Three Parameters Estimation of Log-Logistic Distribution Using Algorithm of Percentile Roots

Doaa Abd El-Shafi Abd El-Rahman1,2, Mohammed Mohammed El Genidy3

Abstract This study provided a new algorithm to estimate the three parameters of log-logistic distribution called algorithm of percentile roots, AOPR. The effectiveness of AOPR was tested by the Anderson-Darling test which accepted the results at a high level of significance. The estimated parameters by AOPR were very close to EasyFit program's results. In addition, they were somewhat more accurate than EasyFit program's ones. The analyzed data is a data set of daily global solar exposure for a year in Queensland, Australia.

Keywords: Three-Parameter Log-Logistic Distribution, Algorithm of Percentile Roots, Anderson-Darling Test, Daily Global Solar Exposure

1 Introduction

Data analysis is a wide field that plays a major role in any decision-making process based on a numerical survey. That made statisticians always look forward to the most accurate modeling of data. Out of this, the distribution by which the data modeled should be estimated too accurate to fit the data almost perfectly. The factors that have a real impact on data modeling are the chosen distribution and the parameters estimation method. Some researchers chose three-parameter log-logistic distribution (LLD3p) to model data of various phenomena and fields. Kantam et al. (2007) presented LLD3p in a model of collected life of given items to decide accepting or rejecting the product. Log-normal and log-logistic distributions were analyzed lifetime data in a study by Dey and Kundu (2009). Others used generalized logistic distribution (GLD) in many fields too. Asgharzadeh et al. (2013) analyzed a real data set of log times to breakdown of a fluid insulating in an accelerated test by GLD and maximum likelihood estimation. While, Shin et al. (2012) applied generalized extreme value (GEV) distribution and GLD to data of annual maximum rainfall in Korea. Likewise, GLD was fit to data in the economic sphere, as Huang et al. (2017) performed it on a data set of expected shortfall and value-at-risk in resource sector.

Parameters Estimator also affects substantially on data analysis, that's why many studies have focused on the estimation of the parameters for different distributions. For instance, Ashkar and Mahdi (2006) estimated two-parameter log-logistic distribution by generalized moments and quantiles estimators. Asgharzadeh (2006) presented maximum likelihood (ML) and approximate ML estimators for GLD. Just like Arora et al. (2010) who applied GLD with ML estimation. And another, ML, approximate ML and Bayes estimation methods were used along with Weibull distribution by Valiollahi et al. (2013). Besides, El Genidy (2019) introduced Exponentiated Gumbel Maximum distribution that was estimated by quartiles-moments method to analyze maximum temperature and solar radiation data. Lindley distribution was estimated using the methods of uniformly minimum variance unbiased and ML by Maiti and Mukherjee (2017).

1Department of Basic Science, Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt.
2Corresponding author, E-mail Address: doaa.abdelshafi@gmail.com.
3Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt. Email: drmng2016@yahoo.com.
Moreover, Khan et al. (2015) compared moments, energy pattern factor, ML, and empirical methods as estimators for determination of Weibull parameters to model daily wind speed data in Jiwani town along ten years. The comparison showed that the best results were those that obtained by ML. Another comparison was made by Kaoga et al. (2014) between five numerical estimators. They were the empirical estimation, ML method, modified ML method, the graphical estimator, and energy pattern factor methods. The last one led to the best fit to the analyzed data. Rocha et al. (2012) discussed seven estimation methods that were ML, modified ML, graphical, moments, equivalent Energy, the empirical, and energy pattern factor methods.

Based on the foregoing, the accurate estimation of parameters is very important to get the best model of any data set by a chosen probability distribution. This research introduced a new parameters estimation method, called algorithm of percentile roots (AOPR), that is commensurate to calculate the shape, scale and location parameters of LLD3p. This estimator could combine percentile equations with the measures of central tendency, that made the estimation more accurate than EasyFit program estimation. The data set that was used to apply LLD3p with AOPR was daily global solar exposure in Queensland throughout one year. AOPR's results were checked by Anderson-Darling test and they were significant.

