Journal of AL-Rafidain University College for Sciences (2023); Issue 54; 547 - 556

AL- Rafidain University College

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

Journal of AL-Rafidain University College for Sciences

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

JRUCS

Journal of AL-Rafidain University College for Sciences

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

Hanaa S. Mohammed Shebib	Zahraa K. Jaafar
hanas.mohammed@uokufa.edu.iq	stat.post09@qu.edu.iq
Department of Statistics - College of Administration and Economics - University of Kufa, AL-Najaf, Iraq.	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

Article Information

Article History:
Received: January, 1, 2023
Accepted: March, 3, 2023
Available Online: December, 31, 2023

Keywords:
Lindley Distribution, Fitting Data, Maximum Likelihood Method, Akaike's Test.

Correspondence:
Zahraa K. Jaafar
stat.post09@qu.edu.iq

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* $P=\frac{ba}{ba+1}$*, 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 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 mixture 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.*

https://doi.org/10.55562/jrucs.v54i1.623

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-Samirnov, 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]

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

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

$$f_1(t) = be^{-bt} \qquad ; t > 0 \tag{2}$$

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

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

Then,

$$f(t;a,b) = pbe^{-bt} + (1-p)b^2te^{-bt} \quad (4)$$

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:

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

where is b scale parameter, a 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:

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

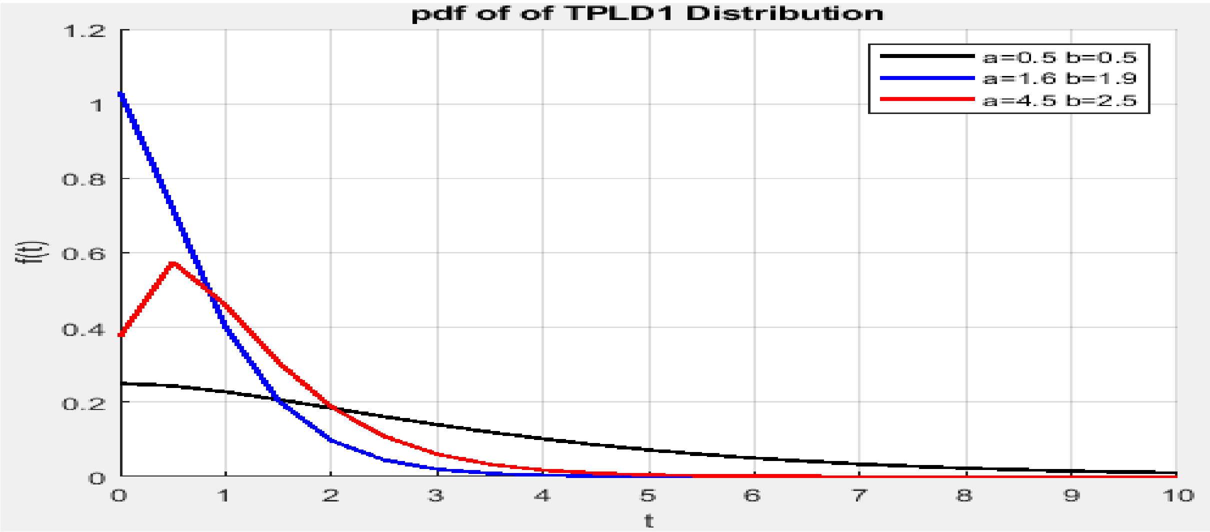

Figure (1): The behavior of the probability density function curve for the distribution of (TPLD1)

