Annual Peak Flood Frequency Analysis by Log-Normal Distribution for the River Krishnai of Assam

Shonahar Ali[1], Manash Pratim Kashyap[2]

Research Scholar, Department of Statistics, Assam downtown University; Guwahati, Assam, India[1]

Associate Professor, Department of Statistics, Assam downtown University, Guwahati, Assam, India[2]

Abstract

This study shows the analysis of 46 years annual peak flood frequency data by Log-Normal Distribution model at the Velterghat Gauging station for the Krishnai river of Assam. Parameters are estimated by using different methods and the Maximum Likelihood Estimation (MLE) is found as the best method. As goodness of fit criteria, Empirical and theoretical densities, Empirical and theoretical CDFs, Q-Q plot, P-P plot are plotted using R. Anderson Darling test was performed, which shows the Log-Normal Distribution fits the data. The return period for 2, 5, 10, 25, 50,100 and 200 years are estimated by this distribution model for the river. Kolmogorov- Smirnov test was performed to check the goodness of the fitted model.

Key word: Log-normal, Parameter, MLE, Return Period, Plot.

I.INTRODUCTION

Krishnai River is a South bank tributary of the river Brahmaputra. It is originated from Garo Hills of Meghalaya at an elevation of about 280 meter above mean sea level and connected to the Dudhnoi River at Tomuni about 12 km north from Dudhnoi town at an elevation about 150 meter above mean sea and finally follows towards the River Brahmaputra. The catchment area lying between 25° 35/ to 26° 2/ north latitude and 90° 20/ to 90° 45/ east longitude is in Garo hills of Meghalaya in the north slope of the state adjoining Assam.

II.METHODOLOGY

A. Data collection

The annual peak discharge flow data in cumecs, at Velterghat Gausing station is collected from 1972 to 2018 for 46 years from Goalpara Investigation Division (Irrigation), Goalpara. Here the annual peak flood flow data are assumed to be independent of each other.

B. Log- Normal Distribution

The log-normal distribution occurs in practice whenever we encounter a random variable which is such that its logarithm has a normal distribution. Its probability density is given by

$$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\{-\frac{1}{2}(\frac{lnx-\mu}{\sigma})^2\} \quad ; -\infty <x<+\infty \qquad (1)$$

Whereμ: continuous location parameter,

σ: Continuous scale parameter.

In hydrology, the probability of exceedence, is the probability that a flow is greater than, or equal to a particular value. The relative frequency is the probability of the flow being less than a particular value. The return period is termed for the average recurrence interval as giver below.

If P(X),F(X) and T(X) are the exceedence probability, relative frequency and average recurrence interval then P(X)= 1-F(X) and $T(X)=\frac{1}{P(X)}$=1/{1-F(X)} (2)

Here Weibull's method is used to calculate the return period. In this method, the annual peak discharge series is ranked in order of magnitude and the return period is calculated as

$$T(x) = Tr = (n+1)/m \quad (3)$$

Where m is the rank of an individual flood event X within the data series and n is the size of the series.

To calculate the above mentioned parameters the following methods are used.

In log-normal distribution, the series is normally distributed with mean and standard deviation and the coefficient of skewness is assumed to be zero. This is a popular method to study about flood. The estimated peak discharge is given by

$$Z = Z_{ave} + K_0\sigma_z \quad (4)$$

Where, Z_{ave} is the mean of logarithm series and σ_z is the standard deviation and K_0 is the frequency factor, which is the function of return period with coefficient of skewness as zero as per determined by the frequency table.

C. Methods of parameter estimation: Maximum Likelihood Estimator (MLE), Moment Matching (MME), matching Quantile (MQE) and maximum goodness-of-fit (gof) are used to analyses.

Kolmogorov-Smirnov test (KS): The KS test compares an empirical and a theoretical model by computing the maximum absolute difference between the empirical and theoretical distribution functions: $D = \max_x |F^\wedge(x) - F(x)|$ (5)

Where F^(x) is the empirical distribution function (the relative frequency of observations that are smaller than or equal to x), and $F(x) = P(X \geq x)$ is the theoretical distribution function.

III. ANALYSIS, RESULTS AND DISCUSSION

A.Table1.The following table shows the summary calculation for Mean, Variance, SD, Return period and Exceedence Probability.

