Choose Best Formula for Lindley Distribution for Modeling of Rainfall Data in Iraq in 2020

Hanaa S. Mohammed Shebib  
hanas.mohammed@uokufa.edu.iq  
Department of Statistics - College of Administration and Economics - University of Kufa, AL-Najaf, Iraq.

Zahraa K. Jaafar  
stat.post09@qu.edu.iq  
General Directorate of Al-Qadisiyah Education, Al-Qadisiyah, Iraq.

Bashar K. Ali  
bashar.alhallaq@mizan.edu.iq  
Department of Planning and Human Resources - Division of Health and Vital Statistics - Babylon Health Directorate - Ministry of Health & Environment, Babylon, Iraq

Abstract

In this article, we will used vary formulas of Lindley distribution which are two parameter Lindley distribution under mixture parameter \( p = \frac{b}{b + a} \), two parameter Lindley distribution under mixture parameter \( P = \frac{b}{b + a + 1} \), and two parameter Lindley distribution under mixture parameter \( p = \frac{b}{b + a} \) by shrinkage negative exponential and negative gamma distribution for fitting rainfall data in Iraq by estimating the parameters of these distribution by using maximum likelihood method and comparing among distributions by using Akaike's test, Bayes Akaike's test, Consistent Bayes Akaike's test, and Hannan-Quinn information Criterion. We concluded that the two parameter Lindley distribution for fitting rainfall data is and two parameter Lindley distribution under parameter \( p = \frac{b}{b + a} \) by shrinkage negative exponential and negative gamma distribution. There is a high probability of a decrease in the amount of rainfall in Iraq within the framework of the fitting distribution.

1. Introduction

The study of these phenomena and the knowledge of the probability distributions that follow them are crucial to know and diagnose their random behavior. The study of these phenomena is related to the study of probability distributions and what possible results to it. Statistics sheds light on the study of vital phenomena with random behavior that have an important impact on the life of the organism and society. In order to understand the behavior of this natural phenomenon, as well as to calculate the precise probability of precipitation and plan for the completion of projects and services that lessen the effects of precipitation, it is very helpful to study the statistical distribution of the precipitation series. Rainfall is one of the most crucial climatic variables since it provides residential, agricultural, and groundwater supplies. The findings of frequency analysis are used to
estimate typical hydrologic engineering designs, such as dam height, design flows, etc. Hydrologic frequency analysis is a method for evaluating the likelihood of hydrologic occurrences, which are averaged out from a statistical point of view. Numerous studies on the statistical distribution of rainfall have been conducted. In (2016) P. P. Dabral and et al. used the time series models to predict with climate in india[1]. In (2020) S. Kaur, and M. Rakshit used the Seasonal and Periodic Autoregressive Time Series Models Used for Forecasting Analysis of Rainfall Data[2]. In (2019) Y. Sun, D. Wendi, and D.E. Kim Deriving intensity–duration–frequency (IDF) curves using downscaled in situ rainfall assimilated with remote sensing data[3]. In (2018) M. Ombadi used Intensity-Duration-Frequency IDF rainfall curves, for data series and climate projection in African cities[4]. In (2013) Y. A. Hamaamin Developing of Rainfall Intensity-Duration-Frequency Model for Sulaimani City[6]. In (2008) Ghitany, M. E., Atieh, B., & Nadarajah, S., use simulation to evaluation Lindley distribution and its application[7]. In (2016) Shanker, R & Sharma, S. DERIVE two parameter Lindley distribution and Its Applications to model Lifetime data[8]. In (2017) Shanker, R., Kamlesh, K. K., & Fesshaye, H develop a two parameter lindley distribution: Its properties and applications[9]. R&Sharma,S (2016) discussed the two-parameter Lindley distribution (TPLD), they extracted the mathematical properties of the distribution such as the moment generating function, mean deviations, ordered statistics, Lorenze & Bonferroni curves, Renyi entropy function and stress-durability reliability [11]. In 2016, Amin M. T., Rizwan M., and Alazba A. A. utilized normal, log-normal, log-Pearson type-III, and the Gumbel max distribution to rainfall data in Pakistan. The normal distribution was discovered to be the best-fitting probability distribution based on the results of the goodness of fit tests [14]. Lindley, 1958 presented Lindley distribution, which was named after him, and studied its properties and its relationship to other life time distributions, and its role in modeling waiting times data as an alternative to the exponential distribution and the gamma distribution where estimated the parameters of the model using the method of maximum Likelihood method and the method of moments, and compared the Lindley distribution with two parameters with the exponential distribution and the Lindley distribution with one parameter from set of real data using the (Kolmogorov-Smirnov, AIC, BIC, AICc, -2lnL)[15]. In this paper we used vary formulas of Lindley distribution which are two parameter Lindley distribution under mixture parameter $p = \frac{b}{b+a}$, two parameter Lindley distribution under mixture parameter $P = \frac{ba}{ba+1}$, and two parameter Lindley distribution under mixture parameter $p = \frac{b}{b+a}$ by shrinkage negative exponential and negative gamma distribution for fitting rainfall data in Iraq by estimating the parameters of these distribution by using maximum likelihood method, Length-biased Moments Method, and percentiles methods, and comparing among distributions by using Akaike's test, Bayes Akaike's test, Consistent Bayes Akaike's test, and Hannan-Quinn information Criterion.

