Indian J. Met. Hydrol. Geophys. (1976), 27,4,431-435

551.577 : 556.16 (541.1)

A Markov chain model for the occurrences of dry and wet days

J. MEDHI

Department of Statistics, Gauhati University Gauhati,

(Received 19 November 1975)

ABSTRACT. Study of rainfall data in India began as early as 1889 with the work of Blanford. His work was extended by Sir Gilbert Walker. Mahalanobis (1940), in his classic work, broke entirely new grounds with sound statistical reasonings in his study of the rainfall, runoff and other meteorological features of the river basins of Orissa. Models based on stochastic processes and time series are now being applied in meteorological studies. Gabriel and Neumann (1962) used a Markov chain model in a study of weather conditions in Tel Aviv. We use here a similar Markov Chain model for the occurrence of dry and wet days in Gauhati (Airport) and test the statistical significance of the model.

1. Introduction

Here the patterns of occurrences of dry and wet days during two periods have been studied. The days are recorded under two categories : dry and wet. A day is called dry if the rainfall during the 24 hours commencing from 0830 IST on that day is either *nil* or trace; otherwise, it is termed wet. One of the two periods is from 1 March to 31 May and the other is from 1 June to 30 September. The period March-May is pre-monsoon season and the period June-September is monsoon season, the monsoon in Gauhati normally commencing in the first week of June and withdrawing in the beginning of October (Das 1968). Two sets of data of the pre-monsoon season (March-May) are considered: set I covering a total of $4\times92 = 368$ days in the 4 years, 1967-70 and set II covering a total of 92 days in the year 1975. Two other sets of data relating to the monsoon (June-September) are taken : set III covers a total of $4\times122=488$ days in the four years 1967—70 and set IV, a total of 122 days in the year 1975. The data are taken from the records maintained by the Meteorological Centre Gauhati Airport for rainfall at the Gauhati Airport.

The total annual rainfall (in mm) and the amount of rainfall (in mm) during the monsoon months (June-September) during 1964-70 are given in Table 1

It may be seen from the above that the variability of annual rainfall is low ; except for the year 1967, rainfall for other years lie within 10 per cent of the average amount 1686·37 mm. This is in agreement with the observation of Parthasarathy (1958).

2. Markov Chain model

As considered by Gabriel and Neumann (1962), a two-state Markov chain model has been used to describe the occurrences of dry and wet days at Gauhati. Assume that the probability of dry (or wet) days depend on the condition (dry or wet) of the previous day.

Let. P_r [actual day is wet, given that precedin day is dry]$=p_0$ and

P_r [actual day is dry, given that preceding day is wet]$=p_1$

Denote dry day by state 0 and wet day by state 1; the occurrences of dry and wet days can then be denoted by an irreducible Markov chain with two ergodic states 0 and 1 and transition probability matrix

$$A = \begin{array}{c} \\ 0 \\ 1 \end{array}\begin{array}{c} \quad 0 \qquad\quad 1 \\ \begin{pmatrix} 1-p_0 & p_0 \\ p_1 & 1-p_1 \end{pmatrix} \end{array}$$

The n-step transition probabilities are given by the elements of A^n.

where

$$A^n = \frac{1}{p_0+p_1}\begin{pmatrix} p_1 & p_0 \\ p_1 & p_0 \end{pmatrix} + \frac{(1-p_0-p_1)^n}{p_0+p_1}\begin{pmatrix} p_0 & -p_0 \\ -p_1 & p_1 \end{pmatrix}$$

Further, $\lim_{n\to\infty} A^n = \begin{pmatrix} v_0 & v_1 \\ v_0 & v_1 \end{pmatrix}$

where $v_0 = p_1/(p_0+p_1)$ and $v_1 = p_0/(p_0+p_1)$

TABLE 1

Rainfall	1964	1965	1966	1967	1968	1969	1970
			ANNUAL				
Amount (mm)	1716·3	1850·8	1661·1	1462·5	1625·1	1648·8	1840·0
			JUNE—SEPTEMBER				
Amount (mm)	1072·1	1166·7	1209·9	869·3	1019·8	1314·4	1372·0
As % of annual rainfall	62·47	63·04	72·84	59·44	62·75	79·72	74·57

