Parameter Estimation for Re-Parameterized Length-Biased Inverse Gaussian Distribution

Kamon Budsaba¹, Wikanda Phaphan²

¹Department of Mathematics and Statistics
Faculty of Science and Technology
Thammasat University Rangsit Campus
Pathum Thani, Thailand

²Department of Applied Statistics
Faculty of Applied Science
King Mongkuts University of Technology North Bangkok
Bangkok, Thailand

e-mail: wikanda.p@sci.kmutnb.ac.th

(Received June 11, 2021, Revised July 13, 2021, Accepted July 16, 2021)

Abstract

We study the re-parameterized length-biased inverse Gaussian distribution based on [1]. We propose an alternative to estimators with the method of moments and the maximum likelihood method in a closed-form expression. We compare the effectiveness of estimators with the method of moments and the maximum likelihood method using mean squared error (MSE), and Bias. Moreover, we use the R package “ELBIG” for the parameter estimation for re-parameterized length-biased inverse Gaussian distribution with two estimation methods: the maximum likelihood method, and the method of moments. Furthermore, we illustrate with an example the real data for the proposed estimators. The results show that the parameter estimation of re-parameterized length-biased inverse Gaussian distribution using the method of moments and the maximum likelihood method produces a consistent estimator and the maximum likelihood estimators

Key words and phrases: Length biased inverse Gaussian distribution, maximum likelihood method, method of moments, ELBIG, twoCrack.

AMS (MOS) Subject Classifications: 62F10, 62E17, 62-04.
ISSN 1814-0432, 2022, http://ijmcs.future-in-tech.net
are more precise than the method of moment estimators. In addition, the proposed estimators are more efficient than the maximum likelihood estimators via “nlminb” function in the R program.

1 Introduction

The inverse Gaussian distribution first became well-known when it was presented by Schrodinger [2] as the first passage time distribution of Brownian motion with positive drift. Also, Tweedie [3], and [4] named this distribution the Inverse Gaussian Distribution due to the cumulant generating function of this distribution. It was the inverse with the cumulant generating function of Gaussian distribution. The length-biased version of the inverse Gaussian distribution was studied by Khattree [6]. It is actually a special weighted distribution, proposed by Patil and Rao [5]. The other name of the length-biased inverse Gaussian distribution is the complementary reciprocal of the inverse Gaussian distribution. The inverse Gaussian distribution and the length-biased inverse Gaussian distribution are suitable for the biometric sampling plans and the survival analysis. The research of Patil and Rao [7], Blumenthal [8], Scheaffer [9], and Simon [10] demonstrated the use of the length-biased sampling in various areas. The research of Cnaan [11] described the meaning of the length-biased distribution in random sampling for studying coronary artery disease. Sen [12] studied mathematical properties, arithmetic, geometric and harmonic means, and described the characteristics of length-biased sampling. Gupta and Kirmani [13] examined the relationship between the length-biased and the former random variable in an article on experiments of service life with reliability measurement. Akman and Gupta [14] compared the uniformly minimum variance unbiased estimator (UMVUE) and the maximum likelihood estimator (MLE) of the parameter \( \mu \) where the data have the inverse Gaussian distribution \( IG(\mu, c\mu^2) \) and the length-biased inverse Gaussian distribution \( LBIG(\mu, c\mu^2) \). Gupta and Akman [15] used some results of Sen [12] for developing the inverse moments, asymptotic tests, and confidence intervals for the mean and the coefficient of variation of the inverse Gaussian distribution based on the length biased data. Phaphan and Pongsart [16] carried out a study into the Fisher’s information matrix to construct the asymptotic confidence ellipses of parameters for the length-biased inverse Gaussian distribution in the parameter proposed by Chhikara and Folk [17]. Other researchers, including Pudprommarat [18] and Simmachan et al. [19], conducted a study of length-biased distribution. In this paper, we deal with the study of re-parameterized length-biased inverse Gaussian
distribution according to Phaphan [1]. The probability density function of the re-parameterized length-biased inverse Gaussian distribution is given by the formula:

