lower and upper limits for θ are solved from

$$1-\alpha = P(-z_{1-\alpha/2} \leqslant Z_2 \leqslant z_{1-\alpha/2}),$$

leading to a $(1-\alpha)100\%$ confidence interval:

$$CI_{\mathrm{mm}} = \left(\hat{\theta}_{\mathrm{mm}} - z_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_{\mathrm{mm}})}, \hat{\theta}_{\mathrm{mm}} + z_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_{\mathrm{mm}})}\right), \tag{15}$$

where $\widehat{Var}(\hat{\theta}_{\mathrm{mm}}) = (1-c^2)/n\hat{\lambda}_{\mathrm{mm}}$ is the estimated variance of $\hat{\theta}_{\mathrm{mm}}$ and $c = \Gamma(3/2)$.

Method of variance estimate recovery

The method of variance of estimates recovery (MOVER), or the closed-form method of variance estimation, was proposed by Zou and their collages [18–20]. This approach is used to construct the confidence interval for a function of parameters expressed in forms such as $\theta_1+\theta_2$, $\theta_1-\theta_2$, and θ_1/θ_2. The procedure is based on a large-sample approximation concept. If the confidence intervals for single parameter, namely θ_1 and θ_2, are available, the confidence interval for the function of these parameters is obtained in the specified closed-form solution. The MOVER is used in several papers for constructing the confidence intervals [21–24].

In this paper, we let $\theta=\theta_1+\theta_2$ be the population mean of a two-parameter RL distributed variable and suppose that $\theta_1=c/\sqrt{\lambda}$ and $\theta_2=\mu$, where $c=\Gamma(3/2)$. If the estimators for θ_1 and θ_2 are obtained from ML estimation, the MOVER confidence interval for θ will be identical to CI_{ml}, as given in the previous section. We apply the normal equation from the method of moment: $E(X)=\bar{X}$ to derive the estimator for the single parameter. It can be rewritten as $c/\sqrt{\lambda}+\mu=\bar{X}$, or $\theta_1=\bar{X}-\mu$. Under the invariant property, $\hat{\mu}_{\text{ml}}$ is applied to estimate μ. The estimators for θ_1 and θ_2 become $\hat{\theta}_1=\bar{X}-\hat{\mu}_{\text{ml}}$ and $\hat{\theta}_2=\hat{\mu}_{\text{ml}}$, respectively. Next, the estimated variances of $\hat{\theta}_1$ and $\hat{\theta}_2$ are considered. Using the information given in the second section, we yield

$$\widehat{Var}(\hat{\theta}_1)=\frac{2(1-c^2)}{n\hat{\lambda}_{\text{ml}}} \quad \text{and} \quad \widehat{Var}(\hat{\theta}_2)=\frac{1-c^2}{n\hat{\lambda}_{\text{ml}}}.$$

The Wald-type confidence intervals for parameters θ_1 and θ_2 can be defined using the following form:

$$(l_i,u_i)=\left(\hat{\theta}_i-z_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_i)},\hat{\theta}_i+z_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_i)}\right),$$

for $i=1,2$. Regarding the closed-form solution introduced in Zou and Donner [19] and the information provided above, a $(1-\alpha)100\%$ MOVER confidence interval for $\theta=\theta_1+\theta_2$ is given by

$$CI_{\text{mov}}=\Big(\bar{X}-\sqrt{(\hat{\theta}_1-l_1)^2+(\hat{\theta}_2-l_2)^2},\ \bar{X}+\sqrt{(\hat{\theta}_1-u_1)^2+(\hat{\theta}_2-u_2)^2}\Big), \quad (16)$$

where the confidence limits for θ_1 are of the form

$$(l_1,u_1)=\left(\hat{\theta}_1-z_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_1)},\hat{\theta}_1+z_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_1)}\right),$$

similarly for (l_2,u_2) which is the confidence limits for θ_2.

Bootstrap approach

Bootstrap [25] is a computationally intensive method that is highly beneficial in assessing the precision of an estimate. It is often used when the sampling error of sample statistics cannot be expressed in a simple formula and to estimate the distribution of a statistic without using normality assumptions or making assumptions about underlying distributions. Applications of bootstrapping methods are described in papers [26–29]. This current work implements two bootstrap methods for constructing the confidence intervals for θ. These are based on the percentile bootstrap and bootstrap-t methods.

Percentile bootstrap method

The percentile bootstrap confidence interval is defined as the interval between the $(\alpha/2)100\%$ and $(1-\alpha/2)100\%$ percentiles of the distribution for the estimates of parameter of interest acquired from resampling with replacement [30]. In our case, $X=(X_1,X_2,\dots,X_n)$ is assumed to be a random sample of size n from a $RL(\mu,\lambda)$ distribution. The interested parameter is θ given in (3). To create the percentile bootstrap confidence interval for θ, the processes are based on Algorithm 1.

Algorithm 1 *Percentile bootstrap interval*

1. Draw a bootstrap sample of size n from X_1, X_2, $\dots, X_n$ using sampling with replacement, denoted as $X_1^*, X_2^*, \dots, X_n^*$, and compute the bootstrap ML estimate, namely, $\hat{\theta}_{\text{ml}}^{*(b)}$.
2. Repeat Step 1 B times to get $\hat{\theta}_{\text{ml}}^{*(b)}$, for $b=1$, 2, $\dots, B$.
3. Estimate the overall mean of $\hat{\theta}_{\text{ml}}^{*(1)}$, $\hat{\theta}_{\text{ml}}^{*(2)}$,…, $\hat{\theta}_{\text{ml}}^{*(B)}$ from
$$\hat{\theta}_{\text{boot}}=\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{\text{ml}}^{*(b)}. \quad (17)$$
4. Compute the $(\alpha/2)$-th and $(1-\alpha/2)$-th quantiles of the bootstrap distribution for $\hat{\theta}_{\text{ml}}^{*(1)}$, $\hat{\theta}_{\text{ml}}^{*(2)}$, $\dots$, $\hat{\theta}_{\text{ml}}^{*(B)}$ to obtain a $(1-\alpha)100\%$ two-sided bootstrap confidence interval for θ, defined by
$$CI_{\text{pct.boot}}=\Big(L_{\hat{\theta}_{\text{ml}}^{*(1)},\hat{\theta}_{\text{ml}}^{*(2)},\dots,\hat{\theta}_{\text{ml}}^{*(B)}}(\alpha/2),\ U_{\hat{\theta}_{\text{ml}}^{*(1)},\hat{\theta}_{\text{ml}}^{*(2)},\dots,\hat{\theta}_{\text{ml}}^{*(B)}}(1-\alpha/2)\Big), \quad (18)$$
where $L(\cdot)$ and $U(\cdot)$ denote the lower and upper bounds of the confidence interval.