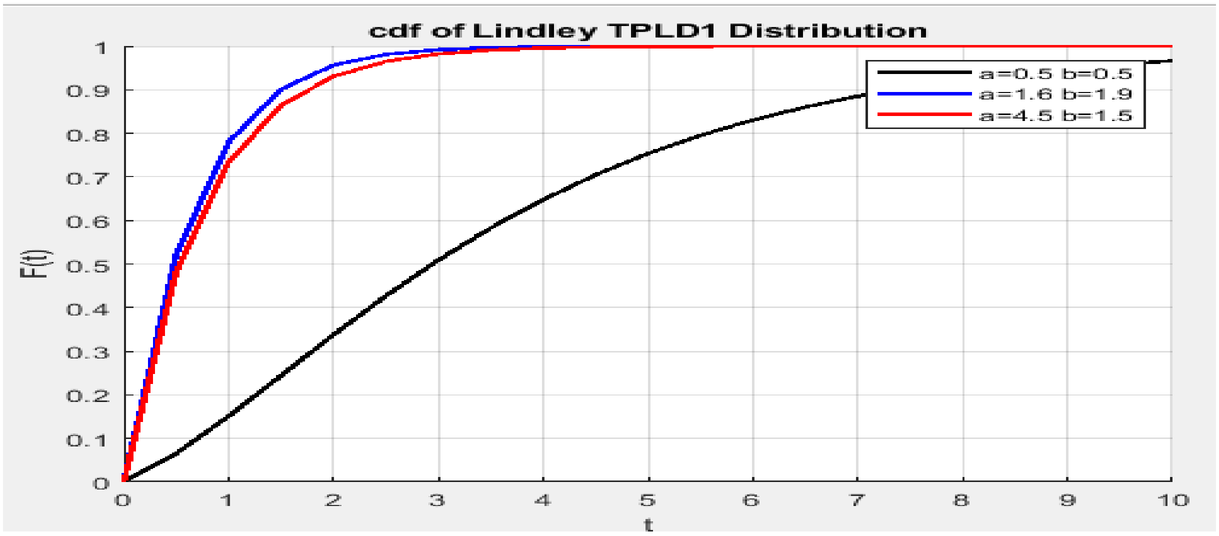

Figure (2): The behavior of the Cumulative probability density function curve for the distribution of (TPLD1)

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 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:

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

$$= \frac{b^2}{ab+1} (a+t)e^{-bt} \quad ; t > 0, b > 0, ab > -1 \qquad (7)$$

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.

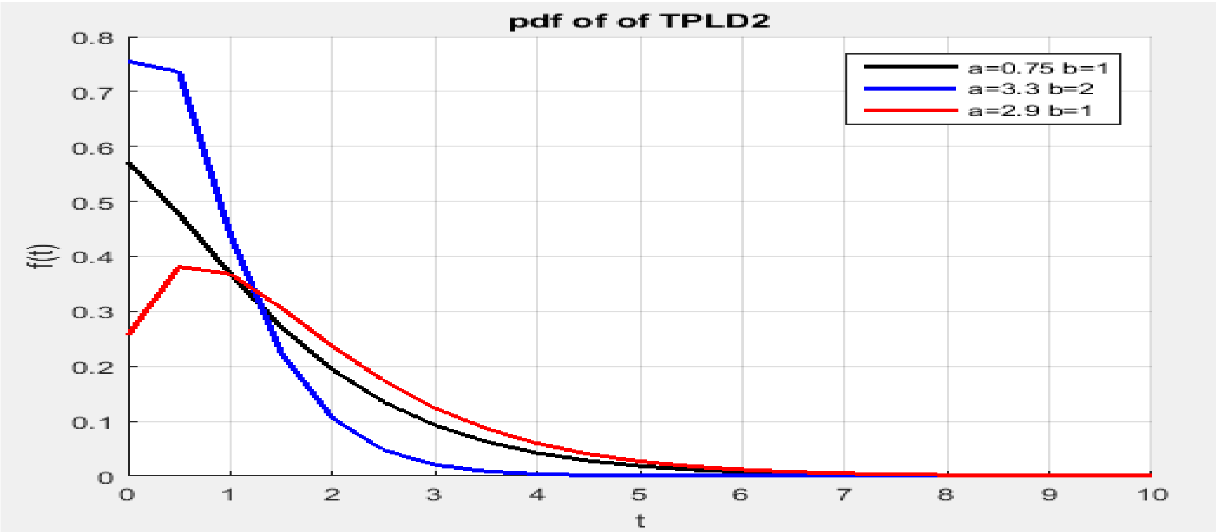

Figure (3): The behavior of the probability density function curve for the distribution of (TPLD2)

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

$$F(t) = P_r(T \leq t)$$

$$= 1 - \left[\frac{1+ab+bt}{ab+1}\right] e^{-bt} \quad ; t > 0, b > 0, ab > -1 \qquad (8)$$