\[
f_{LBIG}(x; \lambda, \theta) = \begin{cases} 
\frac{1}{\theta \sqrt{2\pi}} \left( \frac{\theta}{x} \right)^{3/2} \exp \left[ -\frac{1}{2} \left( \sqrt{\frac{\theta}{x}} - \lambda \sqrt{\frac{\theta}{x}} \right)^2 \right] ; & x > 0 \\
0 ; & \text{otherwise},
\end{cases}
\]  

(1.1)

where \( \lambda > 0, \theta > 0 \) represent the parameters of the length-biased inverse Gaussian distribution \( LBIG(\lambda, \theta) \).

According to the literature review of the re-parameterized length-biased inverse Gaussian distribution, the results lead us to conclude that the study of constructing estimators of re-parameterized length-biased inverse Gaussian distribution based on [1] by employing the method of moments and the maximum likelihood method in a closed-form expression have not been carried out in a study. Consequently, we are not only interest in the topic mentioned above but also in using the R language package for parameter estimation for practical applications in medical science and engineering.

## 2 Parameter Estimation

### 2.1 Method of Moment

**Theorem 2.1.** Given \( X \sim LBIG(\lambda, \theta) \) and \( x_1, \ldots, x_n \) be a positive random sample of size \( n \). Let \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \), \( s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \). Then the method of moments estimators of the parameters \( \lambda \) and \( \theta \) are

\[
\hat{\lambda}_{MME} = \frac{\bar{x}^2 - 2s^2 + \bar{x} \sqrt{\bar{x}^2 + 4s^2}}{2s^2},
\]

\[
\hat{\theta}_{MME} = -\bar{x} + \sqrt{\bar{x}^2 + 4s^2}.
\]  

(2.2)

**Proof.** The characteristic function of the LBI\( G \) distribution [21] is

\[
\phi_{LBIG}(t) = (1 - 2i\theta t)^{-1/2} \exp \left\{ \lambda \left[ 1 - (1 - 2i\theta t)^{1/2} \right] \right\}.
\]  

(2.3)

Then the logarithm of the characteristic function is

\[
\ln \phi_{LBIG}(t) = -\frac{1}{2} \ln(1 - 2i\theta t) + \lambda \left[ 1 - (1 - 2i\theta t)^{1/2} \right] = -\frac{1}{2} \ln(1 - 2i\theta t) - \lambda(1 - 2i\theta t)^{1/2} + \lambda.
\]  

(2.4)
Using MacLaurin’s series expansion, from equation (2.4), we obtain
\[
\ln \varphi_{LBIG}(t) = \frac{it}{1!(1-\lambda)\theta} + \frac{(it)^2}{2!(2+\lambda)\theta^2} + \frac{(it)^3}{3!(8+3\lambda)\theta^3} + \frac{(it)^4}{4!(48+15\lambda)\theta^4} + O(t^5).
\] (2.5)

This expansion (2.5) gives us the central population moments of LBIG distribution (see Lisawadi [21]).

\[
K_1 = \mu = \theta(1+\lambda), \quad K_2 = \sigma^2 = \theta^2(2+\lambda),
\]
\[
K_3 = \mu_3 = \theta^3(8+3\lambda), \quad K_4 = \mu_4 - 3\sigma^4 = \theta^4(48+15\lambda). \tag{2.6}
\]

We will estimate the equations for two parameters of LBIG distribution when

\[
m_1 = \bar{x}, \quad m_2 = \mu_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2, \tag{2.7}
\]

where \(m_1\) is the first central sample moment and \(m_2\) is the second central sample moment. From equations (2.6) and (2.7), we get

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = K_1 = \theta(1+\lambda), \tag{2.8}
\]
\[
s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = K_2 = \theta^2(2+\lambda). \tag{2.9}
\]