2 Materials and Methods

This study applied a new parameter estimator called AOPR with LLD3p. The algorithm proved its efficiency compared to EasyFit program as clarified by Anderson-Darling test.

2.1 The Analyzed Data set

The data set that used to apply the new algorithm was on the daily global solar exposure in Queensland for a year, 2015 (Figure 1.). It was actual and reliable data, as it had been recorded and published by the Bureau of Meteorology in Australian government. It was posted online in http://www.bom.gov.au/climate/data/index.shtml.

![Figure 1. Data set of Daily Global Solar Exposure during 2015 at Queensland, Australia.](image-url)
2.2 Three-Parameter Log-Logistic Distribution

Let $X$ be a random variable that presented a data set of the daily global solar exposure in Queensland throughout 2015. This random variable of the same data set was matched with two-parameter Weibull distribution in a research by El Genidy and Abd El-Rahman (2019). But here, $X$ has been studied to have LLD3p with the following form of probability density function (PDF):

$$f(x) = \frac{\left(1 + \frac{\beta}{\sigma(x-\mu)}\right)^{-\frac{1}{\beta}}}{\sigma\left(1 + \left(1 + \frac{\beta}{\sigma(x-\mu)}\right)^{-1}\right)^{\frac{1}{\beta}}}; \sigma > 0, \ \beta > 0 \ \text{and} \ x > a \ \text{S.T} \ a = \mu - \frac{\sigma}{\beta} \quad (1)$$

Such that $\beta$ is the shape parameter, $\sigma$ is the scale parameter and $\mu$ is the location parameter.

Since the cumulative distribution function (CDF) is given by:

$$F(x) = \int_a^x f(y) \, dy \quad (2)$$

Then, it would be:

$$F(x) = \left(1 + \left(1 + \frac{\beta}{\sigma(x-\mu)}\right)^{-1}\right)^{-\frac{1}{\beta}} \quad (3)$$

The mean is the expected value of $x$, and it is obtained by the equation:

$$M = E(x) = \int_a^\infty x \, f(x) \, dx \quad (4)$$

Which implies:

$$M = \mu + \frac{\sigma}{\beta} (\pi \beta \csc (\pi \beta) - 1) \quad (5)$$

And the median is the value of $x$ that satisfies:

$$F(x) = 0.5 \quad (6)$$

$$\left(1 + \left(1 + \frac{\beta}{\sigma(x-\mu)}\right)^{-1}\right)^{-\frac{1}{\beta}} = 0.5$$

$$\left(1 + \frac{\beta}{\sigma(x-\mu)}\right)^{\frac{1}{\beta}} = 1$$
\[ \frac{\beta}{\sigma} (x-\mu) = 0 \]

And \( \frac{\beta}{\sigma} \) couldn't be zero, which means the median would be:

\[
\text{Median} = x = \mu \quad (7)
\]

### 2.3 Algorithm of Percentile Roots

The algorithm, which was created in this research, AOPR, depended on solving three equations together to get approximated values for the three parameters of LLD3p. The equations were CDF \( y_i \) at any point \( x_i \), the mean and the median. This technique combined the percentile roots with some important common measures of central tendency, such as mean and median.

By using Eq.7, the median of the used data set gave the value of the location parameter \( \mu \), as \( \mu = \text{the median} \). From Eq.5, the used data set's mean was used to simplify it as follows:

\[
\sigma = \frac{\beta(M-\mu)}{(\pi\beta \csc(\pi\beta)-1)} \quad (8)
\]

Then, Eq.5 was used with an arbitrary percentile equation of the used data set at any point \( x_i \), \( y_i = F(x_i) \), to get the value of the shape parameter \( \beta_i \) as follows:

\[
y_i = \left( 1+\left(1+\frac{(x-\mu)(\pi\beta_i \csc(\pi\beta_i)-1)}{M-\mu}\right)^{-\frac{1}{\beta_i}} \right)^{-1} \quad (9)
\]

Finally, the scale parameter \( \sigma_i \) could be calculated by Eq.5 after obtaining \( \beta_i \).