Bootstrap-t method

We consider another bootstrap approach that allows estimation of the sampling distribution of almost any statistic using random sampling methods. This method is extended from the percentile bootstrap; however, it relies on the computation of a data-driven t distribution. This is referred to as the bootstrap-t, or studentized bootstrap approach. The main concept of the method is that it surpasses the Student's t-test and is then used to perform the confidence interval. Since the bootstrap-t considers the fluctuation of the standard error, the method is accurate and has a small bias, especially in small sample contexts [31, 32]. From the literature, the bootstrap-t method is also noted to have good performance in estimating a parameter, as is the bias-corrected and accelerated (BCa) bootstrap interval, which can deal with bias in estimation [33, 34]. Therefore, the bootstrap-t confidence interval is the focus of this paper.

Suppose that $\delta(\hat{\theta}_{\text{ml}})$ is the standard error of $\hat{\theta}_{\text{ml}}$, which is unknown. $\hat{\theta}_{\text{ml}}$ is the estimated parameter for θ and $\hat{\delta}(\hat{\theta}_{\text{ml}})$ is the estimated standard error for $\delta(\hat{\theta}_{\text{ml}})$. On the basis of a random sample $X=(X_1, X_2, \dots, X_n)$, we define a statistic as $T=(\hat{\theta}_{\text{ml}}-\theta)/\hat{\delta}(\hat{\theta}_{\text{ml}})$, which follows a t-distribution with $n-1$ degrees of freedom, or $T\sim t_{df=n-1}$. The next step involves applying the classical t-statistic and bootstrap method to construct the bootstrap-t interval.

Under bootstrapping, we can find an estimate of $\delta(\hat{\theta}_{\text{ml}})$, denoted as $\hat{\delta}^*(\hat{\theta}^*_{\text{ml}})$, by creating B bootstrap samples. Algorithm 2 provides more details for bootstrap sampling and computation. Hence, the estimate of $\delta(\hat{\theta}_{\text{ml}})$ can be computed by

$$\hat{\delta}^*(\hat{\theta}^*_{\text{ml}}) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}^{*(b)}_{\text{ml}} - \hat{\theta}_{\text{boot}}\right)^2},$$

where $\hat{\theta}^{*(b)}_{\text{ml}}$ is the estimate for θ in the b-th bootstrap sample and $\hat{\theta}_{\text{boot}}$ is the overall estimate for θ. It can be seen that we have only one value of T from this process. So, we need to rely on $T\sim t_{df=n-1}$. This paper suggests using the percentiles of T statistics, derived from B bootstrap samples, to create the bootstrap distribution of T:

$$T^{*(b)} = \frac{\hat{\theta}^{*(b)}_{\text{ml}} - \hat{\theta}_{\text{boot}}}{\hat{\delta}^{*(b)}(\hat{\theta}^*_{\text{ml}})},$$

for $b=1,2,\dots,B$. A large number B of independent replications is required to give the estimated percentiles of the t-distribution. We introduce the procedure of the bootstrap-t confidence interval for θ in Algorithm 2.

Algorithm 2 *Bootstrap-t interval*

1. Sample a bootstrap sample of size n from X_1, X_2, $\dots$, X_n using sampling with replacement, denoted as X^*_1,X^*_2, $\dots$, X^*_n, and compute the ML estimate $\hat{\theta}^*_{\text{ml}}$.

2. Create another R bootstrap samples from X^*_1, X^*_2, $\dots$, X^*_n, denoted as
$$\begin{array}{c} X^{**}_{11}, X^{**}_{21}, \dots, X^{**}_{n1} \\ X^{**}_{12}, X^{**}_{22}, \dots, X^{**}_{n2} \\ \vdots \\ X^{**}_{1R}, X^{**}_{2R}, \dots, X^{**}_{nR} \end{array}$$
and compute the ML estimate of θ from each sample, namely $\hat{\theta}^{**(r)}$, for $r=1,2,\dots,R$.

3. Calculate the mean of $\hat{\theta}^{**(1)}, \hat{\theta}^{**(2)}, \dots, \hat{\theta}^{**(R)}$ from $\hat{\theta}^{**}=\sum_{r=1}^{R}\hat{\theta}^{**(r)}/R$, and compute the empirical standard error
$$\hat{\delta}^{**} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}^{**(r)} - \hat{\theta}^{**}\right)^2}.$$

4. Create a bootstrap statistic
$$T^* = \frac{\hat{\theta}^*_{\text{ml}} - \hat{\theta}^{**}}{\hat{\delta}^{**}}. \tag{19}$$

5. Repeat Steps 1 to 4 B times to get the estimated percentiles of the t-distribution: $T^{*(1)}$, $T^{*(2)}$, $\dots$, $T^{*(B)}$.

6. Find the $(\alpha/2)$-th and $(1-\alpha/2)$-th quantiles of $T^{*(1)}$, $T^{*(2)}$, $\dots$, $T^{*(B)}$, denoted as $t^*(\alpha/2)$ and $t^*(1-\alpha/2)$, respectively.

7. Compute a $(1-\alpha)100\%$ bootstrap-t confidence interval for θ from
$$\begin{aligned} CI_{\text{boot.t}} = \Big(&\hat{\theta}_{\text{ml}} + t^*(\alpha/2)\sqrt{\widehat{Var}(\hat{\theta}_{\text{ml}})}, \\ &\hat{\theta}_{\text{ml}} + t^*(1-\alpha/2)\sqrt{\widehat{Var}(\hat{\theta}_{\text{ml}})}\Big), \end{aligned} \tag{20}$$
where the estimated variance $\widehat{Var}(\hat{\theta}_{\text{ml}}) = (4-3c^2)/4n\hat{\lambda}_{\text{ml}}$ and $c=\Gamma(3/2)$.

SIMULATION STUDY

The performance of the proposed methods is conducted using simulations. All computations are generated under the R programming [35]. The random samples used in this study are derived based on the inverse transform method under the following process:

- Let U be a continuous uniform variable on the interval (0,1).