From equation (2.8), we obtain

\[
\frac{1}{n} \sum_{i=1}^{n} x_i = \theta(1+\lambda),
\]
\[
\tilde{\lambda} = \frac{\bar{x}}{\theta} - 1. \tag{2.10}
\]

From equation (2.9),

\[
\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \theta^2(2+\lambda). \tag{2.11}
\]
Substituting (2.10) into (2.11), we obtain
\[
\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \theta^2 \left( 2 + \frac{\bar{x}}{\theta} - 1 \right),
\]
\[
\theta^2 + \bar{x} \theta - s^2 = 0.
\]
(2.12)

Solving equation (2.12)
\[
\tilde{\theta} = \frac{-\bar{x} \pm \sqrt{\bar{x}^2 + 4s^2}}{2}.
\]
(2.13)

Since \( \tilde{\theta} \) is positive parameter,
\[
\tilde{\theta} = \frac{-\bar{x} + \sqrt{\bar{x}^2 + 4s^2}}{2}.
\]
(2.14)

Next, substitute (2.14) into (2.10).
\[
\tilde{\lambda} = \frac{\bar{x}}{-\bar{x} + \sqrt{\bar{x}^2 + 4s^2}} - 1,
\]
\[
= \frac{\bar{x}^2 + \bar{x} \sqrt{\bar{x}^2 + 4s^2}}{2s^2} - 1,
\]
\[
= \frac{\bar{x}^2 - 2s^2 + \bar{x} \sqrt{\bar{x}^2 + 4s^2}}{2s^2}.
\]
(2.15)

Hence, the method of moments estimators for LBIG distribution is
\[
\tilde{\lambda}_{MME} = \frac{\bar{x}^2 - 2s^2 + \bar{x} \sqrt{\bar{x}^2 + 4s^2}}{2s^2},
\]
\[
\tilde{\theta}_{MME} = \frac{-\bar{x} + \sqrt{\bar{x}^2 + 4s^2}}{2}.
\]

2.2 Maximum Likelihood Method

**Theorem 2.2.** Let \( X \sim LBIG(\lambda, \theta), x_1, \ldots, x_n \) be a positive random sample of size \( n \), \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \), \( T = \sum_{i=1}^{n} \frac{1}{x_i} \). Then the maximum likelihood estimators of the parameters \( \lambda \) and \( \theta \) are
\[
\hat{\lambda}_{MLE} = \frac{n}{T \bar{x} - n},
\]
\[
\hat{\theta}_{MLE} = \frac{T \bar{x} - n}{T}.
\]
(2.16)
Proof. The LBIG distribution has the density function:

\[
f_{LBIG}(x; \lambda, \theta) = \frac{1}{\theta \sqrt{2\pi}} \left( \frac{\theta}{x} \right)^{\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( \sqrt{\frac{x}{\theta}} - \lambda \sqrt{\frac{\theta}{x}} \right)^2 \right],
\]

(2.17)

The likelihood function is

\[
L(x; \lambda, \theta) = \frac{1}{\theta^{\frac{n}{2}} (\sqrt{2\pi})^n \prod_{i=1}^{n} x_i^{\frac{1}{2}}} \exp \left[ -\frac{n}{2} \frac{\theta}{x} + n\lambda - \frac{\sum_{i=1}^{n} x_i}{\theta} \right],
\]

(2.18)

and the logarithm likelihood function is

\[
g(\lambda, \theta) = \ln L(x; \lambda, \theta) = n \ln 1 - \frac{n}{2} \ln \theta - n \ln \sqrt{2\pi} - \sum_{i=1}^{n} \ln x_i^{\frac{1}{2}} - \frac{\lambda^2 \theta}{2} \sum_{i=1}^{n} \frac{1}{x_i} + n\lambda - \frac{\sum_{i=1}^{n} x_i}{\theta}.
\]

(2.19)