2. Two parameter Lindley distribution formula 1(TPLD1)
The two parameters Lindley distribution was proposed by (Shanker et al, 2013) as a model for the study of environment studies. It is one of the continuous probability distributions resulting from mixing the Gamma (2,b) distribution with the Exponential (b) distribution using the following mixing formula:[10]

$$f(t; a, b) = p f_1(t) + (1 - p) f_2(t)$$  \hspace{1cm} (1)

Where $f_1(t)$ is the probability density function of the exponential distribution with the following form:

$$f_1(t) = be^{-bt} \quad ; t > 0$$  \hspace{1cm} (2)

and that $f_2(t)$ is the probability density function of the (2,b) gamma distribution with the following form:

$$f_2(t) = b^2te^{-bt} \quad ; t > 0$$  \hspace{1cm} (3)

Then,
\[ f(t; a, b) = pbe^{-bt} + (1 - p)b^2te^{-bt} \]  \hspace{1cm} (4)

Assuming that the mixing parameter is \( p = \frac{b}{b+a} \)

The probability density function of a Lindley distribution can be defined with two parameters in the following formula:

\[ f(t; a, b) = \frac{b^2}{b+a}(1 + at)e^{-bt}; t > 0, b > 0, a > -b \]  \hspace{1cm} (5)

where \( b \) is the scale parameter, \( a \) is the shape parameter.

If \( a = 1 \), we reduce the distribution to the Lindley distribution with one parameter. And if \( a = 0 \) we will get the distribution of the exponential distribution with parameter \( b \).

The cumulative distribution function (CDF) of the distribution is known by the following formula:

\[ F(t; b) = 1 - \left[ \frac{b + a + abt}{(b + a)} \right] e^{-bt}; t > 0, b > 0, a > -b \]  \hspace{1cm} (6)

Figure (1): The behavior of the probability density function curve for the distribution of (TPLD1)

Figure (2): The behavior of the Cumulative probability density function curve for the distribution of (TPLD1)

3. Two parameter Lindley distribution formula 2(TPLD2):

(Shanker, R & Sharma, S, 2016) [8] proposed a new developed formula for the Lindley distribution with the two parameters by assuming the mixing parameter (P) and defined by the
Choose Best Formula for Lindley Distribution for Modeling of Rainfall Data in Iraq in 2020

Hanaa S. Mohammed Shebib; Zahraa k. jaafar and Bashar K. Ali

formula \( P = \frac{ba}{ba+1} \) as shown in the following: [11],[12]. It was possible to reach a new formula for the Lindley distribution with two parameters, as shown below:

\[
f(t; a, b) = \frac{ab}{ab+1} b e^{-bt} + \left(1 - \frac{ab}{ab+1}\right) b^2 t e^{-bt}
= \frac{b^2}{ab+1} (a + t) e^{-bt}; t > 0, b > 0, ab > -1
\]

(7)

The formula (7) represents the probability density function for the Lindley distribution with two parameters, the most common in recent times. Where \( a \) shape parameter , \( b \) scale parameter.