- We define the two-parameter RL variable as $X = F^{-1}(U)$, where F^{-1} is the inverse function of the cumulative distribution function F.
- Since X is distributed as F, we have $U = F(x) = 1-\exp[-\lambda(x-\mu)^2]$.
- Solving the previous equation, the generated data are computed by

$$x = \mu + \sqrt{-\frac{1}{\lambda}\log(1-u)},$$

where u is sampled value from a uniform distribution on interval (0,1), and λ and μ are the parameters of probability model (1).

In simulation, we set the values of population mean $\theta = 5$, 20, and 100, reflecting small to large effects of population mean, and parameter $\lambda = 0.05$, 0.5, and 2, reflecting various different shapes of the data as shown in Fig. 1. The location parameter is then computed by $\mu = \theta - \Gamma(3/2)/\sqrt{\lambda}$. The sample sizes are considered to $n = 5$, 10, 20, 30, 50, and 100. From each scenario, the simulation is repeated $H = 1{,}000$ times. On average, the bias and mean squared error (MSE) of the estimator are approximated by

$$\text{Bias}(\hat{\theta}) = \frac{1}{H}\sum_{h=1}^{H}\hat{\theta}_h - \theta \quad \text{and} \quad \text{MSE}(\hat{\theta}) = \frac{1}{H}\sum_{h=1}^{H}(\hat{\theta}_h - \theta)^2,$$

respectively, where $\hat{\theta}_h$ is the estimate of θ in the h-th replication. The performance of the confidence interval for θ is evaluated in terms of coverage probability (CP) and expected length (EL). These are estimated by

$$\text{CP} = \frac{\#(L_h \leqslant \theta \leqslant U_h)}{H} \quad \text{and} \quad \text{EL} = \frac{1}{H}\sum_{h=1}^{H}(U_h - L_h),$$

where $\#(L_h \leqslant \theta \leqslant U_h)$ is the number that θ lies within the lower and upper limits. Generally, an estimator with a bias close to zero and a small standard error is preferred. For interval estimation, we wish the confidence interval to cover the true parameter with reasonable coverage probability, close to the given coverage level ($1-\alpha = 0.95$), and a short expected length. Below are the major findings from this simulation.

We start by comparing the performance of $\hat{\theta}_{\text{ml}}$, $\hat{\theta}_{\text{mm}}$, and $\hat{\theta}_{\text{boot}}$ in terms of bias and MSE. These are shown in Fig. 2. The bootstrap estimate is obtained using 500 bootstrap samples under ML estimation. Obviously, $\hat{\theta}_{\text{mm}}$ has a bias and MSE closer to zero than $\hat{\theta}_{\text{ml}}$ and $\hat{\theta}_{\text{boot}}$ in all cases in the study. A decrease or increase in bias of $\hat{\theta}_{\text{mm}}$ cannot be observed for all sample sizes, showing accuracy and unbiasness in estimation. Consequently, $\hat{\theta}_{\text{ml}}$ and $\hat{\theta}_{\text{boot}}$ provide similar performance in terms of bias and MSE. They are overestimates, as their biases significantly deviate from zero due to the small sample sizes. The biases of $\hat{\theta}_{\text{ml}}$ and $\hat{\theta}_{\text{boot}}$ tend to decrease for large n. Furthermore, when λ is increased but θ and n are fixed, the biases of these two estimators decrease. Therefore, the advantage of MM estimation is clearly visible from the results. We conclude that $\hat{\theta}_{\text{mm}}$ performs the best to estimate the population mean in the two-parameter RL distribution.

The performance of the five confidence intervals (CIs) is discussed and presented in Fig. 3. In all cases, CI_{mm} and $CI_{\text{pct.boot}}$ have the coverage probabilities lower than the nominal coverage level of 0.95. This indicates that they perform poorly when estimating the population mean in our settings. CI_{mov} has coverage probabilities greater than the other estimators and much larger than the target level of 0.95. For CI_{ml}, its coverage probability is satisfied when $n \geqslant 20$. The behaviour of CI_{mm}, CI_{mov}, and $CI_{\text{pct.boot}}$ depends on n. If n is increased, the coverage probabilities of these confidence intervals increase and become stable when $n \geqslant 20$. Meanwhile, $CI_{\text{boot.t}}$ works well and has the coverage probabilities greater than and close to the target level for all parameter settings in the study. It is also performed well for all sample size determination. In general, the expected lengths of the five confidence intervals decrease, when n gets large. For $n \leqslant 10$, CI_{mov} and $CI_{\text{boot.t}}$ have interval lengths wider than the comparators; however, only $CI_{\text{boot.t}}$ outperforms in terms of coverage probability as described above. Overall, the bootstrap-t confidence interval is a more efficient method in almost all situations. The MOVER confidence interval can control the coverage level for a small sample size and could be applied for $n = 5$.

REAL DATA ILLUSTRATION

In this section, we use the three real data sets on natural phenomena and environmental pollution to illustrate the purpose of the study. Importantly, sea surface temperature is one of the key climate change indicators used to describe conditions at the boundary between the atmosphere and the oceans. If the oceans absorb more heat from the atmosphere, sea-surface temperatures are expected to increase. This is known as an El Niño event. If it is colder than usual, we face a La Niña event. The first data set used here is the average sea surface temperature over the extrapolar global ocean for July 1979 to 2023 (45 observations). Similarly, the second data set is the average sea surface temperature across the North Atlantic. These data are obtained from the Copernicus Climate Change Service [36], originally reported by the fifth-generation ECMWF reanalysis (ERA5). The sea surface temperature in degrees Celsius (°C) are shown by the histogram, given in Fig. 4(a,b). The third example is the air pollution from particulate matter with an aerodynamic diameter less than 2.5 µm