LogX	(LogX-Ave(logX))^2	(LogX-Ave(LogX))^3	Return PeriodTr =(n+1)/m	Exceedence pro(1/Tr)	Non exceedance pro.
2.968655698	0.19908078	0.088827	47	0.021276596	0.9787234
2.921160843	0.15895359	0.063373	23.5	0.042553191	0.9574468
2.920957706	0.15879165	0.063276	15.666667	0.063829787	0.9361702
2.85985852	0.11383034	0.038405	11.75	0.085106383	0.9148936
2.843550438	0.10309201	0.033101	9.4	0.106382979	0.893617
2.810360233	0.08288021	0.02386	7.8333333	0.127659574	0.8723404
2.80740648	0.08118823	0.023133	6.7142857	0.14893617	0.8510638
2.794167261	0.07381886	0.020056	5.875	0.170212766	0.8297872
2.7857568	0.06931941	0.018251	5.2222222	0.191489362	0.8085106
2.77555948	0.06405378	0.016211	4.7	0.212765957	0.787234
2.754333043	0.05376001	0.012465	4.2727273	0.234042553	0.7659574
2.74804841	0.05088517	0.011479	3.9166667	0.255319149	0.7446809
2.743352951	0.04878884	0.010777	3.6153846	0.276595745	0.7234043
2.715251029	0.03716414	0.007165	3.3571429	0.29787234	0.7021277
2.666826744	0.02083858	0.003008	3.1333333	0.319148936	0.6808511
2.66093185	0.01917141	0.002654	2.9375	0.340425532	0.6595745
2.622763022	0.01005849	0.001009	2.7647059	0.361702128	0.6382979
2.595165415	0.00528448	0.000384	2.6111111	0.382978723	0.6170213
2.57591491	0.00285625	0.000153	2.4736842	0.404255319	0.5957447

2.575511135	0.00281326	0.000149	2.35	0.425531915	0.5744681
2.538498329	0.00025688	4.12E-06	2.2380952	0.446808511	0.5531915
2.538272036	0.00024967	3.95E-06	2.1363636	0.468085106	0.5319149
2.530468647	6.3962E-05	5.12E-07	2.0434783	0.489361702	0.5106383
2.524525937	4.2228E-06	8.68E-09	1.9583333	0.510638298	0.4893617
2.481342285	0.00169157	-7E-05	1.88	0.531914894	0.4680851
2.472419997	0.0025051	-0.00013	1.8076923	0.553191489	0.4468085
2.455240878	0.00451989	-0.0003	1.7407407	0.574468085	0.4255319
2.454738178	0.00458774	-0.00031	1.6785714	0.595744681	0.4042553
2.422245025	0.01004525	-0.00101	1.6206897	0.617021277	0.3829787
2.3931363	0.01672746	-0.00216	1.5666667	0.638297872	0.3617021
2.360025089	0.02638867	-0.00429	1.516129	0.659574468	0.3404255
2.346470335	0.03097623	-0.00545	1.46875	0.680851064	0.3191489
2.346392098	0.03100378	-0.00546	1.4242424	0.70212766	0.2978723
2.319251841	0.04129803	-0.00839	1.3823529	0.723404255	0.2765957
2.309949384	0.04516544	-0.0096	1.3428571	0.744680851	0.2553191
2.305093296	0.04725307	-0.01027	1.3055556	0.765957447	0.2340426
2.298809427	0.0500245	-0.01119	1.2702703	0.787234043	0.212766
2.296467739	0.05107747	-0.01154	1.2368421	0.808510638	0.1914894
2.28141974	0.05810571	-0.01401	1.2051282	0.829787234	0.1702128
2.271632537	0.06291993	-0.01578	1.175	0.85106383	0.1489362
2.221987702	0.09029021	-0.02713	1.1463415	0.872340426	0.1276596
2.19926138	0.10446446	-0.03376	1.1190476	0.893617021	0.106383
2.177853809	0.11876101	-0.04093	1.0930233	0.914893617	0.0851064
2.176669933	0.11957838	-0.04135	1.0681818	0.936170213	0.0638298
2.111060782	0.16925837	-0.06963	1.0444444	0.957446809	0.0425532
2.054919327	0.21860457	-0.10221	1.0217391	0.978723404	0.0212766
Average=2.522471391	Sum=2.6624510 5	Sum=0.02276 7			

