Logistic Probability Approach for discharge level data at Sathanur dam, Tamil Nadu, India

B. Polaiah[1], K. Pushpanjali[2], SK.KhadarBabu[3]
[1]*Research Scholar, Department of Statistics, Sri Krishnadevaraya University, Anantapur, Andhra Pradesh*
[2]*Professor, Department of Statistics, Sri Krishnadevaraya University, Anantapur, Andhra Pradesh, India*
[3]*Associate Professor, Department of Mathematics, VIT,Vellore,TN,India*

Abstract - In Water Resource Engineering, choosing a probability distribution for discharge level, rainfall flow and windspeed data is expected a big task for data analysis. Actually, every distribution having the standard parameters, that can be useful to estimate the standard measures of the experiments. In the present article it is proposed to estimate the parameters of the discharge level data, which are collected from the Sathanur Dam,Tiruvannamalai District of Tamil Nadu, India. Using the method of moments and take them to logistic distribution for data analysis. Usually, the logistic distribution is the best fit for discharge level data for finding expected frequencies for the corresponding observed ones. We applied RMSE approach for the validation of the Logistic approach for water level discharge datasets.

Keywords: Probability density function, Gaussian probability Approach, goodness of fit,
Root mean square approach etc.,

I. INTRODUCTION

Logistic distribution plays an important role in Probability theory and Statistical analysis. The Logistic distribution and the S-shaped pattern of its cumulative distribution function and quantile function have been extensively used in many different areas. One of the most common applications is in logistic regression, which is used for modelling categorical dependent variables (yes-no choices or a choice of three or four possibilities) much as standard linear regression is used for modelling continuous variables. And also, logistic regression is used in various fields, including machine learning, most medical fields, and social sciences. Specifically, logistic regression models can be phrased as latent variable models with error variables following a logistic distribution. This phrasing is common in the theory of discrete choice models, where the logistic distribution plays the same role in logistic regression as the normal distribution does in probit regression. Indeed, the logistic and normal distributions have a quite similar shape. However, the logistic distribution has heavier tails, which often increases the robustness of analyses based on it compared with using the normal distribution.

In hydrology the distribution of long duration river discharge and rainfall (e.g., monthly and yearly totals, consisting of the sum of 30 respectively 365 daily values) is often thought to be almost normal according to the central limit theorem. The normal distribution, however, needs a numeric approximation. As the logistic distribution, which can be solved analytically, is similar to the normal distribution, it can be used instead.

The Logistic distribution is popular in demographic and economic modelling because it is similar to the Normal distribution but somewhat more peaked. It does not appear often in risk analysis modelling

II. REVIEW OF LITERATURE:

Sunitha et.al (2019), proposed a method for the estimation of the pre-defined parameters using standard robust statistical technique under the fuzzy environment. Some of the distributions like Gaussian, Log -Gaussian and Gumbel are useful to estimate the parameters for the pre- planned study area. Chitrasen Lairenjam etal (2016), applied standard probability distributions on 25 stations of NER and given best fitted distribution for the time series data. Chin-Yu Lee (2005), fitted the normal, log-normal, extreme value type I distribution and pearson typeIII distribution and log-pearson type III distribution on annual rainfall data collected from 178 stations at chia-Nan and Tainan city. Swami Shivprasad T etal (2013), identified among all probability distributions, the best fitted distribution for rainfall data collected from Satara and Kolhapur districts of western Maharashtra.Oseni. B. Azeez et al (2012), identified the best probability distribution for 30 years of rainfall data collected from Forestry Research Institute of Nigeria using goodness of fit tests.T.Mayooran and A.Laheetharan(2014) , collected the daily rainfall data of 110 years at Meteorology station ,Columbo,SriLanka and propose the best fitted probability distribution using MLE method.S.R .Bakar et al(2008),observed that the best suitable probability distribution for the 35years rainfall data collected from Central Soil and water Conservation Research centre,Kota.TariqMahgoub Mohamad and Abbas Abd Allah Ibrahim(2016), studied about the best fitted probability distribution for every station in Sudan and five distributions fitted for the data, by using statistical test for goodness of fit applied and identified best fitted probability distribution.

III. METHODS AND DISCUSSIONS

Let us consider the data, Zi, i=1,2, 3,…, n observations in a data set, for these values, by using the method of moments, to estimate the parameters of the probability distributions.

At present for present data, enumerate the standard values like mean, standard deviation using the following procedure.
Actually, in method of moments, broadly divided in to two types, one is moments about origin/any number and another is moments about mean, the first one called raw moments and the second one is central moments. In theory of estimation, we have the methods like Maximum likelihood method of estimation and the method of least squares are also plays a pivotal role at the time of data analysis. Out of these, moments method is considered to estimate the pre-defined parameters and which are helpful to explain the complete structure of the data, which are already collected from a study area.
The r th moment about origin is given by