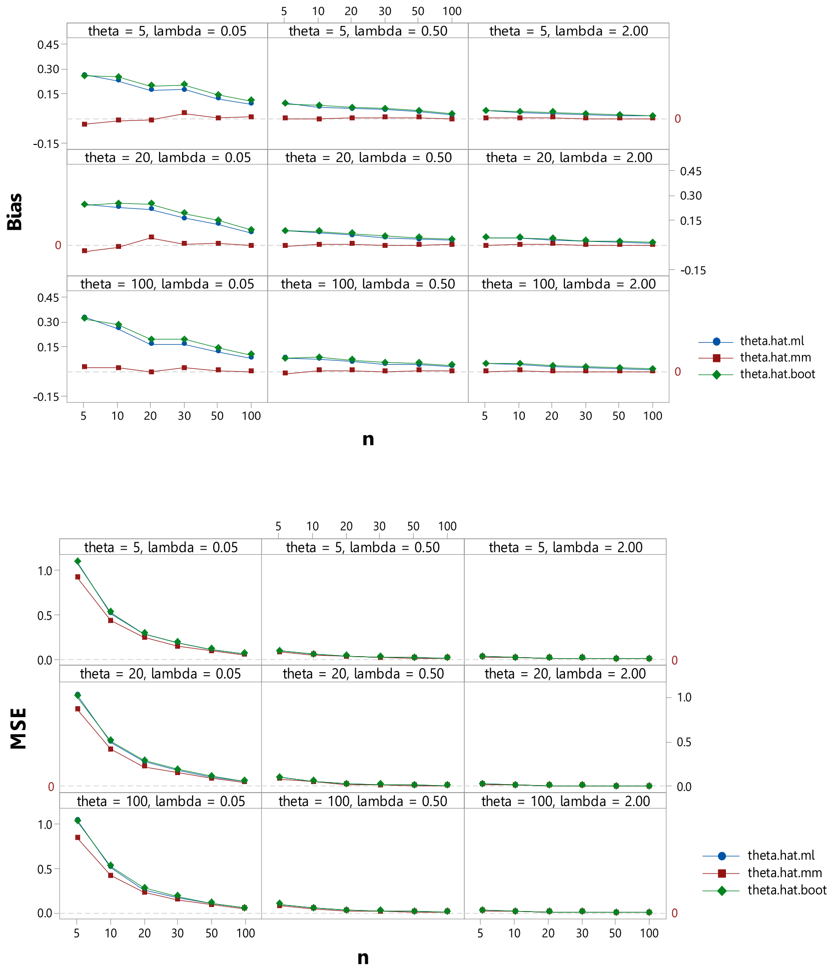

Fig. 2 Simulation results for bias and MSE of the three estimators for population mean θ: $\hat{\theta}_{\mathrm{ml}}$, $\hat{\theta}_{\mathrm{mm}}$, and $\hat{\theta}_{\mathrm{boot}}$.

(PM2.5). PM2.5 is the most dangerous pollutant encountered in many countries. It can penetrate the lung barrier and enter the blood system, causing cardiovascular and respiratory disease, as well as cancer. For calculation, we use a PM2.5 data set from the capital city of Thailand. They are obtained from the Environmental Department, Bangkok Metropolitan Administration [37]. The data on PM2.5 reported in $\mu g/m^3$ from 84 stations collected in the Bangkok area between 9.00 and 10.00 am on January 22, 2024, are shown in Fig. 4(c).

According to the histograms in Fig. 4, the distribution of each data set is skewed to the right. To find a suitable probability model for the data, distribution fitting is considered. We compare the two-parameter RL distribution to the one-parameter RL, one- and two-parameter exponential, Gamma, and Weibull distributions. Model selection is based on the Akaikes information criterion (AIC), Bayesian information criterion (BIC), and log-likelihood (Log-L). The model with the minimum AIC (or BIC) and maximum Log-L will be selected as the best model for these data. Table 1 presents the ML estimates of the parameters together with the log-likelihood, AIC, and BIC values for the six probability models. Note that the notations *1-Rayleigh* and *2-Rayleigh* mean the one-parameter and two-parameter Rayleigh distributions, respectively. From the results under these criteria, we conclude that the two-parameter RL distribution is better suited for fitting the sea surface temperatures and Bangkok PM2.5 data sets than the compared distributions. We also use the chi-square goodness of fit (GoF) statistic to test the following hypotheses:

- H_0: the data follow a two-parameter Rayleigh distribution
- H_1: the data do not follow the two-parameter Rayleigh distribution.

The p-values corresponding to the GoF test statistics are given in Table 1. The sea surface temperatures from the two oceans and Bangkok PM2.5 data have two-parameter Rayleigh distributions at the 0.01

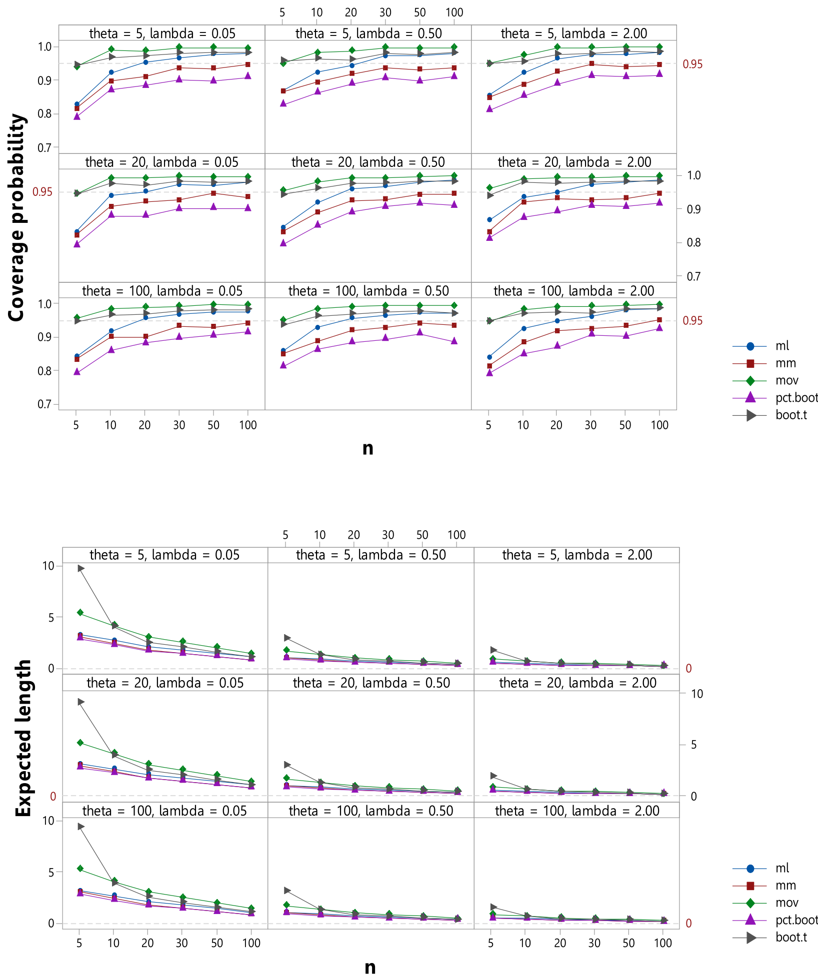

Fig. 3 Simulation results for coverage probability and expected length of the 95% confidence intervals for population mean θ: CI_{ml}, CI_{mm}, CI_{mov}, $CI_{\text{pct.boot}}$, and $CI_{\text{boot.t}}$.