Variance=2.66245105/(46-1)= 0.05916558 and SDσlogX=sqrt 0.05916558=0.24323976

Skew coefficient = (46*0.022767)/(46-1)*(46-2)*0.022767=0.023232~0

B. Table 2: Estimation of Peak Discharge by Log-Normal Distribution

Return period T	Coefficient of skewness,Cs	K=f(T,Cs)	XT=X bar +kσx	Peak discharge in cumescQ=10^Q
2		0	2.522471	333.02052
5	0	0.842	2.727278878	533.67748
10		1.282	2.834304372	682.81707
25		1.751	2.94838382	887.94041
50		2.054	3.022085467	1052.1689
100		2.326	3.088246682	1225.312
200		2.576	3.149056622	1409.4725

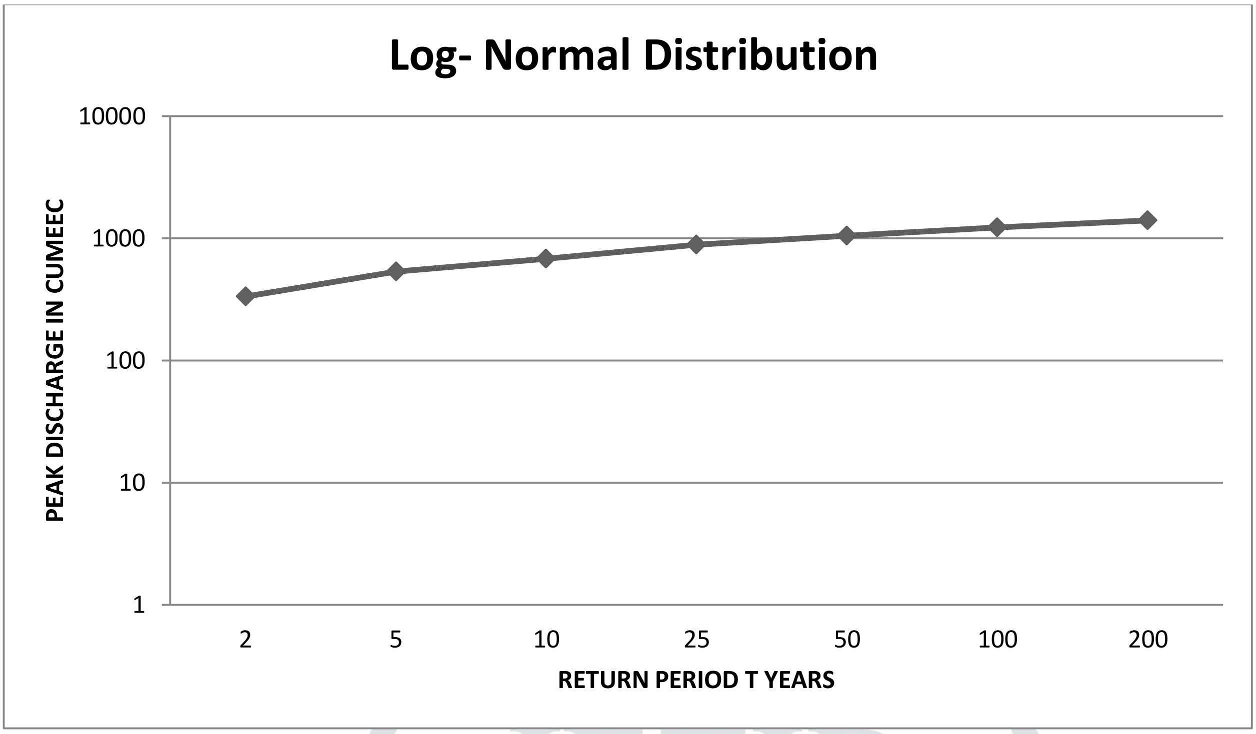

Fig.1: The graph of Return period and estimated peak discharge in cumesc is shown in the graph paper

C.Table4: As a goodness of fit criterion Anderson-Darling test was performed is summarized in the following table

Test name	Value of Statistic	P-value	Decision
AD	0.51825	0.17869>0.05	Accepted

IV. Fitting of the Log-normal distribution by Maximum Likelihood Estimator (MLE), Moment Matching (MME), matching Quantile (MQE) and maximum goodness-of-fit (gof):

A. The estimated parameters and their corresponding plots for Empirical and theoretical densities, Empirical and theoretical CDFs,Q-Q plot, P-P plot are given below respectively using R software.

	estimate	Std. Error
meanlog	5.808205	0.08167679
sdlog	0.553959	0.05775337

Loglikelihood: -305.278 AIC: 614.5561 BIC:618.2133

Correlation matrix:

	meanlog	sdlog
meanlog	1	0
sdlog	0	1

B. Construction of Confidence Interval (CI):

The 95% CI for the parameter estimate mean is given by 5.808205 $\pm$ 1.96*(0.08167679); which (the mean) could be anywhere in the interval (4.207339916, 5.9682915084).