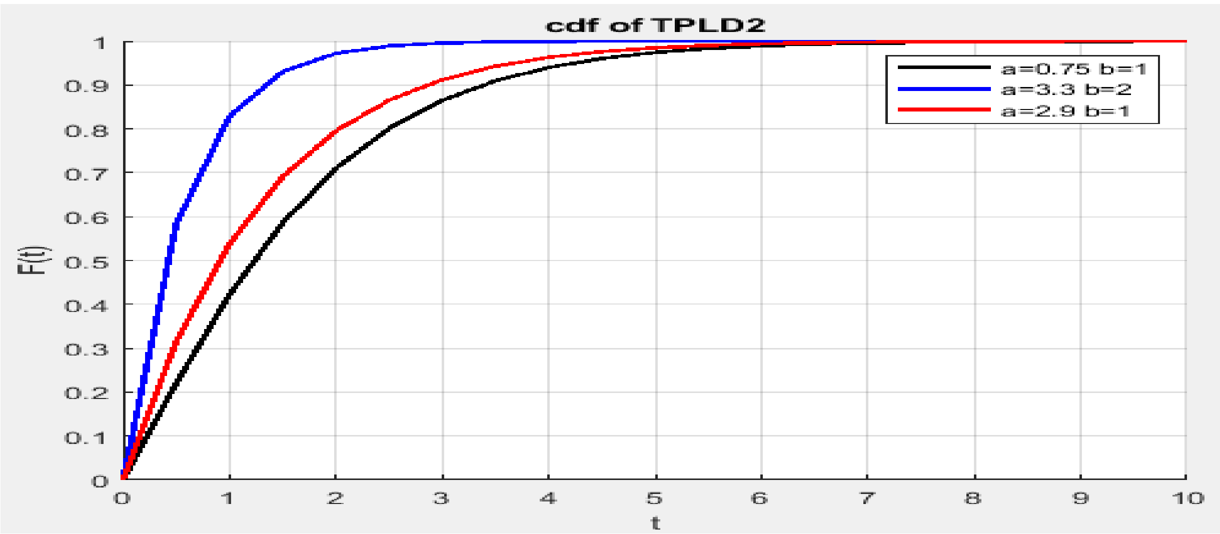

Figure (4): The behavior of the cumulative probability density function curve for the distribution of (TPLD2)

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:

$$f_3(t) = \frac{1}{b} e^{-\frac{t}{b}} , \quad t > 0 \tag{9}$$

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

$$f_4(t) = \frac{1}{b^2} t \ e^{-\frac{t}{b}} \quad , \quad t > 0 \tag{10}$$

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

$$f_{NTPLD}(t; a, b) = \frac{1}{ab+1} \left(a + \frac{t}{b^2}\right) e^{-\frac{t}{b}} \quad , \ t > 0 \tag{11}$$

Where a Shape Parameter , b Scale Parameter

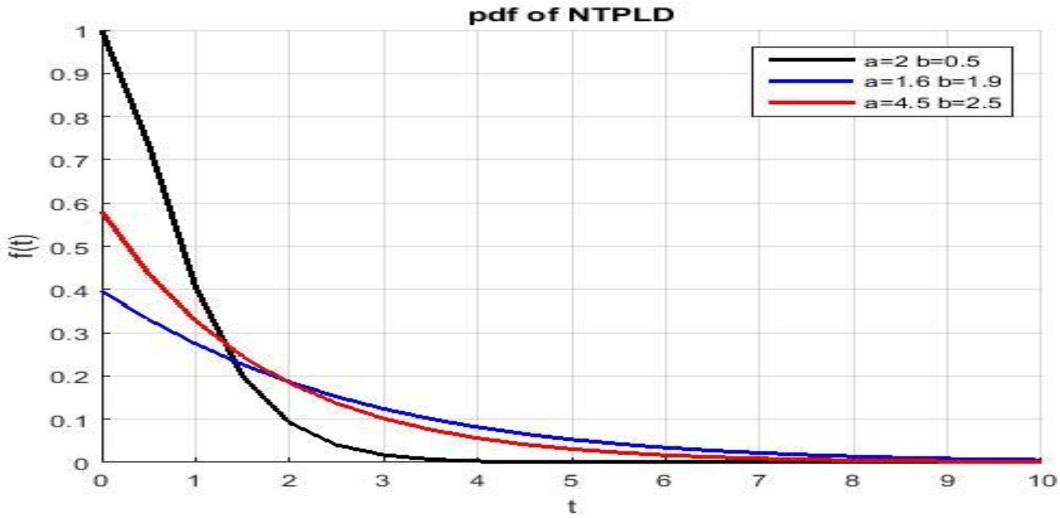

Figure (5): The behavior of the probability density function curve for the distribution of (NTPLD)