This operation would be iterated arbitrary number of times \( N \), such that \( i=1,2,3...N \), so that \( N \) values of the shape and the scale parameters would be obtained. The median could be calculated for the \( N \) values of each parameter to get the final estimation.

### 2.4 Algorithm Statement

**Input:**
- \( N \) the number of percentile equations, \( X[0..N-1] \) array of x values, \( Y[0..N-1] \) array of y values, \( M \) the mean, \( \mu \) the median

**Output:**
- A the median of \( \beta \) values, B the median of \( \sigma \) values

**Procedure:**
- \( N \gets \text{read()} \)
- for \( i \gets 0 \) to \( N-1 \) do
  - \( X[i] \gets \text{read()}, Y[i] \gets \text{read()}, M \gets \text{read()}, \mu \gets \text{read()} \)
\[
\text{FindRoot}\left(\left(1+\left(1+\frac{x-18.1}{0.2}\right)^{\left(\frac{1}{\pi\beta}\right)}\right)^{-1}\right) = y,\{\beta,0.01}\right)
\]

\[
\sigma = \frac{\beta(M-\mu)}{(\pi\beta csc(\pi\beta)-1)}
\]

\[
J[i] \leftarrow \beta \quad K[i] \leftarrow \sigma
\]

A = median(J)  
B = median(K)  
print A, B

2.5 Goodness-of-Fit Test

The fitness of AOPR's results with the used data set was measured by Anderson-Darling test. It's defined as:

\[
A^2 = -\sum_{i=1}^{n} \left(\frac{2i-1}{n} \left[\ln(F(x_i)) + \ln(1-F(x_{n+1-i}))\right]\right) - n
\]  \hspace{1cm} (10)

2.6 Software

The programs carried out in this study were EasyFit professional version 5.5 and Mathematica 8 version 8.0.1. The first was released in February 2010 and available from MathWave Technologies, http://www.mathwave.com. And the else was released in March 2011 and available from Wolfram Mathematica, http://www.wolfram.com.

3 Results

3.1 Estimation method

Calculating the measures of central tendency of the cited data set of daily global solar exposure in Queensland along 365 days. Therefore, from Eq. 5,7 and 8, the mean and the median were as follows:

\[
M=\mu+\sigma\frac{\pi\beta\csc(\pi\beta)-1}{\beta} \quad M=\mu+\sigma\frac{\pi\beta\csc(\pi\beta)-1}{\beta} = 18.3
\] \hspace{1cm} (11)

\[
\text{Median} = \mu = 18.1
\] \hspace{1cm} (12)
\[ y_i = \left(1 + \left(1 + \frac{(x-18.1)(\pi\beta \csc(\pi\beta) - 1)}{0.2}\right)^{-1}\right)^{-1} \]  

(13)

By solving Eq.13 using Mathematica program at percentile equations \( y_1 = 0.005, y_2 = 0.01, y_3 = 0.015, y_3 = 0.02, y_3 = 0.025 \ldots, y_{99} = 0.995 \) and there corresponding values \( x_1, x_2, x_3, \ldots, x_{99} \) as follows:

\[
\text{FindRoot}\left[\left(1 + \left(1 + \frac{(x-18.1)(\pi\beta \csc(\pi\beta) - 1)}{0.2}\right)^{-1}\right)^{-1} = y_i,\{\beta, 0.01\}\right]
\]

Hence, 197 values of the parameter \( \beta \) could be obtained, excluding \( y_{100} = 0.5 \) as it's the median equation, and then their medians were calculated as shown in Table1.