C. The plots of Empirical and theoretical, Q-Q plot, Empirical and theoretical CDFs and P-P plot by MLE are given below.

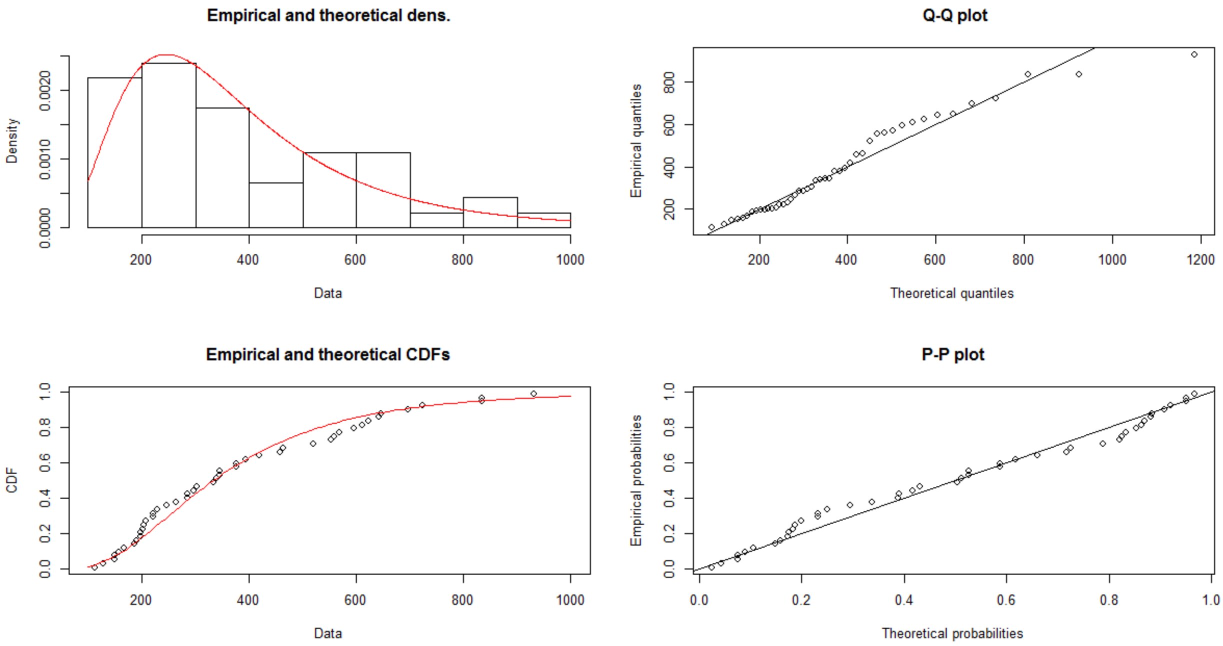

Fig.2 shows the plots of Empirical and theoretical, Q-Q plot, Empirical and theoretical CDFs and P-P plot by MLE