Define a wet spell W (dry spell D) of length r as a sequence of r wet (dry) days preceded and followed by dry (wet) days. It follows that the distributions of W and D are geometric with parameters p_1 and p_0 respectively, *i.e.*

$$Pr\{W=r\}=p_1\,(1-p_1)^{r-1}\ ,\ r=1,2,3,\ldots$$

and $Pr\{D=r\}=p_0\,(1-p_0)^{r-1}\ ,\ r=1,2,3,\ldots$

so that the mean $E(W)=1/p_1$ and mean $E(D)=1/p_o$

The central limit theorem for dependent variables gives that the distribution of the number of wet days Y_n in a sequence of n transitions (or out of n days) is asymptotically normal with

$$\text{mean}\,E\,(Y_n)=\frac{n\,p_0}{p_0+p_1} \quad \text{and}$$

$$\text{var}\,(Y_n)=\frac{np_0\,p_1\,(2-p_0-p_1)}{(p_0+p_1)^3}$$

3. Data points for Gauhati

The actual transitions corresponding to data points of set I (pre-monsoon 1967-70) are as given below :

		Actual day		
		Dry (0)	Wet (1)	Total
Preceding day	Dry (0)	175	49	224
	Wet (1)	48	96	144
	Total	223	145	368

The maximum likelihood estimates of p_0 and p_1 are

$$\hat{p}_0=49/224=0\cdot21875 \qquad \text{and}$$

$$\hat{p}_1=48/144=0\cdot33333$$

The estimated transition probability matrix for the pre-monsoon seasons 1967-70 is

$$P=(\hat{p}_{ij})=\begin{pmatrix}0.78125 & 0.21875\\ 0.33333 & 0.66667\end{pmatrix}$$

We can easily find the n step transition probabilities from the matrix $P^n=(p_{ij}{}^{n})$ where $p_{ij}(^n)$ denotes the probability of transition from state to state j after n days $(i,j=0,1,)$. The matrix P^n for some values of n are given below :

$$P^2=\begin{pmatrix}0\cdot68327 & 0\cdot31673\\ 0\cdot48264 & 0\cdot51736\end{pmatrix},\ P^5=\begin{pmatrix}0\cdot61091 & 0\cdot38909\\ 0\cdot59289 & 0\cdot40711\end{pmatrix}$$

$$P^{10}=\begin{pmatrix}0\cdot60390 & 0\cdot39610\\ 0\cdot60358 & 0\cdot39642\end{pmatrix}$$

For large n, $(n\geqslant 20)$

$$P^n=\begin{pmatrix}0\cdot60377 & 0\cdot39623\\ 0\cdot60377 & 0\cdot39623\end{pmatrix}$$

From P^2, we find that the probability that 5 April is a wet day given that 3 April was a dry day is $0\cdot31673$ and so on. From P^{10}, the probability that 18 May is a dry day given that 8 May was wet day is $0\cdot60358$ and so on. The limiting matrix P^n as $n\rightarrow\infty$, indicates that the probability of a dry day is $0\cdot60377$ and that of a wet day is $0\cdot39623$.

The actual transitions for set II consisting of 92 data points (for pre monsoon 1975) are as given below :

		Actual day		
		0	1	Total
Proceeding day	0	52	9	61
	1	11	20	31
	Total	63	29	92

The estimated transition probability matrix is

$$\hat{Q}=(\hat{q}_{ij})=\begin{pmatrix}0\cdot85246 & 0\cdot14754\\ 0\cdot35484 & 0\cdot64516\end{pmatrix}$$

The transitions for set III consisting of $122\times4=488$ data points (for monsoon 1967—70) are as given below :

		Actual day		
		0	1	Total
Preceding day	0	72	81	153
	1	82	253	335
	Total	154	334	488

The estimated transition probability matrix is

$$\hat{R}=(\hat{r}_{ij})=\begin{pmatrix}0\cdot47059 & 0\cdot52941\\ 0\cdot24478 & 0\cdot75522\end{pmatrix}$$

The transitions for set IV consisting of 122 data points (for monsoon 1975) are as given below:

		Actual day		
		0	1	Total
Preceding day	0	19	20	39
	1	21	62	83
	Total	40	82	122

The estimated transition probability matrix is

$$\hat{S}=(\hat{s}_{ij})=\begin{pmatrix}0\cdot48718 & 0\cdot51282\\ 0\cdot25301 & 0\cdot74699\end{pmatrix}$$

4. Adequacy of the model

The adequacy of the model can be tested by the tests of goodness of fit proposed by Anderson and Goodman (1957). We consider the following tests.