**Table 1. Numerical values of the shape and the scale parameters obtained by AOPR.**

<table>
<thead>
<tr>
<th>( x_i )</th>
<th>( y_i )</th>
<th>( \beta_i )</th>
<th>( \sigma_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.982</td>
<td>0.005</td>
<td>0.03845</td>
<td>3.15403</td>
</tr>
<tr>
<td>3.792</td>
<td>0.01</td>
<td>0.03594</td>
<td>3.37521</td>
</tr>
<tr>
<td>3.9</td>
<td>0.015</td>
<td>0.0334</td>
<td>3.6331</td>
</tr>
<tr>
<td>4.336</td>
<td>0.02</td>
<td>0.03227</td>
<td>3.76033</td>
</tr>
<tr>
<td>5.3</td>
<td>0.025</td>
<td>0.03275</td>
<td>3.70469</td>
</tr>
<tr>
<td>5.852</td>
<td>0.03</td>
<td>0.03258</td>
<td>3.7238</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>30.516</td>
<td>0.97</td>
<td>0.03622</td>
<td>3.34898</td>
</tr>
<tr>
<td>30.7</td>
<td>0.975</td>
<td>0.03786</td>
<td>3.20369</td>
</tr>
<tr>
<td>31.072</td>
<td>0.98</td>
<td>0.03935</td>
<td>3.08195</td>
</tr>
<tr>
<td>31.362</td>
<td>0.985</td>
<td>0.04184</td>
<td>2.89754</td>
</tr>
<tr>
<td>31.672</td>
<td>0.99</td>
<td>0.0457</td>
<td>2.65217</td>
</tr>
<tr>
<td>31.918</td>
<td>0.995</td>
<td>0.05369</td>
<td>2.25525</td>
</tr>
</tbody>
</table>

**Median**  
0.02862  
4.24075

Therefore, AOPR estimated the parameters as 18.1, 0.02862 and 4.24075 for the location, the shape and the scale parameters, respectively. Formula of the CDF of LLD3p became:

\[
F(x) = \left(1 + \left(1 + \frac{0.02862}{4.24075}(x-18.1)\right)^{-1}\right)^{-1}
\]

(14)

**3.2 Checking the validity of AOPR**

Anderson-Darling test was performed in this research to measure the validity of AOPR's results, as it's considered one of the most effective statistical tests. Applying Eq. 10, the test statistic value of AOPR's results was \( A^2 = 1.32095 \) which was less than the EasyFit program's one.
A²=1.4144 (Table 2.). This means AOPR gave an estimated LLD3p that was closer to the actual data than the other one produced by EasyFit program.

Table 2. Estimated parameters by AOPR and EasyFit program and their and Anderson-Darling test values.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>EasyFit Program</th>
<th>AOPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>μ</td>
<td>18.21</td>
<td>18.1</td>
</tr>
<tr>
<td>β</td>
<td>0.01374</td>
<td>0.02954</td>
</tr>
<tr>
<td>σ</td>
<td>3.85913</td>
<td>4.1085</td>
</tr>
<tr>
<td>A²</td>
<td>1.4144</td>
<td>1.32095</td>
</tr>
</tbody>
</table>

So, AOPR proved its ability of estimating accurate parameters for LLD3p as well as the other traditional known estimation methods.

4 Discussion

This study is concerned about finding a new estimator of LLD3p. This estimator, AOPR, is characterized by its ease of use, accuracy and its fitness with any three-parameter distribution. According to Anderson-Darling test, AOPR produced significant results when it was used to model real data of daily global solar exposure with LLD3p. Moreover, the results were slightly more accurate than results obtained by EasyFit program when using the same data, distribution and statistical test.

