International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Probability Distribution Analysis of Rainfall Data for Western Maharashtra Region

Swami Shivprasad T.[1], Anandrao Deshmukh[2], Ganesh Patil[3], Sagar Kahar[4]

[1]Department Civil Engineering, Rajarshi Shahu College of Engineering, Tathawade, Pune, India

[2]Assit.Professor, Department Civil Engineering, Rajarshi Shahu College of Engineering, Tathawade, Pune, India

[3]Department Civil Engineering, Rajarshi Shahu College of Engineering, Tathawade, Pune, India

[4]Department Civil Engineering, Rajarshi Shahu College of Engineering, Tathawade, Pune, India

Abstract: ***In India rainfall is the major source for the most sought after commodity of water all over the world. India experiences a monsoon season of 4 months making it more precious. Every attempt is therefore made to study and analyse rainfall data in order to understand its distribution, quantity and frequency. The present work is such an attempt where in various probability distributions are be applied to estimate one month and two to seven consecutive cumulative monthly rainfall. As many as 16 different types of continuous probability distributions were tested using goodness of fit tests such as Kolmogorov Smirnov, Anderson Darling and Chi-Square for Western Maharashtra's districts , India to check which distribution is best. The following analysis of rainfall data is prepared with the help of EASY FIT and MICROSOFT EXCEL software's. Monthly rainfall data for 35 years (1966-2000) measured by Indian Meteorological Department, Pune was used. It was found that, in Ahmednagar and Solapur district Beta distribution is the best fitted distribution for one month rainfall dataset, Dagum distributions for consecutive cumulative two months rainfall dataset, Dagum and Gen.Extreme distribution for consecutive cumulative three months rainfall dataset, Log Logistic and Gen.Extreme distribution for consecutive cumulative four and five months rainfall dataset, Log Logistic and Nakagami distributions for consecutive cumulative six months rainfall dataset and Log Logistic and Johnson SU distribution for consecutive cumulative seven months rainfall, as per above the best fitted distributions for other districts of western Maharashtra are as follows***

Keywords: Rainfall, Probability distribution, goodness-of-fit tests

Month	Distribution in Districts	Column1	Column2	Column3	Column4	Column5
	Pune	**Ahmednagar**	**Solapur**	**Sangli**	**Satara**	**Kolhapur**
1	Dagum	Beta	Beta	Dagum	Beta	
2		Dagum		Chi squared	Dagum	Kumaraswamy
3	Gen.pareto	Dagum	Gen.Extreme value	Gen.Extreme	Dagum	
4	Error	Log Logistic	Gen.Extreme value	Gen.Gamma	Dagum	
5	Gen.Extreme	Log Logistic	Gen.Extreme value	Burr	Gumbel Min.	
6	Gen Gamma	Log Logistic	Nakagami	Gen.Extreme	Laplace	Gumbel Min.
7		Log Logistic	Johnson SU	Burr	Cauchy	

1. Introduction

Analysis of rainfall data strongly depends on its distribution pattern. It has long been a topic of interest in the fields of meteorology in establishing a probability distribution that provides a good fit to monthly rainfall. Several studies have been conducted in India and abroad on rainfall analysis and best fit probability distribution functions.

Various probability distributions and transformations can be applied to estimate one day and two to five consecutive days annual maximum rainfall of various return periods. Patel and Shete (2008) predicted consecutive days maximum rainfall by probabilistic approach for Sabarkantha district of Gujarat. Patel and Shete (2007) also predicted consecutive days rainfall from one day rainfall using regression models for the same district. Three commonly used probability distributions (viz: Normal, Log Normal and Gamma distributions) were tested comparing the Chi-square value by Bhakar et al. (2006). Gamma distribution was found to be the best fit for Banswara, Rajasthan. Kuller et al. (1999) carried out the goodness of fit of three frequency model distribution models viz. Lognormal 3-parameters, Lognormal 2 parameters and Extreme Value Type 1 and concluded that Lognormal 3-parameters distribution model was the best distribution model for all the three duration of rainfall (viz. Maximum 1-day, maximum 2-day and maximum 3-day) for Punjab. Although hydrological variables are of continuous type, they are discretized and used as a discrete series. It is generally assumed that a hydrological variable has a certain distribution type. Probability distribution arises from the experiments where the outcome is subject to chance. The nature of the experiment dictates which probability distribution may be appropriate for modelling the resulting outcomes. There are two types of probability distribution – continuous and discrete. Using a probability model does not allow one to predict the result of any individual experiment but one can determine the probability that a given outcome

will fall inside a specific range of values. The data for present analysis is continuous type. Therefore, sixteen continuous probability distributions are checked for the rainfall dataset (1966-2000) in districts of Western Maharashtra.

