Three – State Markov Chain Approach On the Behavior Of Rainfall

Vivek Kumar Garg and Jai Bhagwan Singh
School of Basic and Applied Sciences, Shobhit University, Meerut, Uttar Pradesh, 250110, India
Email: vg284@yahoo.co.in

Abstract: An analytical procedure for testing the independence of behavior of rainfall using three-state Markov Chain approach has been used in the present study. The developed analytical procedure was applied for daily rainfall data of 42 years (1961 – 2002) observed from IMD approved meteorological observatory, Pantnagar, India. The whole year was divided into three different periods viz. Pre-monsoon (Jan 1-May 31), Monsoon (June 1-Sep 30) and Post-monsoon (Oct 1-Dec 31) for the analysis of daily and weekly rainfall data. A day/week was taken as dry if the rainfall was below 2.5mm/17.5mm and day/week was taken as wet if the rainfall was between (2.5mm to 5mm)/(17.5mm to 35mm) respectively, otherwise it was taken as a rainy day/week. Based on three conditions of rainfall, during each period, it was concluded that consecutive day/week are not independent and expected length of dry, wet, rainy spells, and weather cycles of all the three periods has been computed.
[Vivek Kumar Garg and Jai Bhagwan Singh. Three – State Markov Chain Approach On the Behavior Of Rainfall. New York Science Journal 2010;3(12):76-81]. (ISSN: 1554-0200).

Key words: Three -State Markov Chain, Bernoulli trials, Rainfall occurrence pattern.

1. Introduction

Rainfall analysis is not only important for agricultural production but also for other administrative purposes. The distribution pattern of rainfall rather than the total rainfall during the entire period of time is more important for studying the pattern of rainfall occurrence. The occurrence of a sequence of wet and dry spells can be regarded as a series of Bernoulli trials.

Fisher (1924) studied the influence of rainfall on the yield of wheat and showed that during a season, the distribution of rainfall rather than total amount influence the crop yield. Gabriel and Neuman (1962); Green (1965, 1970); Weiss (1964); Wiser (1965); Feyerham and Bark, (1967); Stern et. al. (1984) developed rainfall models. Aneja & Srivastava(1986,1999) developed two-state with two parameters and three-state with five independent parameters Markov Chain model to study the pattern of rainfall occurrence. Srikanthan et. at. (2001) present a review on stochastic generation of annual, monthly and daily climate data. In this review firstly they studied traditional time series and after that more complex models using pseudo-cycles in the data.

Purohit et.al. (2008) used two –state Markov chain model to find the probabilities of occurrence of dry &wet weeks and also did weekly analysis of rainfall at Bangalore. Garg & Singh (2010) studied the pattern of rainfall occurrence at Pantnagar using two – State Markov chain approach. In the present paper, the two – state Markov chain approach is being tried to extend further for three – states Markov chain approach dividing a day/week into dry, wet and rainy (day/week).

2. Analytical Procedure

Markov assumption is that future evaluation only depends on the current state. A three – state Markov chain approach is applied on the pattern of occurrence of rainfall. In this present study, we consider three types of rainfall condition viz: a day is a dry day if rainfall is less than 2.5mm, wet day if rainfall lies between 2.5mm and less than 5mm and rainy day if rainfall is 5mm or more.

Based on the daily rainfall observation accordingly for weekly analysis, if total rainfall of seven days is less than 17.5mm , it is classified as a dry week, if lies between 17.5mm and less than 35mm, wet week and rainy week if rainfall in 35mm or more. So three possible states for each (day/week) are dry, wet, rainy and three – state Markov Chain approach has been applied to study the pattern of occurrence of dry, wet and rainy (days/weeks) at Pantnagar during Pre-monsoon, Monsoon and Post-monsoon. The expected lengths of dry, wet, rain spells and weather cycle have been derived for each period of daily and weekly rainfall data.