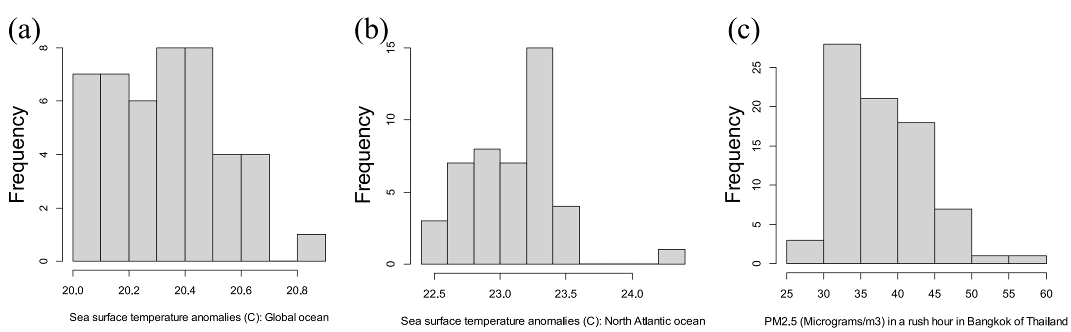

Fig. 4 Histograms of sea surface temperatures from (a) Global ocean, (b) North Atlantic ocean, and (c) PM2.5 of Bangkok.

significance level. Therefore, the proposed methods introduced in this paper are reasonable to use for estimating the population mean of these data.

Table 2 shows the estimated values of θ in the two-parameter RL distribution. We use the ML, MM, MOVER, percentile bootstrap, and bootstrap-t approaches. The results from the real-data examples match the results obtained from the simulation study, as the percentile bootstrap confidence interval has the shortest interval length. The methods that rank from the smallest interval length to the largest length are the percentile bootstrap, the large-sample method based

Table 1 Maximum likelihood estimation for parameters and performance criteria under six models using the real datasets.

Data	Probability model	Estimated parameter	Log-likelihood estimate	AIC value	BIC value
Sea surface temperature from Global ocean (°C)†	1-Exponential	Rate = 0.05	−176.54	355.10	356.88
	1-Rayleigh	Rate = 14.38	−146.05	294.11	295.90
	2-Gamma	Shape = 10547.92 Rate = 518.65	8.82	−13.64	−10.07
	2-Weibull	Shape = 96.62 Scale = 20.44	3.23	−2.46	1.10
	2-Exponential	Location = 20.02 Scale = 0.32	6.60	−9.20	−5.63
	2-Rayleigh	Location = 20.01 Scale = 6.84	9.95	−15.89	−12.32
Sea surface temperature from North Atlantic (°C)‡	1-Exponential	Rate = 0.04	−182.17	366.35	368.13
	1-Rayleigh	Rate = 16.34	−151.69	305.38	307.16
	2-Gamma	Shape = 4960.48 Rate = 214.64	−13.40	30.80	34.37
	2-Weibull	Shape = 60.99 Scale = 23.28	−21.04	46.08	49.65
	2-Exponential	Location = 22.56 Scale = 0.55	−17.46	38.92	42.49
	2-Rayleigh	Location = 22.51 Scale = 2.19	−12.22	28.44	32.01
Bangkok PM2.5 concentration (µg/m^3)§	1-Exponential	Rate = 0.03	−362.02	726.04	728.39
	1-Rayleigh	Rate = 27.27	−310.51	623.02	625.38
	2-Gamma	Shape = 46.50 Rate = 1.22	−244.38	492.77	497.48
	2-Weibull	Shape = 6.68 Scale = 40.65	−251.87	507.74	512.45
	2-Exponential	Location = 28.1 Scale = 10.04	−257.91	519.82	524.54
	2-Rayleigh	Location = 27 Scale = 0.01	−242.06	488.13	492.84

The chi-square goodness of fit test statistics for the two-parameter Rayleigh distribution, using the three sets of data, give p-values of 0.0427†, 0.0214‡, and 0.4788§.

on the MM estimator, the Wald method using the ML estimator, bootstrap-t, and the MOVER method. According to the performance of the estimator from simulations, the MM estimator ($\hat{\theta}_{\mathrm{mm}}$) and bootstrap-t confidence interval ($CI_{\mathrm{boot.t}}$) are recommended. Therefore, we conclude that the average sea surface temperatures were 20.33 °C (95% CI: 20.27–20.41) over the global extrapolar ocean and 23.10 °C (95% CI: 22.95–23.23) over the North Atlantic ocean. The ERA5 reported that the sea surface temperature climates for the global ocean and North Atlantic, relative to the 1991–2020 reference period and used as a standard diagnostic for climate monitoring, were 20.38 °C and 23.16 °C, respectively. Hence, the sea surface temperatures in each ocean had an estimated upper limit greater than the normal value. High sea surface temperatures were observed. This could be the development of an El Niño event. For the air pollution data, we concluded that Bangkok PM2.5 in the study period was 38 µg/m^3 (95% CI: 36.41–39.97). For the standard value of Thailand in the atmosphere by 24-h mean, PM2.5 was 37.5 µg/m^3 [38], while the World Health Organization stated that 24-h average exposures should not exceed 15 µg/m^3 [39].

CONCLUDING REMARKS

Parameter estimation for the mean plays a crucial role in continuous data. Maximum likelihood estimation and method of moment estimation are two basic methods often used to estimate the parameter of the probability model. The bootstrap method is an extremely useful alternative to the traditional method. It is usually used in situations where the distribution of the statistic that we need to measure is limited. In this paper, we investigate that the MM estimator ($\hat{\theta}_{\mathrm{mm}}$) for the population mean of the two-parameter Rayleigh distribution has a simple solution and performs well in estimating the parameter. By simulation, its performance in terms of bias and mean squared error outperforms the ML ($\hat{\theta}_{\mathrm{ml}}$) and bootstrap ($\hat{\theta}_{\mathrm{boot}}$) estimates. A crucial reason that $\hat{\theta}_{\mathrm{ml}}$ has a low efficiency in estimation is that the method needs two stages to estimate parameters μ and λ for performing $\hat{\theta}_{\mathrm{ml}}$. We

Table 2 Estimated mean for θ in the two-parameter RL distribution using the data examples.