Figure (3): The behavior of the probability density function curve for the distribution of (TPLD2)

The cumulative distribution function (CDF) of the two-parameter Lindley distribution (TPLD2) is:

\[
F(t) = Pr(T \leq t)
= 1 - \left[\frac{1+ab+bt}{ab+1}\right] e^{-bt}; t > 0, b > 0, ab > -1
\]

(8)

Figure (4): The behavior of the cumulative probability density function curve for the distribution of (TPLD2)
4. **A new Two parameter Lindley distribution formula 3 (NTPLD) :**[13]

We proposed another formula for the Lindley distribution with the two parameters by assuming the same mixing formula previously defined by $\frac{ba}{ba+1}$, but assuming that the probability density function in the exponential distribution is defined by the formula:

$$f_3(t) = \frac{1}{b} e^{-\frac{t}{b}}, \quad t > 0$$

(9)

And the probability density function of the gamma distribution with the two parameters $(2, b)$ is defined by the following formula:

$$f_4(t) = \frac{1}{b^2} t e^{-\frac{t}{b}}, \quad t > 0$$

(10)

An alternative formula for the Lindley distribution with the two parameters can be proposed, which is shown in the following formula:

$$f_{NTPLD}(t; a, b) = \frac{1}{ab+1} \left( a + \frac{t}{b^2} \right) e^{-\frac{t}{b}}, \quad t > 0$$

(11)

Where $a$ Shape Parameter, $b$ Scale Parameter.

---

**Figure (5): The behavior of the probability density function curve for the distribution of (NTPLD)**

**Figure (6): The behavior of the cumulative probability density function curve for the distribution of (NTPLD)**

Cumulative Distribution Function of Distribution (NTPLD):
Choose Best Formula for Lindley Distribution for Modeling of Rainfall Data in Iraq in 2020

Hanaa S. Mohammed Shebib; Zahraa k. jaafar and Bashar K. Ali

\[ F_{NTPLD}(t) = 1 - \frac{[b + t + ab^2]}{b(ab + 1)} e^{-\frac{t}{b}} \]  

(12)

5. Maximum Likelihood Method:
If we have a random sample \((t_1, t_2, ..., t_n)\) from the (TPLD) distribution with a probability density function of TPLD 1 TPLD2 , NTPLD , then the likelihood function can be written in the following forms:

\[ L = \left(\frac{b^2}{b+a}\right)^n \prod_{i=1}^{n} \left(1 + at\right) e^{-b \sum_{i=1}^{n} t_i} \]  

(13)

\[ L = \left(\frac{b^2}{ab+1}\right)^n \prod_{i=1}^{n} (a + t) e^{-b \sum_{i=1}^{n} t_i} \]  

(14)

\[ L = \left(\frac{1}{ab+1}\right)^n \prod_{i=1}^{n} \left(\frac{a + t_i}{b^2}\right) e^{-\frac{\sum_{i=1}^{n} t_i}{b}} \]  

(15)

By taking the natural logarithm of both sides of the equations (13, 14, 15), it results:

\[ \ln L = 2n \ln b - n \ln (b + a) + \sum_{i=1}^{n} \ln(1 + at) - b \sum_{i=1}^{n} t_i \]  

(16)

\[ \ln L = 2n \ln b - n \ln(ab + 1) + \sum_{i=1}^{n} ln(a + t) - b \sum_{i=1}^{n} t_i \]  

(17)

\[ \ln L = n \ln \left(\frac{1}{ab + 1}\right) + \sum_{i=1}^{n} \ln \left(\frac{a + t_i}{b^2}\right) - \frac{\sum_{i=1}^{n} t_i}{b} \]  

(18)

To obtain the estimators of \(b\) and \(a\), we take the first partial derivative of the formulas (16, 17, 18) with respect to \((b, a)\) and set it equal to zero, we get the estimators of ML.