The data observed over the sequence of time (days/ weeks) can be regarded as a three-state Markov chain with state space S = {d, w, r} as the current (day's/week's) rainfall was supposed to depend only on the preceding (days/weeks) rainfall and the transition probability matrix is defined as

$P = (P_{ij}) = P(j \mid i)$ Where $i, j \in S$

and is given in Table 1.

Table 1. 3x3 transition probability matrix

			Current (Day/week)		
			dry (d)	wet (w)	rainy (r)
P=	Previous (Day/ week)	dry (d)	P_{dd}	P_{dw}	P_{dr}
		wet (w)	P_{wd}	P_{ww}	P_{wr}
		rainy (r)	P_{rd}	P_{rw}	P_{rr}

Where
$P_{dd} = P(d|d)$: Probability of a dry (day/week) preceded by a dry (day/week).
$P_{dw} = P(w|d)$: Probability of a wet (day/week) preceded by a dry (day/week)
$P_{dr} = P(r|d)$: Probability of a rainy (day/week) preceded by a dry (day/week).
and so on.

Subject to the condition that the sum of probabilities of each row is one i.e.

$$P_{dd} + P_{dw} + P_{dr} = 1$$
$$P_{wd} + P_{ww} + P_{wr} = 1$$
$$P_{rd} + P_{rw} + P_{rr} = 1$$

For three–state Markov chain approach, we required the values of P_{ij} $(i, j = d, w, r)$ and these values can be estimated from the 3x3 order observed frequency table 2 given below,

Table 2. 3x 3 order observed frequency table

		Current (Day/week)			
Previous (Day/ week)		dry (d)	wet (w)	rainy (r)	Total
	dry (d)	n_{dd}	n_{dw}	n_{dr}	$n_{d\bullet}$
	wet (w)	n_{wd}	n_{ww}	n_{wr}	$n_{w\bullet}$
	rainy (r)	n_{rd}	n_{rw}	n_{rr}	$n_{r\bullet}$

Where n_{ij} $(i, j = d, w, r)$ is observed frequency i.e.
n_{dd} : Number of dry (days/weeks) preceded by dry (days/weeks)
n_{dw}: Number of wet (days/weeks) preceded by dry (days/weeks)
n_{dr} : Number of rainy (days/weeks) preceded by dry (days/weeks)
and so on.

$n_{d\bullet} = n_{dd} + n_{dw} + n_{dr}$ i.e. Total number of dry (days/weeks)
$n_{w\bullet} = n_{wd} + n_{ww} + n_{wr}$ i.e. Total number of wet (days/weeks)
$n_{r\bullet} = n_{rd} + n_{rw} + n_{rr}$ i.e. Total number of rainy (days/weeks)

The Maximum likelihood estimators of P_{ij} $(i, j = d, w, r)$ are given by

$$\hat{P}_{ij} = \frac{n_{ij}}{n_{ig}} \qquad (\text{where } i, j = d, w, r)$$

3. Test of Goodness of fit:

In this section, we test the validity of the three-state Markov chain approach i.e. one-day dependence could be considered as explaining the behavior of rainfall. This can be achieved by applying traditional Chi-square test and the test suggested by Wang and Maritz in 1990. For applying these tests, we have

H_0 = Rainfall on consecutive day/week is independent

Against the alternative hypothesis

H_1 = Rainfall on consecutive day/week is not independent.

The traditional Chi-square test statistic is

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

Where
$O_i (i = 1, 2, ...n)$ are observed frequencies,
$E_i (i = 1, 2, ...n)$ are corresponding expected frequencies.
i = number of observations $(1, 2, ...n)$

The critical region for testing null hypothesis is $\chi_c^2 \geq \chi^2_{(n-1,\, \alpha)}$ where χ_c^2 is the calculated value of Chi-square and $\chi^2_{(n-1,\, \alpha)}$ is the tabulated value of Chi-square for (n–1) degree of freedom and α level of significance.