Data	Method	Estimated value	95% CI	Length of interval
Sea surface temperature from Global ocean (°C)	ML estimation	20.34	(20.27, 20.42)	0.15
	MM estimation	20.33	(20.27, 20.39)	0.12
	MOVER	–	(20.23, 20.43)	0.20
	Percentile bootstrap	20.34	(20.30, 20.40)	0.10
	Bootstrap-t	–	(20.27, 20.41)	0.14
Sea surface temperature from North Atlantic (°C)	ML estimation	23.11	(22.98, 23.24)	0.26
	MM estimation	23.10	(23.00, 23.20)	0.20
	MOVER	–	(22.93, 23.27)	0.66
	Percentile bootstrap	23.11	(23.02, 23.21)	0.19
	Bootstrap-t	–	(22.95, 23.23)	0.28
Bangkok PM2.5 concentration (µg/m^3)	ML estimation	38.01	(36.25, 39.77)	3.52
	MM estimation	38.00	(36.72, 39.28)	2.56
	MOVER	–	(35.79, 40.21)	4.42
	Percentile bootstrap	38.17	(36.93, 39.50)	2.57
	Bootstrap-t	–	(36.41, 39.97)	3.56

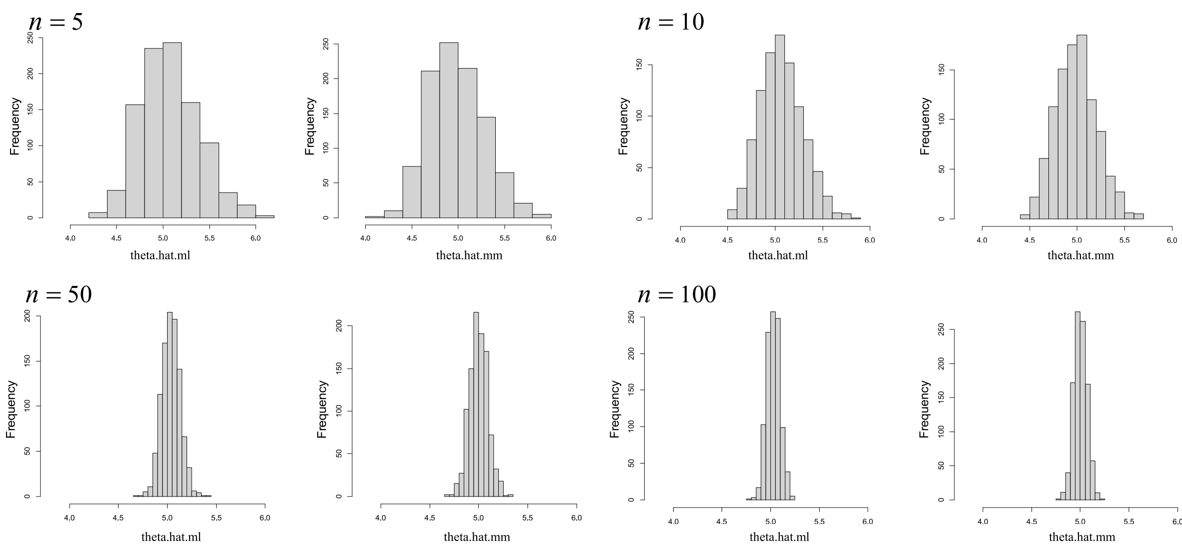

Fig. 5 Sampling distributions of $\hat{\theta}_{\mathrm{ml}}$ and $\hat{\theta}_{\mathrm{mm}}$ under simulated data from two-parameter RL distributions with $\lambda = 0.5$, $\mu = 3.75$, $\theta = 5$, and various sample sizes n.

note that bias can occur in each estimation step. So that it provides more bias in estimation compared to the MM estimator that uses a single-stage computation. $\hat{\theta}_{\mathrm{boot}}$ uses the ML estimate. Its performance is then similar to that of $\hat{\theta}_{\mathrm{ml}}$. As a result, we recommend using $\hat{\theta}_{\mathrm{mm}}$ as a point estimator to estimate the population mean in the two-parameter RL distribution.

Furthermore, we present the confidence interval for the population mean in this paper. The methods are based on parametric and nonparametric approaches. For the parametric approach, we provide the approximate confidence intervals based on the normality properties of $\hat{\theta}_{\mathrm{ml}}$ and $\hat{\theta}_{\mathrm{mm}}$. The nonparametric confidence intervals are constructed based on the percentile bootstrap and bootstrap-t over the ML estimator. Our in-depth simulation study shows that the confidence interval from the bootstrap-t method ($CI_{\mathrm{boot.t}}$) gives coverage rates satisfactorily close to a nominal coverage level for all situations. Especially, its performance does not depend on the sample size or parameter in the two-parameter RL model. The confidence intervals derived using the normality properties of $\hat{\theta}_{\mathrm{mm}}$ and $\hat{\theta}_{\mathrm{ml}}$ are not as satisfactory as they should be, particularly for small samples.

Before reaching our final conclusion, the distributions of $\hat{\theta}_{\mathrm{ml}}$ and $\hat{\theta}_{\mathrm{mm}}$ are discussed. We simulate the data and show the distributions of these estimators in Fig. 5. It can be seen that the empirical distribution of each estimator presents a symmetric curve. Generally, a statistical theory based on the normal distribution can be applied to these estimators. However, as shown in Fig. 5, although $\hat{\theta}_{\mathrm{ml}}$ and $\hat{\theta}_{\mathrm{mm}}$ have symmetric distributions, their distributions change in accordance with the sample size. In such a case, it is more closely related to the t distribution. As we expected, our bootstrap-t method is then remarkably more accurate than the classical approaches based on the normal distribution. According to the simulation results, $CI_{\mathrm{boot.t}}$ performs better than the Wald-type, large-sample, and MOVER approaches in terms of coverage probability. In conclusion, the proposed bootstrap-t confidence interval should be used as a routine for estimating the mean parameter of the two-parameter Rayleigh distribution.

Appendix: Supplementary data

Supplementary data associated with this article can be found at https://dx.doi.org/10.2306/scienceasia1513-1874.2025.006.

Acknowledgements: My thanks go to the editor and referees for the valuable suggestions to improve this paper. The author gratefully acknowledges the financial support provided by Faculty of Science and Technology, Thammasat University, Contract no. SciGR 8/2567.