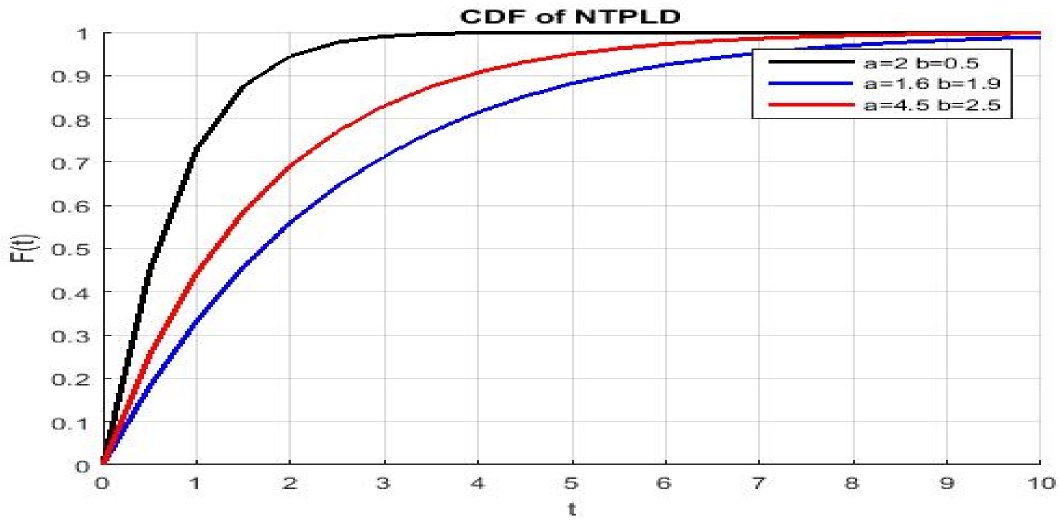

Figure (6): The behavior of the cumulative probability density function curve for the distribution of (NTPLD)

Cumulative Distribution Function of Distribution (NTPLD):

$$F_{NTPLD}(t) = 1 - \left[\frac{b+t+ab^2}{b(ab+1)}\right] e^{-\frac{t}{b}} \quad (12)$$

5. Maximum Likelihood Method:

If we have a random sample $(\mathbf{t_1}, \mathbf{t_2}, \dots, \mathbf{t_n})$ 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:

$$L = \left(\frac{b^2}{b+a}\right)^n \prod_{i=1}^{n} \left((1+at)\right) e^{-b\sum_{i=1}^{n} t_i} \quad (13)$$

$$L = \left(\frac{b^2}{ab+1}\right)^n \prod_{i=1}^{n} (a+t)\, e^{-b\sum_{i=1}^{n} t_i} \quad (14)$$

$$L = \left(\frac{1}{ab+1}\right)^n \prod_{i=1}^{n} \left(a + \frac{t_i}{b^2}\right) e^{-\frac{\sum_{i=1}^{n} t_i}{b}} \quad (15)$$

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

$$\ln L = 2n \ln b - n\ln(b+a) + \sum_{i=1}^{n} ln(1+at) - b\sum_{i=1}^{n} t_i \quad (16)$$

$$\ln L = 2n \ln b - n\ln(ab+1) + \sum_{i=1}^{n} ln(a+t) - b\sum_{i=1}^{n} t_i \quad (17)$$

$$\ln L = n \ln\left(\frac{1}{ab+1}\right) + \sum_{i=1}^{n} ln\left(a + \frac{t_i}{b^2}\right) - \frac{\sum_{i=1}^{n} t_i}{b} \quad (18)$$