In 1990, Wang and Maritz observed that WS test statistic for independence of three–state Markov Chain is better than the traditional χ^2-test method. According to Wang and Maritz, the test statistic, for testing the above null hypothesis, is

$$WS = \frac{\hat{A} + \hat{B} - 1}{\sqrt{V(\hat{A} + \hat{B} - 1)}} \xrightarrow{D} N(0,1)$$

The variance of maximum likelihood estimator is given by

$$V\left(\hat{A}+\hat{B}-1\right)=2\pi_1\pi_2\pi_3\left(\frac{1}{n_{d\bullet}n_{w\bullet}}+\frac{1}{n_{w\bullet}n_{r\bullet}}+\frac{1}{n_{r\bullet}n_{d\bullet}}\right)$$

where

$$\hat{A}=P_{dd}+P_{ww}+P_{rr}$$

$$\hat{B}=P_{rd}P_{dr}+P_{wr}P_{rw}+P_{dw}P_{wd}-P_{dd}P_{ww}-P_{dd}P_{rr}-P_{ww}P_{rr}$$

(π_1,π_2,π_3) are stationary probabilities which are calculated as

$$\pi_1=\left[(1+p)+(1+s)p/q\right]^{-1}$$

$$\pi_2=\left[r+ps/q\right]\pi_1 \qquad \pi_3=\left[p/q\right]\pi_1$$

$$p=\left[P_{dr}+\frac{P_{wr}(1-P_{dd})}{P_{wd}}\right]\left(\frac{1}{1-P_{rr}}\right)$$

$$r=\left(\frac{P_{dw}}{1-P_{ww}}\right) \qquad q=1+\left[\frac{P_{wr}P_{rd}}{P_{wd}(1-P_{rr})}\right]$$

$$s=\left(\frac{P_{rw}}{1-P_{ww}}\right)$$

Critical region is $(WS)_c \geq Z_\alpha$ at $'\alpha'$ level of significance i.e. we reject null hypothesis, if $|WS|>Z_\alpha$, where Z_α is the $100(1-\alpha)$ lower percentage point of a standard normal distribution.

4. Expected length of different spells:

(a) A dry spell of length ‘d’ is defined as sequence of consecutive dry (days/week) preceded and followed by wet or rainy (days/week) and the probability of a sequence of ‘d’ dry (days/week) is

$$P(d)=(P_{dd})^{d-1}(1-P_{dd})$$

and expected length of dry spell is

$$E(D)=\frac{1}{(1-P_{dd})}$$

where $(1-P_{dd})$ is the probability of a day being wet or rainy day.

(b) A wet spell of length ‘w’ is defined as sequence of consecutive wet (days/week) preceded and followed by dry or rainy (days/week) and the probability of a sequence of ‘w’ wet (days/week) is

$$P(w)=(P_{ww})^{w-1}(1-P_{ww})$$

and expected length of dry spell is

$$E(W)=\frac{1}{(1-P_{ww})}$$

where $(1-P_{ww})$ is the probability of a (days/week) being dry or rainy (days/week).

(c) Similarly, for rainy spell of length ‘r’, we have

$$P(r)=(P_{rr})^{r-1}(1-P_{rr})$$

and expected length of rain spell is

$$E(R)=\frac{1}{(1-P_{rr})}$$

where $(1-P_{rr})$ is the probability of a (days/week) being dry or wet day and

(d) Weather cycle (WC) is

$$E(WC)=E(D)+E(W)+E(R)$$

(e) The number of days/weeks (N) after which the equilibrium state is achieved is equal to the number of times the matrix ‘P’ is powered till the elements of a column of the matrix $P^{(N)}$ become equal i.e. for 3x3 matrix ‘P’

$$P^{(N)}=\begin{bmatrix}\pi_1 & \pi_2 & \pi_3\\ \pi_1 & \pi_2 & \pi_3\\ \pi_1 & \pi_2 & \pi_3\end{bmatrix}$$

where

π_1 : Probability of a day/week being a dry day/dry week

π_2 : Probability of a day/week being a wet day/wet week

π_3 : Probability of a day/week being a rainy day/rainy week

5. Results:

In the present study, the daily rainfall data for 42 years from Jan.1, 1961 to Dec31, 2002, was collected from the meteorological observatory, Pantnagar, India located at 29N latitude, 79.3E longitude and altitude 243.84m above mean sea level. On an average the region has a humid subtropical climate having hot summers (40-42^{0}C) and cold winters (2-4^{0}C) and rain is received from south-west monsoon during June to September. The rainfall occurrence, based on daily rainfall and weekly rainfall, was studied for three different periods, viz. Pre-monsoon, Monsoon and Post-monsoon.