REFERENCES

1. Rayleigh JWS (1880) On the resultant of a large number of vibrations of the some pitch and of arbitrary phase. *Phil Mag* **10**, 73–78.
2. Asgharzadeh A, Azizpour M (2016) Bayesian inference for Rayleigh distribution under hybrid censoring. *Int J Syst Assur Eng Manag* **7**, 239–249.
3. Dey S, Das MK (2007) A note on prediction interval for a Rayleigh distribution: Bayesian approach. *Am J Math Manag Sci* **1**, 43–48.
4. Kotb MS, Raqab MZ (2018) Bayesian inference and prediction of the Rayleigh distribution based on ordered ranked set sampling. *Commun Stat Simul Comput* **47**, 905–923.
5. Prakash G (2015) Progressively censored Rayleigh data under Bayesian estimation. *Int J Intell Technol Appl Stat* **8**, 257–273.
6. Teawpongpan T, Sangnawakij P (2023) Confidence interval for the population mean of Rayleigh distribution with application to shear wave velocity of soils in Thailand. *J Appl Sci Emerg Technol* **22**, 1–9.
7. Wu SJ, Kus C (2009) On estimation based on progressive first failure-censored sampling. *Comput Stat Data Anal* **53**, 1–12.
8. Dey T, Dey S, Kundu D (2016) On progressively type-II censored two-parameter Rayleigh distribution. *Commun Stat Simul Comput* **11**, 189–212.
9. Liu S, Gui W (2020) Estimating the parameters of the two-parameter Rayleigh distribution based on adaptive type II progressive hybrid censored data with competing risks. *Mathematics* **8**, 1–16.
10. Mkolesia AC, Shatalov M (2017) Exact solutions for a two-parameter Rayleigh distribution. *Global J Pure Appl Math* **13**, 8039–8051.
11. Ullah E, Shahzad MN (2016) Transmutation of the two parameters Rayleigh distribution. *Int J Adv Stat Prob* **4**, 95–101.
12. Casella G, Berger RL (2002) *Statistical Inference*, Duxbury Press, Pacific Grove, USA.
13. Sangnawakij P (2023) Alternative confidence interval estimation for the mean and coefficient of variation in a two-parameter exponential distribution. *J Stat Comput Simul* **93**, 2936–2955.
14. Asgharzadeh A, Fernández AJ, Abdi M (2017) Confidence sets for the two-parameter Rayleigh distribution under progressive censoring. *Appl Math Model* **47**, 656–667.
15. Shang J, Wenhao G (2019) Inference on the lifetime performance index for the Gompertz distribution with the progressively first-failure censored samples. *Int J Innov Comput Inf Control* **15**, 2093–2052.
16. Awwad RRA, Bdair OM, Abufoudeh GK (2021) Bayesian estimation and prediction based on Rayleigh record data with applications. *Stat Transit* **23**, 59–79.
17. Krishnamoorthy K, Waguespack D, Hoang-Nguyen-Thuy N (2020) Confidence interval, prediction interval and tolerance limits for a two-parameter Rayleigh distribution. *J Appl Stat* **25**, 160–175.
18. Donner A, Zou GY (2010) Closed-form confidence intervals for functions of the normal mean and standard deviation. *Stat Methods Med Res* **21**, 347–359.
19. Zou GY, Donner A (2008) Construction of confidence limits about effect measures: A general approach. *Stat Med* **27**, 1693–1702.
20. Zou GY, Huang W, Zhang X (2009) A note on confidence interval estimation for a linear function of binomial proportions. *Comput Stat Data Anal* **53**, 173–196.
21. Nam J, Kwon D (2017) Inference on the ratio of two coefficients of variation of two lognormal distributions. *Commun Stat Theory Methods* **46**, 8575–8587.
22. Panichkitkosolkul W, Budsaba K (2020) Methods for testing the difference between two signal-to-noise ratios of log-normal distributions. In: Huynh VN, Entani T, Jeenanunta C, Inuiguchi M, Yenradee P (eds) *Integrated Uncertainty in Knowledge Modelling and Decision Making*, 8th International Symposium, pp 11–13.
23. Sangnawakij P, Niwitpong S (2017) Confidence intervals for coefficients of variation in two-parameter exponential distributions. *Commun Stat Simul Comput* **46**, 6618–6630.
24. Tang Y (2022) MOVER confidence intervals for a difference or ratio effect parameter under stratified sampling. *Stat Med* **41**, 194–207.
25. Efron B (1988) Bootstrap confidence intervals: Good or bad?. *Psychol Bull* **104**, 293–296.
26. Morton SC (1990) Bootstrap confidence intervals in a complex situation: A sequential paired clinical trial. *Commun Stat Theory Methods* **19**, 181–195.
27. Sangnawakij P, Böhning D (2024) On repeated diagnostic testing in screening for a medical condition: How often should the diagnostic test be repeated?. *Biom J* **6**, 2300175.
28. Sudjai N, Duangsaphon M (2020) Liu-type logistic regression coefficient estimation with multicollinearity using the bootstrapping method. *Sci Eng Health Stud* **14**, 59–79.
29. Zhao S, Peimin R (2011) Weighted bootstrap confidence intervals for generalized Behrens-Fisher problems. *Commun Stat Theory Methods* **40**, 3780–3788.

30. Rousselet GA, Pernet CR, Wilcox RR (2021) The percentile bootstrap: A primer with step-by-step instructions in R. *Adv Methods Pract Psychol Sci* **4**, 1–10.
31. Hoyle SD, Cameron DS (2003) Confidence intervals on catch estimates from a recreational fishing survey: A comparison of bootstrap methods. *Fish Manag Ecol* **10**, 97–108.
32. Zhou XH, Dinh P (2005) Nonparametric confidence intervals for the one- and two-sample problems. *Biostatistics* **6**, 187–200.
33. DiCiccio TJ, Efron B (1996) Bootstrap confidence intervals. *Stat Sci* **11**, 189–212.
34. Hall P (1988) Theoretical comparison of bootstrap confidence intervals. *Ann Stat* **16**, 927–953.
35. R Core Team (2023) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
36. The Copernicus Climate Change Service (2023) *Global Sea Surface Temperature Reaches a Record High*, Copernicus. Available at: https://climate.copernicus.eu/global-sea-surface-temperature-reaches-record-high.
37. Bangkok Metropolitan Administration (2024) *Previous Average Data 24 Hour*, Bangkok Air Quality. Available at: https://airquality.airbkk.com/PublicWebClient/#/Modules/Aqs/HomePage.
38. Manomaiphiboon K (2023) *New PM2.5 Standard will Alarm Public*, Bangkok Post. Available at: https://www.bangkokpost.com/opinion/opinion/2675628.
39. World Health Organization (2021) *WHO Air Quality Guidelines*, C40 Knowledge Hub. Available at: https://www.c40knowledgehub.org/s/article/WHO-Air-Quality-Guidelines?language=en_US.