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

Basrah		Rutba		Baghdad		Mosul		Month
The Monthly Rate	2020	The Monthly Rate	2020	The Monthly Rate	2020	The Monthly Rate	2020	
26.2	1.9	13.9	19.0	24.3	36.1	61.3	85.3	January
17.6	4.7	23.2	10.8	17.1	6.4	54.7	65.1	February
18.3	1.0	13.5	25.3	16.2	22.9	58.7	120.5	March
12.0	0.0	8.7	3.0	15.6	3.2	45.6	60.8	April
3.7	0.5	7.8	0.0	3.1	0.0	17.0	3.6	May
0.0	0.0	0.1	0.0	0.0	0.0	1.2	0.0	June
0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0	July
0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	August
0.0	0.0	0.5	0.0	0.1	0.0	0.6	0.0	September
6.4	0.0	12.7	0.0	7.5	0.0	12.4	0.0	October
18.9	63.2	18.0	61.2	24.0	84.2	40.6	22.0	November
24.4	12.7	11.5	19.0	16.6	2.8	61.4	38.9	December
127.5	**84.0**	**110.1**	**138.3**	**124.5**	**155.6**	**353.7**	**396.2**	**Total**

Table (2): Chi-square sets

Distribution	χ^2_c	χ^2_t	Sig.
TPLD1	0.941	7.82	0.511
TPLD2	0.831		0.511
NTPLD	0.564		0.97

Table (3): Compression criteria for Lindley Distributions

Distribution	AIC	AICc	BIC	HQIC
TPLD1	165.89	166.82	165.89	12.69
TPLD2	165.66	166.41	165.55	12.33
NTPLD	**164.22**	**165.11**	**165.13**	**12.12**

Where AIC is Akiaki's Information Criteria , AICc is corrected Akiaki'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

City	Distribution	Parameters estimation	MSE
Baghdad	TPLD1	a=0.67 b=0.72	0.038
	TPLD2	a=0.66 b=0.66	
	NTPLD	a=0.57 b=0.59	
Mosul	TPLD1	a=0.74 b=0.65	0.037
	TPLD2	a=0.66 b=0.32	
	NTPLD	a=0.54 b=0.52	
Rutba	TPLD1	a=0.63 b=0.59	0.038
	TPLD2	a=0.63 b=0.56	
	NTPLD	a=0.54 b=0.59	
Basrah	TPLD1	a=0.56 b=0.59	0.012
	TPLD2	a=0.55 b=0.56	
	NTPLD	a=0.51 b=0.53	

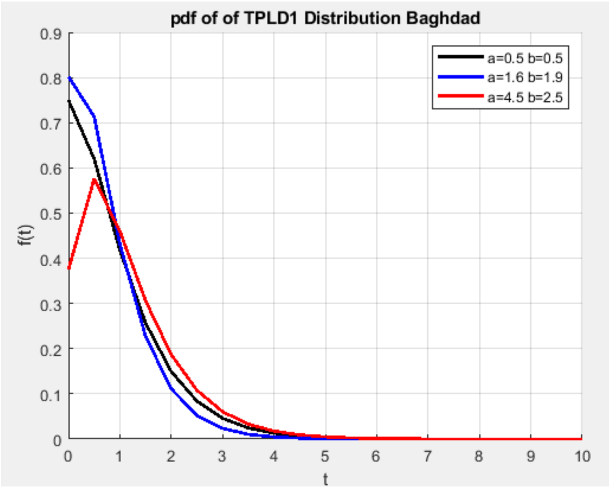

Figure (7): The probability density function curve of TPLD1 under different parameters for rainfall data in Baghdad city

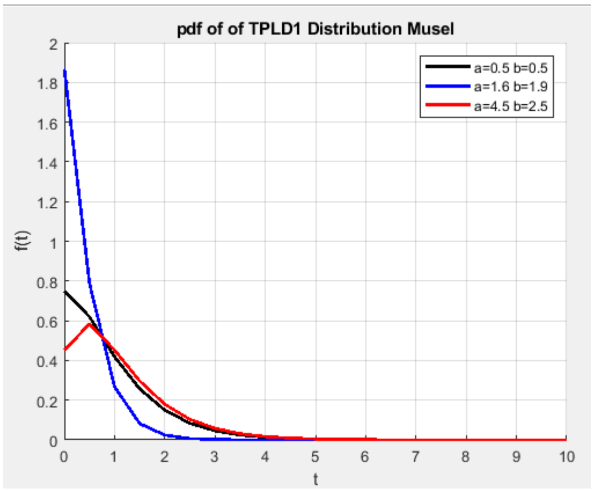

Figure (8): The probability density function curve of TPLD1 under different parameters for rainfall data in Mosul city