6. Applied side
We used data representing rainfall amounts in the cities of Baghdad, Mosul, Rutba, and Basrah from Ministry of Transport / General Authority for Meteorology and Seismic Monitoring for the year 2020 as in the table (1), the data were fitted by using the chi-square test under Lindley distribution where showed that significant all formulas of Lindley distribution under real data. Three models were used for the default values of the distribution, which are \((a=0.5, 1.6, 4.5)\) and \((b=0.5, 1.9, 2.5)\) the results conducted by using Matlab 2022b.
Table (1): Monthly averages of the amount of rainfall (ml) by provinces and month for 2020

<table>
<thead>
<tr>
<th>City</th>
<th>Basrah</th>
<th>Rutha</th>
<th>Baghdad</th>
<th>Mosul</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2020</td>
<td>2020</td>
<td>2020</td>
<td>2020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Monthly Rate</td>
<td>The Monthly Rate</td>
<td>The Monthly Rate</td>
<td>The Monthly Rate</td>
<td></td>
</tr>
<tr>
<td>Baghda</td>
<td>26.2</td>
<td>1.9</td>
<td>13.9</td>
<td>19.0</td>
<td>January</td>
</tr>
<tr>
<td>Mosul</td>
<td>17.6</td>
<td>4.7</td>
<td>23.2</td>
<td>10.8</td>
<td>February</td>
</tr>
<tr>
<td>Rutba</td>
<td>18.3</td>
<td>1.0</td>
<td>13.5</td>
<td>25.3</td>
<td>March</td>
</tr>
<tr>
<td>Basrah</td>
<td>12.0</td>
<td>0.0</td>
<td>8.7</td>
<td>3.0</td>
<td>April</td>
</tr>
<tr>
<td></td>
<td>3.7</td>
<td>0.5</td>
<td>7.8</td>
<td>0.0</td>
<td>May</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>June</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>July</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>August</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>September</td>
</tr>
<tr>
<td></td>
<td>6.4</td>
<td>0.0</td>
<td>12.7</td>
<td>0.0</td>
<td>October</td>
</tr>
<tr>
<td></td>
<td>18.9</td>
<td>63.2</td>
<td>18.0</td>
<td>61.2</td>
<td>November</td>
</tr>
<tr>
<td></td>
<td>24.4</td>
<td>12.7</td>
<td>11.5</td>
<td>19.0</td>
<td>December</td>
</tr>
<tr>
<td>Total</td>
<td>127.5</td>
<td>84.0</td>
<td>110.1</td>
<td>138.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>124.5</td>
<td>155.6</td>
<td>353.7</td>
<td>396.2</td>
<td></td>
</tr>
</tbody>
</table>

Table (2): Chi-square sets

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$\chi^2$</th>
<th>$\chi_t^2$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPLD1</td>
<td>0.941</td>
<td></td>
<td>0.511</td>
</tr>
<tr>
<td>TPLD2</td>
<td>0.831</td>
<td>7.82</td>
<td>0.511</td>
</tr>
<tr>
<td>NTPLD</td>
<td>0.564</td>
<td></td>
<td>0.97</td>
</tr>
</tbody>
</table>

Table (3): Compression criteria for Lindley Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>AIC</th>
<th>AICc</th>
<th>BIC</th>
<th>HQIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPLD1</td>
<td>165.89</td>
<td>166.82</td>
<td>165.89</td>
<td>12.69</td>
</tr>
<tr>
<td>TPLD2</td>
<td>165.66</td>
<td>166.41</td>
<td>165.55</td>
<td>12.33</td>
</tr>
<tr>
<td>NTPLD</td>
<td>164.22</td>
<td>165.11</td>
<td>165.13</td>
<td>12.12</td>
</tr>
</tbody>
</table>

Where AIC is Akaike's Information Criteria, AICc is corrected Akaike's Information Criteria, BIC Bayesian Information Criteria and HQIC is Hannan Queen Information Criteria which is criteria to comparing with distribution and the less criteria is best.