The present paper is an effort to determine the best fittest probability distribution for the given rainfall dataset in order to determine the rainfall characteristic of that region. To assist the above mentioned objective, Kolmogorov Smirnov, Anderson Darling, Chi-Square tests are used to measure the goodness of fit of an estimated statistical distribution function.

2. Study Area

- Satara

The study area comprises of Satara and Kolhapur districts of Western Maharashtra, India. Satara is one of the districts of western part of the state of Maharashtra. It is situated between 17° 05' and 18° 11' N latitudes and 73 ° 33' and 74° 54' E longitude. The total area of the district is 10,480 sq. km. Satara district. The average annual rainfall is 1282.5 mm considering the rainfall dataset from 1901-2000.

- Kolhapur

Kolhapur is also one of the districts of western part of the state of Maharashtra. It is situated in the western part of the state and lies between 15° 43' and 17° 10' N latitudes and 73 ° 40' and 74° 42' E longitude. The total area of the district is 7,685 sq. km. The average annual rainfall is 1658.3 mm considering the rainfall dataset from 1901-2000.

- Sangli: It is located in western Maharashtra. Its latitude and longitude for sagli 16.86®N,74.57®E Total area=8578km²

- Solapur: Solapur is located in 17.10 to 18.32°N latitude, the district is situated on the south east fringe of Maharashtra state and lies entirely in the Bhima Seena basin. The geographical area for solapur is 14895Km².the city receives an annual rainfall 545mm per year.

- Pune:Most of the 722 mm of annual rainfall of city falls between june and September and July is the wettest month of the year. The total geographical area of pune district is 15642Km²

- Ahmednagar:The average annual rainfall for a.nagar if 578.8 mm.The total geographical area for ahmednagar district is 17048Km²

3. Distributions

A. Beta distribution

The PDF and CDF of Beta distribution with continuous shape parameters α1 and α2 (α1, α2 > 0) is given by,

$$F(x) = \frac{1}{B(\alpha 1,\alpha 2)} \frac{(x-a)^{\alpha_1-1}(b-x)^{\alpha_2-1}}{(b-a)^{\alpha_1+\alpha_2-1}}$$

$$F(x) = Iz(\alpha_1, \alpha_2)$$

where, $z=\frac{x-a}{b-a}$ B is the beta function, Iz is the regularized incomplete beta function and a ≤ x ≤ b.

B. Cauchy distribution

The PDF and CDF of Cauchy distribution with continuous scale parameter σ (σ > 0) and continuous location parameter μ is given by,

$$f(x)= \{\pi\sigma[1+(\frac{x-\mu}{\sigma})^2\ ]\}^{-1}$$

$$F\ x = \frac{1}{\pi}\arctan(\frac{x-\mu}{\sigma}) +0.5$$

Where, -∞ < x < +∞.

C. Dagum distribution

The PDF and CDF of Dagum distribution (three parameter) with continuous shape parameter k (k > 0), continuous shape parameter α (α > 0) and continuous scale parameter β (β > 0) is given by,

$$f(x)=\frac{\propto k(\frac{x}{\beta})^{\propto k-1}}{\beta\{1+(\frac{x}{\beta})^{\propto}\}^{k+1}}$$

$$F(x)=\{1+\left(\frac{x}{\beta}\right)-\propto\}^{-k}$$

where, γ ≤ x < +∞

D. Gumbel Min. distribution

The PDF and CDF of Gumbel Min. distribution with continuous scale parameter σ (σ > 0) and continuous location parameter μ is given by,

$$f(x)= \frac{1}{\sigma}\exp\{z-\exp(z)\}$$

$$F\ x =1-\exp\{-\exp(z)\}$$

where, $z=\frac{x-\mu}{\sigma}$ and -∞ < x < +∞.