D. Parameter Estimation by Matching Moments and their corresponding plots are given below.

	Estimate
Meanlog	5.8273117
sdlog	0.5122819

Loglikelihood: -305.6067 AIC: 615.2134 BIC: 618.8707

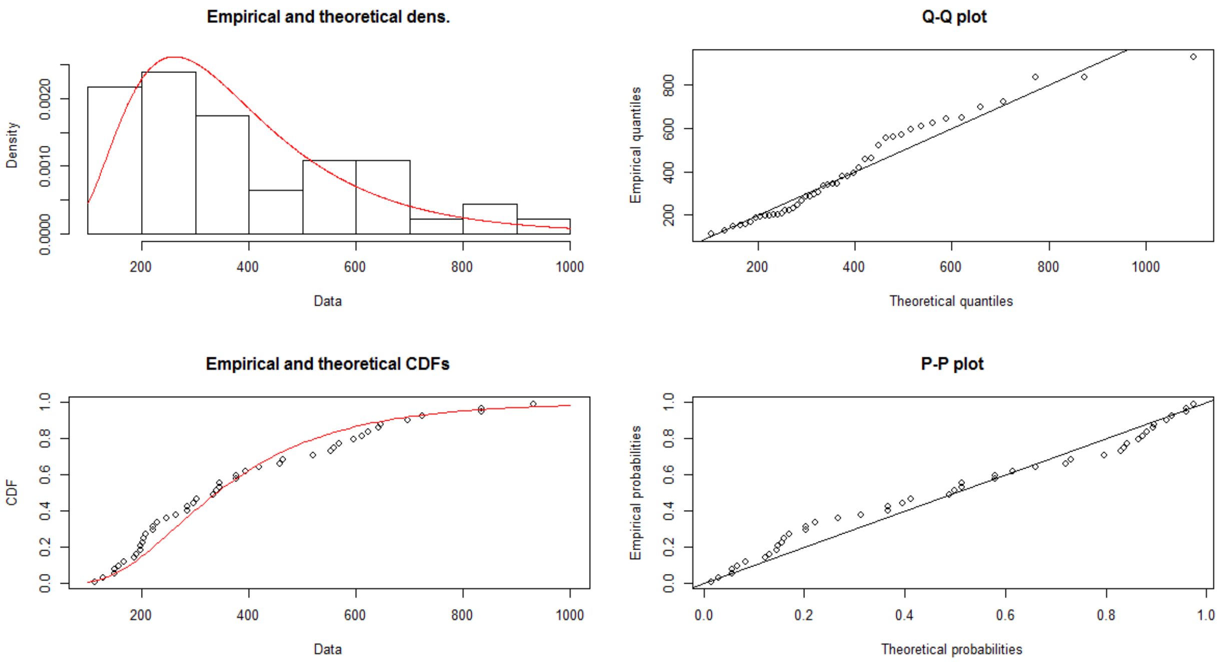

Fig.3 shows the plots of Empirical and theoretical, Q-Q plot, Empirical and theoretical CDFs and P-P plot by MME.

E. Parameter Estimation by Matching quantiles and their corresponding plots are given below.

	Estimate
meanlog	5.8245948
sdlog	0.7417846

Loglikelihood: -308.5469 AIC: 621.0937 BIC: 624.751

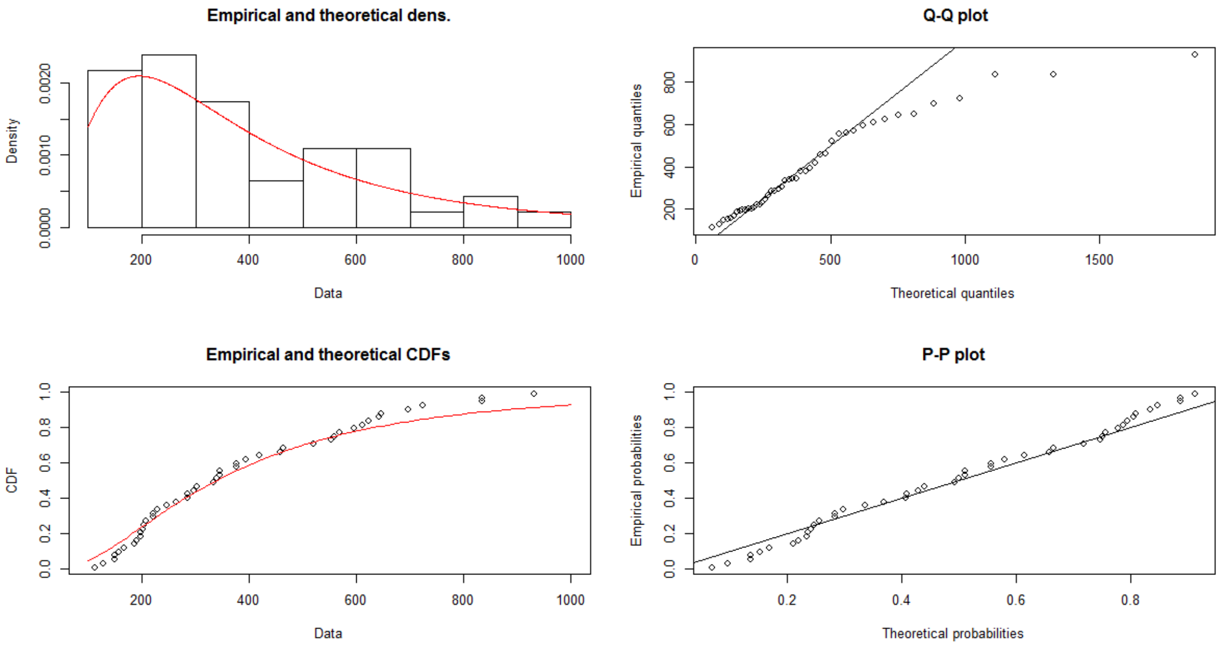

Fig.4 shows the plots of Empirical and theoretical, Q-Q plot, Empirical and theoretical CDFs and P-P plot by MQE