(i) The yearly maximum daily rainfall based on 42 years is presented in Figure 1 and the weekly maximum rainfall based on 42 years data for 52 weeks is also presented in Figure 2.

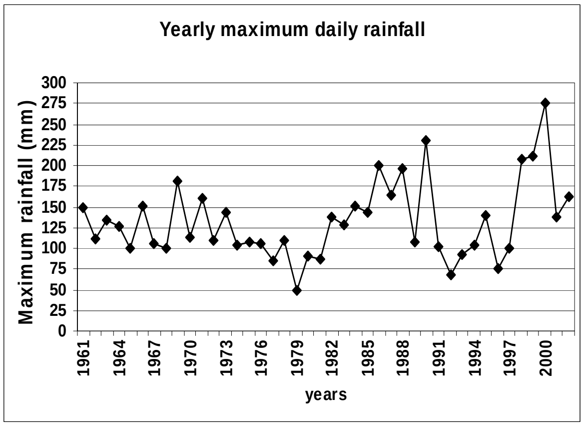

Figure 1. Yearly maximum daily rainfall

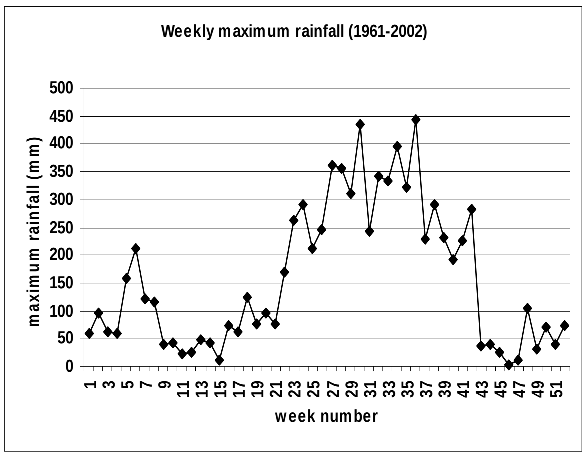

Figure 2. Weekly maximum rainfall

(ii) Based on daily rainfall data, the calculated values of traditional Chi-square statistic and WS test statistic, for three different periods are reported in table 3.

Table 3. Calculated values of Chi-square and WS test statistic

Chi-square test statistic	Pre-monsoon	Monsoon	Post-monsoon
	390.50	559.14	382.09

WS test statistic	Pre-monsoon	Monsoon	Post-monsoon
	87.24	16.79	17.58

(iii) Based on weekly rainfall data, the calculated values of traditional Chi-square statistic and WS test statistic, for three different periods are reported in table 4.

Table 4. Calculated values of Chi-square and WS test statistic

Chi-square test statistic	Pre-monsoon	Monsoon	Post-monsoon
	59.16	45.76	42.24

WS test statistic	Pre-monsoon	Monsoon	Post-monsoon
	8.60	19.52	13.55

(iv) The estimated transition probability matrices for daily rainfall data are reported in table 5a (Pre–monsoon), table 5b (Monsoon) and table 5c (Post–monsoon) respectively

Table 5a. Pre–monsoon based on daily rainfall

Transition probability matrix			
	d	w	r
d	0.948	0.015	0.037
w	0.780	0.110	0.110
r	0.696	0.048	0.256

Table 5b. Monsoon based on daily rainfall

Transition probability matrix			
	d	w	r
d	0.755	0.037	0.208
w	0.544	0.077	0.379
r	0.409	0.090	0.501

Table 5c. Post–monsoon based on daily rainfall

Transition probability matrix			
	d	w	r
d	0.977	0.006	0.017
w	0.774	0.032	0.194
r	0.636	0.081	0.283

(v) The estimated transition probability matrices for weekly rainfall data are reported in table 6a (Pre–monsoon), table 6b (Monsoon) and table 6c (Post–monsoon) respectively.