In the previous studies, many types of logistic distributions had been performed to fit real data in various fields with different estimators of parameters. Due to its relative simplicity and flexibility, GLD was performed side by side with Weibull, gamma, and log-normal distributions with ML estimation to analyze data of light, deep and rapid eye movement sleep stages by Saat and Jemain (2009). Besides, Gupta and Kundu (2010) discussed some properties of the proportional reversed hazard logistic distributions and the skew logistic distribution with proposing data of the measured strength in grade point average for single carbon fibers of a specific length. Ahmed (2018) modeled a data set of the monthly closing values of indices for the entire composite index using GLD with two different estimation methods. In another study by Kantam et. al (2007), LLD3p was estimated to model the life of a specific product in order to accepting or rejecting it. Also, LLD3p was also able to represent well solar exposure data in the present study.

Turning to the role of parameters estimation method, several estimators are considered good generators of parameters, as ML method, method of moments, least squares method and the methods depends on percentiles. According to Soukissian and Tsalis (2015), data of extreme wind speed is analyzed by applying GEV distribution with nine different estimation methods such as moments, ML, ordinary entropy and quantile least squares methods. El Genidy and Ali (2016) applied method of moments with GEV distribution on data of pollutants ozone amount. Further, Khan and King (2013) estimated the parameters of modified Weibull distribution in which the estimation of parameters was done by ML estimation as well as least squares estimation.
Moreover, Alkasasbeh and Raqab (2009) estimated shape and scale parameters of GLD by the methods of ML, moments, least squares, weighted least squares, L-moment. They conducted that the best fit was obtained by least squares estimations. Ahmed (2018) made a comparison between the estimation of GLD extreme quantiles by moments and probability weighted moments methods, which conducted that moments method gave a higher estimation. Later on, El Genidy and Abd El-Rahman (2019) presented nested percentiles algorithm as a higher accurate estimator with Weibull distribution. Many other studies, that made by Ahmed et all. (2012); El Genidy (2012); Agboola et al. (2018) and Wasinrat et al. (2013), dealt with ML method as a good estimator with various distributions.

Overall, ML is considered a common and widely used estimator, even though it is quite hardly applied compared to AOPR. However, AOPR derived satisfactory results. It deals with too many values (200) of actual data, unlike some other methods which deal with limited values just as moments method. Also, AOPR combines the percentiles values and two impactful central tendency measures like mean and median, while nested percentiles algorithm solves only the percentiles equations. As documented, AOPR is an appropriate estimator for any distribution that has three parameters.

Since the goodness of fit test is the main measure of the accuracy of any estimator, there have been many developments on statistical tests. For example, Razali and Yap (2011) had ranked Anderson-Darling test second after Shapiro-Wilk test in a study compared between certain statistical tests. In another study, four goodness-of-fit tests, including Anderson-Darling test, were carried out to fit WD after predicting the parameters by Poudel and Cao (2013). Likewise, Shin et al. (2012) conducted modified Anderson–Darling test to have the highest efficiency with GLD and GEV distribution compared to other traditional tests such as Chi-Square, Kolmogorov–Smirnov, and Cramer Von Mises tests. It was used to fit GEV distribution and GLD with annual maximum rainfall data. Selectively, in this study, Anderson-Darling test was applied for fitting AOPR estimator with the analyzed data set. It has been concluded that AOPR is a suitable estimator which is easy to apply and gives significant results.

Finally, it's recommended to simulate AOPR on other three-parameter probability distributions, such as gamma, log-normal and GEV distributions with other data sets in different fields. Programmers can implement the algorithm to generalize its usage in the statistical analysis.

5 Conclusion

This study suggested a method, called Algorithm of percentile roots, that works as a three parameters estimator. This algorithm proved its ability to compete the other known estimation methods. As resulted from Anderson-Darling test, AOPR’s outcomes were quite better suited than software's results such as EasyFit program. AOPR dealt with percentile equations along with some important measures of the central tendency such as mean and median, which gave the results more accuracy and realism. It can be applied to estimate any other three-parameter distributions in relevant fields.

Acknowledgment
The authors are thankful of Australian Bureau of Meteorology for providing online climate data of Queensland, Australia. They also thank contributing researchers in the field of study.

References