From (2.19), taking derivative and setting it equal to zero, we obtain

\[
\frac{\partial}{\partial \theta} g(\lambda, \theta) = -\frac{n}{2\theta} - \frac{\lambda^2}{2} \sum_{i=1}^{n} \frac{1}{x_i} + \frac{\sum_{i=1}^{n} x_i}{2\theta^2} = 0,
\]

\[
\lambda^2 \theta^2 \sum_{i=1}^{n} \frac{1}{x_i} + n\theta - \sum_{i=1}^{n} x_i = 0. \tag{2.20}
\]

Denoting \( T = \sum_{i=1}^{n} \frac{1}{x_i} \) and \( n\bar{x} = \sum_{i=1}^{n} x_i \), we have

\[
\lambda^2 T \theta^2 + n\theta - n\bar{x} = 0. \tag{2.21}
\]

Thus,

\[
\hat{\theta} = -\frac{n \pm \sqrt{n^2 + 4\lambda^2 T n\bar{x}}}{2\lambda^2 T}. \tag{2.22}
\]

Since \( \theta \) is positive,

\[
\hat{\theta} = -\frac{n + \sqrt{n^2 + 4\lambda^2 T n\bar{x}}}{2\lambda^2 T}. \tag{2.23}
\]

Next,

\[
\frac{\partial}{\partial \lambda} g(\lambda, \theta) = -\lambda \theta \sum_{i=1}^{n} \frac{1}{x_i} + n = 0,
\]

\[
\lambda \theta T - n = 0. \tag{2.24}
\]
Parameter Estimation for Re-Parameterized Length-Biased...

Since \( \hat{\theta} = \frac{-n + \sqrt{n^2 + 4\lambda^2 T \bar{x}}}{2\lambda T} \),

\[
\lambda \left( \frac{-n + \sqrt{n^2 + 4\lambda^2 T \bar{x}}}{2\lambda T} \right) T - n = 0,
\]

\[
\frac{-n + \sqrt{n^2 + 4\lambda^2 T \bar{x}}}{2\lambda} - n = 0,
\]

\[
\sqrt{n^2 + 4\lambda^2 T \bar{x}} = 2n\lambda + n,
\]

\[
\left( \sqrt{n^2 + 4\lambda^2 T \bar{x}} \right)^2 = (2n\lambda + n)^2,
\]

\[
n\lambda = 0, \quad (\lambda \bar{x} - n\lambda - n) = 0.
\]

Hence, \( \hat{\lambda} = 0 \). The root \( \hat{\lambda} = 0 \) can not used because \( \hat{\lambda} \) is a positive parameter. Then \( \hat{\lambda} = \frac{n}{T \bar{x} - n} \).

In the expression (2.23)

\[
\hat{\theta} = \frac{-n + \sqrt{n^2 + 4\lambda^2 T \bar{x}}}{2\lambda T},
\]

substitute \( \hat{\lambda} = \frac{n}{T \bar{x} - n} \) to get

\[
\hat{\theta} = \frac{-n + \sqrt{n^2 + 4(\frac{n}{T \bar{x} - n})^2 T \bar{x}}}{2 \left( \frac{n}{T \bar{x} - n} \right)^2 T},
\]

\[
= \frac{-n(T \bar{x} - n)^2 + \sqrt{n^2(T \bar{x} - n)^4 + 4n^2(T \bar{x} - n)^2T \bar{x}}}{2n^2 T},
\]

\[
= \frac{-n(T \bar{x} - n)^2 + n(T \bar{x} - n)\sqrt{(T \bar{x} - n)^2 + 4T \bar{x}}}{2n^2 T},
\]

\[
= \frac{-n(T \bar{x} - n)^2 + n(T \bar{x} - n)(T \bar{x} + n)}{2n^2 T},
\]

\[
= \frac{-n(T \bar{x} - n)^2 + n(T^2 \bar{x}^2 - n^2)}{2n^2 T},
\]

\[
= \frac{-n(T \bar{x} - n)^2 + n(T^2 \bar{x}^2 - n^2)}{2n^2 T},
\]

\[
= \frac{-T^2 \bar{x}^2 + 2T \bar{x} \bar{n} - n^2 + T^2 \bar{x}^2 - n^2}{2nT},
\]

\[
= \frac{T \bar{x} - n}{T}.
\]