E. Kumaraswamy distribution

The PDF and CDF of Kumaraswamy distribution with continuous shape parameters α1 and α2 (α1, α2 > 0) is given by,

$$f(x)=\frac{\propto 1\propto 2z^{\propto 1-1}(1-z^{\propto 1})^{\propto 2-1}}{(b-a)}$$

$$F\ x =1-(1-z^{\propto 1})^{\propto 2}$$

where, $z=\frac{x-a}{b-a}$ and a ≤ x ≤ b.

F. Laplace distribution

The PDF and CDF of Laplace distribution with continuous inverse scale parameter λ (λ > 0) and continuous location parameter μ is given by,

$$f(x)= \frac{\lambda}{2}\exp(-\lambda|x-\mu|)$$

$$f(x)= \{\frac{1}{2}\exp\{-\lambda(\mu-x)\},\ x<\mu$$

$$f(x) = \{1-\frac{1}{2}\exp\{-\lambda(x-\mu)\},\ x>\mu$$

where, -∞ < x < +∞.

Three goodness-of-fit tests were conducted at 5% level of significance. Note that X denotes the random variable and; n, the sample size. The tests are as follows:

1) Anderson-Darling test:

The Anderson-Darling statistic (A2) is defined as

$$A^2=-n-\frac{1}{n}\sum_{i=1}^{1}(2i-1)[\ln F(X)i\ +\ln(1-F(Xn-i+1))]$$

The Anderson-Darling procedure is a general test to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. This test gives more weight to the tails than the Kolmogorov-Smirnov test.

2) Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov statistic (D) is based on the largest vertical difference between the theoretical and the empirical cumulative distribution function, $D=\max_{1\le i\le n}(F(x_i)-\frac{i-1}{n})$

3) Chi-Square test:

The Chi-Squared statistic is defined as,

$$\chi^2=\sum_{i=1}^{k}\frac{(Oi-Ei)^2}{Ei}$$

where; Oi is the observed frequency Ei is the expected frequency which is calculated by,

$Ei=F(x_2)-F(x_1)$

where, F is CDF of probability distribution being tested and x1, x2 are the limits.

Although there is no optimal choice for the number k, there are several formulae which can be used to calculate this number based on the sample size N.

$k=1+\log_2 N$

This test is for continuous sample data only and is used to determine if a sample comes from a population with a specific distribution. The three goodness of fit test mentioned above were fitted to the consecutive cumulative monthly rainfall data treating different data set. The test statistic of each test were computed and tested at (α=0.05) level of significance. Accordingly the ranking of different probability distributions were marked from 1 to 16 based on minimum test statistic value. The distribution holding the first rank was selected for all the three tests independently. The common highest rank of the entire three tests were summarized to identify the best fit distribution obtained.

4. Methodology

For the present study, monthly rainfall data from Satara and Kolhapur districts for 35 years (1966-2000) was obtained from the Indian Meteorological Department, Pune. The monthly rainfall is the sum of daily rainfall considering all the days in a month. If rainfall data of even a single day in the month is missing, the monthly rainfall data of that whole month is not considered and is left blank. The monthly rainfall data is in mm. The monthly rainfall and cumulative monthly rainfall is used for analysis of probability distribution. 16 different probability distributions presented in Table I are applied to the dataset of one, consecutive cumulative 2, 3, 4, 5, 6 and 7 months rainfall. The goodness of fit using Kolmogorov Smirnov, Anderson Darling and Chi-Square are tested.