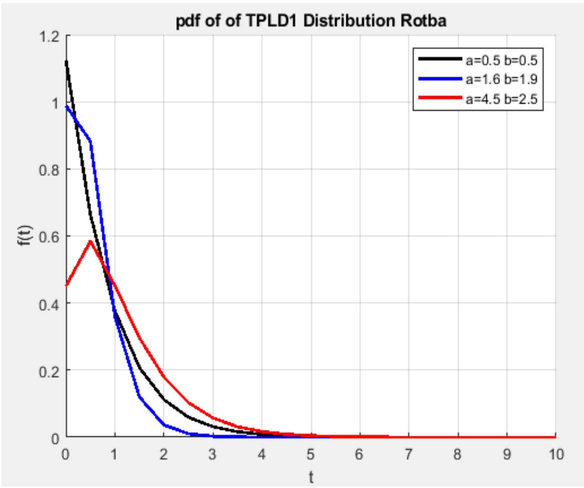

Figure (9): The probability density function curve of TPLD1 under different parameters for rainfall data in Rutba city

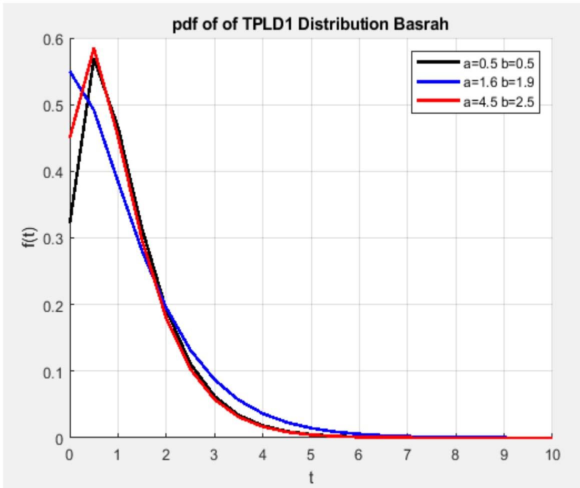

Figure (10): The probability density function curve of TPLD1 under different parameters for rainfall data in Basrah city

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

[1] P. P. Dabral, T. Saring and D. Jhajharia, "Time series models of monthly rainfall and temperature to detect climate change for Jorhat (Assam), India". Global NEST International Journal, Vol. 18 Issue 3, pp. 494-507, 2016.

[2] S. Kaur, and M. Rakshit, "Seasonal and Periodic Autoregressive Time Series Models Used for Forecasting Analysis of Rainfall Data", International Journal of Advanced Research in Engineering and Technology, Vol. 10, Issue: 1, pp. 230-242, 2020.

[3] Y. Sun, D. Wendi, and D.E. Kim, "Deriving intensity–duration–frequency (IDF) curves using downscaled in situ rainfall assimilated with remote sensing data", Geosciences Letters, Vol. 6, Issue 17, 2019.

[4] M. Ombadi, P. Nguyen, S. Sorooshian and K. Hsu, "Developing intensity-duration-frequency (IDF) curves from satellite-based precipitation: methodology and evaluation", Water Resour Res. Vol. 54, pp.7752–7766, 2018.

[5] F. De Paola, M. Giugni and M.E. Topa, "Intensity-Duration-Frequency IDF rainfall curves, for data series and climate projection in African cities", Springer Plus Vol. 3, 133, 2014.

[6] Y. A. Hamaamin, "Developing of Rainfall Intensity-Duration-Frequency Model for Sulaimani City" Journal of Zankoy Sulaimani, Vol. 19, Issue: 3-4, pp.93-101, 2013.

[7] Ghitany, M. E., Atieh, B., & Nadarajah, S., (2008). "Lindley distribution and its application". Mathematics and computers in simulation, Vol. 78, No.4,pp. 493-506.