(2.25)
Checking the property of estimator that is maximum by the Hessian test, where $H(\lambda, \theta)$ is the Hessian matrix of $g$.

$$H(\lambda, \theta) = \begin{pmatrix} -\theta T & -\lambda T \\ -\lambda T & n^2 + \frac{n^2}{\theta^2} \end{pmatrix},$$

(2.26)

Since the determinant of Hessian matrix is

$$det \left( H(\hat{\lambda}, \hat{\theta}) \right) = \frac{-T^2 n}{2(T \bar{x} - n)} + \frac{T^3 n \bar{x}}{(T \bar{x} - n)^2} - \frac{T^2 n^2}{(T \bar{x} - n)^2} < 0,$$

(2.27)

and

$$\frac{\partial^2}{\partial \theta^2} g(\hat{\theta}, \hat{\lambda}) = -\frac{nT^2}{2(T \bar{x} - n)^2} + T^3 \frac{n \bar{x}}{(T \bar{x} - n)^3} < 0,$$

(2.28)

and

$$\frac{\partial^2}{\partial \lambda^2} g(\hat{\theta}, \hat{\lambda}) = -\left( \frac{T \bar{x} - n}{T} \right) T < 0.$$ 

(2.29)

Hence, $(\hat{\theta}, \hat{\lambda})$ is a local maximum for $g$ and thus for $L$ (since $\ln$ is an increasing function). Therefore, the estimators for the two parameters of LBUG distribution by the maximum likelihood method are

$$\hat{\lambda}_{MLE} = \frac{n}{T \bar{x} - n},$$

$$\hat{\theta}_{MLE} = \frac{T \bar{x} - n}{T}.$$ 

\[\square\]

3 Monte Carlo Simulations

The research generated random numbers of re-parameterized length-biased inverse Gaussian distribution through the composition method by using the “twoCrack” package [22] and the simulations were repeated 1,000 times for each model. Also, determining sample of size $n = 5, 30, 50, 500$, parameters $\lambda = 0.5, 1, 2, 3, 4, 5, 10$, and $\theta = 0.5, 1, 3, 5, 10$. All of the experiments were run on the R program version 4.0.5. Regarding the simulation results in Figures 1 and 2, we observe that the proposed maximum likelihood estimators and the method of moments estimators worked well. The average estimates of $\lambda$ and $\theta$ were close to the actual value given, especially the large samples. Moreover, the estimator of $\lambda$ by the maximum likelihood method and the
method of moments had slightly more values than the actual values in every case, and the bias of parameter estimation $\lambda$ increased when the value of $\lambda$ was increased. In addition, the bias of maximum likelihood estimators was slightly less than the method of moments estimators for both parameters.

4 A New R Package for Parameter Estimation of the Re-parameterized Length-biased Inverse Gaussian Distribution

This section presents a new package for R named “ELBIG”. The package ELBIG provides functions for parameter estimation for re-parameterized length-biased inverse Gaussian distribution with two estimation methods: the maximum likelihood method, the method of moments. The ELBIG package was published only on GitHub at https://github.com/wikanda-phaphan/ELBIG and requires a version of R $\geq$ 3.6.1. The steps for installing the package ELBIG on R or RStudio from GitHub are as follows:

```r
install.packages("devtools")
library(devtools)
install_github("wikanda-phaphan/ELBIG")
library(ELBIG)
```

After installing the ELBIG package, users can use “? ELBIG” command to get to the user’s manual. This new package includes the three functions: 1. MME(X) gives the value of parameter estimates by the method of moments in Eq. (2.2), 2. MLE(X) gives the value of parameter estimates by the maximum likelihood method in Eq. (2.16). 3. MillVibration gives the vibration of the vertical roller mill in 60 minutes, and contains the time (a.m.) and values of mill vibration (um), collected on February 10, 2019 from [22]. In the calling sequence for using the functions, $X$ denotes data with a positive value.