The formulae for distributions and goodness of fit tests are taken from the manual of Easy Fit 5.5. It is a data analysis and simulation application allowing to fit probability distributions to sample data, select the best model, and apply the analysis results to make better decisions. It can be used as a stand-alone Windows application or with Microsoft Excel and other third party Excel-based simulation tools, leaving the complex technical details behind the scenes and enabling you to focus on your business goals. EasyFit 5.5 combines the classical statistical analysis methods and innovative data analysis techniques, making it a tool of choice for anyone dealing with probability data. [6]

Sr. No.	Name of the Continuous Distribution	Sr. No.	Name of the Continuous Distribution
1	Beta	9	Kumaraswamy
2	Cauchy	10	Laplace
3	Dagum	11	Logistic
4	Exponential	12	Log-Logistic
5	Gamma	13	Lognormal
6	Generalized Pareto	14	Normal
7	Gumbel Min.	15	Rayleigh
8	Inverse Gaussian	16	Weibull

The probability density functions (PDF) and cumulative distribution functions (CDF) for the six distributions which are drawn from the results and analysis out of sixteen distributions are given as follows. The remaining distributions can be checked from the same manual. Note that x is the random variable representing the monthly rainfall intensity.

5. Result and Statistics

The summary of statistics mean, mode, variance, standard deviation, skewness coefficient and coefficient of variation values of consecutive cumulative monthly rainfall is presented in Table 2. Using the rainfall data of cumulative 1 and cumulative consecutive 2, 3, 4, 5, 6 and 7 months, the probability distribution with their ranking based on Kolmogorov Smirnov, Anderson Darling and Chi-Square tests are given in Table 3.

6. Conclusion

The Kolmogorov Smirnov, Anderson Darling, Chi-Square tests give the optimum probability function for the given dataset. It is concluded that Beta distribution is best for cumulative 1 month rainfall for Satara and Kolhapur districts; Dagum distribution is best for cummulative consecutive 2 months rainfall for Satara district and Kumaraswamy distribution is best for cumulative consecutive 2 months rainfall for Kolhapur district; Dagum distribution is best for cumulative consecutive 3 and 4 months rainfall for Satara and Kolhapur districts; Gumbel Min. distribution is best for cumulative consecutive 5 months rainfall for Satara and Kolhapur districts; Laplace distribution is best for cumulative consecutive 6 months rainfall for Satara district and Gumbel Min. distribution is best for cummulative consecutive 6 months rainfall for Kolhapur district and Cauchy distribution is best for cummulative consecutive 7 months rainfall for Satara and Kolhapur districts. for other western Maharashtra district's are tabulated below.

Table 3: Conclusion

Month	Distribution in Districts					
	Column 1		Column 2	Column 3	Column 4	Column 5
	Pune	Ahmednagar	Solapur	Sangli	Satara	Kolhapur
1	Gen.Gamma	Beta		Dagum	Beta	
2	Dagum			Chi squared	Dagum	Kumaraswamy
3	Gen.pareto	Dagum	Gen.Extreme value	Gen.Extreme	Dagum	
4	Error	Log Logistic	Gen.Extreme value	Gen.Gamma	Dagum	
5	Gen.Extreme	Log Logistic	Gen.Extreme value	Burr	Gumbel Min.	
6	Gen Gamma	Log Logistic	Nakagami	Gen.Extreme	Laplace	Gumbel Min.
7	Log Logistic		Johnson SU	Burr	Cauchy	