Table 6a. Pre–monsoon based on weekly rainfall

Transition probability matrix			
	d	w	r
d	0.904	0.050	0.046
w	0.632	0.263	0.105
r	0.830	0.064	0.106

Table 6b. Monsoon based on weekly rainfall

Transition probability matrix			
	d	w	r
d	0.450	0.130	0.420
w	0.237	0.175	0.588
r	0.201	0.134	0.664

Table 6c. Post–monsoon based on weekly rainfall

Transition probability matrix			
	d	w	r
d	0.947	0.030	0.022
w	0.700	0.200	0.100
r	0.783	0.043	0.174

(vi) Based on daily rainfall data, equilibrium state probabilities, expected length of different spells, weather cycles and number of days required for the system to achieve the equilibrium state (N), are reported in table 7.

(vii) Based on weekly rainfall data, equilibrium state probabilities, expected length of different spells, weather cycles and number of weeks required for the system to achieve the equilibrium state (N), are reported in table 8.

6. Conclusions:

(i) The data under study was collected for 42 years (1961-2002) from IMD approved meteorological observatory, Pantnagar. From figure 1, it can be seen that yearly maximum daily rainfall is lowest (50mm) in year 1979 and highest (275mm) in year 2000 over 42 years. Similarly, over the same period it can be seen from figure 2 that weekly maximum rainfall is lowest (0 mm) in 46th week (Nov, 12-18) and highest (443mm) in 36th week (Sep, 3-9). These values indicating very large fluctuation during the period of study.

(ii) Before applying Three-state Markov Chain approach, we tested H_0 against H_1 by applying traditional Chi-square test and WS test, for all three periods (based on daily and weekly rainfall data). From table 3 and table 4, it is clear that, according to traditional Chi-square and WS test statistic the null hypothesis of independence of rainfall on consecutive day/week is rejected at 5% level of significance. This indicates that current day/week rainfall depends on preceding day/week rainfall.

(iii) Because current day/week rainfall depends on preceding day/week rainfall so three-state Markov Chain approach is used to study the behavior of rainfall. Estimated transition probability matrices are reported in table 5a (Pre–monsoon), table 5b (Monsoon) and table 5c (Post–monsoon) for daily rainfall data and in table 6a (Pre–monsoon), table 6b (Monsoon) and table 6c (Post–monsoon) for weekly rainfall data respectively.

(iv) Based on daily rainfall data, equilibrium state probabilities π_1, π_2 and π_3 of a day being dry, wet or rainy are given in table 7 and on the basis of these values it can be said that.

(a) On the average for every wet day there are about 47, 13 and 97 dry days and about 3, 6 and 2 rainy days for all three periods.

(b) Similarly, on the average for every rainy day there are about 19, 2 and 49 dry days and about 1, 1 and 1 wet days for all three periods.

(v) Based on weekly rainfall data, equilibrium state probabilities π_1, π_2 and π_3 of a week being dry, wet or rainy are given in table 8 and on the basis of these values it can be said that

(a) On the average for every wet week there are about 13, 2 and 23 dry weeks and about 1, 4 and 1 rainy weeks for all three periods.

(b) Similarly, on the average for every rainy week there are about 18, 1 and 31 dry weeks and about 1, 1 and 1 wet weeks for all three periods.