F. Parameters Estimation and their corresponding plots by Maximum goodness- of-fit for Cramer von Mises (CvM), Kolmogorov Smirnov (KS) and Anderson Darling (AD) respectively given below.

Parameters estimation by CvM and their plots are given below.

	Estimate
meanlog	5.8020347
sdlog	0.6393415

Log likelihood: -306.1412 AIC: 616.2824 BIC: 619.9397

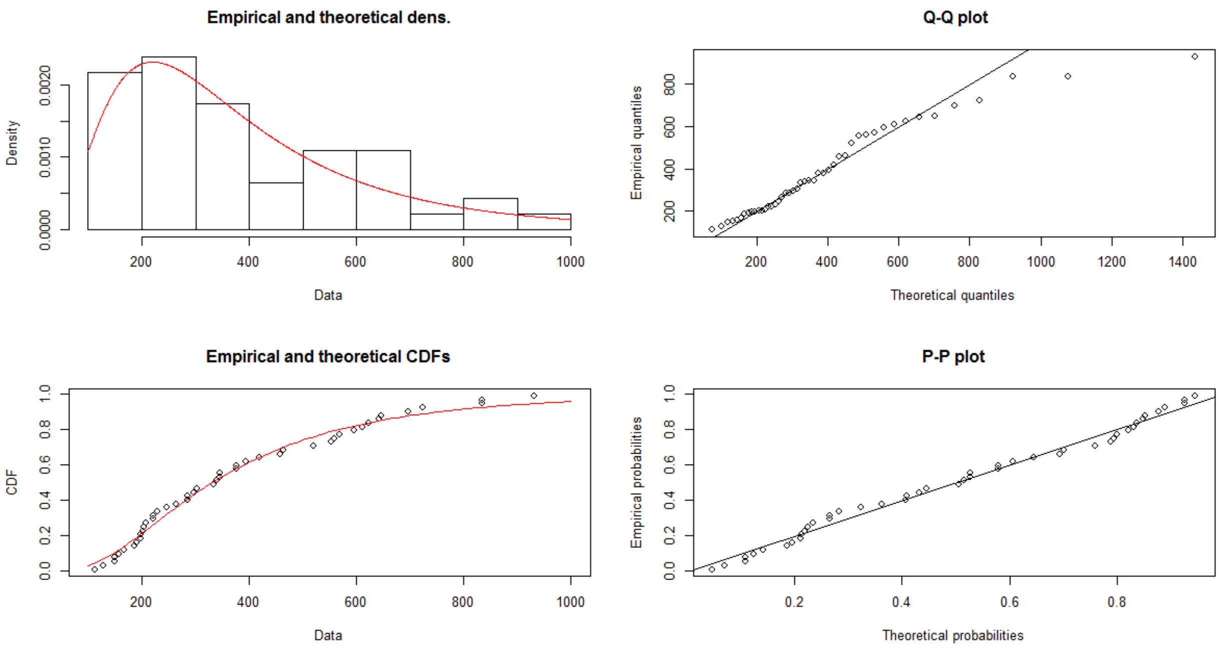

Fig.5 shows the plots of Empirical and theoretical, Q-Q plot, Empirical and theoretical CDFs and P-P plot by gof(CvM)

G. Parameters Estimation by KS and their plots are given below.

	Estimate
meanlog	5.8086127
sdlog	0.6482479

Log likelihood: -306.3042 AIC: 616.6085 BIC: 620.2658

Fig.6 shows the plots of Empirical and theoretical, Q-Q plot, Empirical and theoretical CDFs and P-P plot by gof(KS)