Table 4: Result statistics

Data	Month	Distribution	Min.	Max.	Mode	Mean	Variance	Std. Dev.	Coef. of Var.	Skewness	Kurtosis
Satara	1	Beta	-1E-14	671.5	-1E-14	128.53	21514	146.88	1.1412	1.3316	1.0084
	2	Dagum	0	8	0	300.44	63818	252.62	0.84083	0.68358	-0.41401
	3	Dagum	0	8	0	479.33	104600	323.42	0.67473	0.35615	-0.77493
	4	Dagum	0	8	893.21	689.28	119520	345.71	0.50156	0.00708	-0.7618
	5	Gumbel Min.	8	8	1043.2	895.51	107690	328.17	0.36646	-1.1395	2.4
	6	Laplace	8	8	1056	1056	82960	288.03	0.27275	0	3
	7	Cauchy	8	8	1140.7	--	--	--	--	--	--
Kolhapur	1	Beta	0.1	988.5	0.1	186.77	49888	223.36	1.1959	1.3769	1.0749
	2	Kumaraswamy	0.3	1671.3	0.3	432.28	162850	403.55	0.93352	0.94674	-0.04414
	3	Dagum	0	8	0	707.96	256170	506.13	0.71492	0.37599	-0.92431
	4	Dagum	0	8	1370.6	992.54	288180	536.82	0.54086	0.02314	-0.95694
	5	Gumbel Min.	8	8	1541.6	1300.7	286420	535.18	0.41144	-1.1395	2.4
	6	Gumbel Min.	8	8	1721.8	1493.2	257970	507.9	0.34014	-1.1395	2.4
	7	Cauchy	8	8	1691.5	--	--	--	--	--	--
Pune	1	Gen.Gamma	0.1	8	0.1	106.34	16239	127.43	1.1993	1.4698	1.7336
	2	Dagum	0	8	0	259.75	53209	230.67	0.88803	0.76992	-0.27854
	3	Gen.pareto	-27.523	1100.5	-27.523	426.92	87909	296.5	0.69459	0.32984	-0.96101
	4	Error	8	8	606.43	606.43	105390	324.64	0.53533	0	-0.76139
	5	Gen.Extreme	8	1509.7	844.59	475.71	104030	322.54	0.41581	-0.39209	-0.10859
	6	Gen.Gamma	-1.4E+10	8	986.57	906.84	143290	378.54	0.41743	452240	-4.98E+13
	7	Log Logistic	-2E+10	8	994.61	994.61	169070	411.18	0.41341	-47453000	4.504E+15
A.NAGAR	1	Beta	0	1	0.33333	0.42857	0.05442	0.23328	0.54433	0.22268	-0.86014
	2	Dagum	0	8	0	136.88	13606	116.64	0.85219	1.5343	7.8533
	3	Dagum	0	8	1.291	0.9452	N/A	N/A	N/A	N/A	N/A
	4	Log Logistic	0	8	0.7937	1.2092	0.95624	0.97787	0.80879	N/A	N/A
	5	Log Logistic	0	8	0.7937	1.2092	0.95624	0.97787	0.80869	N/A	N/A
	6	Log Logistic	-259.02	8	406.99	459.88	38570	196.39	0.42705	1.4776	7.5604
	7	Log Logistic	-142.97	8	458.24	511.29	35669	188.84	0.36934	1.5846	8.7065
Solapur	1	Beta	0.1	395	0.1	68.311	5913.2	76.897	1.1257	1.4098	1.4244
	2	Dagum	0	8	0	158.26	15852	125.91	0.79555	1.3156	5.2792
	3	Gen.Extreme value	-2	8	-0.36701	1.5449	N/A	N/A	N/A	N/A	N/A
	4	Gen.Extreme value	8	2212.7	280.42	345.48	32534	180.37	0.52209	0.72874	0.81884
	5	Gen.Extreme value	-2	8	-0.36701	1.5449	N/A	N/A	N/A	N/A	N/A
	6	Nakagami	0	8	0.86603	0.93999	0.11643	0.34121	0.363	0.4037	0.0549
	7	Johnson SU	8	8	--	-0.59048	0.42339	0.65068	-1.102	-0.56901	2.587
Sangli	1	Dagum	0	8	0	75.508	5129.3	71.619	0.9485	1.6993	9.2151
	2	Chi squared	-6095	8	167.02	169.02	125.28	111.93	0.66221	0.03575	0.00192
	3	Gen.Extreme	8	1084	224.04	261.74	19845	140.87	0.5382	0.45152	0.16577
	4	Gen.Gamma	-167.96	8	330.93	364.52	24018	154.98	0.42516	0.41751	0.17915
	5	Burr	0	8	0.37796	0.58905	0.15302	0.39118	0.66409	1.9086	9.4635
	6	Gen.Extreme	8	1273.6	519.85	549.99	27388	165.49	0.3009	0.25525	-0.11877
	7	Burr	0	8	0.37796	0.58905	0.15302	0.39188	0.66409	1.9086	9.4635