(I) Test the hypothesis H_0	Observations at successive time points are statistically independent.
against alternative hypothesis H_1	Observations are from a Markov chain of order one.
(II) Test the hypothesis H_0	Observations are from Markov chain of order one.
against alternative hypothesis H_1	Observations are from Markov chain of order two,
(III) Test the hypothesis H_0	Observations are from Markov chain of order two.
against alternative hypothesis H_1	Observations are from a Markov chain of order three,

It appears that the tests (I) and (II) demonstrate quite well the adequacy of the Markov chain model of order one for the data of set III (monsoon seasons, 1967-70) while test (III) is also considered for demonstrating the adequacy for the Markov chain model of the data of set I (pre-monsoon 1967-70). For applying the tests we break the data of sets I and III further as in table 2.

TABLE 2

	Preceding days			*Actual days*	
	Third	Second	First	0	1
		For set I Pre-moon-soon season 1967-70.			
(*a*)		0	0	142	35
		1	0	33	14
		0	1	20	29
		1	1	28	67
(*b*)	0	0	0	119	22
	1	0	0	23	13
	0	1	0	14	6
	1	1	0	19	8
	0	0	1	13	22
	1	0	1	7	7
	0	1	1	8	21
	1	1	1	20	46
	For set III	Monsoon season 1967-70			
(*a*)		0	0	33	41
		1	0	39	40
		0	1	21	58
		1	1	61	195

Application of test I to data of set I, gives the value of $\chi^2=24\cdot76$ for 1 d.f., which is very highly significant. So the null hypothesis is rejected in favour of the alternative. Further test II gives $\chi^2=4\cdot05$ at 2 d.f. so that the null hypothesis is not rejected, the probability of the difference arising due to chance being $0\cdot10$—$0\cdot20$. Test III gives $\chi^2=11\cdot08$ at 6 d.f (probability$=0\cdot05$—$0\cdot10$). Thus we find that Markov chain of order 1 gives an adequate fit for the observations of the premonsoon period-(1967-70).

Applications of the test I to the observations of set III gives $\chi^2=73\cdot65$ for 1 d. f. which is very very highly significant and the null hypothesis is rejected in favour of the alternative. Test II gives $\chi^2=0\cdot60$ for 2 d.f. so that the null hypothesis is accepted with probability $0\cdot60$-$0\cdot70$. Thus the two tests are considered sufficient for testing the adequacy of the Markov chain model of order 1 for abservations of the monsoon seasons 1967-70.

We now test the matrix Q for significant difference from P by means of the statistic.

$$\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{2} n_i\,(\hat{q}_{ij}-\hat{p}_{ij})^2/\hat{p}_{ij}$$

TABLE 3

Length of spell/ Frequencies	Premonsoon 1967-70				Monsoon 1967-70			
	Dry spell		Wet spell		Dry spell		Wet sepll	
	O	E	O	E	O	E	O	E
1	15	11·861	19	15·559	41	42·882	22	19·012
2	12	9·156	8	10.296	27	20·180	11	14·378
3	4	12·523	6	6·814	4	13·966	14	10·874
4	3		2	7·488	4		11	8·223
5	5	4·211	4		2		3	10·923
6	3	5·747	2		1	3·971	4	
7	2		3	5·843	2		1	
8	8	8·503	2				3 1	9·632
>8							8	4·958
Total	52	52·001	46	46·000	81	80·999	78	78·000
Estimated mean		4·38462		2·95652		1·88889		4·10256
Estimated parameter of geometric distribution		0·22807		0·33824		0·52941		0·24375
χ^2		4·839		3·224		5·203		8·599
d.f.		4		3		2		
Probability		·30-·50		·30-·50		·05-·10		·10-·20

The calculated value of $\chi^2=1\cdot87$ at 2 d.f. The null hypothesis of Q not being different from P is not rejected, the probability of the difference arising due to chance being 0·30-0·50. This implies that on the basis of evidence there is no reason to assume that the patterns of occurrences of dry and wet days in the two pre-monsoon periods (1967-70) and 1975 differ significantly.