[8] Shanker, R & Sharma, S, (2016). "On two parameter Lindley distribution and Its Applications to model Lifetime data". Biometrics & Biostatistics International Journal, Vol. 3, No.1,

[9] Shanker, R., Kamlesh, K. K., & Fesshaye, H., (2017). A two parameter lindley distribution: Its properties and applications. Biostatistics and Biometrics Open Access Journal, Vol. 1, No.4, pp.85-90.

[10] Shanker, R., Sharma, S., & Shanker, R., (2013b). "A two-parameter Lindley distribution for modeling waiting and survival times data", Applied Mathematics, Vol.4, No.2.

[11] Shanker, R & Sharma, S, (2016). "On two parameter Lindley distribution and Its Applications to model Lifetime data". Biometrics & Biostatistics International Journal, Vol. 3, No.1, pp.1-8.

[12] Shanker, R., & Mishra, A., (2013a). "A two-parameter Lindley distribution". Statistics in Transition new series, Vol. 14, No.1, pp.45-56.

[13] Zahra Khalled Jafar, Al-Sharoot, Mohammed Habeeb, (2021), "Estimation of the Reliability Function of a two-parameter Lindley distribution For patients infected with COVID-19", Thesis, College of Administration & Economics\ University of Al-Qadisiyah.

[14] M. T. Amin , M. Rizwan and A. A. Alazba, (2016), "A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan", Open Life Sciences, Vol. 11, No. 1, https://doi.org/10.1515/biol-2016-0057

[15] Lindley, D. V., (1958), "Fiducial Distributions and Bayes' Theorem", Journal of the Royal Statistical Society. Series B (Methodological), Vol. 20, No. 1 (1958), pp. 102-107

مجلة كلية الرافدين الجامعة للعلوم (2023)؛ العدد 54؛ 547 - 556

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

مجلة كلية الرافدين الجامعة للعلوم

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

JRUCS

Journal of AL-Rafidain University College for Sciences

AL- Rafidain University College

اختيار أفضل صيغة لتوزيع ليندلي لنمذجة بيانات هطول الأمطار في العراق

هناء سعد محمد شبيب	**زهراء خالد جعفر**
hanas.mohammed@ uokufa.edu.iq	stat.post09@qu.edu.iq
قسم الاقتصاد - كلية الادارة والاقتصاد - جامعة الكوفة، الكوفة، العراق	المديرية العامة لتربية القادسية، القادسية، العراق

بشار خالد علي

bashar.alhallaq@mizan.edu.iq

قسم التخطيط وتنمية الموارد - شعبة الاحصاء الصحي والحياتي - دائرة صحة بابل - وزارة الصحة، بابل، العراق

المستخلص

في هذا البحث ، تم استعمال صيغًا مختلفة لتوزيع Lindley وهما توزيع Lindley ذي المعلمتين في ظل معلمة الخلط $p = \frac{b}{b+a}$ ، وتوزيع Lindleyذي المعلمتين بمعلمة الخلط $P = \frac{ba}{ba+1}$ ، وتوزيع Lindley ذي المعلمتين في ظل معلمة الخلط $p = \frac{b}{b+a}$ والناتج من خلط التوزيع الأسي السالب وتوزيع كاما المعكوس لملائمة بيانات هطول الأمطار في العراق عن طريق تقدير معلمات هذا التوزيع باستعمال طريقة الامكان الاعظم والمقارنة بين التوزيعات باستخدام اختبارات (Akaike ، Bayes Akaike، Consistent Bayes Akaike، Hannan-Quinn information Criterion). تم التوصل إلى أن توزيع Lindley ذي المعلمتين في ظل معلمة الخلط $p = \frac{b}{b+a}$ والناتج من خلط التوزيع الأسي السالب وتوزيع كاما المعكوس اكثر ملائمة لبيانات هطول الأمطار. وهنالك احتمالية عالية لانخفاض كميات الامطار في العراق في اطار التوزيع الملائم.

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث: 2023/1/1

تاريخ قبول البحث: 2023/3/3

تاريخ رفع البحث على الموقع: 2023/12/31

الكلمات المفتاحية:

توزيع ليندلي، ملائمة بيانات، بيانات هطول الامطار، طريقة الامكان الاعظم، اختبار Akaike.

للمراسلة:

زهراء خالد جعفر

stat.post09@qu.edu.iq

https://doi.org/10.55562/jrucs.v54i1.622