5 Illustrative Examples

To illustrate the use of the proposed estimators with real data, we use the vibration of the vertical roller mill data set in 60 minutes. This data set contains the time (a.m.) and values of mill vibration (um), collected on February 10, 2019 from [22]. As this data was right-skewed, we chose the
Figure 1: Line graph for the bias of estimator $\lambda$ and $\theta$
Figure 2: Line graph for the MSE of estimator $\lambda$ and $\theta$
Table 1: The MLE of the model parameters for the mill vibration, and AIC measure

<table>
<thead>
<tr>
<th>Fitting Dist.</th>
<th>Estimate parameters</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>λ</td>
<td>θ</td>
</tr>
<tr>
<td>TCR</td>
<td>19.73067</td>
<td>0.1</td>
</tr>
<tr>
<td>BS</td>
<td>0.001</td>
<td>0.1</td>
</tr>
<tr>
<td>IG</td>
<td>3</td>
<td>0.01</td>
</tr>
<tr>
<td>LBIG</td>
<td>130.9168</td>
<td>0.5669015</td>
</tr>
</tbody>
</table>

two-parameter crack (TCR), inverse Gaussian (IG), re-parameterized length-biased inverse Gaussian (LBIG), and Birnbaum-Saunders (BS) distribution to be a candidate distribution for the mill vibration data [22]. We estimate the parameters of the re-parameterized length-biased inverse Gaussian distribution by using the proposed estimators in Eq. (2.16), and estimate the parameters of other distributions by maximizing the likelihood function using the “nlminb” function, the result is in Table 1.

Regarding the results in Table 1, we observe that the re-parameterized length-biased inverse Gaussian distribution has an AIC value less than another distribution and so the re-parameterized length-biased inverse Gaussian distribution fits the data better than other distributions. Finally, we fitted the re-parameterized length-biased inverse Gaussian distribution to the mill vibration data set so the thickness of the machine element was 130.9168 and the nominal treatment pressure on the machine element was 0.5669015.

6 Conclusion

This paper dealt with the re-parameterized length-biased inverse Gaussian distribution based on [1]. We introduced an alternative estimator with the method of moments and the maximum likelihood method in a closed-form expression. We experimented with an R program using Monte Carlo techniques. The sample sizes were 5, 30, 50, 500, and repeated 1,000 times in each case. We compared the effectiveness of estimators with the method of moments and the maximum likelihood method by considering mean squared error and Bias. The results showed that the parameter estimation of re-parameterized length-biased inverse Gaussian distribution through the method of moments
and the maximum likelihood method produce a consistent estimator. Moreover, the maximum likelihood estimators were more precise than the method of moment estimators. Furthermore, we provided an R language package for parameter estimation where the data have the re-parameterized length-biased inverse Gaussian distribution and showing the example of applying the proposed estimators with the vertical roller mill vibration data set to 60 minutes [22]. Phaphan’s research [22] shows that the two-parameter crack distribution fits the vertical roller mill vibration data better than other distributions by maximizing the likelihood function using the nlminb function. Our research indicates that the re-parameterized length-biased inverse Gaussian distribution fits the vibration data better than other distributions by using the maximum likelihood estimators in a closed-form expression which differs from the result of Phaphan [22]. Hence, we can conclude that the proposed estimators are more efficient than the maximum likelihood estimators via nlminb function in the R program.

Acknowledgment. This research was funded by the Faculty of Applied Science, King Mongkut’s University of Technology North Bangkok. Contract no. 641048.

References