The calculated value of χ^2 at 2 d.f. for testing the matrix S for significant difference from the matrix R is found to be 0·073. The hypothesis of S not being different from R is not rejected, the probability of the difference arising due to chance lying between 0·95-0·98. It shows that the transition probability matrix obtained on the basis of one season only (whether pre-monsoon or monsoon) is not significantly different from that based on some other season. Thus, it can be taken that the data of one season only may be used for determining the structure of the Markov chain.

The matrix R is also tested for significant difference from P by means of a similar statistic. The calculated value of $\chi^2=98\cdot22$ at 2 d.f., is highly significant. The occurrences of dry and wet days during pre-monsoon and monsoon seasons do not follow the same pattern. In other words, though each of the transitions of set I and set III can be described by a Markov chain of order one, the corresponding transition probability matrices are significantly different, as is to be expected.

5. Dry and wet spells and number of wet days

Table 1 gives the observed (O) and expected (E) frequencies of lengths of dry and wet spells, the estimated parameter of the corresponding geometric distributions and the calculated value of χ^2 for testing the goodness of fit of the geometric distributions.

On the basis of available data there is no reason to reject the hypothesis that the distribution are geometric. Here it is to be noted that some bias may have crept in labelling a day as dry or wet in view of fact that a day with rainfall trace has been called dry. Further, a wet (dry) spell more than half of which, lies in a particular month, is being included in that month. This may have also caused some bias. Nevertheless, the geometric distributions give good fit and the estimated values of the parameters of the fitted distributions agree very well with the elements of the corresponding transition probability matrices. From the estimated mean length of dry and wet spells we get the estimated mean length of weather cycle for pre-monsoon season to be 7·34 days (the same estimated from $1/\hat{p}_0+1/\hat{p}_1$ gives 7·57 days) and that for the monsoon period to be 5·99 days (the same estimated from $1/\hat{p}_0+1/\hat{p}_1$ being 5·97 days).

The distribution of the number Y_n of wet days is asymptotically normal. For pre-monsoon season of $n=92$ days, it is found that $E(Y_n)=36{\cdot}45$ and var $(Y_n)=57{\cdot}722$. Thus the number of wet days in pre-monsoon seasons would lie in the range $(36{\cdot}45\pm2\times7{\cdot}60)=(21{\cdot}25,\ 51{\cdot}65)$ for 95% of the years. In each of the years 1967—70, it was around 40, *i.e.*, well within this range. For monsoon seasons of $n=122$ days we have $E(Y_n)=83{\cdot}42$ and var $(Y_n)=41{\cdot}769$ so that the number of wet days in monsoon seasons in 95% of the years would lie within the range $(83{\cdot}42\pm2\times6{\cdot}46)=(70{\cdot}50,\ 96{\cdot}34)$. In each of the years 1967-70, it was between 80 to 90, which is well within the range.

6. Conclusions

We thus conclude that there seems to be good ground for assuming that the weather conditions of Gauhati for each of pre-monsoon and monsoon seasons can be adequately described by a Markov chain model of order one, and that the wet and dry spells follow geometric distributions.

Acknowledgement

Thanks are due to Shri P. K. Mishra and Shri B. K Sridhar of the Meteorological Centre Gauhati Airport for kindly supplying the data.

REFERENCES

Anderson, T. W. and Goodman, L. A.	1957	*Ann. Math. stat.*, **28**, pp. 89-110.
Cox, D. R. and Miller, H. D.	1965	*The theory of Stochastic Processes*, Methuen & Co. Ltd., London.
Das, P. K.	1968	*The Monsoon*, National Book Trust, India.
Gabriel, K. R. and Neumann, J.	1957	*Quart. J. R. met. Soc.*, **83**, pp. 375-380.
	1962	*Ibid.*, **88**, pp. 90-95.
Mahalanobis, P. C.	1940	*Sankhya*, **5**, pp. 1-20.
Parthasarathy, K.	1958	*Monsoons of the World*, India met. Dep., New Delhi.