Table 5: Goodness of fit

Column1	Column2	Column3	Column4	Column5	Column6
Data	Month	Distribution	Goodness of Fit		
			Kolmogorov Smirnov	Anderson Darling	Chi-Square
Pune	1	Gen.Gamma	4	1	1
	2	Dagum	1	1	3
	3	Gen.pareto	4	2	1
	4	Error	1	1	2
	5	Gen.Extreme	1	1	2
	6	Gen.Gamma	1	1	4
	7	Log Logistic	1	10	7
A.NAGAR	1	Beta	2	8	1
	2	Dagum	1	1	4
	3	Dagum	3	1	1
	4	Log Logistic	2	1	1
	5	Log Logistic	1	2	1
	6	Log Logistic	1	1	1
	7	Log Logistic	1	1	1
Solapur	1	beta	1	7	1
	2	Dagum	1	1	3
	3	Gen.Extreme value	2	1	11
	4	Gen.Extreme value	1	1	2
	5	Gen.Extreme value	2	1	4
	6	Nakagami	1	20	39
	7	Johnson SU	21	1	17
Sangli	1	Dagum	1	3	1
	2	Chi squared	1	7	1
	3	Gen.Extreme	1	1	4
	4	Gen.Gamma	1	2	1
	5	Burr	2	1	1
	6	Gen.Extreme	1	4	1
	7	Burr	7	1	4
Satara	1	Beta	1	1	1
	2	Dagum	3	1	1
	3	Dagum	1	1	2
	4	Dagum	1	1	1
	5	Gumbel Min.	1	4	1
	6	Laplace	1	1	2
	7	Cauchy	1	3	1
Kolhapur	1	Beta	1	1	1
	2	Kumaraswamy	1	1	1
	3	Dagum	1	1	1
	4	Dagum	3	1	1
	5	Gumbel Min.	1	1	3
	6	Gumbel Min.	2	1	1
	7	Cauchy	4	3	1

References

[1] N. R. Patel and D. T. Shete, "Predicting consecutive days rainfall from one day rainfall using regression models for Sabarkantha district of North Gujarat region, India," Proc. XXVI National Seminar on Hydrology, Shillong, 26th-28th October 2007.

[2] N. R. Patel and D. T. Shete, "Probability distribution analysis of consecutive days rainfall data for Sabarkantha district of North Gujarat region, India," ISH Journal of Hydraulic Engineering, Vol. 14, No. 3, pp. 43-55, 2008.

[3] S. R. Bhakar, A. K. Bansal, N. Chhajed and R. C. Purohit, "Frequency Analysis of consecutive days maximum rainfall at Banswara, Rajasthan, India," ARPN Journal of Engineering and Applied Sciences, Asian Research Publishing Network (ARPN), Vol. 1, No. 3, October, pp. 64-67, 2006.

[4] N. K. Kullar, A. Soni, J. Singh and A. K. Jaggi, "Selection of rainfall frequency distribution model for Punjab," Journal of Applied Hydrology, Vol. XII, No. 4, pp. 22-28, 1999.

[5] Easy Fit Manual, Math Wave Technologies, 2010.

[6] N. R. Patel and D. T. Shete, "Predicting consecutive days maximum rainfall by probabilistic approach for Sabarkantha district of North Gujarat region, India," Proc. Shri Vasantrao Naik Memorial National Agriculture Seminar on "Livelihood Security Through Rainwater Management", Nagpur, 22nd-23rd January 2008

Author Profile

Swami Shivprasad T. DOB: 24/08/1992 is Student of Civil Engineering at Pune University. 2012-2015. Now he is in B.E. and interested in study and inventing new concepts in Civil Engineering.

Anandrao Deshmukh 'DOB-21/6/73 BE Civil 1995-Govt.College ofEngg.Karad with 66.32% M.E.Civil(hydraulics) -Bharati Vidyapeeths College of Engg.PUNE -2006 with Distn Designation-Associ. Prof.CIVIL deptt. Rajarshi Shahu C.O.E.Pune Ph.D.Pursuing-pre Ph.D.comprehensive viva complete at V.T.U.Belgaum

Ganesh Patil N DOB: 24/11/1993, is Student of Civil Engineering at Pune University, 2012-2015

Sagar Kahar, DOB, 06/12/1992 Student of Civil Engineering at Pune University. 2012-2015