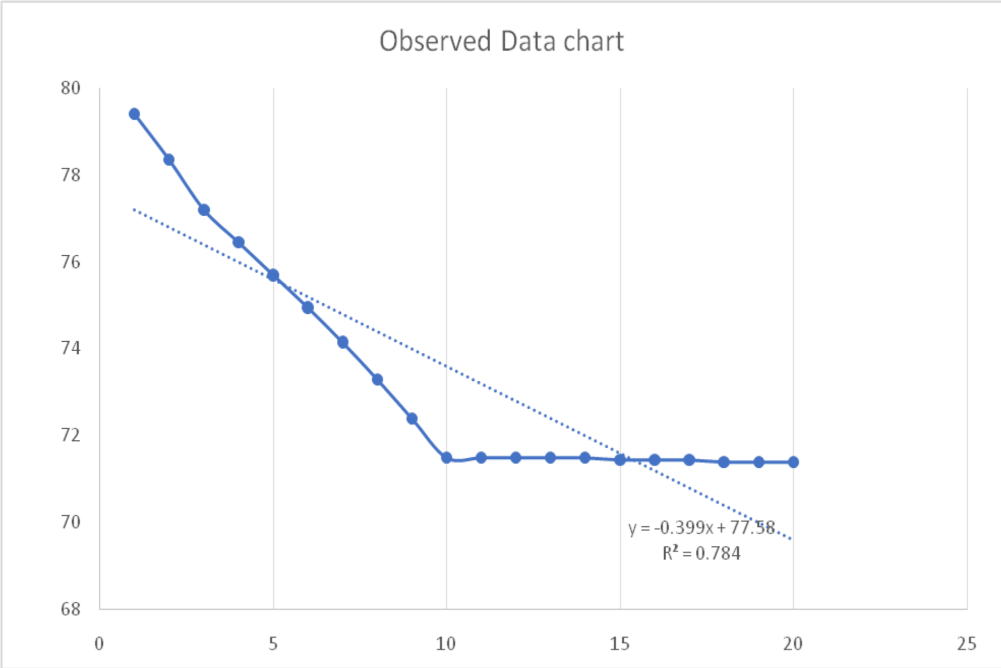
Kr $=\sum_1^n Z_i^r$, if r=1, K1= Mean and Variance = K2 - K12,these are standard ones for logistic probability distribution.
We can find different types of moments by using the assumed /mean values also. The article proposes to the slope and altitude of the water flow at the time of releasing the water for special purposes like for drinking, agriculture etc.,
The probability density function of the logistic probability distribution is as follows:

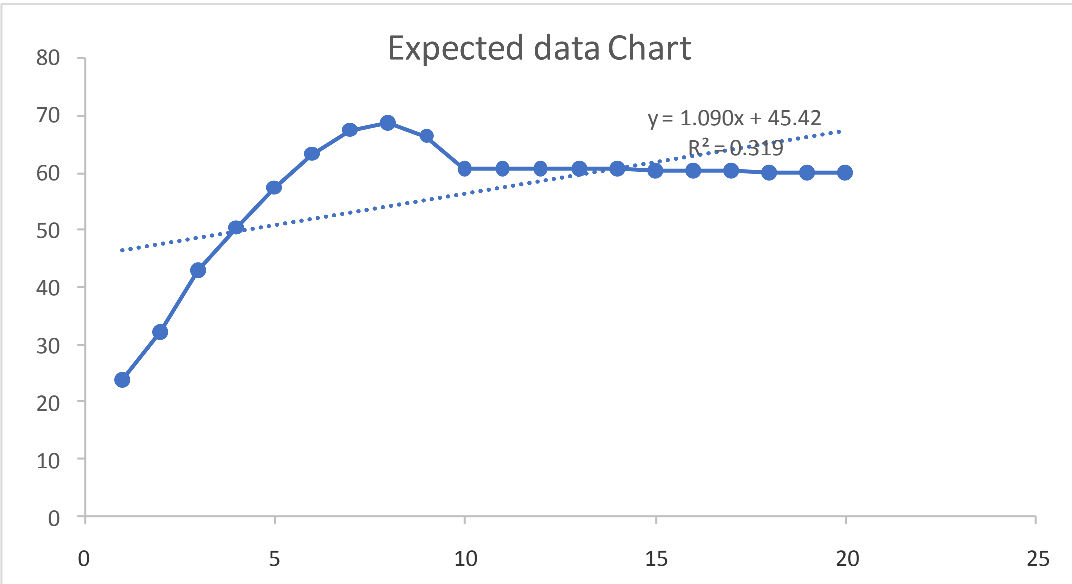
f(y) = $\frac{e^{-y}}{(1+e^{-y})^2}$, $-\infty < y < \infty$,

Let y = $\frac{Z-\alpha}{\beta}$, and it is the standard logistic variate
Mean and variance of the data set are the standard parameters, which are calculated using the method of moments.
By using the above pdf, we find all possible expected frequencies of the data, then apply the chi-square test for validation and also calculated RMSE value for error estimation of the data.