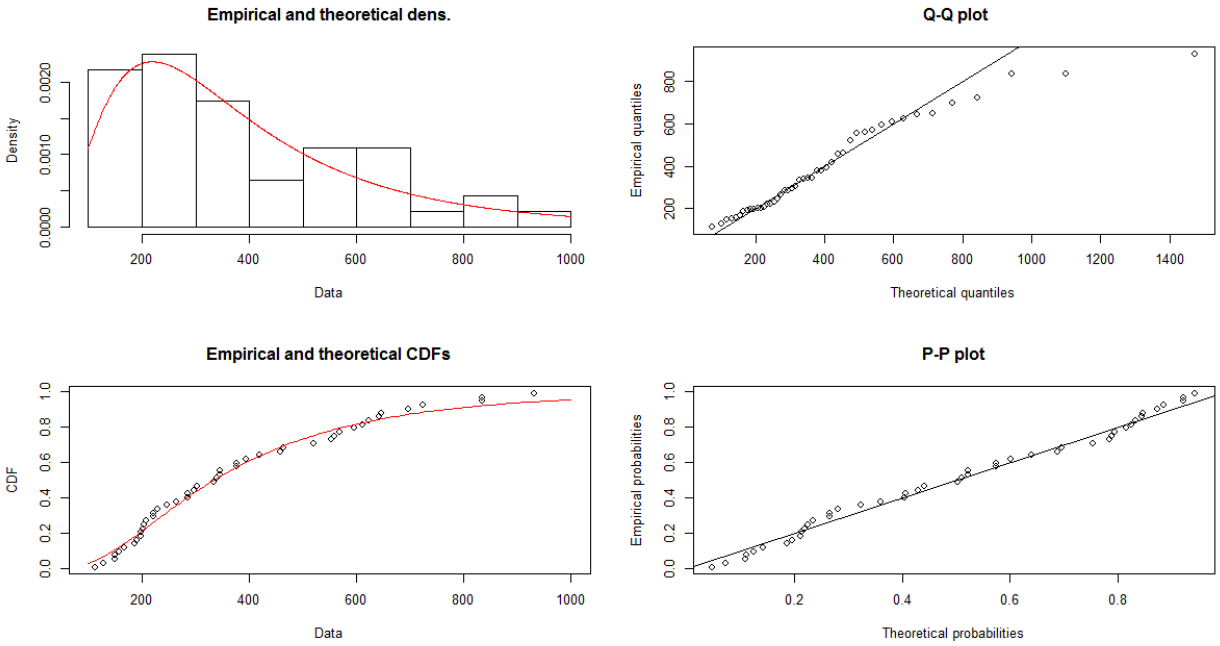

Fig.6 shows the plots of Empirical and theoretical, Q-Q plot, Empirical and theoretical CDFs and P-P plot by gof(KS)

H. Parameters Estimation by AD and their plots are given below.

	Estimate
meanlog	5.8058507
sdlog	0.6095129
Log likelihood:	-305.673 AIC: 615.346 BIC: 619.0032