Table (4): Parameter estimation and Mean square error for parameters for formulas of Lindley Distributions

<table>
<thead>
<tr>
<th>City</th>
<th>Distribution</th>
<th>Parameters estimation</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baghdad</td>
<td>TPLD1</td>
<td>a=0.67 b=0.72</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>TPLD2</td>
<td>a=0.66 b=0.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NTPLD</td>
<td>a=0.57 b=0.59</td>
<td></td>
</tr>
<tr>
<td>Mosul</td>
<td>TPLD1</td>
<td>a=0.74 b=0.65</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>TPLD2</td>
<td>a=0.66 b=0.32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NTPLD</td>
<td>a=0.54 b=0.52</td>
<td></td>
</tr>
<tr>
<td>Rutba</td>
<td>TPLD1</td>
<td>a=0.63 b=0.59</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>TPLD2</td>
<td>a=0.63 b=0.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NTPLD</td>
<td>a=0.54 b=0.59</td>
<td></td>
</tr>
<tr>
<td>Basrah</td>
<td>TPLD1</td>
<td>a=0.56 b=0.59</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>TPLD2</td>
<td>a=0.55 b=0.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NTPLD</td>
<td>a=0.51 b=0.53</td>
<td></td>
</tr>
</tbody>
</table>

553
Choose Best Formula for Lindley Distribution for Modeling of Rainfall Data in Iraq in 2020

Hanaa S. Mohammed Shebib; Zahraa K. Jaafar and Bashar K. Ali

The test results are in table (2) showed that all distributions used were fitted for the real data, but distribution NTPLD is the most fitting because it achieved the lowest values of the criteria. Also, a comparison was made between the distributions by using Akaike's test, Bayes Akaike's test, Consistent Bayes Akaike's test, and Hannan-Quinn information Criterion as in table (3), table (4) showed the parameter estimates under Maximum Likelihood method for real data which are as close to the default parameters. the pdf values under real data for the cities under all formulas of Lindley distributions.

7. Conclusions

The results was showed an increase in the values of the probability density function in the Mosul city and a decrease in the cities (Baghdad - Rutba - Basrah). And Basrah city recorded the lowest probability values for pdf. There is a high probability of a decrease in the amount of rainfall in Iraq within the framework of the fitting distribution. The results showed that the proposed Lindley distribution is more suitable for the amounts of rainfall, and thus the results given by this distribution are more accurate in estimating the probability of rainfall.
8. Recommendations

We recommend that the data should be adapted to any phenomenon before performing the statistical distribution process. As well as the need to use the Lindley distribution to model environmental phenomena, especially the amounts of rainfall, and we also recommend using the fuzzy principle in dealing with such data to reach more accurate results in the estimate.

References

Choose Best Formula for Lindley Distribution for Modeling of Rainfall Data in Iraq in 2020

Hanaa S. Mohammed Shebib; Zahraa k. jaafar and Bashar K. Ali

PISSN: (1681-6870); EISSN: (2790-2293)

Available online at: https://www.jrucs.iq

JRUCS
Journal of AL-Rafidain University College for Sciences

.Abstract

In this study, different formulas of the Lindley distribution were used under the mixed parameter, namely the two-parameter Lindley distribution, the two-parameter Lindley distribution under the mixed parameter, and the distribution resulting from mixing the negative exponential distribution and the inverse gamma distribution. The aim was to determine which distribution is the best suited for modeling rainfall data in Iraq in 2020, using the maximum likelihood method and comparing the distributions using the Akaike, Bayes Akaike, Consistent Bayes Akaike, and Hannan-Quinn information criteria. The research findings indicate that the two-parameter Lindley distribution under the mixed parameter is the most suitable for modeling rainfall data in Iraq, with a high probability of a decrease in rainfall in the future.

Key words: Lindley distribution, model fitting, rainfall data, maximum likelihood method, Akaike test.

Zahraa K. Jaafar
Zahraa.kjaafar@mizan.edu.iq
Stat. and Econ. Dept., College of Business Administration and Economics, University of Kufa, Kufa, Iraq

Bashar K. Ali
bashar.alhallaq@mizan.edu.iq
Planning and Development of Resources, Health and Vital Statistics Branch, Health Directorate of Babylon, Ministry of Health, Babylon, Iraq

References