Table: Predicted and Observed Logistic distributed values with parameters

Xi	$y_i = \frac{(x_i - \mu)}{\sigma}$	e^{-y_i}	$(1+e^{-y_i})^2$	$2\sigma(1+e^{-y_i})^2$	$\frac{e^{-y_i}}{2\sigma(1+e^{-y_i})^2}$	$N\frac{e^{-y_i}}{2\sigma(1+e^{-y_i})^2}$
79.4	2.2521	0.1052	1.2214	6.510605937	0.0162	23.6984
78.35	1.8582	0.1560	1.3362	7.122673385	0.0219	32.1212
77.2	1.4267	0.2401	1.5378	8.197355422	0.0293	42.9685
76.45	1.1453	0.3181	1.7375	9.261385225	0.0344	50.3916
75.7	0.8639	0.4215	2.0207	10.7711834	0.0391	57.4091
74.95	0.5825	0.5585	2.4289	12.94715752	0.0431	63.2820
74.15	0.2823	0.7540	3.0766	16.39941629	0.0460	67.4503
73.3	-0.0366	1.0373	4.1504	22.12342819	0.0469	68.7805
72.4	-0.3743	1.4539	6.0217	32.09812768	0.0453	66.4493
71.5	-0.7119	2.0379	9.2291	49.19480117	0.0414	60.7719
71.5	-0.7119	2.0379	9.2291	49.19480117	0.0414	60.7719
71.5	-0.7119	2.0379	9.2291	49.19480117	0.0414	60.7719
71.5	-0.7119	2.0379	9.2291	49.19480117	0.0414	60.7719
71.5	-0.7119	2.0379	9.2291	49.19480117	0.0414	60.7719
71.45	-0.7307	2.0765	9.4651	50.45263681	0.0412	60.3790
71.45	-0.7307	2.0765	9.4651	50.45263681	0.0412	60.3790
71.45	-0.7307	2.0765	9.4651	50.45263681	0.0412	60.3790
71.4	-0.7495	2.1159	9.7086	51.75062415	0.0409	59.9793
71.4	-0.7495	2.1159	9.7086	51.75062415	0.0409	59.9793
71.4	-0.7495	2.1159	9.7086	51.75062415	0.0409	59.9793





From the data fitted the trend line for two data sets like observed and predicted values using logistic approach with two parameters as location and altitude of the depth of the dam levels, which are at the time of discharge data. The Dam data follows a standard pattern of discharge level data sets in Sathanur Region. The present method for fitting the Logistic distribution is also acceptable for statistical Analysis. The Method of moments for big data sets is exactly appropriate for the parameters of the distributions. The present paper conclude that the method adopted for finding the parameters are perfectly fit for the data. The Scientific results clearly established that the analytical procedure devised and tested in this study may suitably applied for the identification of the best fit probability distribution of discharge level data sets.

IV. CONCLUSIONS

From the above calculations, we conclude that the logistic probability approach approximately suitable for the data taken from Sathanur Dam, Tamil Nadu, India. Mainly observed that the Observed data in this case a standard probability distribution pattern at the time discharge the data, which follows a logistic distribution is also suitable for the data collected from the dam. The squared R values also acceptable for the study area, and Logistic distribution is best fitted one for the data taken from sathanur Dam, Tamil Nadu India.

The Dam data follows a standard pattern of discharge level data sets in Sathanur Region. The present method for fitting the Logistic distribution is also acceptable for statistical Analysis. The Method of moments for big data sets is exactly appropriate for the parameters of the distributions. The present paper concludes that the method adopted for finding the parameters perfectly fits for the data. The Scientific results clearly established that the analytical procedure devised and tested in this study may suitably applied for the identification of the best fit probability distribution of discharge level data sets.

V. REFERENCES

[1] M.Sunitha, B. Polaiah,S.K. KhadarBabu, K. Pushpanjali, M.V. Ramanaiah (2019), Normal Optimisation Technique for Hydrological Data Sets under Fuzzy Environment, International Journal of Pure and Applied Biosciences,7(6),264-269.

[2] ChitrasenLairenjam, ShivaraniHuidrom, Arnab Bandyopadhyay and Aditi Bhadra (2016), Assessment of Probability Distribution of Rainfall of North East Region(NER) of India, Journal of Research in Environmental and Earth Science,2(9),pp12-18.

[3] Chin-Yu Lee(2005), Application of Rainfall Frequency Analysis on Studying Rainfall Distribution Characteristics of Chia-Nan Plain Area in Southern Taiwan, Crop, Environment &Bioinformatics,(2), pp 31-38.

[4] Swami Shivprasad, Anandrao Deshmukh, Ganesh Patil, Sagar Kahar (2013), Probability Distribution Analysis of Rainfall Data for Western Maharashtra Region, International Journal of Science and Research, 2319-7064.

[5] Oseni, B. Azeez and Femi J. Ayoola(2012), Fitting the Statistical Distribution for daily Rainfall in IBADAN, based on Chi-Squared and Kolmogrov-Smirnov Goodness of fit tests, European Journal of Business and Management,4(17),pp 62-70.

[6] T. Mayooran and A. Laheetharan(2014), The Statistical Distribution of Annual Maximum Rainfall in Colombo District, The Sri Lankan Journal of Applied Statistics,15(2)

[7] S.R. Bhakar, Mohammed Iqbal, Mukesh Devanda, Neeraj Chhajed and Anil K. Bansal(2008), Probability Analysis of Rainfall at KOTA, Indian Journal of Agricultural Research,42(3),pp 201-206