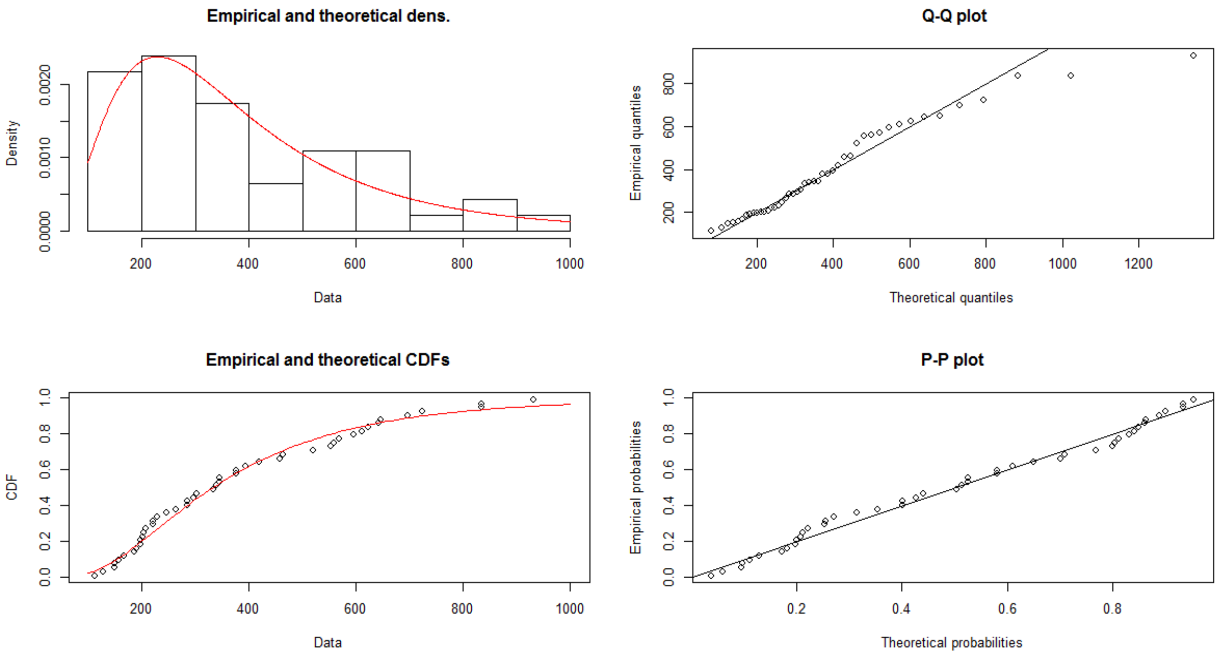

Fig.7 shows the plots of Empirical and theoretical, Q-Q plot, Empirical and theoretical CDFs and P-P plot by gof(AD)

IV.A.Table5: The Comparisons table for Log-likelihood, AIC and BIC of the given methods respectively.

	Log-likelihood	AIC	BIC
MLE	-305.278	614.5561	618.2133
MME	-305.6067	615.2134	618.8707
MQE	-308.5469	621.0937	624.751
gof(CvM)	-306.1412	616.2824	619.9397
gof(KS)	-306.3042	616.6085	620.2658
gov(AD)	-305.673	615.346	619.0032

From the above Table5 it is obvious that the maximum of Log-likelihood value is -305.278 but the minimum values of AIC is 614.5561 and BIC is 618.2133; which are provided by MLE compare to other given values in the table. So, MLE is the best estimate method compare to all other given methods which is also visualized by all the pictorial methods like Empirical and theoretical densities, Empirical and theoretical CDFs, Q-Q plot, P-P plot.

B. One-sample Kolmogorov-Smirnov test based on MLE as found the best estimator.

	Estimate	Std. Error
Mean log	5.808205	0.08167679
Sd log	0.553959	0.05775337

D = 0.10069, p-value = 0.7017

C. Table 6: Shows the KS test

Test name	Value of Statistic	p-value	Decision
KS	D=0.10069	0.7017>0.05	Accepted

CONCLUSION

MLE is found as the best parameter estimation method based on the criterion Log-likelihood, AIC and BIC. From Anderson- Darling test (given in Table4) it is found that the P-Value is greater than 0.05 and from Table6 by Kolmogorov-Smirnov test, the p-value is highly greater than0.05 which implies that the lognormal distribution model fit the data well. . Not only by Q-Q plots by MLE method but even by other given methods also fits the data as the data points are approximately falls around the line. Hence the Log-normal distribution model can be used to analyze the annual peak discharge flood of the river.

References

1.Adamowski, K.(2000) Regional analysis of maximum and partial duration flood data by nonparametric andL-moment method. *Journal of Hydrology 229 (3-4):219-231,*

2.Alila, Y. (1999) A hierarchical approach for the regionalization of precipitation annual maximum in Canada.J *Geophys Res-Atmos 104 (D24):31645-31655.*

3.Burn, D.H. (1997) Catchment similarity for regional flood frequency analysis using seasonality measures. *Jornal of Hydrology 202 (1-4):212-230.*

4.Chow, V.T.(1964) Handbook of Applied Hydrology.Mc Graw Hill,New York.

5.Haan C.T. Statistical Methods in Hydrology, lowa State University Press, Ames.lowa,USA,1977.

6.Das L.M.and Qureshi Z. H.S Flood Frequency Analysis for Jiya Dhol River of Brahmaputra Valley.IJSBAR(2014) Volume14. No2.pp14-24

7.Ali S. and Kashyap M.P. Flood Frequency Analysis by Gumble Distribution for Krishnai River of Assam. IARJSET(2018) Vol.5,Issue10,October2018

8. Suhartanto E., Limantara M., Noviadriana D., Hatra F.I., and Aryani D. Estimation of Design of Flood With Four Frequency Analysis Distributions.AJAST(2018) Volume 2, Issue1,pp 13-27,2018