(vi) The number of days required for the system to achieve the equilibrium state (N) are 7, 10 and 9 for daily rainfall data which indicates that after 7 days from Jan 1, 10 days from June 1 and 9 days from Oct 1, the probability of the day being dry, wet or rainy is independent of the initial weather condition during the pre-monsoon, monsoon and post-monsoon respectively.
Similarly, for weekly data analysis, the number of weeks required for the system to achieve the equilibrium state (N) are 7, 7 and 8 which indicates that after 7 weeks from Jan 1, 7 weeks from June 1 and 8 weeks from Oct 1, the probability of the week being dry, wet or rainy is independent of the initial weather condition during the pre-monsoon, monsoon and post-monsoon respectively.

Table 7. Equilibrium state probabilities, expected length of different spells, weather cycles and number of days for equilibrium state

	π_1	π_2	π_3	Expected length of				N	Total days
				Dry run	Wet run	Rain run	Cycle		
Pre monsoon	0.93	0.02	0.05	19.0	1.0	1.0	21.0	7	151
Monsoon	0.64	0.05	0.31	4.0	1.0	2.0	7.0	10	122
Post monsoon	0.97	0.01	0.02	43.0	1.0	1.0	45.0	9	92

Table 8. Equilibrium state probabilities, expected length of different spells, weather cycles and number of weeks for equilibrium state

	π_1	π_2	π_3	Expected length of				N	Total weeks
				Dry run	Wet run	Rain run	Cycle		
Pre monsoon	0.88	0.07	0.05	10.0	1.0	1.0	12.0	7	22
Monsoon	0.27	0.14	0.59	1.0	1.0	2.0	6.0	7	17
Post monsoon	0.93	0.04	0.03	19.0	1.0	1.0	21.0	8	13

References:

1. Aneja D.R., Srivastava O.P. Markov chain model for rainfall occurrence. Journal Indian Society Agriculture Statistics. 1999;52(2): 169 – 175.
2. Aneja D.R., Srivastava O.P. A study technique for rainfall pattern. Aligarh Journal of Statistics. 1986; vol.6:26 – 31.
3. Feyerherm A. M., Bark L. D. Goodness of fit of a Markov Chain Model for Sequences of wet and dry days. Journal of Applied Meteorology. 1967; vol.6: 770–773.
4. Fisher R.A. The influence of the rainfall on the yield of wheat at Rothamsted. Philosophical transaction of the Royal Society of London. 1924; Series B, Vol.: 213.
5. Gabriel, K R., Neumann, J. A Markov chain Model for study for daily rainfall occurrence at Tel–Aviv. Quarterly Journal Roy. Met. Society. 1962; 88: 90–95.
6. Garg V.K., Singh J. B. Markov Chain Approach on the behavior of Rainfall. International Journal of Agricultural and Statistical Sciences. 2010; vol.6: No.1.
7. Green J. R. Two probability models for sequences of wet or dry days. Monthly Weather Review. 1965; 93:155–156.
8. Green J. R. A generalized Probability Model for Sequences of Wet and Dry Days. Monthly Weather Review. 1970; vol.98: No.3. 238 – 241.
9. Purohit R.C., Reddy G. V. S., Bhaskar S. R., Chittora A. K. Markov Chain Model Probability of Dry, Wet Weeks and Statistical Analysis of Weekly Rainfall for Agricultural Planning at Bangalore. Karnataka Journal of Agricultural Science. 2008; 21 (1):12-16.
10. Srikanthan, R., McMahon, T. A. Stochastic Generation of Annual, Monthly and Daily Climate Data: A Review. Hydrology and Earth System Sciences. 2001; 5(4):653–670.
11. Stern R.D., Coe R. A model fitting analysis of daily rainfall data. Journal of Royal Statistical Society. 1984; 147(1):1- 34.
12. Wang D. Q., Martiz J. S. Note on testing a three-state Markov Chain for independence. Journal Statistics Computation and Simulation. 1990; vol. 37: 61 – 68.
13. Weiss L. L. Sequences of wet or dry days described by a Markov Chain probability Model. Monthly Weather Review. 1964; 92: 169 – 176.
14. Wiser E. H. Modified Markov probability models of sequences of precipitation events. Monthly Weather Review. 1965: 511–516.

09/17/2010