APPENDIX: PROPERTY OF ESTIMATOR

In this section, the proofs of variance and expected value of the maximum likelihood estimator ($\hat{\theta}_{\mathrm{ml}}$) referred to in the second and the third sections of this article are presented. We first consider the second-derivative of $\log L(\mu,\lambda;x)$ shown in (2). It is given by

$$\frac{\partial^2}{\partial\lambda^2}\log L(\mu,\lambda;x) = -\frac{n}{\lambda^2}.$$

Hence, the expected Fisher information of λ is

$$I(\lambda) = -E\left(\frac{\partial^2}{\partial\lambda^2}\log L(\mu,\lambda;x)\right) = \frac{n}{\lambda^2}.$$

For $n\to\infty$, the variance of $\hat{\lambda}_{\mathrm{ml}}$ is approximated by using the inverse of $I(\lambda)$. Therefore, it is given as

$$Var(\hat{\lambda}_{\mathrm{ml}}) = \frac{1}{I(\lambda)} = \frac{\lambda^2}{n}.$$

Then, we will find the mean of $\hat{\lambda}_{\mathrm{ml}}$. It is simple to start with the expected value of $1/\hat{\lambda}_{\mathrm{ml}}$. As $1/\hat{\lambda}_{\mathrm{ml}} = \sum_{i=1}^{n}(X_i-\hat{\mu}_{\mathrm{ml}})^2/n$, the expected value of $1/\hat{\lambda}_{\mathrm{ml}}$ is obtained from

$$\begin{aligned}E\left(\frac{1}{\hat{\lambda}_{\mathrm{ml}}}\right)&\\ &= \frac{1}{n}E\left(\sum_{i=1}^{n}(X_i-\mu)^2 - 2n(\hat{\mu}_{\mathrm{ml}}-\mu)(\bar{X}-\mu) + n(\hat{\mu}_{\mathrm{ml}}-\mu)^2\right)\\ &= \frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2 - 2n\hat{\mu}_{\mathrm{ml}}\bar{X} + n\hat{\mu}_{\mathrm{ml}}^2\right)\\ &= \frac{1}{n}\left(\sum_{i=1}^{n}E(X_i^2) - 2nE(\hat{\mu}_{\mathrm{ml}})E(\bar{X}) + nE(\hat{\mu}_{\mathrm{ml}}^2)\right).\end{aligned}$$

Since $E(X_i^2) = \frac{1}{\lambda} + \frac{2\mu c}{\sqrt{\lambda}} + \mu^2$, $E(\hat{\mu}_{\mathrm{ml}}) = \frac{c+\mu}{\sqrt{n\lambda}}$, $E(\bar{X}) = \frac{c}{\sqrt{\lambda}} + \mu$, and $E(\hat{\mu}_{\mathrm{ml}}^2) = \frac{1}{n\lambda} + \frac{2c\mu}{\sqrt{n\lambda}} + \mu^2$, we have

$$E\left(\frac{1}{\hat{\lambda}_{\mathrm{ml}}}\right) = \frac{1}{\lambda}\left(1+\frac{1}{n}-\frac{2c^2}{\sqrt{n}}\right).$$

To estimate the mean of $\sqrt{1/\hat{\lambda}_{\mathrm{ml}}}$, the delta method based on the first-order Taylor series expansion is applied. This is given by

$$E\left(\sqrt{\frac{1}{\hat{\lambda}_{\mathrm{ml}}}}\right) \approx \sqrt{E\left(\frac{1}{\hat{\lambda}_{\mathrm{ml}}}\right)} = \sqrt{\frac{1}{\lambda}\left(1+\frac{1}{n}-\frac{2c^2}{\sqrt{n}}\right)}.$$

For the expected value of $\hat{\mu}_{\mathrm{ml}}$, it can be obtained from

$$\begin{aligned}E(\hat{\mu}_{\mathrm{ml}}) &= E(X_{(1)})\\ &= \int_0^\infty n x_{(1)}[1-F_X(x_{(1)})]^{n-1} f_X(x_{(1)})\,\mathrm{d}x_{(1)}\\ &= \int_0^\infty 2n\lambda x_{(1)}(x_{(1)}-\mu)\exp[-n\lambda(x_{(1)}-\mu)^2]\,\mathrm{d}x_{(1)},\end{aligned}$$

where $f_X(x_{(1)})$ is the density function of $X_{(1)}$. Again, using the delta method, the variance of $1/\hat{\lambda}_{\mathrm{ml}}$ is approximated by

$$\begin{aligned}Var\left(\sqrt{\frac{1}{\hat{\lambda}_{\mathrm{ml}}}}\right) &\approx [g'(\lambda)]^2 Var(\hat{\lambda}_{\mathrm{ml}})\\ &= \left(-\frac{1}{2\lambda^{3/2}}\right)^2\frac{\lambda^2}{n} = \frac{1}{4n\lambda}.\end{aligned}$$

Using the above information, we have the expected value of $\hat{\theta}_{\mathrm{ml}}$:

$$\begin{aligned}E(\hat{\theta}_{\mathrm{ml}}) &= E\left(\frac{c}{\sqrt{\hat{\lambda}_{\mathrm{ml}}}} + \hat{\mu}_{\mathrm{ml}}\right) = cE\left(\sqrt{\frac{1}{\hat{\lambda}_{\mathrm{ml}}}}\right) + E(\hat{\mu}_{\mathrm{ml}})\\ &\approx c\sqrt{\frac{1}{\lambda}\left(1+\frac{1}{n}-\frac{2c^2}{\sqrt{n}}\right)} + \frac{c}{\sqrt{n\lambda}} + \mu.\end{aligned}$$

Finally, the proposed variance of $\hat{\theta}_{\mathrm{ml}}$ given in (11) is proved. It is derived by

$$\begin{aligned}Var(\hat{\theta}_{\mathrm{ml}}) &= Var\left(\frac{c}{\sqrt{\hat{\lambda}_{\mathrm{ml}}}} + \hat{\mu}_{\mathrm{ml}}\right) = \frac{c^2}{4n\lambda} + \frac{1}{n\lambda}(1-c^2)\\ &= \frac{c^2+4-4c^2}{4n\lambda} = \frac{4-3c^2}{4n\lambda},\end{aligned}$$

where $c = \Gamma(3/2)$.